



ERNEST ORLANDO LAWRENCE BERKELEY NATIONAL LABORATORY

NMR Studies of DNA Oligomers and Their Interactions with Minor Groove Binding Ligands

P.A. Fagan
Structural Biology Division

May 1996
Ph.D. Thesis

RECEIVED
SEP 23 1996
OSTI



DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor The Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or The Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof, or The Regents of the University of California.

Ernest Orlando Lawrence Berkeley National Laboratory
is an equal opportunity employer.

**NMR Studies of DNA Oligomers and Their Interactions
with Minor Groove Binding Ligands**

Patricia A. Fagan
Ph.D. Thesis

Department of Chemistry
University of California, Berkeley

and

Structural Biology Division
Ernest Orlando Lawrence Berkeley National Laboratory
University of California
Berkeley, CA 94720

May 1996

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED
RB

This work was supported by the National Institutes of Health (NIH) Grant No. GM43129 through the U.S.
Department of Energy under Contract No. DE-AC03-76SF00098.

MASTER

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

NMR Studies of DNA Oligomers and Their Interactions
with Minor Groove Binding Ligands

by

Patricia A. Fagan

B.S. (Purdue University) 1989

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Chemistry

in the

GRADUATE DIVISION

of the

UNIVERSITY of CALIFORNIA, BERKELEY

Committee in charge:

Professor David E. Wemmer, Chair

Professor Ignacio Tinoco, Jr.

Professor Stuart Linn

1996

Abstract

NMR Studies of DNA Oligomers and Their Interactions with Minor Groove Binding Ligands

by

Patricia A. Fagan

Doctor of Philosophy in Chemistry

University of California, Berkeley

Professor David E. Wemmer, Chair

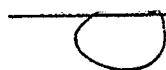
The cationic peptide ligands distamycin and netropsin bind noncovalently to the minor groove of DNA. The binding site, orientation, stoichiometry, and qualitative affinity of distamycin binding to several short DNA oligomers were investigated by NMR spectroscopy. The oligomers studied contain A,T-rich or I,C-rich binding sites, where I = 2-desaminodeoxyguanosine. I•C base pairs are functional analogs of A•T base pairs in the minor groove. The different behaviors exhibited by distamycin and netropsin binding to various DNA sequences suggests that these ligands are sensitive probes of DNA structure. For sites of five or more base pairs, distamycin can form 1:1 or 2:1 ligand:DNA complexes. Cooperativity in distamycin binding is low in sites such as AAAAAA which have narrow minor grooves, and is higher in sites with wider minor grooves such as ATATAT. The distamycin binding and base pair opening lifetimes of I,C-containing DNA oligomers suggest that the I,C minor groove is structurally different from the A,T minor groove.

Molecules which direct chemistry to a specific DNA sequence could be used as antiviral compounds, diagnostic probes, or molecular biology tools. I studied two ligands in which reactive groups were tethered to a distamycin to increase the sequence specificity

of the reactive agent. The DNA cleaving agent netropsin-diazene binds at the predicted DNA site, but the presence of the diazene reduces the affinity and the orientational preference of the ligand compared to distamycin. A tethered ligand containing a 2,3-bis(hydroxymethyl) pyrrole crosslinked to a DNA duplex exhibits many similarities to mitomycin C:DNA complexes. The tether unexpectedly bridges one base pair. These results are important considerations in ligand design, since sometimes ligand binding behavior cannot be predicted from the behavior of the parent molecule.

The pairing geometry of cytosine with the highly mutagenic base analog 2-aminopurine has been debated for years. With the use of NMR spectroscopy and selectively ^{15}N -labeled DNA, the AP•C mispair was found to form a wobble geometry. The structure and stability of this base mispair depend upon the local base sequence.

Approved:

A handwritten signature, possibly "J. D. Smith", written in dark ink over a horizontal line.

Date

5.2.96

This dissertation is dedicated to

my parents

Tom and Denna Fagan

and

my husband

Frank Jones

Table of Contents

List of Figures	ix
List of Tables	xiii
Acknowledgments	xiv

CHAPTER 1

Introduction

1.1 Summary	1
1.2 DNA Structure	2
1.2.1 B-form DNA	2
1.2.2 Non-Standard DNA Structures	5
1.2.3 A-Tract DNA	5
1.3 Binding of Ligands to DNA	8
1.3.1 Minor Groove Binding Ligands	8
1.3.2 Minor Groove Binding Proteins	11
1.4 Nuclear Magnetic Resonance Spectroscopy	12
1.4.1 General Principles	13
1.4.2 The Basic NMR Experiment	15
1.4.3 Spectral Density	16
1.4.4 The Density Matrix	16
1.4.5 T_1 and T_2 Relaxation	18
1.4.6 Dipolar Relaxation	19
1.4.7 Nuclear Overhauser Effect	19
1.4.7.1 General Principles	19
1.4.7.2 The Solomon Equation	22
1.4.7.3 Distance-Dependence of the NOE	23
1.4.7.4 Two-Dimensional NOE Spectroscopy	26
1.4.7.5 NOESY Data Analysis	27
1.4.8 Chemical Exchange	30
1.4.8.1 Time Scales	32
1.4.8.2 Measurement of Exchange Kinetics by NMR	34
1.4.8.3 Exchange Effects in 2D NOESY	35

1.5 Molecular Modeling	38
1.5.1 Forcefield	38
1.5.2 Energy Minimization	40
1.5.3 Molecular Dynamics	41
1.6 Application of NMR and Molecular Modeling Techniques to DNA and its Complexes	42

CHAPTER 2

The Binding Modes of Distamycin A to A,T- and I,C-Rich DNA Oligomers

2.1 Summary	44
2.2 Introduction	45
2.3 Materials and Methods	47
2.3.1 NMR Sample Preparation	47
2.3.2 NMR Experiments	49
2.3.3 NMR Data Processing and Analysis	49
2.3.4 Distance Restraints	50
2.3.5 Molecular Modeling	50
2.3.6 Base Pair Opening Lifetime Measurements	51
2.3.7 Job's Plots	52
2.4 Distamycin and Netropsin Binding to Four Base Pair Sites	53
2.4.1 Distamycin Binding to AAAA:TTTT	53
2.4.2 Netropsin Binding to AAAA:TTTT,	61
2.4.3 Distamycin Binding to IICC	68
2.4.4 Base Pair Opening Kinetics for IICC	71
2.5 Distamycin Binding to Five and Six Base Pair Sites	74
2.5.1 Distamycin Binding to IIICC:IICCC	74
2.5.2 Distamycin Binding to Other I,C-Containing Sites	81
2.5.3 Distamycin Binding to TTATAA	90
2.5.4 Distamycin Binding to ATATAT	93
2.6 Conclusions	101

CHAPTER 3

The Binding Modes of a Rationally Designed Photoactivated DNA Nuclease: Netropsin-Diazene Bound to d(CGCAAAAGGC):d(GCCTTTTGCG)

3.1 Summary	103
3.2 Introduction	104
3.3 Materials and Methods	109
3.3.1 Synthesis of Ligands and Oligonucleotides	109
3.3.2 NMR Sample Preparation	109
3.3.3 NMR Experiments	110
3.3.4 Distance Restraints	110
3.3.5 Model Refinement	112
3.4 Results	112
3.4.1 Titration of the AAAA Site with Netropsin-Diazene	112
3.4.2 NMR Resonance Assignments	112
3.4.3 Intermolecular Contacts	119
3.4.4 Molecular Modeling	122
3.5 Discussion	123
3.5.1 DNA Cleavage	123
3.5.2 Binding Site	125
3.5.3 Complex Stoichiometry	126
3.5.4 Ligand Orientation	126
3.5.5 Binding Affinity	127
3.5.6 Molecular Design	128
3.6 Conclusions	129

CHAPTER 4

NMR Characterization of [d(CGCGAATTCGCG)]₂ Containing an Interstrand Cross-Link Derived from a Distamycin-Pyrrole Conjugate

4.1 Summary	130
4.2 Introduction	130

4.3 Materials and Methods	135
4.3.1 Synthesis and Purification of Crosslinked Oligonucleotides	135
4.3.2 NMR Sample Preparation	136
4.3.3 NMR Experiments and Resonance Assignments	137
4.3.4 Distance Restraints and Model Refinement	137
4.4 Results	138
4.4.1 Crosslinked Sample Preparation	138
4.4.2 Titration of the AATT Site with XL-Dst	138
4.4.3 One-Dimensional NMR of Crosslinked DNA	138
4.4.4 NMR Resonance Assignments of DNA	139
4.4.5 NMR Resonance Assignments of the Ligand	139
4.4.6 Ligand-DNA Contacts	143
4.4.7 Molecular Modeling of the Crosslink	143
4.5 Discussion	148
4.5.1 Site of Crosslink	148
4.5.2 Distortion of the DNA Due to Crosslink	149
4.5.3 Overall Binding Site	151
4.5.4 Crosslinking Mechanism	151
4.5.5 Comparison to Mitomycin C	153
4.5.6 Molecular Design	154
4.6 Conclusions	155

CHAPTER 5

An NMR Study of the Conformation of the 2-Aminopurine•Cytosine Mismatch in DNA

5.1 Summary	157
5.2 Introduction	158
5.3 Materials and Methods	163
5.3.1 Synthesis and Purification of Oligonucleotides	163
5.3.2 NMR Sample Preparation	164
5.3.3 NMR Experiments and Resonance Assignments	164
5.4 Results	165
5.4.1 Nonexchangeable Proton Resonance Assignments	165

5.4.2	Resonance Assignment of Mismatch Protons	165
5.4.3	Comparison of NOE Patterns in the Mismatch Region	169
5.4.4	¹⁵ N NMR Resonance Assignments	170
5.4.5	¹ H NMR pH Titrations	170
5.4.6	¹⁵ N-filtered pH and Temperature Series	173
5.4.7	One-Dimensional NOE Measurements	173
5.5	Discussion	177
5.5.1	Structure of the AP•C Mispair	177
5.5.1.1	Imino vs. Amino Tautomers	177
5.5.1.2	Watson-Crick Geometry	178
5.5.1.3	Wobble Geometry	180
5.5.2	Sequence Dependence of Structure	182
5.5.3	Implications for Misinsertion and Repair of AP	183

CHAPTER 6

Summary and Conclusions	186
Bibliography	189

List of Figures

1-1	Stereo view of B-form DNA	3
1-2	The two standard Watson-Crick base pairs	4
1-3	Noncovalent DNA minor groove binding ligands related to distamycin A	9
1-4	Spectral densities plotted as a function of $\log(\omega)$ for molecules in the slow, intermediate, and fast tumbling limits	17
1-5	Energy level diagram of a dipolar coupled two spin system	20
1-6	Dependence of the maximum homonuclear nuclear Overhauser effect on $\omega\tau_c$..	24
1-7	A two-dimensional NOESY spectrum for a hypothetical three spin system	28
1-8	A one-dimensional ^1H NMR spectrum of a DNA decamer in H_2O	29
1-9	^1H NOE connectivity pathway used in assigning DNA proton resonances	31
1-10	The effects of chemical exchange on the appearance of the NMR spectrum	33
1-11	The AMBER target potential function	39
2-1	Distamycin A and netropsin	46
2-2	1D NMR titration of d(CGCAAAGGC):d(GCCTTTGCG) with distamycin	56
2-3	The aromatic proton to H1' region of a 2D NOESY spectrum of the 1:1 complex of distamycin with d(CGCAAAGGC):d(GCCTTTGCG)	58
2-4	Summary of intermolecular ligand-DNA contacts in the 1:1 complex of d(CGCAAAGGC):d(GCCTTTGCG) with distamycin	59
2-5	Molecular model of the 1:1 complex of distamycin with d(CGCAAAGGC): d(GCCTTTGCG)	60
2-6	Schematic representation of the flip-flop exchange undergone by distamycin in complex with d(CGCAAAGGC):d(GCCTTTGCG)	62
2-7	1D NMR titration of d(CGCAAAGGC):d(GCCTTTGCG) with netropsin ..	63
2-8	Expansions of a 2D NOESY spectrum of the complexes present at 1:1 stoichiometry of netropsin and d(CGCAAAGGC):d(GCCTTTGCG)	64
2-9	Diagram of A•T, G•C and I•C base pairs	69
2-10	1D NMR titration of [d(CGCGIICCGCG)] ₂ with distamycin	70

2-11	Plot of exchange lifetime vs. $1/[\text{NH}_3]$ for the I_5 and I_6 imino protons in the free duplex $[\text{d}(\text{CGCGIICCCGCG})]_2$	73
2-12	1D NMR titration of $\text{d}(\text{CGCIHCCGGC})$: $\text{d}(\text{GCCIHCCCGCG})$ with distamycin	76
2-13	Expansions of a 2D NOESY spectrum of the 2:1 complex of distamycin with $\text{d}(\text{CGCIHCCGGC})$: $\text{d}(\text{GCCIHCCCGCG})$	77
2-14	Summary of intermolecular ligand-DNA contacts in the 2:1 complex of distamycin with $\text{d}(\text{CGCIHCCGGC})$: $\text{d}(\text{GCCIHCCCGCG})$	79
2-15	Molecular model of the 2:1 complex of distamycin with $\text{d}(\text{CGCIHCCGGC})$: $\text{d}(\text{GCCIHCCCGCG})$	80
2-16	1D NMR titrations of several inosine-containing analogs of $\text{d}(\text{CGCAAATTGGC})$: $\text{d}(\text{GCCAATTTGCG})$ with distamycin	82
2-17	Intensities of resonances corresponding to free DNA, 1:1 complex and 2:1 complex during the 1D NMR titration of $\text{d}(\text{CGCAAITTGGC})$: $\text{d}(\text{GCCAACTTGCG})$ with distamycin	88
2-18	Schematic representation of 2:1 complexes of distamycin with various I,C-containing analogs of $\text{d}(\text{CGCAAATTGGC})$: $\text{d}(\text{GCCAATTTGCG})$	89
2-19	1D NMR titration of $[\text{d}(\text{CGCTTTAAAGCG})]_2$ with distamycin	91
2-20	Job's plot of distamycin binding to $[\text{d}(\text{CGCTTTAAAGCG})]_2$	92
2-21	1D NMR titration of $[\text{d}(\text{CGCATATATGCG})]_2$ with distamycin	95
2-22	Expansions of a 2D NOESY spectrum of the 2:1 complex of distamycin with $[\text{d}(\text{CGCATATATGCG})]_2$	96
2-23	Schematic representation of the sliding model of 2:1 complexes formed by distamycin with the sequence $[\text{d}(\text{CGCATAtATGCG})]_2$	98
2-24	Schematic representation of 2:1 complexes of distamycin with various A,T-rich sites	100
3-1	Representatives of the enediyne family of antitumor antibiotics	105
3-2	Radical species used to cleave DNA, diagram of netropsin-diazene	106
3-3	Synthetic conjugated ligands consisting of a DNA cleaving agent and a lexitropsin	108
3-4	1D NMR titration of $\text{d}(\text{CGCAAAAGGC})$: $\text{d}(\text{GCCTTTTGCG})$ with netropsin-diazene	113

3-5	Expansion of the aromatic proton to H1' region of a NOESY spectrum of the 1:1 complex of netropsin-diazene with d(CGCAAAAGGC): d(GCCTTTTGCG)	114
3-6	Expansion of the upfield diagonal regions of NOESY and TOCSY spectra of the 1:1 complex of netropsin-diazene with d(CGCAAAAGGC): d(GCCTTTTGCG)	118
3-7	The two enantiomers of netropsin-diazene	120
3-8	Schematic of selected intermolecular ligand-DNA NOEs in the 1:1 complex of netropsin-diazene with d(CGCAAAAGGC):d(GCCTTTTGCG)	121
3-9	Expansion of the aromatic proton to methylene proton region of a NOESY spectrum of the 1:1 complex of netropsin-diazene with d(CGCAAAAGGC): d(GCCTTTTGCG)	122
3-10	Stereo view of a molecular model of the 1:1 complex of netropsin-diazene with d(CGCAAAAGGC): d(GCCTTTTGCG)	124
4-1	Structure of the 2,3-bis-(hydroxymethyl)pyrrole-distamycin conjugate	132
4-2	Structures of mitomycin C and 2,3-bis-(hydroxymethyl)pyrrole	133
4-3	Several designed ligands that have been synthesized which contain an alkylating group tethered to a distamycin analog	134
4-4	Expansion of the amide and aromatic proton to H1' region of a NOESY spectrum of the XL-Dst:[d(CGCGAATTCGCG)] ₂ complex	142
4-5	Schematic of the ligand-DNA NOE contacts used in modeling the interstrand crosslinked XL-Dst:DNA complex	146
4-6	Stereo view of a molecular model of the XL-Dst:[d(CGCGAATTCGCG)] ₂ complex	147
4-7	View down the helical axis of standard B-form DNA compared with the same view of the two crosslinked base pairs in the XL-Dst:DNA complex	150
4-8	Schematic of the XL-Dst:CGAATT complex and the distamycin:AATT complex	152
5-1	Several mismatches between natural DNA bases which have been characterized by structural methods	159

5-2	Examples of mutagenic base analogs which have been shown to mispair in DNA oligomers	160
5-3	Four proposed structures for the 2-aminopurine•cytosine mispair	162
5-4	Expansion of the amino and aromatic proton to H1' region of a ^{15}N -decoupled ^1H - ^1H NOESY spectrum of d(GATCPCTAC):d(GTAGCGATC) where P = 2-aminopurine and C = 4- $^{15}\text{NH}_2$ -cytosine	166
5-5	^{15}N NMR spectrum of d(GATGPGTAC):d(GTACCCATC) where P = 2-aminopurine and C = 4- $^{15}\text{NH}_2$ -cytosine	171
5-6	^1H NMR spectra of both duplexes d(GATGPGTAC):d(GTACCCATC) and d(GATCPCTAC):d(GTAGCGATC) during a pH titration	172
5-7	1D ^{15}N -filtered ^1H NMR spectra of d(GATCPCTAC):d(GTAGCGATC), where P = 2- $^{15}\text{NH}_2$ -purine and C = 4- $^{15}\text{NH}_2$ -cytosine, acquired during a pH titration	174
5-8	1D ^{15}N -filtered ^1H NMR spectra of d(GATCPCTAC):d(GTAGCGATC), where P = 2- $^{15}\text{NH}_2$ -purine and C = 4- $^{15}\text{NH}_2$ -cytosine, acquired at different temperatures	175
5-9	1D NOE experiment of d(GATCPCTAC):d(GTAGCGATC) where P = 2- $^{15}\text{NH}_2$ -purine and C = 4- $^{15}\text{NH}_2$ -cytosine, with ^{15}N decoupling during acquisition	176

List of Tables

2-1	Nonexchangeable proton chemical shift assignments of the d(CGCAAAAGGC):d(GCCTTTTGCG) duplex, free and in the 1:1 distamycin:DNA complex	54
2-2	Ligand-DNA intermolecular contacts for two observed 1:1 netropsin complexes with d(CGCAAAAGGC):d(GCCTTTTGCG)	67
2-3	Nonexchangeable proton chemical shift assignments of the d(CGCIICCGGC):d(GCCIICCGCG) duplex, free and in the 2:1 distamycin:DNA complex	75
3-1	Ligand-DNA restraints used in modeling the netropsin-diazene complex with d(CGCAAAAGGC):d(GCCTTTTGCG)	111
3-2	Nonexchangeable proton chemical shift assignments of the d(CGCAAAAGGC):d(GCCTTTTGCG) duplex, free and in the 1:1 netropsin-diazene:DNA complex	115
3-3	Chemical shift assignments of netropsin-diazene complexed with d(CGCAAAAGGC):d(GCCTTTTGCG)	116
4-1	Nonexchangeable proton chemical shift assignments of the d(CGCGAATTCGCG) ₂ duplex, free and in the XL-Dst:DNA complex	140
4-2	More nonexchangeable proton chemical shift assignments of the d(CGCGAATTCGCG) ₂ duplex, free and in the XL-Dst:DNA complex	141
4-3	Chemical shift assignments of XL-Dst in the interstrand crosslinked complex with [d(CGCGAATTCGCG)] ₂	144
4-4	Ligand-DNA contacts for the XL-Dst:[d(CGCGAATTCGCG)] ₂ interstrand crosslinked complex	145
5-1	Nonexchangeable proton chemical shift assignments of d(GATGPGTAC): d(GTACCCATC) and d(GATCPCTAC):d(GTAGCGATC)	167

Acknowledgments

I would like to thank Professor David E. Wemmer for accepting me as his student and for his support, guidance, patience and excellent advice throughout my graduate career. He is without question an excellent scientist and mentor and I feel very fortunate to have worked with him.

I would like to acknowledge the contributions of H. Peter Spielmann, who contributed significantly to my research experience both personally and professionally. Professors Paul Hopkins, R. Daniel Little, Ramon Eritja and Myron Goodman were valued collaborators. I would also like to thank Professors Ignacio Tinoco, Jr. and Stuart Linn for providing valuable perspectives and advice as part of my dissertation committee.

I would like to acknowledge the members of the Wemmer lab, all of whom have been outstanding colleagues and valued friends. I have benefited enormously from interacting with the members of the DNA group past and present, including Tim Cole, Tammy Dwyer, Rachel Leheny, Rafael Pelaez, Stephanie Robertson, Christian Seel, Qing Zhang and in particular Bernhard Geierstanger and Jeff Pelton. I would to thank the other members of the Wemmer group for making my graduate experience more enjoyable and productive: Ho Cho, Valérie Copié, Mark Coy, Fred Damberger, Nat Gordon, Jonathan Heller, Dorothee Kern, Werner Kremer, Mark Kubinec, Andrew Lee, Corey Liu, Mike Nohaile, Marty Pagel, Ann Caviani Pease and Beth Snowden-Ifft. I also acknowledge the many contributions of members of the Hearst, Kim, Tinoco, and Sauer labs and the staff of the Structural Biology Division of LBNL with whom I have interacted throughout the years. (Please forgive any inadvertent omissions.)

I wish to extend special thanks to my parents, Tom and Denna Fagan, and to my husband, Frank Jones, for their love, support and encouragement.

Chapter 1

Introduction

1.1 Summary

DNA carries the genetic information for the cell. Transmission of this information to the rest of the cell is critical for its proper functioning. The process of information transfer requires the recognition of DNA sequence and structure by various parts of the cellular machinery. Double-stranded DNA must maintain a regular structure which can be recognized by general cellular apparatus such as polymerases and histones. Interestingly, double-stranded DNA must also somehow be recognized at very specific sequences by proteins such as restriction enzymes, transcriptional activators and repressors in order to properly regulate gene expression. At the same time, repair enzymes must be able to distinguish properly base paired, normal DNA from improperly base paired or chemically modified DNA. How can DNA be recognized simultaneously and efficiently in so many different ways?

The main objective of this work is to use the small molecule distamycin A and its analogs to gain insight into DNA recognition. Distamycin consists of a positively charged end group and three pyrrole rings linked by peptide bonds. Chapter 2 focuses on the binding behavior (stoichiometry, directionality and qualitative binding affinity) of distamycin A binding to the minor groove of DNA. The insights revealed can assist us in understanding the subtleties of sequence-dependent DNA structure, and may extend to other ligand-DNA and protein-DNA interactions. Additionally, these insights can be used in drug development. My studies of distamycin binding behavior, along with those of J. G. Pelton and B. H. Geierstanger, have led to the design of ligands which can deliver chemistry to a particular DNA sequence. The binding of these designed ligands is

described in Chapters 3 and 4. Such a system has great potential in the development of anticancer and antiviral compounds.

Chapter 5 describes the structural characterization of a DNA mispair containing a highly mutagenic base analog, 2-aminopurine. The pairing geometry of this mispair has been the subject of some controversy for many years. The use of selectively isotopically labeled DNA made it possible to unambiguously determine the geometry of this mispair, and to resolve previously conflicting conclusions by other researchers. This result will contribute to the overall understanding of polymerase fidelity and repair processes in mutagenesis, and has implications for researchers who use 2-aminopurine as a fluorescent probe in various biochemical studies.

1.2 DNA Structure

1.2.1 B-form DNA. The DNA conformation most commonly found in biological systems is termed "B-form". B-form DNA is a right-handed double helix with the two strands stabilized as a double helix by hydrogen bonding between bases on opposite strands and by base stacking interactions (Figure 1-1). The two standard Watson-Crick base pairs are shown in Figure 1-2.

B-form DNA is generally characterized by 2'-endo sugar pucker, 10.4 base pairs (bp) per helical turn, 3.4 Å rise per bp, with the glycosidic bonds in *anti* conformation (Saenger, 1983). B-form DNA has two grooves: a wide, shallow major groove and a deep, narrow minor groove (Figure 1-1). X-ray crystallography studies have demonstrated, however, that this general characterization does not fully describe DNA structure. Subtle, local structural features occur as a function of the exact DNA sequence (Dickerson & Drew, 1981).

DNA structure can be quantified by a set of local translational and rotational helical parameters including base pair tilt, slide and propeller twist (see Dickerson, 1989 for a complete description). Another helical parameter discussed extensively in Chapter 2 is

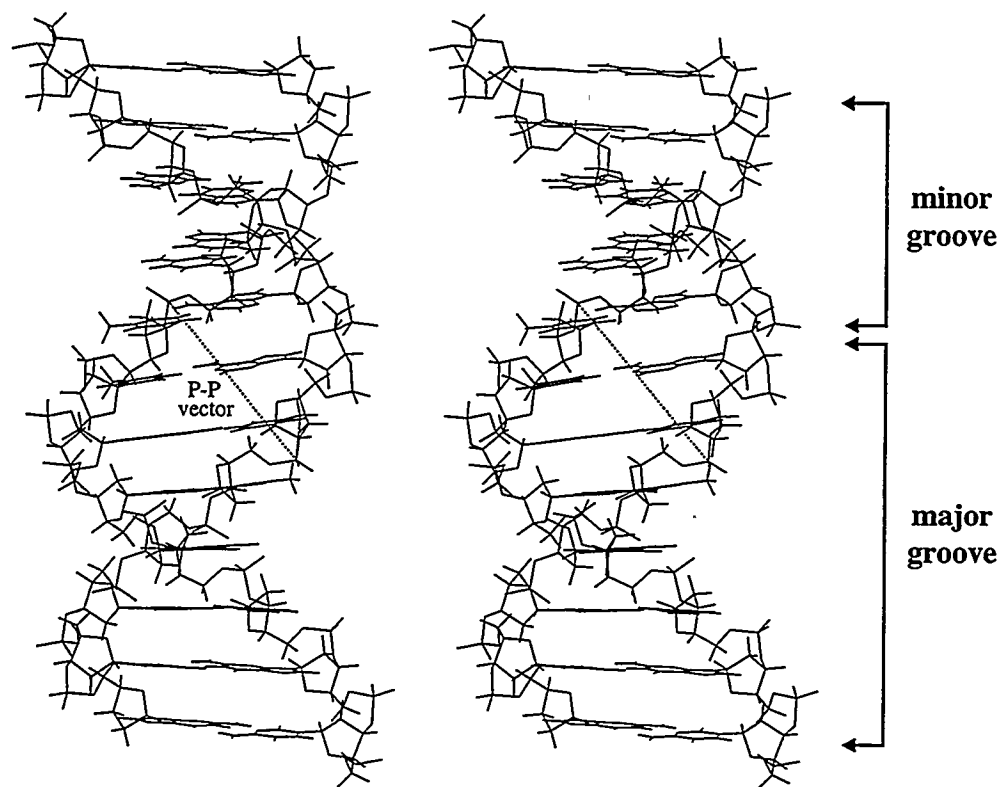


Figure 1-1. Stereo view of B-form DNA. The phosphate-phosphate vector defining the minor groove width is shown.

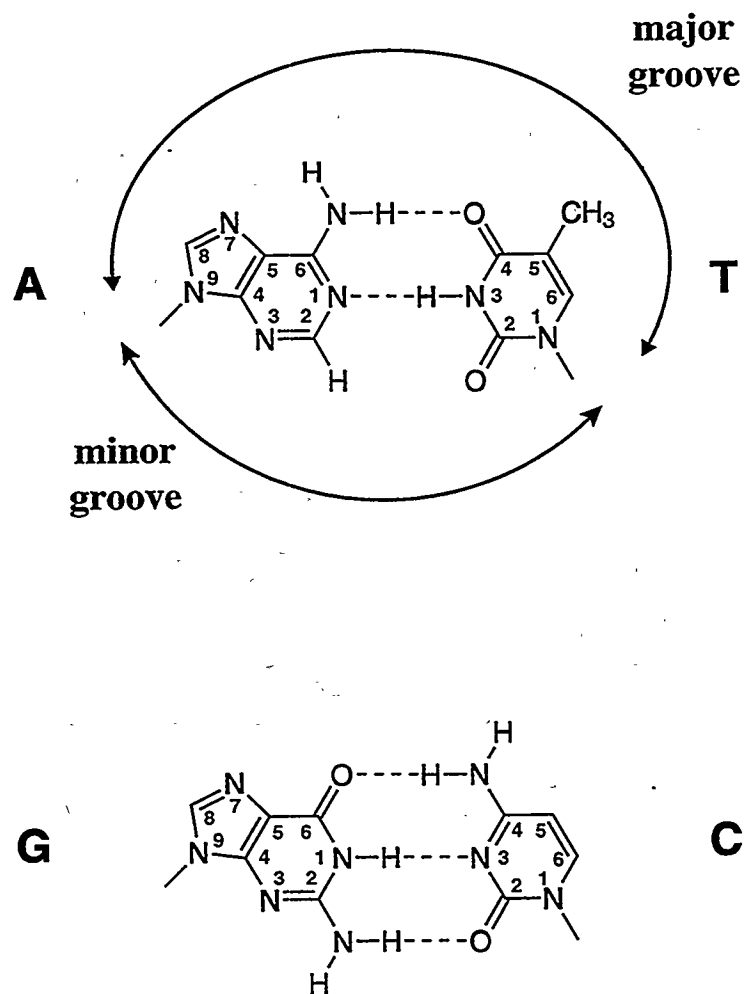


Figure 1-2. The two standard Watson-Crick base pairs, where A = adenine, T = thymine, G = guanine and C = cytosine. The numbering scheme for the atoms is shown.

minor groove width. The average minor groove width expected in B-form DNA from fiber diffraction is 5.7 Å (phosphate-phosphate distance ≈ 11.5 Å minus two phosphate group radii of 2.9 Å each) (Yoon et al., 1988). The minor groove width is determined by a phosphate-phosphate vector which connects a phosphorus atom on one strand and the phosphorus of the third nucleotide to the 3' side of its base pairing partner on the opposite strand (Figure 1-1).

1.2.2 Non-Standard DNA Structures. Although the standard B-form DNA duplex composed of Watson-Crick A•T and G•C base pairs is by far the most prevalent in biological systems, the rare occurrence of nonstandard DNA structures merits study. Nonstandard DNA structures may arise accidentally, for example through damage to the DNA by UV light or chemical mutagens, or through replication errors (Voet & Voet, 1990). "Normally occurring" nonstandard DNA conformations can also be found in many biological systems. For example, quadruplex DNA is found at the ends of chromosomes in telomeres (Smith et al., 1995). Triple-stranded DNA is proposed to provide a mechanism for recombination of homologous DNA strands (Kohwi & Panchenko, 1993). Modified bases also play important biological roles. For example, prokaryotic cells contain methyltransferases which methylate cytosines and adenines to protect the DNA from degradation, and in eukaryotic cells methylation of cytosines is believed to regulate transcription (Voet & Voet, 1990). The G•U wobble base pair is a classic case of a mispair found normally in the acceptor stem of some yeast tRNA molecules (Voet & Voet, 1990). Certain sequences of Watson-Crick based paired double helical DNA have been shown to have unusual structural properties, and will be discussed in the next section. The study of these unusual DNA structures is important in order to understand how cells distinguish between "normal" structures or base modifications and those which are unwanted and are in need of repair.

1.2.3 A-Tract DNA. Because most of the work described in the following chapters focuses on ligand binding to A,T-rich B-form DNA, its distinct properties are

described. The first indication that so-called "A-tracts" adopt a distinct structure was in studies by Englund and coworkers (Marini et al., 1982), in which it was observed that kinetoplast DNA migrated anomalously during polyacrylamide gel electrophoresis. Since that time, experiments including gel electrophoresis, DNA cyclization kinetics, DNase I cleavage, and hydroxyl radical cleavage have shown that regions of DNA take on a curved structure (for a review, see Hagerman, 1990b). The sequence requirement for the "A-tract" structure is a run of four to eight consecutive A•T base pairs, either consecutive adenines as in the sequence AAAA:TTTT, or containing one 5'-AT-3' step as in the sequence (AATT)₂, with the tract flanked by G•C pairs on either side. 5'-TA-3' steps have been found to disrupt the bend, resulting in normal electrophoretic migration (Hagerman, 1986).

Several DNA oligomers ranging from ten to twelve base pairs in length and containing central A,T-rich sequences, flanked by G,C-rich ends, have been characterized structurally by X-ray crystallography (for a review see Dickerson, 1990 and Dickerson, 1992). These structures have shown that A-tracts have a characteristically narrow minor groove (3-4 Å wide), with high propeller twist and negative base pair tilt. NMR studies also suggest that the minor groove is narrow in A-tract DNA (Kintanar et al., 1987; Katahira et al., 1988; Katahira et al., 1990). In several cases, it is believed that bifurcated hydrogen bonds between adenine N6 amino protons and thymine O3 carbonyl oxygens in the major groove may stabilize the narrow minor groove observed in A-tracts (Nelson et al., 1987; Yoon et al., 1988). This type of bifurcated hydrogen bond is only possible in sequences of consecutive adenines, as in 5'-AA-3'. Alternating sequences such as 5'-ATAT-3' cannot form such putative hydrogen bonds, consistent with their somewhat wider minor grooves as seen in crystal structures (Yuan et al., 1992). Based on hydroxyl radical footprinting (Burkhoff & Tullius, 1987) and NMR data (Nadeau & Crothers, 1989; Katahira et al., 1990), the minor groove is believed to be compressed gradually from 5' to 3' in the A-tract.

X-ray crystal structures of DNA oligomers have also shown the presence of a spine of hydration in the narrow minor grooves of A-tracts (e.g. Dickerson, 1990; Edwards et al., 1992), and NMR data has shown the presence of water molecules with relatively long residence times in the minor groove of the oligomer [d(CGCGAATTCGCG)]₂ (Kubinec & Wemmer, 1992; Liepinsh et al., 1992). A model has been proposed by Chuprina and coworkers which suggests that this spine of hydration actually stabilizes the narrow minor groove in A-tracts through hydrogen bonding and van der Waals contacts (Chuprina, 1985; Chuprina, 1987; Teplukhin et al., 1992).

In general, DNA shows sequence-specific variance in modification by chemical probes, suggesting that functional groups are not equally available for modification (Drew & Travers, 1984; McCarthy et al., 1990; Price & Tullius, 1992; Nightingale & Fox, 1993; Price & Tullius, 1993). Consistent with crystal data, nuclease and chemical digestion of several A,T-rich DNA dodecamers suggests that the sequences AATT and AAATTT have narrow minor grooves, while sequences containing 5'-TA-3' steps or alternating ATAT-type sequences have somewhat wider minor grooves, although still narrow relative to standard B-form DNA (Fox, 1992).

Additional evidence which distinguishes A-tract DNA from standard B-form DNA comes from NMR measurements of individual base pair opening rates in DNA oligomers containing A,T-rich regions (Guéron et al., 1990a; Guéron et al., 1990b). For oligomers containing "A-tracts" the opening lifetimes of the A-T base pairs are anomalously long. In sequences containing 5'-TA-3' steps, the A-T base pairs exhibit significantly shorter opening lifetimes, following closely the pattern observed for narrow minor groove width. Taken together, this database of information strongly suggests that A-tract DNA has unique structural characteristics.

1.3 Binding of Ligands to DNA

1.3.1 Minor Groove Binding Ligands. Small molecules which have been found to bind DNA noncovalently can be divided into two main classes: intercalators and minor groove binders. Intercalators insert between stacked bases in the double helix, and generally recognize very short sequences of DNA (two base pairs). Minor groove binders, however, bind longer sites on the DNA, and have the potential to be utilized as probes of DNA structure and models for protein-DNA interactions. The following discussion and the major part of this dissertation will focus on minor groove binders.

One of the best characterized minor groove binders is distamycin A (Figure 1-3). The preference of distamycin A for A,T-rich DNA binding sites has been recognized for many years (Dervan, 1986; Zimmer & Wähnert, 1986). NMR was the first structural technique to show unambiguously that the closely related ligand netropsin bound in the minor groove of B-form DNA (Patel, 1979; Patel, 1982). Both NMR and crystallographic studies of complexes of the minor groove binders distamycin A (Klevit et al., 1986; Coll et al., 1987; Pelton & Wemmer, 1989), netropsin (Kopka et al., 1985b; Kopka et al., 1985a; Patel & Shapiro, 1986; Coll et al., 1989; Tabernero et al., 1993), Hoescht 33258 (Pjura et al., 1987; Teng et al., 1988; Parkinson et al., 1990; Fede et al., 1993), berenil (Brown et al., 1992; Hu et al., 1992) and SN-6999 (Chen et al., 1992; Gao et al., 1993) have been carried out (Figure 1-3). In all of these studies to date, the ligands are bound deep in the minor groove of an A,T-rich region of DNA. Where the ligand is bound, the structure of the minor groove closely matches the shape and width of the ligand molecule. These molecules bind with submicromolar affinity in the minor groove at sites containing at least four successive A•T or T•A base pairs (Dervan, 1986; Klevit et al., 1986; Zimmer & Wähnert, 1986; Breslauer et al., 1987; Coll et al., 1989; Wemmer et al., 1990). A more detailed description of distamycin binding to different sites can be found in Chapter 2 and in (Wemmer et al., 1994).

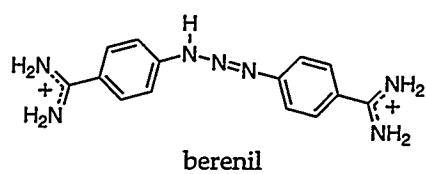
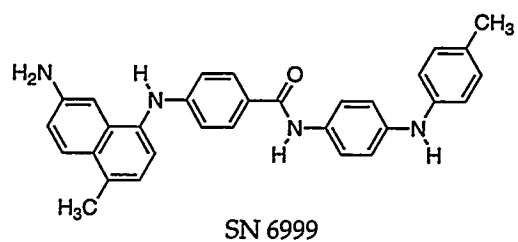
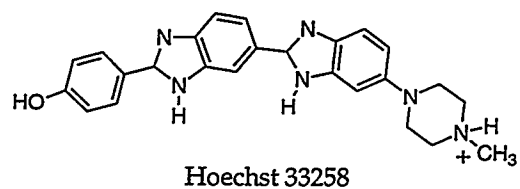
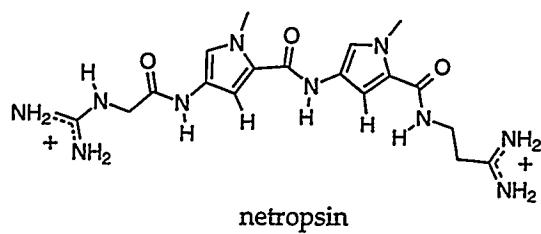
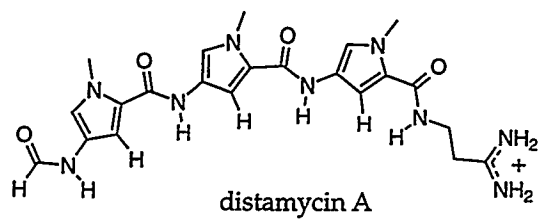


Figure 1-3. Noncovalent DNA minor groove binding ligands related to distamycin A.

The affinity of these minor groove binding ligands for DNA is derived from a combination of several interactions. Hydrogen bonding occurs between ligand amide protons and adenine N3 or thymine O2 along the floor of the minor groove. Favorable electrostatic interactions occur between the positively charged ligand and the negatively charged DNA, with the region of highest negative electrostatic potential in DNA located in the minor groove of A,T-rich regions (Pullman & Pullman, 1981). Additionally, van der Waals interactions occur between the flat aromatic pyrrole rings and the walls of the minor groove (Zakrzewska et al., 1983; Marky & Breslauer, 1987; Dabrowiak et al., 1990). The relative contributions of these various effects to binding affinity and specificity are not fully understood. A related molecule, SN 18071, cannot form hydrogen bonds along the floor of the minor groove, but still exhibits selectivity for A,T-rich sites (Pullman, 1989). A positive charge on the ligand is not required for binding as evidenced by studies on uncharged netropsin analogs which demonstrate lowered A,T specificity (Thurston & Thompson, 1990). A recent study comparing the binding energies of several benzimidazole derivatives suggested that van der Waals interactions make the largest contribution to the binding energy of minor groove binding ligands (Czarny et al., 1995). It appears that part of the specificity for A,T-rich regions arises from the presence of the guanine 2-amino group in the minor groove of G-C base pairs, which sterically blocks the ligand from binding deep in the groove.

In complexes of minor groove binding ligands with DNA, the relative contributions of enthalpic and entropic effects to the overall binding energy differ according to the DNA sequence. Enthalpic effects dominate in poly(dA):poly(dT) and entropic effects dominate in poly[d(AT)]:poly[d(AT)] (Marky & Breslauer, 1987; Marky & Kupke, 1989), resulting in nearly the same overall binding energy in either case. The positive entropic effect associated with binding of netropsin to poly(dA):poly(dT) is attributed to the release of ordered water molecules which solvate the unbound ligand and DNA. However, solvation effects on binding affinity are not well understood.

In the Wemmer lab, distamycin has been selected as a starting point for developing analogs which specifically bind target DNA sequences. Distamycin was chosen for ligand design due to its unique capacity to bind DNA in an antiparallel, side-by-side 2:1 ligand:DNA mode, which was first characterized with the DNA oligomer d(CGCAAATTGGC): d(GCCAATTTGCG) by NMR methods (Pelton & Wemmer, 1989; Pelton & Wemmer, 1990b). There has been significant success in the development of distamycin analogs in which one or more of the pyrrole rings are substituted with an imidazole ring (lexitropsins) (Lown et al., 1986; Wade & Dervan, 1987; Wade et al., 1993; Lown, 1994). Lexitropsins utilize the side-by-side 2:1 binding mode to target the minor groove of mixed G,C- and A,T-containing sites, with the imidazole nitrogen targeting the guanine 2-amino group (Mrksich et al., 1992; Geierstanger et al., 1994a; Geierstanger et al., 1994b). The design of sequence-specific minor groove binding ligands requires consideration of sequence-dependent groove width as a major determinant of binding motif, as well as specific hydrogen bonds to the floor of the minor groove by pyrrolicarboxamides and imidazolecarboxamides. Several lexitropsin:DNA complexes have been characterized structurally, and have been found to bind in the predicted fashion to their target DNA sites (e.g. Dwyer et al., 1992; Mrksich et al., 1992; Geierstanger et al., 1994a). Advances made in the design of lexitropsins were based in part on the foundation laid by studies on the 2:1 distamycin binding mode by J. G. Pelton and B. H. Geierstanger, and by the studies presented here in Chapter 2. Further development of this new class of minor groove binding ligands is in progress since there are several features which need to be examined, including the length of the targeted site, different types of linkers, and the ability of ligands to distinguish A•T from T•A base pairs in the minor groove.

1.3.2 Minor Groove Binding Proteins. Although the vast majority of protein-DNA interactions characterized thus far have occurred in the major groove of DNA, more and more examples of proteins which have important contacts in the minor groove are

appearing in the scientific literature. Some of the best characterized examples are the TATA-binding protein (TBP) (Kim et al., 1993a; Kim et al., 1993b), LEF-1 (for lymphoid enhancer-binding factor) (Love et al., 1995), SRY (for sex-determining region Y) (Werner et al., 1995), and purine repressor (PurR) (Schumacher et al., 1994). The wide range of sequence specificity, or lack of it, exhibited by these and other minor groove-binding proteins can be better understood in the context of our knowledge of small molecule minor groove binders. In fact, it has been suggested that the binding of SRY is reminiscent of lexitropsin binding (Travers, 1995), and it has been shown by NMR that the binding of a short peptide taken from HMG-I (a member of the high mobility group class of proteins) to the minor groove of DNA is strikingly similar to that of netropsin (Geierstanger et al., 1994c).

1.4 Nuclear Magnetic Resonance Spectroscopy (NMR)

Solution NMR spectroscopy is useful for studying biomolecular complexes for several reasons. NMR can be used to determine the structure of a molecule in both the free and the bound states. In obtaining the structure of the complex, the ligand binding site and mode are identified, and the interactions between the ligand and DNA can be analyzed. Comparative analysis can reveal structural changes induced by complex formation. In these aspects, NMR is quite similar to X-ray crystallography, with the exception that crystallographers study molecules in crystalline form.

NMR and X-ray crystallography do not produce entirely duplicate information, however, since each technique has its own strengths and limitations. NMR is size-limited, with X-ray crystallography better suited for large systems. Currently, complete NMR structure determination is possible for molecules up to about 40 kDa. An advantage of NMR is that biomolecules can be studied in solution which may better approximate physiological conditions (although the high concentrations of NMR solutions may not be the ideal model conditions). Crystal packing forces sometimes produce structural artifacts

in crystallographic structures, particularly in flexible molecules. In cases where both solution and X-ray structures of a biomolecule are available, comparison between the two enables us to discern errors in methodology and inherent biases in the techniques.

Some systems are best studied in crystal form due to solubility problems, while others are not readily crystallized. In fact, crystallographic studies have failed to produce 2:1 distamycin:DNA complexes (with one notable exception, see Chen et al., 1994), and evidence suggests that 1:1 complexes are preferentially crystallized over 2:1 complexes (Pelton, 1990). So, we rely primarily on NMR data to give us information about 2:1 ligand:DNA complexes with distamycin or lexitropsins.

NMR is particularly useful for characterizing slowly exchanging equilibrium mixtures of species, since integration of NMR resonances yields the relative populations of all species in solution. Additionally, NMR has a distinct advantage in that kinetic information on several time scales can be obtained. For the case of ligand:DNA complexes, it is possible to estimate with reasonable accuracy the relative (not absolute) equilibrium constants of a mixture of 1:1 and 2:1 ligand:DNA complexes, and to study the exchange processes associated with ligand binding.

1.4.1 General Principles. The following discussion is taken largely from texts by Harris (1986), Neuhaus & Williamson (1989) and Chandrakumar & Subramanian (1987). Nuclear magnetic resonance spectroscopy is made possible because nuclei, like electrons, have an intrinsic spin angular momentum (characterized by the symbol I). For a given value of I , the nuclear spin quantum number m_I runs in integral steps from $-I$ to $+I$. The focus of this introduction will be on the nuclei of several biologically relevant atoms including ^1H , ^{13}C , and ^{15}N , which have $I = 1/2$. For such nuclei, m_I may take the values $+1/2$ and $-1/2$.

The magnetic dipole moment μ of a nucleus is defined to be:

$$\mu = (\gamma h/2\pi) \sqrt{I(I+1)} \quad (\text{Equation 1-1})$$

where γ = the gyromagnetic ratio, a constant of proportionality between the spin and the dipole moment, and h = Plank's constant. In the absence of an external magnetic field, the energy of the nucleus is independent of m_I . If the nucleus is put into an external magnetic field $\mathbf{B}_0$ (with its direction arbitrarily assigned to the +z-axis), the energy of this interaction is given by:

$$E = -\boldsymbol{\mu} \cdot \mathbf{B}_0 = -\mu_z B_{0z} = -(\gamma h/2\pi) m_I B_0 \quad (\text{Equation 1-2})$$

The two states characterized by $m_I = \pm 1/2$ are now nondegenerate, since the resulting energies from Equation 1-2 are of opposite sign. The lower energy state corresponds to $m_I = +1/2$, with the z component of the dipole moment aligned parallel to the $\mathbf{B}_0$ vector, and is referred to as the α state. The higher energy state corresponds to $m_I = -1/2$, with the z component of the dipole moment antiparallel to the $\mathbf{B}_0$ vector, and is referred to as the β state. Therefore, for spin $1/2$ nuclei, the energy difference between the two states is given by:

$$\Delta E = (\gamma h/2\pi) B_0 \quad (\text{Equation 1-3})$$

The frequency of this transition is given by:

$$\nu = \Delta E/h = (\gamma/2\pi) B_0 \quad (\text{Equation 1-4})$$

Transitions between the α and β states can be stimulated by irradiation at this frequency, termed the Larmor frequency. For typical magnetic field strengths and γ values, the Larmor frequency falls into the radiofrequency (rf) range.

The effective $\mathbf{B}_0$ field felt by any nucleus is a sum of the externally applied $\mathbf{B}_0$ field and the much smaller contributions from the local chemical environment. The induced motion of electrons by the external magnetic field slightly reduces the overall $\mathbf{B}_0$ field felt by a nearby nucleus. The relative density of electrons near a nucleus determines the extent of this effect. Therefore, the various nuclei in a molecule will resonate at slightly different Larmor frequencies, resulting in dispersion of the spectrum. Because the spectrum is dispersed, NMR becomes useful for chemists and structural biologists who wish to uniquely assign resonances to particular nuclei in a molecule.

The spins are in contact with the surroundings (the "lattice") which functions as a thermal reservoir, so that at thermal equilibrium the two energy states α and β will be populated according to the Boltzmann distribution:

$$\frac{N_{\beta}}{N_{\alpha}} = e^{\frac{-\Delta E}{kT}} \quad (\text{Equation 1-5})$$

where N_{β} = the population of the higher energy state, N_{α} = population of the lower energy state, ΔE = the energy difference between the two states, k = Boltzmann's constant and T = temperature of the system. In NMR experiments at typical temperatures (300 K), the population difference between the α and β states is very small (parts per million). If we use a sample of sufficient concentration (typically millimolar), when the sample is placed into a magnetic field B_0 , even this small population difference generates a macroscopic net magnetization M_z .

Note that the component of the dipole moment in the direction of B_0 , m_{Iz} , is smaller than the value of μ . The magnetic dipole moment can be represented classically by a vector tilted away from the B_0 vector at an angle determined by m_I . The B_0 field exerts a torque $B \times M$ on the dipole moment μ causing it to rotate, or precess, about B_0 with a rate $\nu_0 = (\gamma/2\pi)B_0$. This precession frequency is equal to the Larmor frequency.

1.4.2 The Basic NMR Experiment. A sample in a B_0 field is perturbed from equilibrium by the application of a weak oscillating magnetic field, B_1 , which is perpendicular to B_0 . In modern NMR experiments, this B_1 field is a short rf pulse which contains a range of frequencies including the transition frequencies of the nucleus of interest. For simplicity, we will consider an isolated spin system and its respective Larmor frequency. If we change our perspective by "jumping" into a coordinate frame which is rotating at the Larmor frequency, the effects of the B_0 field (the precession of the sample magnetization M_0) and the oscillation of the B_1 field are subtracted out. We can then clearly see that the net magnetization M_{0z} feels a new torque due to B_1 which tips M_{0z} into the xy-plane, generating a transverse component of magnetization (M_{xy}). The B_1 field is

then turned off and the system is allowed to freely evolve. A current is induced in the receiver coil of the spectrometer, which is recorded as a free induction decay. The intensity of the induced signal is proportional to the difference in populations of the α and β energy levels.

1.4.3 Spectral Density. The spectral density describes all of the fluctuating magnetic fields generated by random thermal motions in a sample. These thermal motions form a continuum of frequencies. The spectral density $J(\omega)$ takes the form of a Lorentzian function:

$$J(\omega) = 2\tau_c / (1 + \omega^2\tau_c^2) \quad (\text{Equation 1-6})$$

where ω = angular frequency and τ_c = correlation time (a measure of the overall tumbling rate of the molecule). This function tells us the power of motions in the lattice available to induce transitions in the system. The probability W of a particular transition is a function of $J(\omega)$. Spectral densities are plotted in Figure 1-4 for the three general cases of τ_c . Biological macromolecules, with correlation times on the order of nanoseconds, fall into the slow tumbling or spin diffusion limit.

1.4.4 The Density Matrix. In NMR studies we consider our system to be an ensemble of spins coupled to a thermal bath. We can represent the available information on the state of the system by the density matrix $\rho(t)$. $\rho(t)$ contains all of the available statistical information about an ensemble of spins, and takes the form of a matrix of ensemble averaged coefficients. If we want to know the expectation value of a measurable property Q , using the operator $\hat{Q}$ corresponding to that property (Chandrakumar & Subramanian, 1987):

$$\langle Q \rangle = \sum_j \sum_i c_j^* c_i \langle \Psi_j | \hat{Q} | \Psi_i \rangle \quad (\text{Equation 1-7})$$

For an ensemble of spins,

$$\rho(t) = \overline{c_j^* c_i} \quad (\text{Equation 1-8})$$

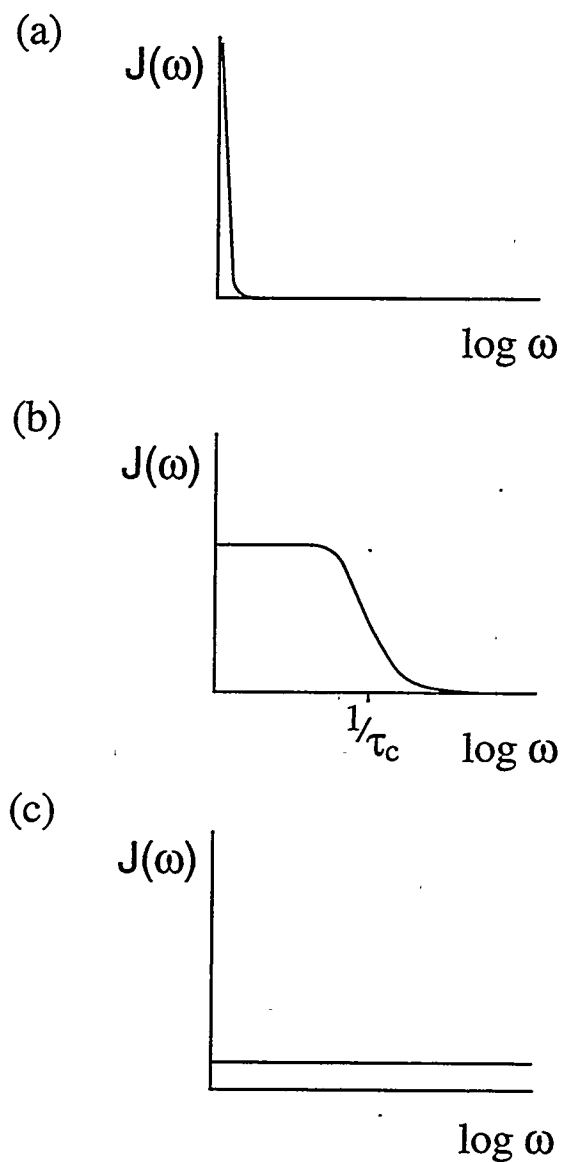


Figure 1-4. Spectral densities plotted as a function of $\log(\omega)$ for molecules with correlation times in the (a) slow tumbling (spin diffusion), (b) intermediate, and (c) fast tumbling (extreme narrowing) limits. All spectral density functions have the same integrated area (these plots are not scaled relative to each other).

and the ensemble averaged expectation value of Q is:

$$\overline{\langle Q(t) \rangle} = \text{Tr}(\rho(t)Q) \quad (\text{Equation 1-9})$$

We wish to solve for $\rho(t)$. The solution to the equation of motion for the density matrix is given by:

$$\rho(t) = e^{(-2\pi i \mathcal{H}t/\hbar)} \rho(0) e^{(+2\pi i \mathcal{H}t/\hbar)} \quad (\text{Equation 1-10})$$

where $\mathcal{H}$ is the Hamiltonian and $\rho(0)$ is the density matrix of the system at some initial time point. In an NMR experiment we can manipulate the effective Hamiltonian operating on the system at different times during the experiment. Different Hamiltonians can arise from the application of rf pulses and from various types of coupling and relaxation processes.

1.4.5 T_1 and T_2 Relaxation. An ensemble of spins which have been excited in some way returns to thermal equilibrium through relaxation. There are several mechanisms by which relaxation can occur. Before a sample is placed into an external magnetic field, the α and β energy levels are degenerate, and the sample has no net magnetization. At the moment the sample is placed into a B_0 field, a net magnetization begins to develop as the spins seek to align themselves parallel or antiparallel with the direction of the field. Equilibrium is reached with a slight excess population of spins in the parallel direction, as described by the Boltzmann distribution. This relaxation process occurs with first order kinetics, where T_1 is the first order time constant. T_1 relaxation is termed longitudinal or spin-lattice relaxation because it occurs by random thermal motions in the lattice which cause transfer of energy from the system to the lattice. T_1 depends upon the correlation time τ_c , since τ_c determines the spectral density, i.e. the power of motions in the lattice available to induce transitions at a given frequency.

In the general description of the NMR experiment given in the previous section, magnetization was generated in the transverse or xy-plane. During free evolution of transverse magnetization, transverse or spin-spin relaxation can occur. Consider the macroscopic transverse magnetization (M_{xy}) derived from an ensemble of isolated spins. Dephasing of M_{xy} occurs with a first order time constant T_2 , so that the intensity of the

NMR signal decays over time and eventually all of the components in the xy-plane cancel each other. In other words, there is a range of Larmor frequencies, an uncertainty in ΔE . After Fourier transformation of the free induction decay, T_2 manifests itself in the line width of the resonance in the frequency domain spectrum.

1.4.6 Dipolar Relaxation. Dipole-dipole coupling arises from a fluctuating local magnetic field generated at the site of one dipole (nucleus) due to the presence of another dipole (nucleus) nearby. For a homonuclear spin pair the dipolar coupling Hamiltonian takes the form:

$$\mathcal{H}_{dd} = \sum_{i < j} \frac{\hbar^2 \gamma^2}{8\pi^2 r_{ij}^3} (1 - 3\cos^2\theta) (3I_{iz}I_{jz} - \mathbf{I}_i \cdot \mathbf{I}_j) \quad (\text{Equation 1-11})$$

where the labels i and j denote two spins which have $I = 1/2$, γ = gyromagnetic ratio, r_{ij} = internuclear distance and θ = the angle between the internuclear vector $\mathbf{r}$ and the $\mathbf{B}_0$ field. Direct effects from dipolar coupling are not observed in the NMR spectrum because isotropic tumbling in liquids averages the $(1 - 3\cos^2\theta)$ term to zero. However, dipole-dipole coupling provides a relaxation mechanism which proves to be very useful in structural NMR studies.

Dipolar relaxation occurs via dipolar coupling of two spins. Note that in Equation 1-11, dipolar coupling strength is inversely proportional to r_{ij}^3 . Dipolar coupling is strictly a through-space interaction, and does not require that the nuclei be connected at all through chemical bonds. This feature of dipolar coupling is important for structural biologists who use the rate of dipolar relaxation to measure interproton distances in biological molecules. Dipolar relaxation will be discussed in more detail in the context of the nuclear Overhauser effect (NOE).

1.4.7 Nuclear Overhauser Effect.

1.4.7.1 General Principles. An energy diagram of two identical spins (I and S) both with $I = 1/2$ which are dipolar coupled but not scalar coupled is shown in

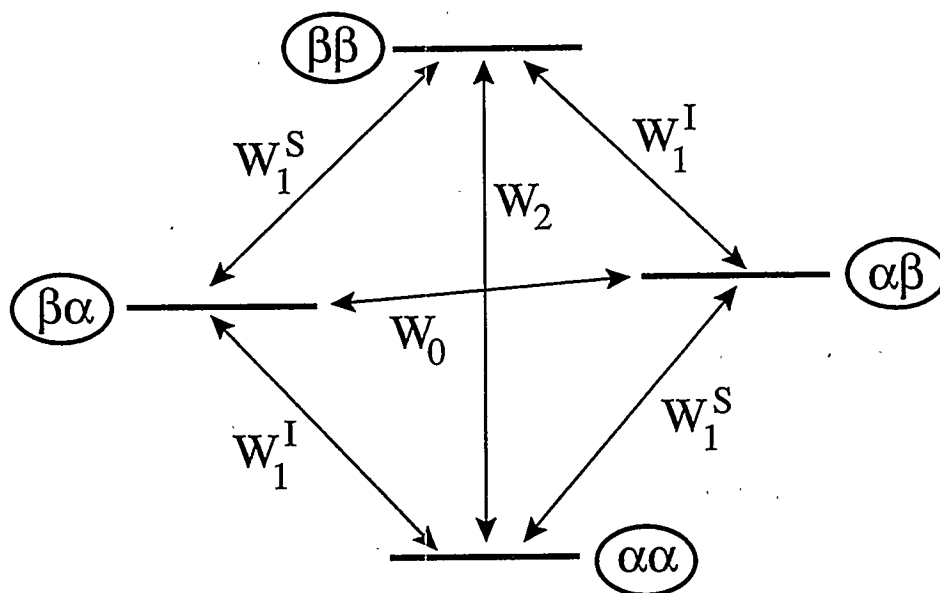


Figure 1-5. Energy level diagram of a two spin system where the two spins I and S are dipolar coupled, but not scalar coupled. The available states for this system are described by the four basic product functions $|\alpha\alpha\rangle$, $|\alpha\beta\rangle$, $|\beta\alpha\rangle$ and $|\beta\beta\rangle$ where the spin states are listed in the order $|IS\rangle$. The transition probabilities W_0 , W_1^I , W_1^S , and W_2 are shown. W_0 (flip-flop) and W_2 (double flip) correspond to cross-relaxation pathways which give rise to the NOE.

Figure 1-5. The full expression of the Hamiltonian $\mathcal{H}_{dd}$ in terms of raising and lowering operators is (Harris, 1986):

$$\begin{aligned}\mathcal{H}_{dd} &= \frac{\hbar^2 \gamma_I \gamma_S}{r_{ij}^3} [A+B+C+D+E+F] \\ A &= (1 - 3\cos^2\theta) I_z S_z \\ B &= 1/4 (1 - 3\cos^2\theta) (I_+ S_- + I_- S_+) \\ C &= -3/2 \sin\theta \cos\theta e^{-i\phi} (I_z S_+ + I_+ S_z) \\ D &= -3/2 \sin\theta \cos\theta e^{i\phi} (I_z S_- + I_- S_z) \\ E &= -3/4 \sin^2\theta e^{-i2\phi} I_+ S_+ \\ F &= 3/4 \sin^2\theta e^{i2\phi} I_- S_-\end{aligned} \quad (\text{Equation 1-12})$$

Zero quantum terms ($I_+ S_-$ and $I_- S_+$) correspond to a frequency $|(\omega_I - \omega_S)|$ and involve only a very small net gain or loss of energy, and are commonly referred to as flip-flop transitions. Double quantum terms ($I_+ S_+$ and $I_- S_-$) involve two spin flips in the same direction, corresponding to a frequency $(\omega_I + \omega_S)$ and resulting in a net gain or loss of energy. Double quantum transitions are commonly called flip-flip transitions. In either of these two cases, one spin flips in concert with another, a mechanism termed cross-relaxation. Single quantum transitions ($I_+ S_z$, $I_- S_z$, $I_z S_+$, $I_z S_-$) involve one spin flip and correspond to a frequency ω_I or ω_S . From the expression above, it is clear that zero quantum, single quantum and double quantum transitions are allowed under the dipolar coupling Hamiltonian.

The NOE arises when the α and β populations of one spin I are altered due to dipole-dipole relaxation by another spin S which is close in space. For the NOE to occur, it requires nonequilibrium α and β populations for spin S. This situation is created by applying a long, low power rf pulse to saturate the S resonance, thereby equalizing the α and β populations for spin S. Referring to Figure 1-5, in this case the populations of the $|\alpha\alpha\rangle$ and $|\alpha\beta\rangle$ states are equalized, as are the $|\beta\alpha\rangle$ and $|\beta\beta\rangle$ populations. Because the system is no longer at thermal equilibrium, it undergoes dipolar relaxation to try to restore

equilibrium populations. Relaxation occurs by the allowed transitions, with the probability of a particular transition specified by W . Note that when spin S has been saturated, the cross relaxation transitions W_0 (zero quantum) and W_2 (double quantum) alter the α and β populations of *both* spins I and S, resulting in altered intensity for the I signal.

1.4.7.2 The Solomon Equation. The transition probabilities W_0 , W^I_1 , W^S_1 and W_2 for the coupled two spin system are noted in Figure 1-5. The general definition for the transition probability between two states i and j is (Neuhaus & Williamson, 1989):

$$W_{ij} = J(\omega) |(\langle i | \mathcal{H}_{rel} | j \rangle)^2|_{average} \quad (\text{Equation 1-13})$$

where $J(\omega)$ is the spectral density at the appropriate frequency ω for the transition and $\mathcal{H}_{rel}$ is the Hamiltonian of the relaxation mechanism. The transition probability functions as a first order rate constant in the rate equations describing the change of the population of each state with time. For example, an equation can be constructed which describes how the population of the $|\alpha\alpha\rangle$ state, $N_{\alpha\alpha}$, changes with time:

$$\frac{dN_{\alpha\alpha}}{dt} = -(W^I_1 + W^S_1 + W_2)N_{\alpha\alpha} + W^I_1N_{\beta\alpha} + W^S_1N_{\alpha\beta} + W_2N_{\beta\beta} + \text{constant} \quad (\text{Equation 1-14})$$

where the constant represents population changes due to other forms of relaxation. The fractional change in the intensity of spin I when spin S is saturated is defined to be:

$$f_I\{S\} = \frac{(I_Z - I_Z^0)}{I_Z^0} \quad (\text{Equation 1-15})$$

where I_Z = intensity of the spin I resonance and I_Z^0 = equilibrium intensity of the spin I resonance. Using rate equations such as 1-14 for all four spin states, an equation describing the rate of change of I_Z with time can be derived (Neuhaus & Williamson, 1989):

$$\frac{dI_Z}{dt} = -(I_Z - I_Z^0) (W_0 + 2W^I_1 + W_2) - (S_Z - S_Z^0) (W_2 - W_0) \quad (\text{Equation 1-16})$$

This is the Solomon equation. Substituting into Equation 1-15 for the steady-state NOE experiment, where $dI_z/dt = 0$ and $S_z = 0$, the fractional change in the intensity of spin I is given by:

$$f_I\{S\} = \frac{\gamma_S}{\gamma_I} \frac{W_2 - W_0}{W_0 + 2W_1^I + W_2} = \frac{\sigma_{IS}}{\rho_{IS}} \quad (\text{Equation 1-17})$$

where σ_{IS} is termed the cross-relaxation rate constant, and ρ_{IS} is termed the dipolar longitudinal rate constant. Recall from the definition of the transition probability W that it is proportional to the spectral density at the frequency which matches the appropriate transition energy. For a rigid, isotropically tumbling molecule, the transition probabilities are (Harris, 1986):

$$\begin{aligned} W_0 &= (1/10) R^2 J(\omega_I - \omega_S) \\ W_1^I &= (3/20) R^2 J(\omega_I) \\ W_2 &= (3/5) R^2 J(\omega_I + \omega_S) \end{aligned} \quad (\text{Equation 1-18})$$

where R is the dipolar interaction constant and is proportional to r_{IS}^{-3} , making σ_{IS} and ρ_{IS} proportional to r_{IS}^{-6} . These transition probability expressions can be substituted into Equation 1-17 to obtain the maximum theoretical NOE $f_I\{S\} = \eta_{\max}$, assuming exclusively dipolar relaxation. $\eta_{\max}$ is plotted as a function of $\omega\tau_c$ in Figure 1-6. For large biomolecules in the slow tumbling limit, $J(\omega_I - \omega_S) \gg J(\omega_I + \omega_S)$ (Figure 1-4a), making W_0 a more efficient relaxation pathway than W_2 . Therefore, according to Equation 1-17, for slow tumbling molecules the NOE will be negative. In other words, upon saturation of spin S, the intensity of a nearby dipolar coupled spin I resonance will decrease.

1.4.7.3 Distance-Dependence of the NOE. Structural biologists are interested in using the distance-dependence of the NOE to obtain pairwise distances between protons in a biomolecule. Although the above discussion is useful in understanding the NOE, note that for the steady-state NOE in the absence of external relaxation there is no distance-dependence (the r_{IS}^{-6} terms in σ_{IS} and ρ_{IS} cancel) (Equation 1-17). So, instead of

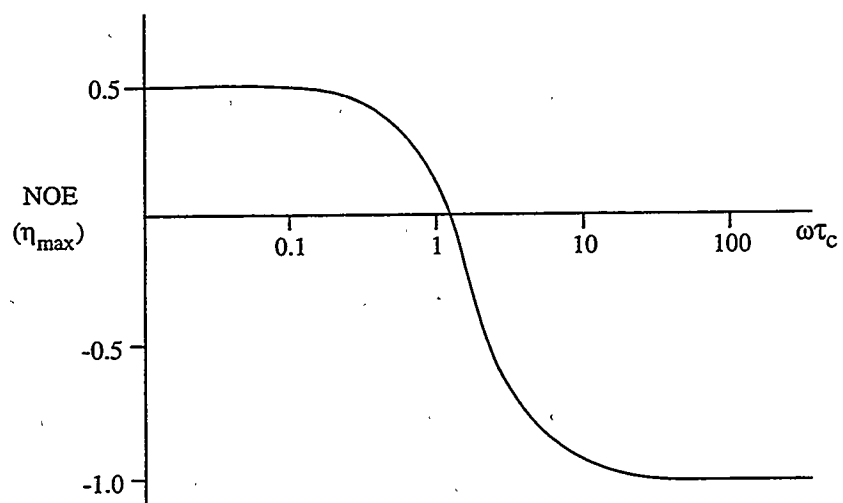


Figure 1-6. Dependence of the maximum homonuclear nuclear Overhauser effect on $\omega\tau_c$.

measuring the absolute steady-state NOE, a kinetic NOE experiment will be used from which the NOE build up rate can be determined.

In the transient NOE experiment, instead of keeping the α and β populations of spin S equalized with a long saturating pulse, the populations are transiently inverted with a very short high power 180° pulse, creating the initial conditions $I_z = I_z^0$ and $S_z = -S_z^0$. The pulse is turned off, and the system is then allowed to freely evolve under the dipolar Hamiltonian (this experimental time delay is called the "mixing" time). Assuming we have a rigid molecule whose tumbling can be described by a single correlation time, the initial homonuclear NOE build up rate is:

$$\left. \frac{dI_z}{dt} \right|_{t=0} = 2\sigma_{IS}S_z^0 \quad (\text{Equation 1-19})$$

where $\sigma_{IS} \propto r_{IS}^{-6}$ from Equations 1-17 and 1-18. Clearly, the NOE build up rate is distance-dependent. This equation holds true for short mixing times. During longer mixing times, the assumption of an isolated spin pair no longer holds. Secondary effects become significant as magnetization is transferred through multiple spin pathways. This effect is called spin diffusion, and is of particular concern for molecules in the slow tumbling limit (also called the spin diffusion limit) because the rate of NOE build up is relatively fast.

Initial NOE build up rates can be converted into internuclear distances with the use of a "ruler" distance, a pair of spins A and B in the molecule which are separated by a fixed distance. Knowledge of this "ruler" distance and measurement of the initial NOE build up rate for the AB spin pair and an IS spin pair in the molecule allows the distance between spins I and S to be calculated:

$$r_{IS}^6 = \frac{f_A\{B\}}{f_I\{S\}} r_{AB}^6 \quad (\text{Equation 1-20})$$

The practical upper limit for measurement of interproton distances by the NOE is 5 Å. For spin pairs farther apart, other relaxation mechanisms dominate. In real systems, errors can arise because large biomolecules can be flexible, with different parts of the

molecule characterized by different correlation times. Additionally other efficient relaxation mechanisms, such as the presence of paramagnetic ions in the sample solution, can affect NOE measurements. Limited signal-to-noise ratios will result in imprecise measurements and loss of long distance information (4 to 5 Å distances), which are at the limit of the sensitivity of the experiment.

1.4.7.4 Two-Dimensional NOE Spectroscopy (2D NOESY). The 2D NOESY (Jeener et al., 1979) is an extension of the one-dimensional transient NOE experiment. The homonuclear NOESY pulse sequence consists of:

$$90 - t_1 - 90 - \tau_M - 90 - t_2$$

where 90 = a 90° high power pulse, t_1 = a delay period during which the resonances are frequency labeled, τ_M = NOE mixing time, and t_2 = acquisition of the free induction decay. From section 1.4.2, we see that a 90°_x pulse (where the subscript x denotes the phase of the pulse) tips the equilibrium z-magnetization for all of the resonances in the molecule 90 degrees onto the y-axis. During t_1 , the resonances fan out as they rotate at their respective Larmor frequencies in the transverse plane. A second 90°_x pulse rotates the transverse magnetization into the xz-plane. During the mixing period τ_M , the z-components of the magnetization undergo dipolar relaxation which results in NOE build up. The final 90° pulse converts the z-magnetization into observable transverse magnetization, which is detected during t_2 . After a delay to allow the spins to equilibrate, the pulse sequence is repeated.

The addition of several free induction decays improves signal-to-noise by a factor equal to the square root of the number of scans. While repeating the pulse sequence, cycling of the pulse and receiver phases allows for suppression of artifacts due to imperfect instrumentation (Hoult & Richards, 1975) and also for quadrature detection (Drobny et al., 1978; States et al., 1982; Marion & Wüthrich, 1983). Another important purpose of phase cycling is to remove observable magnetization resulting from unwanted coherence transfer

(Keeler, 1988). Together, the pulse sequence and the proper phase cycling define this experiment to be a NOESY.

The second dimension arises from repeating this experiment, incrementing t_1 each time while holding τ_M constant. For each t_1 increment a separate FID is detected in t_2 . The raw data is Fourier transformed first in t_2 and then in t_1 , to generate a two dimensional frequency domain spectrum (Figure 1-7). Two-dimensional NMR experiments have been reviewed extensively in Kessler et al., 1988.

The utility of the NOESY experiment is that it provides pairwise distance information. It is preferable to use longer mixing time NOESY experiments with higher signal-to-noise ratios. This is important because longer mixing times allow proton pairs 4 to 5 Å apart to build up enough NOE intensity to be observable. It is often these longer distances which are crucial in calculating accurate three-dimensional structures. However, with longer mixing times, spin diffusion becomes more prevalent. If a reasonable starting structure can be obtained, it is possible to use algorithms which calculate pairwise distances from NOE intensities while taking into consideration spin diffusion pathways available in the starting structure (Borgias & James, 1989; Borgias & James, 1990).

1.4.7.5 NOESY Data Analysis. In a double-stranded DNA dodecamer, there are approximately 250 hydrogen atoms (protons), which generate a very complex and highly overlapped NMR spectrum (Figure 1-8). Before pairwise distances can be extracted from NOESY data, the resonances in the ^1H NMR spectrum must be assigned to specific protons in the molecule. Clearly, it is important to know the exact sequence of the DNA sample before beginning spectral assignment. Note that the different classes of protons fall into characteristic chemical shift ranges. Indeed, even within these ranges base protons exhibit characteristic shifts according to the type of base. For example, in standard B-form duplexes, thymine imino protons typically fall downfield of guanine imino protons. These trends in chemical shifts are structure dependent and although they are not strictly followed, they can be used as qualitative guides when approaching the assignment problem for DNA.

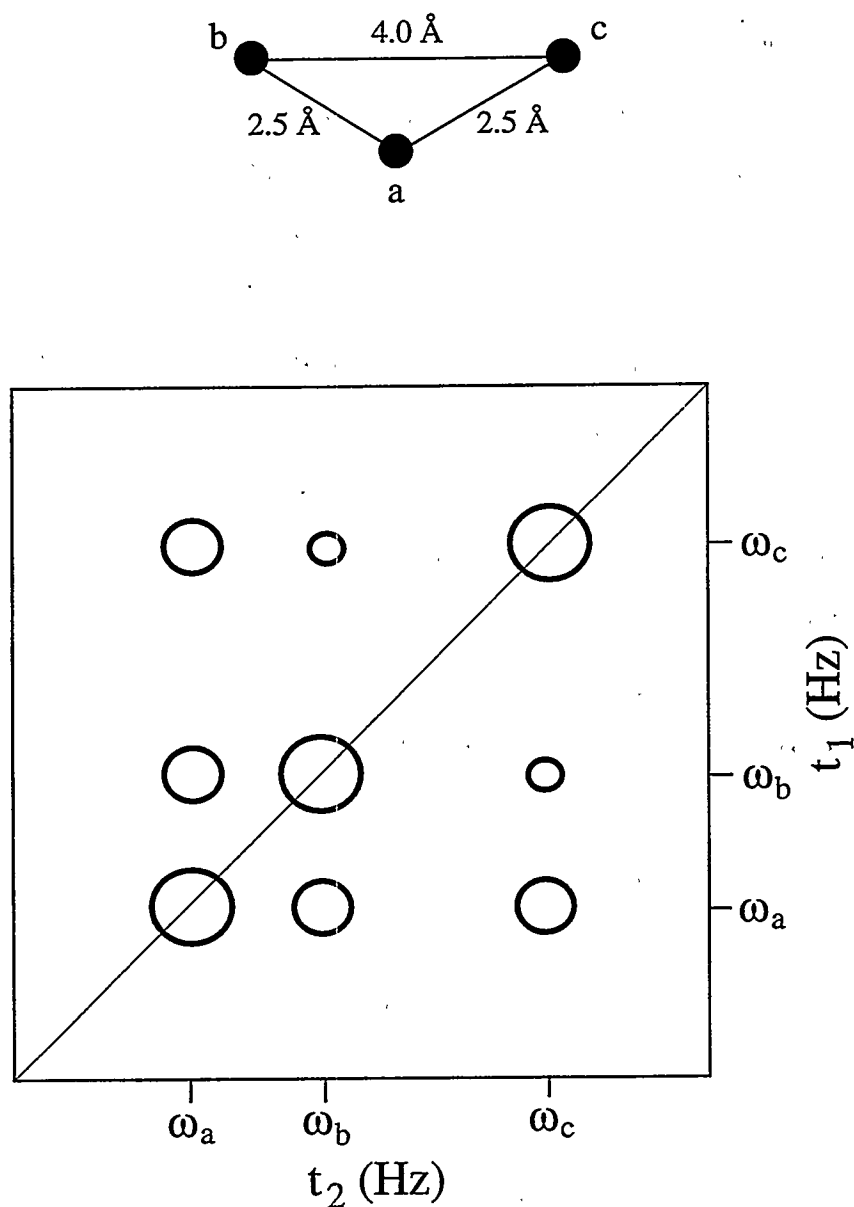


Figure 1-7. A two-dimensional NOESY spectrum for the hypothetical molecule shown. Protons a, b and c resonate at three distinct frequencies. The interproton distances are indicated. The proton pair ab (and likewise for the ac pair) builds up more NOE cross peak intensity than the bc pair because spins a and b are a short distance apart, while b and c are separated by a larger distance.

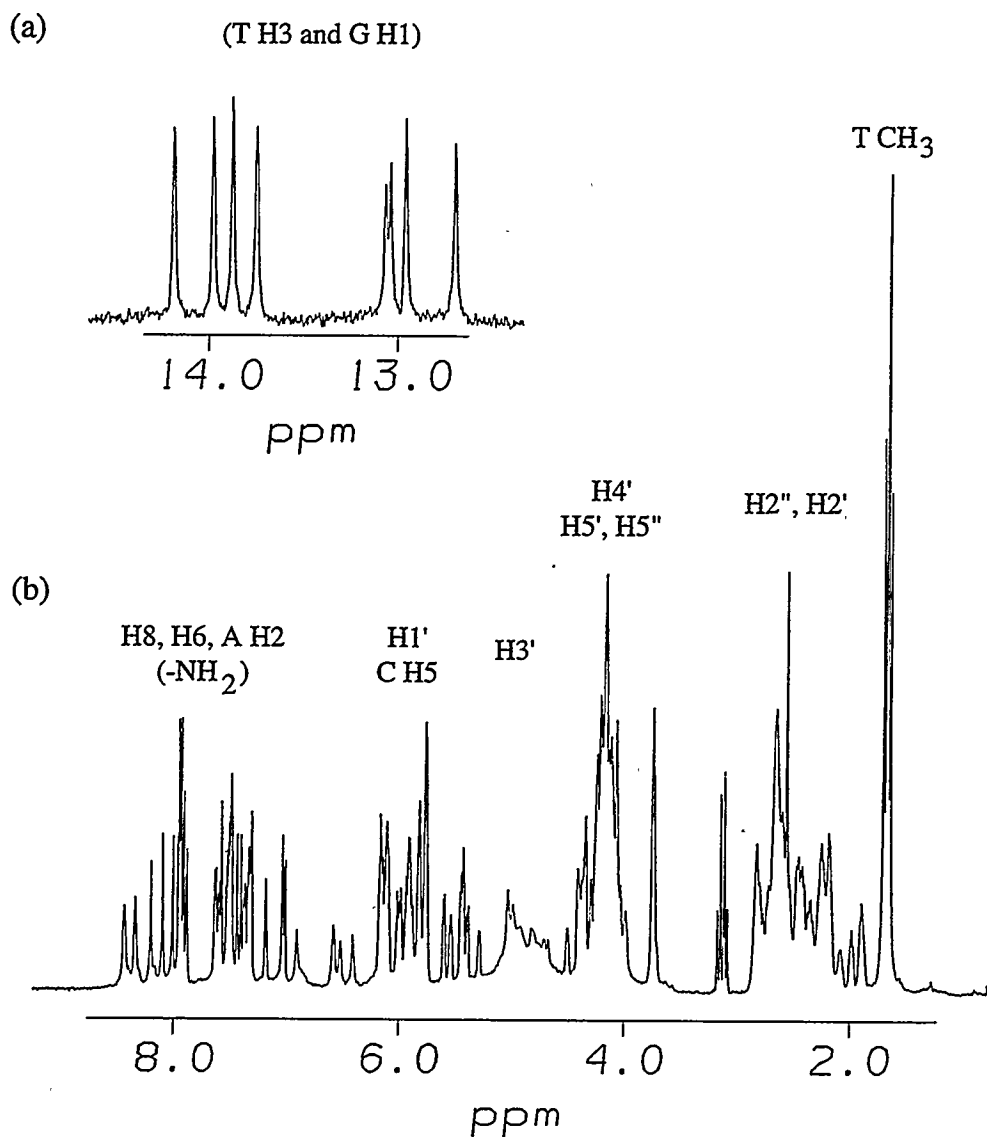


Figure 1-8. A one-dimensional ^1H NMR spectrum of a DNA decamer dissolved in H_2O solution. The general chemical shift ranges for each type of DNA proton are noted, with exchangeable protons listed in parentheses. The downfield region of the spectrum is shown in (a), and the rest of the spectrum is shown in (b). Note that both the vertical and horizontal scaling of (a) is different from (b).

One can take advantage of the rapid exchange rate of nitrogen-bound protons in DNA to simplify the ^1H NMR spectrum for assignment purposes. By dissolving a dried DNA sample in D_2O rather than H_2O , the base amino and imino protons exchange with deuterons and effectively disappear from the spectrum, since deuterons resonate at a very different frequency. This technique also significantly reduces the resonance arising from the solvent protons, simplifying the solvent suppression problem.

Assignment of the NOESY spectrum is facilitated by recognizable NOE patterns which are found in B-form DNA (Wüthrich, 1986; Hosur et al., 1988). In B-form DNA, a deoxyribose H1' proton has a weak NOE cross peak to its own pyrimidine H6 or purine H8 base proton, and another to the (n+1) neighboring base proton, where sequence numbering goes in the 5' to 3' direction. The deoxyribose H2' and H2'' protons display a similar pattern. With this information one can map a NOE connectivity pathway independently along each DNA strand (Figure 1-9). Adenosine H2 resonances (which play an important role in the studies described herein) can sometimes be assigned based on weak NOE cross peaks to the H1' of its (n+1) neighbor and to the H1' of the (n+1) neighbor of the base paired thymine. When the sample is dissolved in H_2O solution, the presence of slowly exchanging imino protons establish base pairing, except for the two terminal imino protons which often undergo rapid exchange with solvent water. Imino-imino NOEs provide sequential assignments for the imino protons.

1.4.8 Chemical Exchange. NMR spectroscopy is a particularly useful method because we can obtain kinetic information in addition to structural information about the system of interest. The NMR spectrum is affected by exchange processes which occur at rates between 1 Hz (faster than T_1 relaxation rates) and 10^6 Hz (slower than molecular tumbling rates) (Neuhaus & Williamson, 1989). Exchange processes play an important role in studies of DNA and ligand:DNA complexes. Examples of exchange processes which occur in such a complex include solvent exchange (of nitrogen bound protons with water), conformational changes (such as rotations of exocyclic amino groups on DNA

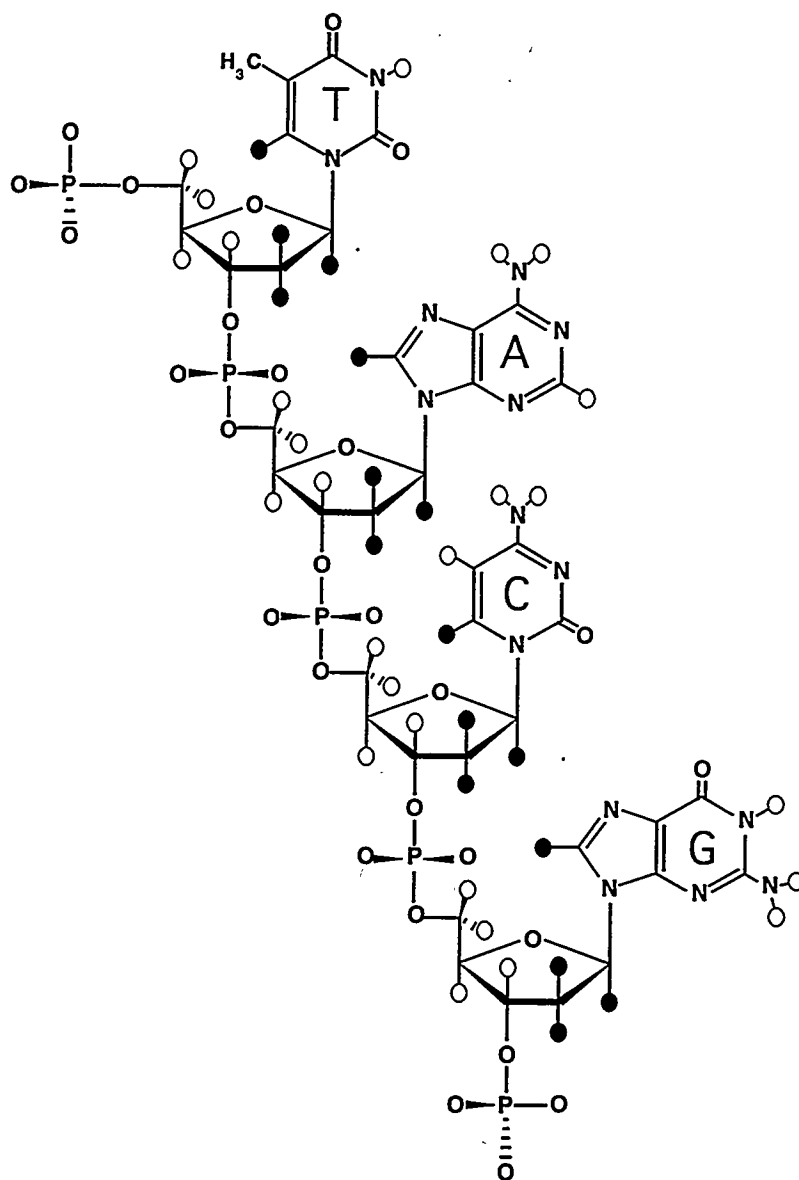


Figure 1-9. ^1H NOE connectivity pathway used in assigning DNA proton resonances. The protons which make up the pathway are indicated by black circles: purine H8, pyrimidine H6, and ribose H1', H2' and H2''. NOE cross peaks are generally observed between a ribose proton and its own base proton as well as the 3' neighboring base proton. Other protons not on the pathway are denoted by white circles.

bases about the N-C bond) and site-to-site exchange (such as the binding of a free ligand to DNA or the reorientation of a ligand on the DNA). Proper interpretation of NMR data requires that we have some understanding of the nature of any exchange processes which may be occurring, their time scales and their effects on the NMR spectrum.

1.4.8.1 Time Scales. Time scales in NMR are defined by relaxation rates and chemical shift differences. When considering exchange effects on NMR spectra, one must compare the rate of the exchange process to these time scales to determine how the effects will appear in the NMR spectrum, if at all. Generally, exchange processes with rates much slower than the overall relaxation rate will have no effect on NMR spectra, since T_1 relaxation will tend to reequilibrate the system before exchange can occur. Therefore, exchange processes must have rates intermediate to fast on the relaxation time scale (or T_1 time scale) to have an observable effect on the NMR spectrum (Sanders & Hunter, 1987).

The simplest case of exchange involves a two-state process:



where k_{AX} and k_{XA} are rate constants for the forward and backward reactions. The relative populations of these two states are determined by the equilibrium constant for the system, $K_{eq} = [X]/[A] = k_{AX}/k_{XA}$. The exchange lifetimes (here, the average time a molecule spends in one state) are defined to be $\tau_A = 1/k_{AX}$ and $\tau_X = 1/k_{XA}$. The assumption of simple two-state exchange is often made in analyzing NMR spectra.

The chemical shift time scale is defined by the difference between the chemical shift of a proton in state A and that of the same proton in state X, $|\omega_A - \omega_X|$ (Figure 1-10). Using this time scale, the slow exchange regime is defined to be the case in which the inverse of the exchange lifetime (we will consider state A) is significantly smaller than the chemical shift difference, or $|\omega_A - \omega_X| (\tau_A) \gg 1$. The one-dimensional NMR spectrum of a system which is slowly exchanging between two states simply consists of a superposition of the two independent 1D spectra. The intermediate exchange regime is defined to be the

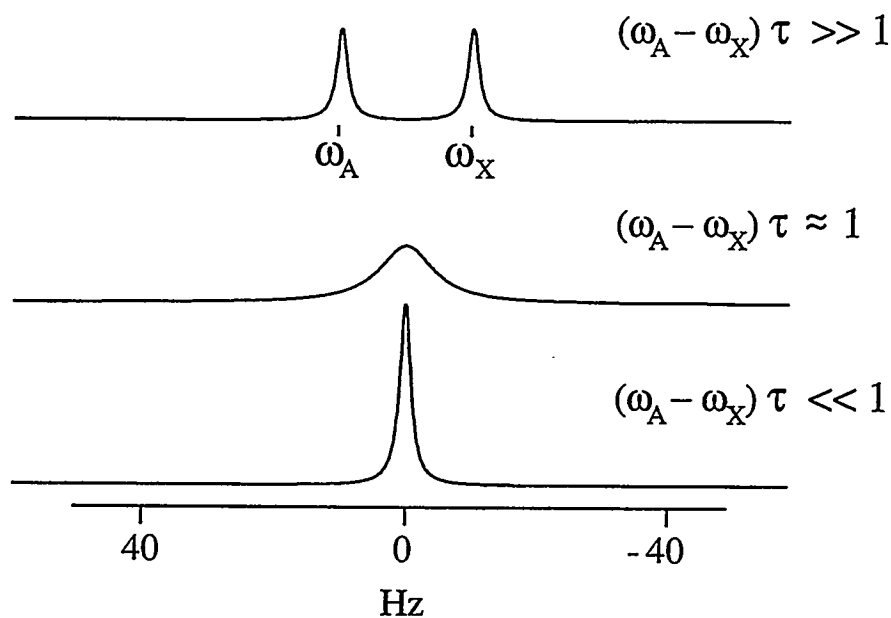


Figure 1-10. The effects of chemical exchange on the appearance of the NMR spectrum. Shown are 1D NMR spectra of one proton which is exchanging between two forms A and X, as in Equation 1-21. In this example the populations of the A and X forms are equal, so that $k_{AX} = k_{XA}$ and $\tau_A = \tau_X = \tau$. The three spectra demonstrate the effects of slow (top), intermediate (center), and fast exchange (bottom).

case in which $|\omega_A - \omega_X|(\tau_A) \approx 1$. Where $|\omega_A - \omega_X|(\tau_A) = 1$, the two resonances coalesce into one broad peak with a flat top. The fast exchange regime is defined to be the case in which $|\omega_A - \omega_X|(\tau_A) \ll 1$. The one-dimensional NMR spectrum of such a system consists of a single resonance which is at the population weighted average chemical shift of the two states A and X.

Note that in real systems, especially for large biomolecules, it is possible for different resonances in the same molecule to be in different exchange regimes. For example, consider the binding of a small molecule to DNA. A DNA proton within the binding site might experience a large chemical shift change upon ligand binding, whereas a proton distant from the binding site might undergo a small chemical shift change in the bound state. Clearly, depending upon the exchange rate of the ligand between the free and bound states, different exchange regimes might apply to the two different protons. Such information can be useful in identifying the location of the binding site on the DNA duplex.

NMR signals decay exponentially with the decay rate characterized by a first order time constant, resulting in a Lorentzian line shape. Under static conditions (no exchange), the line width of a resonance at half height is:

$$\Delta\nu_{1/2} = \frac{1}{(\pi T_2^*)} \quad (\text{Equation 1-22})$$

where T_2^* is the effective transverse relaxation time constant due to intrinsic T_2 relaxation as well as other factors, including magnetic field inhomogeneities. An exchange process which is slow on the chemical shift time scale but is faster than transverse relaxation ($k_{AX} > 1/T_2^*$) will result in additional line broadening of resonances corresponding to state A (Harris, 1986):

$$\Delta\nu_{1/2} = \frac{1}{(\pi\tau_A)} \quad (\text{Equation 1-23})$$

where $\tau_A = (1/k_{AX})$ and is the lifetime of the A state.

1.4.8.2 Measurement of Exchange Kinetics by NMR. Consider a system in slow exchange on the chemical shift time scale (recall that to have an observable effect it still

must be in intermediate to fast exchange on the T_1 time scale). The NMR spectrum of a particular proton in states A and X contains resonances at two distinct chemical shifts, ν_A and ν_X . During the t_1 period of the NOESY experiment, spins in state A are frequency labeled at ν_A . If a proton undergoes chemical exchange from state A to state X during the mixing time τ_M , it will be detected during the t_2 period at a new Larmor frequency, ν_X . Therefore, an off-diagonal peak is generated in the 2D spectrum, not due to cross relaxation, but due to chemical exchange.

In this exchange regime, the 2D NOESY can be used expressly for the purpose of extracting rate constants from exchange cross peaks, analogous to the extraction of distances from NOE cross peaks (reviewed in Perrin & Dwyer, 1990). This experiment is sometimes termed 2D exchange spectroscopy (EXSY), although in practice it is identical to the 2D NOESY. At short mixing times (where little T_1 relaxation and cross relaxation have occurred), the ratio of the intensities of an exchange peak and its diagonal peak will give the fraction of protons which have undergone exchange from state A to state X during τ_M . The exchange lifetime τ_A can be determined by:

$$\frac{I_{AX}(\text{exch})}{I_{AA}(\text{diag})} = \frac{\tau_M}{\tau_A} \quad (\text{Equation 1-24})$$

1.4.8.3 Exchange Effects in 2D NOESY. An exchanging system can create a very complex 2D NOESY spectrum. The equations that govern the rate of NOE build up are affected by the rate of chemical exchange relative to T_1 and τ_C (Neuhaus & Williamson, 1989). Where the exchange rate falls in the chemical shift time scale is not relevant to the rate of NOE build up. However, the appearance of the NOESY spectrum will be affected by the rate of exchange on the chemical shift time scale, which determines the line widths and number of resonances in the spectrum.

First consider exchange processes which are fast relative to T_1 and slow relative to τ_C . The homonuclear NOE build up rate equation is given for the general case of an isotropically tumbling molecule exchanging among i states:

$$\left. \frac{dI_Z}{dt} \right|_{t=0} = 2 \langle \sigma_{IS} \rangle S_Z^0 \quad (\text{Equation 1-25})$$

with
$$\langle \sigma_{IS} \rangle = \sum_i (N^i \sigma_{IS}^i) \quad (\text{Equation 1-26})$$

where $\langle \sigma_{IS} \rangle$ = time average of the cross correlation rate constant, N^i = fractional population of state i , and σ_{IS}^i = cross relaxation rate constant for the IS spin pair in state i . The overall cross relaxation rate and hence the NOE build up rate is the population weighted average over all the states. Extraction of distance information from this build up rate will result in an averaged internuclear distance.

For systems in slow exchange on the chemical shift time scale, the 2D NOESY spectrum will appear quite complex, due to the presence of resonances from the different exchanging forms. If the exchange is fast relative to T_1 , the NOE cross peak intensities build up according to Equation 1-25 above. The only additional consideration is that an extra set of cross peaks appears due to exchange-transferred NOE. These cross peaks arise when the populations of spin S_A are perturbed, giving rise to a direct NOE which builds up on a nearby spin I_A . If I_A then undergoes exchange to I_B , a cross peak will appear in the NOESY from S_A to I_B . These exchange-transferred NOE cross peaks must be distinguished from direct NOE.

For systems in fast exchange on the chemical shift time scale, there are fewer resonances in the spectrum. In fact, it may not be initially apparent that a fast exchange process is occurring, since the number of resonances is the same as for a static system. The complexity arises in the interpretation of the NOE cross peaks observed. For example, consider a ligand which is sliding rapidly (on the chemical shift time scale) between two sites on the DNA, so that the ligand protons resonate at the population weighted average chemical shift of the two exchanging binding modes. Each ligand resonance builds up NOE in both binding sites independently, resulting in cross peaks between the ligand and sets of DNA protons which are inconsistent with one static structure. Quantitative

interpretation of the intensity of these NOE cross peaks is less straightforward than for the slow exchange case. Because the ligand and DNA resonances from both binding modes resonate at the same chemical shift, cross peaks in the NOESY will contain exchange transferred NOE superimposed onto direct NOE contributions.

In biomolecules, which tumble relatively slowly due to their large size (τ_c is on the order of nanoseconds), some exchange processes occur at rates which are faster than the overall tumbling rate of the molecule. Such processes include the rotation of methyl groups and the motion of some protein side chains. These motions occur in local regions of large biomolecules which effectively shorten the correlation time of that region, leading to a spread of τ_c values throughout the molecule (Neuhaus & Williamson, 1989). Because the cross relaxation rate is a function of spectral density, the NOE build up rate will vary for different regions of the molecule. In fact, this observation has led to the use of ^{15}N - ^1H heteronuclear NOE experiments as a method to probe for very fast motions in proteins (Palmer, 1993). Because the amide nitrogen and amide proton are directly bonded and are separated by a fixed distance, significant differences in NOE build up rates are therefore attributed to altered local correlation times. Backbone amides are distributed evenly throughout the protein, making it possible to map their locations on to the three-dimensional structure, thereby revealing the portions of the molecule undergoing dynamics. Clearly then, motions faster than the overall tumbling rate will be a source of error in NOE-derived interproton distances for large molecules.

One more circumstance should be considered with respect to exchange effects on NOE measurements. Recall that from Equation 1-19, the cross relaxation rate of the NOE is proportional to r_{IS}^{-6} , where r_{IS} is considered to be a constant. Suppose that the molecule is not rigid, so that r_{IS} cannot be assumed to be a constant. Consider two protons I and S which are separated by a distance which varies through some range due to internal molecular motions. Because the NOE build up rate falls off quickly with distance, the shorter interproton distances will dominate the dipolar relaxation. The result is that the

NOESY cross peak between the I and S resonances will be more intense than for the rigid case where r_{IS} is a fixed distance. Interpretation of this cross peak intensity under the rigid molecule assumption will result in an interproton distance which is too short. Therefore, a better interpretation of a NOE-derived distance is that it is the minimum allowed distance.

1.5 Molecular Modeling

Once pairwise interproton distances have been extracted from NOESY data, they are used as input restraints for calculations which generate a family of three-dimensional structures for the molecule. The most important aspects of these calculations will be summarized here. The primary reference for the following discussion is the user guide for the molecular modeling program Discover (1992).

1.5.1 Forcefield. The forcefield is an empirically derived analytical expression of the potential energy surface of a system in terms of the nuclear coordinates. One particularly useful forcefield is AMBER, which was developed specifically to describe proteins and nucleic acids (Weiner et al., 1984; Weiner et al., 1986). The functional form and descriptions of the terms in the AMBER energy expression are shown in Figure 1-11. All of the atoms in a system which is modeled under the AMBER forcefield are assigned atom types which specify element type, hybridization, and bonding environment. The forcefield contains idealized equilibrium values for the parameters in the target function, including bond lengths, bond angles and van der Waals radii. As the calculation changes the global structure of the molecule, an energetic penalty is assigned to structures which deviate from idealized equilibrium geometry. The severity of the energetic penalty is determined by both the amount of deviation from equilibrium, and by an adjustable parameter called a force constant. The ability of the forcefield to accurately describe the energetics of the system of interest determines the quality of the calculated results.

Because of the large number of atoms in a nucleic acid or a protein molecule (a DNA dodecamer has approximately 750 atoms), the summation of all interactions in the

$$\begin{aligned}
E_{pot} = & \sum_b K_2 (b - b_0)^2 \\
& + \sum_{\theta} H_{\theta} (\theta - \theta_0)^2 \\
& + \sum_{\phi} \frac{V_n}{2} [1 + \cos(n \phi - \phi_0)] \\
& + \sum \epsilon [(r^*/r)^{12} - 2 (r^*/r)^6] \\
& + \sum \frac{q_i q_j}{\epsilon_{ij} r_{ij}} \\
& + \sum \left[\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^6} \right]
\end{aligned}$$

Figure 1-11. The AMBER energy expression, or target function. The first three terms characterize the energies needed to stretch bonds, bend bond angles, and twist dihedral angles away from their reference equilibrium values. The fourth and fifth terms describe nonbonded interaction energies, van der Waals and electrostatic interactions, which are a function of internuclear distances. The sixth term is a small hydrogen bond term to account for energies not included in the electrostatic terms. The parameters are b = current bond length, b_0 = equilibrium bond length, θ = current bond angle, θ_0 = equilibrium bond angle, ϕ = current dihedral angle, ϕ_0 = equilibrium dihedral angle, ϵ = depth of the Lennard-Jones potential well, r^* = equilibrium internuclear distance, r = current internuclear distance, q = point charge, ϵ_0 = permittivity constant.

molecule becomes enormous, and calculations become unmanageable and very computationally inefficient. Many of the pairwise electrostatic and van der Waals interactions are insignificant because these energies fall off quickly with distance. So, it is customary to set a nonbond cutoff distance which excludes nonbonded atom pairs beyond the cutoff distance from the calculation.

Additional structural restraints obtained experimentally are included as extra terms in the potential function, such as distance restraints obtained from NOESY cross peak intensities or dihedral angle restraints obtained from measured scalar couplings. Distance restraints are input as ranges to allow for experimental error and uncertainty in the measured distance. This term takes the form of a flat-bottomed well with parabolic sides in the target function.

1.5.2 Energy Minimization. The goal of energy minimization is to find a conformation in which the net force on every atom vanishes (a local energy minimum). At an energy minimum if any coordinate is changed, the result is that the energy of the target function increases, and the force applied according to the force constant drives the conformation back into the local energy well. The general procedure for energy minimization consists of two main steps. An initial structure is input and its target function is evaluated. The conformation is then altered slightly and the target function is reevaluated. This process is iterated until the nearest local energy minimum is found.

There are several different algorithms which determine how the molecular conformation is adjusted at each iteration. The most robust algorithm is the method of steepest descents. This method relies on the direction of the energy gradient on the potential energy surface to determine the path of the system on the potential energy surface, and the associated conformational changes. Therefore, it works best when the conformation is far from the energy minimum (i.e. at the beginning of the calculation) and the gradients are large. As the structure approaches an energy minimum, this method becomes less efficient because the energy gradient approaches zero, and this method

continually overcorrects for poor choices of directions in earlier steps. At this point, the conjugate gradients algorithm becomes useful. This method is more computationally intensive than steepest descents when the system is far from the energy minimum, but because of that it works better when the system approaches the energy minimum. Therefore, steepest descents brings the system roughly near the local minimum, and conjugate gradients fine-tunes the structure.

1.5.3 Molecular Dynamics. The goal of the calculation is to find the global energy minimum for the molecule using NOE-derived distance restraints. However, the initial input structure may not be close to the final, "true" structure, the global energy minimum. If the data defines the system well enough, in theory it should be possible to start from any starting structure and arrive at the global energy minimum. Minimization calculations only find the nearest local minimum in the target potential function. It is necessary to search conformational space more effectively by going over energetic barriers on the potential energy surface. Energy minimization *decreases* the value of the target function, when in fact we want to be able to *increase* the energy of the system in order to climb over energetic barriers. Molecular dynamics calculations add kinetic energy to the system so that conformational space can be searched more effectively. The system is computationally "heated" and then slowly "cooled" by a process called simulated annealing. Fast "cooling" will force the system into a nearby local minimum, similar to an energy minimization, whereas slow "cooling" allows the system to search out the lowest accessible energy minimum. It is common to perform an energy minimization step after dynamics to bring the family of conformations further into the energy minimum.

1.6 Application of NMR and Molecular Modeling Techniques to DNA and its Complexes

The following chapters describe applications of NMR spectroscopy and molecular modeling to several ligand:DNA complexes and one DNA mispair containing a mutagenic base analog. What useful information about these systems is obtained by such studies?

NMR studies are a rich source of structural information. In the studies described here, the main body of structural information comes from analysis of NOESY data. Through the use of NOE-derived interproton distances coupled with molecular modeling, structural features of ligand:DNA complexes are determined, including ligand binding stoichiometry, binding site and orientation. Other features of NMR data provide less quantitative, but still useful structural information. For example, chemical shifts of imino and amino protons provide evidence for the presence or absence of hydrogen bonds. Unusual chemical shifts, as well as unusual NOE patterns, can indicate non-standard DNA conformation as seen in the mispair and some of the ligand complexes. Useful structural information is also extracted from analysis of exchange processes. For example, slowly exchanging imino protons appearing in the downfield region of the NMR spectrum indicate that the DNA duplex is stably base paired. Imino protons which are not protected in this way exchange rapidly with solvent, and consequently are often broadened into the baseline. Additionally, the identification of tautomeric forms of bases in the DNA mispair requires the proper assignment of the resonances in the ^1H NMR spectrum, which confirms the chemical structure of the bases.

NMR studies also provide thermodynamic information with respect to ligand binding. Observation of exchange behavior during a titration of a DNA sample with a ligand can give qualitative information regarding ligand binding affinities. The binding constants for the complexes studied are quite high ($K_b \approx 10^4$ to 10^8 M^{-1}). The binding constant is related to the on- and off-rates by $K_b = k_{\text{on}}/k_{\text{off}}$, and the on-rate of the ligand has been found to be very fast. Therefore, the off-rate is much slower and its effects are

apparent in the NMR spectrum. The ligand:DNA complexes characterized in the following chapters exhibit off-rates spanning the range from slow to fast exchange on the NMR chemical shift time scale. The ligand off-rate can be used as a qualitative measure of ligand binding affinity. The absolute binding affinity has been measured for a few distamycin:DNA complexes, and their NMR exchange behavior is known. Qualitative comparisons are made between the NMR exchange behavior of various other complexes and these complexes of known affinity. Additionally, analysis of a slowly exchanging mixture of complexes can reveal more quantitative information. A sample containing an equilibrium of two different ligand:DNA complexes in solution generates a NMR spectrum consisting of two sets of resonances (if the equilibrium is in slow exchange on the chemical shift time scale). The integrated area of resonances corresponding to the two complexes reflects their relative populations, revealing their relative binding affinities. In the case of 1:1 and 2:1 distamycin:DNA complexes in slow exchange, the relative binding affinities of the first and second distamycin molecules are determined.

Lastly, NMR studies on DNA and its complexes can reveal kinetic information about the system. Measurement of T_1 relaxation times in the presence of a proton exchange catalyst reveals the opening lifetimes of individual base pairs in a free DNA duplex. In Chapter 2, the base pair opening rates of a duplex are correlated with its distamycin binding properties.

NMR studies provide critical information on the structural, kinetic and thermodynamic properties of DNA and ligand:DNA complexes. Through these kinds of studies, conclusions can be drawn regarding the structural requirements for sequence specific recognition of DNA. This understanding can then be applied to the rational design of ligands which bind DNA in a predictable fashion.

Chapter 2

The Binding Modes of Distamycin A and Netropsin to A,T- and I,C-rich DNA Oligomers¹

2.1 Summary

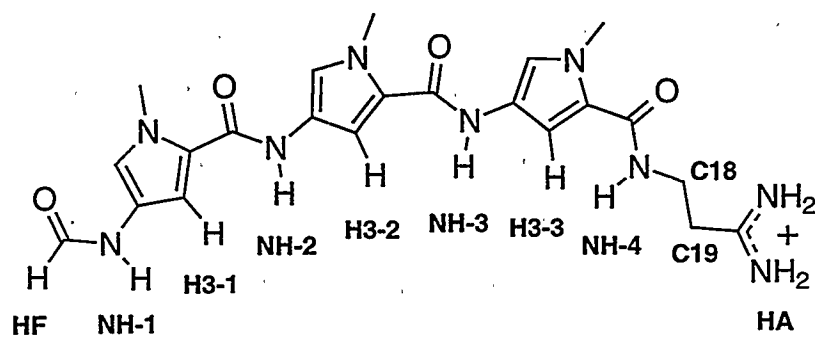
One- and two-dimensional NMR experiments have been used to characterize the binding site, orientation, stoichiometry, and qualitative affinity of distamycin binding to several short DNA oligomers. The binding sites studied contain A•T and/or I•C base pairs, where I = 2-desaminodeoxyguanosine. I•C base pairs are functional analogs of A•T base pairs in the minor groove, although the described distamycin binding and base pair opening lifetime measurements suggest that the groove is structurally different from the narrow minor groove observed in A,T-rich oligomers crystallographically. Sites which are classified as A-tracts, such as AAAA and AATT, have narrow minor grooves while alternating sites, such as TATA, and I,C-containing sites have wider minor grooves. Distamycin binds tightly in a 1:1 ligand:DNA mode to sites with narrow minor grooves, and weakly in the 1:1 mode to sites with wide minor grooves. For sites of five or more A•T or I•C base pairs, distamycin can also form a 2:1 ligand:DNA complex with the two ligands bound antiparallel and side-by-side. In sites with wide minor grooves, although the first ligand binds weakly, the second ligand binds with higher affinity to form the 2:1 complex. The cumulative effects of changes in sequence and substitution of I•C base pairs for A•T base pairs are examined. A comparison between distamycin and netropsin binding to an asymmetric sequence reveals that whereas distamycin binds directionally, netropsin has no orientational preference in binding.

¹ Reprinted in part with permission from Fagan, P. & Wemmer, D.E. (1992) *J. Am. Chem. Soc.* 114, 1080-1081. Copyright 1992 American Chemical Society.

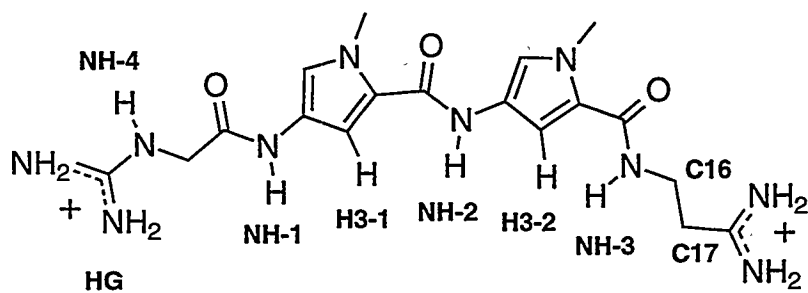
2.2 Introduction

The desire for the development of small molecules which bind to specific DNA sequences has fueled interest in distamycin A (Figure 2-1). The minimal binding site for distamycin is four base pairs, as seen in NMR studies of the distamycin: [d(CGCGAATTCGCG)]₂ complex (Klevit et al., 1986). Distamycin binds in a 1:1 ligand:DNA mode to the minor groove of the AATT site in slow exchange on the NMR time scale, with a binding constant measured by calorimetry to be $2.7 \times 10^8 \text{ M}^{-1}$ (Marky, 1986). Binding studies with the oligomer [d(GGTATACC)]₂ show that distamycin binds weakly to the TATA site with a binding constant of $2 \times 10^5 \text{ M}^{-1}$, obtained by quantitative analysis of footprinting data (Fish et al., 1988). Distamycin has been observed by NMR to bind in intermediate exchange with the TATA site (J. G. Pelton, unpublished results), also indicating a lowered affinity for this site relative to AATT. The distinctly different binding behavior in these two cases suggested that there are subtle differences in the structure of A,T-containing DNA sequences which the ligand recognizes. This hypothesis forms the basis for the studies described in the first part of this chapter, which examine distamycin and netropsin binding to various four base pair sites.

Further motivation for the work described in this chapter comes from NMR studies by J. G. Pelton on two DNA oligomers containing five and six base pair binding sites, d(CGCAAATTGGC): d(GCCAATTTGCG) (Pelton & Wemmer, 1989) and [d(CGCAAATTTGCG)]₂ (Pelton & Wemmer, 1990a). These sites bind a single distamycin at low ligand ratios, analogous to complexes characterized crystallographically. In several of the DNAs studied crystallographically, the groove is found to be equally narrow without the drug present (reviewed in Dickerson, 1990 and Yanagi et al., 1991). However, at higher amounts of added ligand, new complexes form with two distamycin molecules bound side-by-side in the same region of the minor groove. This indicates a significant degree of adaptability in the minor groove width, since an increase in groove width of about 3.5 Å is required to accommodate the second ligand.



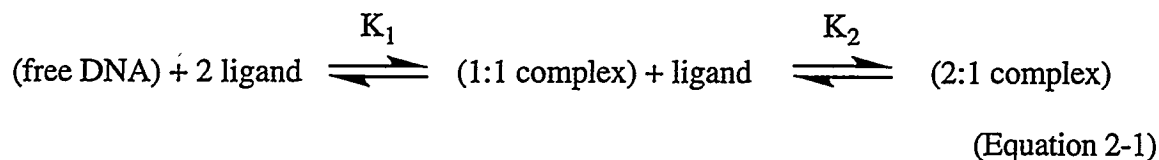
distamycin A



netropsin

Figure 2-1. Distamycin A and netropsin. Numbering schemes are shown.

Analysis of the NMR data shows that the second ligand binds *less* tightly than the first for the AAATT site, and somewhat *more* tightly than the first for the AAATTT site. This binding equilibrium is described by the following set of equations:



Cooperativity in this system is defined as the ratio of the binding affinities K_2/K_1 . The cooperativity of distamycin binding varies with the DNA sequence. AAATT binds distamycin anticooperatively ($K_2/K_1 < 1$), while AAATTT binds distamycin cooperatively ($K_2/K_1 > 1$). What factors determine the extent of cooperativity of distamycin binding with these two binding sites? To explore this question, I have performed experiments on related A,T-rich sites and on analogous sites containing I•C base pairs, which mimic A•T pairs in the minor groove.

The work contained within this chapter examines the recognition of sequence specific variations in DNA structure by distamycin. This chapter examines the binding of distamycin with various A,T-rich DNA sequences, and attempts to correlate DNA sequential and structural features with the observed mode of distamycin binding.

2.3 Materials and Methods

2.3.1 NMR Sample Preparation. The DNA oligomers used were made on an automated DNA synthesizer by standard methods. After HPLC purification using a reversed phase C18 column, samples were desalted by dialysis using a 1000 kDa molecular weight cutoff membrane, or by a Waters Sep-Pak C18 column. Concentrations were determined by UV absorbance at 260 nm at room temperature prior to annealing of single strands, or at 80 °C after annealing. Oligomer extinction coefficients were calculated using standard methods (Warshaw & Cantor, 1970). After annealing, in some cases the duplex

sample was run over a hydroxylapatite ion exchange column (Bio-Rad, DNA grade Bio-Gel HTP) to remove any excess single-stranded DNA.

All DNA oligomers were studied in duplex form, ranging from 10-mers to 12-mers in length. Each nucleotide was numbered, starting from the 5'-end to the 3'-end of the first strand, and then continuing the numbering sequence from the 5'-end to the 3'-end of the second strand. For example, a 10-mer duplex d(CGCAAAGGC):d(GCCTTTTGCG) is numbered in the following way:

5'- C₁ G₂ C₃ A₄ A₅ A₆ A₇ G₈ G₉ C₁₀ -3' (strand 1)

3'- G₂₀ C₁₉ G₁₈ T₁₇ T₁₆ T₁₅ A₁₄ C₁₃ C₁₂ G₁₁ -5' (strand 2)

DNA duplexes which are symmetric, such as [d(CGCGAATTCGCG)]₂, were numbered from 1 through 24 if the ligand complex is nonsymmetric (as is the case for a 1:1 ligand:DNA complex). If the ligand complex is symmetric (as is the case of some 2:1 ligand DNA complexes), then the DNA retains its symmetry and the two strands of DNA are indistinguishable. In such cases, both DNA strands are numbered from 1 through 12:

5'- C₁ G₂ C₃ G₄ A₅ A₆ T₇ T₈ C₉ G₁₀ C₁₁ G₁₂ -3' (strand 1)

3'- G₁₂ C₁₁ G₁₀ C₉ T₈ T₇ A₆ A₅ G₄ C₃ G₂ C₁ -5' (strand 2)

DNA duplexes are referred to throughout the text as "free" or "uncomplexed" DNA when no ligand is present.

Distamycin A and netropsin were purchased from Sigma or Serva and used without further purification. Stock solutions of ligand were typically made by dissolving 1-2 mg ligand into 100 μ L D₂O. Ligand concentrations were determined by UV absorbance using extinction coefficients $\epsilon_{304} = 34,000 \text{ M}^{-1}\text{cm}^{-1}$ for distamycin, and $\epsilon_{296} = 20,200 \text{ M}^{-1}\text{cm}^{-1}$ for netropsin (Liquier et al., 1989). Due to its instability in aqueous solution, fresh distamycin solutions were prepared for each titration.

NMR samples were prepared by dissolving the dried DNA duplex into 99.9% D₂O (Cambridge Isotopes) or a 90% H₂O/10% D₂O mixture to a final volume of 0.5 mL, with

10 mM phosphate buffer (pH 7) and 0.1 mM EDTA. DNA duplex concentrations ranged from 1.0 to 4.0 mM.

2.3.2 NMR Experiments. NMR data were acquired on a 500 MHz General Electric GN or Omega spectrometer or a 600 MHz Bruker AMX spectrometer. For each study, the sequence of experiments generally consisted of a one-dimensional NMR titration followed by two-dimensional NOESY spectra (Jeener et al., 1979) acquired in D₂O solution and/or in H₂O solution. In several cases a 2D NOESY was acquired for the uncomplexed DNA as well. 1D experiments typically consisted of 4096 to 8192 complex points, with 128 to 512 scans. 2D NOESY experiments typically consisted of 1024 complex points in t_2 , 64 scans per block and 512 points in t_1 with a mixing time of 150 to 250 ms. For samples in D₂O, a spectral width of 6024 Hz (at 600 MHz) was used and suppression of the water resonance was achieved by applying a presaturation pulse during the approximately 2 sec recycle delay. For samples in H₂O solution, a spectral width of 13514 Hz (at 600 MHz) was typically used. NOESY spectra of samples in H₂O solution were acquired by replacing the last 90° pulse by a 1-1 jump and return sequence to suppress the solvent signal (Sklenar et al., 1987). Quadrature detection was achieved in t_1 using TPPI (Drobny et al., 1978; Marion & Wüthrich, 1983).

2.3.3 NMR Data Processing and Analysis. The data were processed with FTNMR (Hare Research) on a Vax 4000-300 computer or FELIX (Biosym Technologies) on a Silicon Graphics IRIS/4D workstation. Free induction decays were apodized with a skewed sinebell squared function which was phase shifted by 60 to 90 degrees. For spectra acquired in H₂O solution, the intense solvent resonance was removed using a convolution method. After Fourier transformation, the baseline was flattened using a cubic spline or polynomial correction method.

NOESY spectra of the ligand:DNA complexes in D₂O and/or H₂O enabled assignment of the DNA and ligand resonances (Hare et al., 1983; Wüthrich, 1986; Pelton & Wemmer, 1988; Pelton & Wemmer, 1989; Dwyer et al., 1992). Typically, the DNA

base protons and sugar H1', H2', H2'', and H3' protons were assigned, extending to the H4' in some cases. The ligand pyrrole H3 proton and formyl proton resonances were assigned, with the methylene protons on the propylamidium "tail" assigned as a group. If NOESY experiments were acquired for a complex in H₂O solution, then the ligand amide proton resonances were also assigned.

2.3.4 Distance Restraints. Intermolecular distance restraints for the 1:1 distamycin complex with d(CGCAAAGGC):d(GCCTTTTGCG) and the 2:1 distamycin complex with d(CGCIICCGGC):d(GCCIICCGCG) were generated from the volume integrals of the cross peaks in the D₂O NOESY spectra acquired at a mixing time of 200 msec as described previously (Dwyer et al., 1992; Mrksich et al., 1992). The cross peak volumes were classified semi-quantitatively into three categories: strong (1.8 to 2.8 Å), medium (1.8 to 3.5 Å) or weak (1.8 to 5.0 Å) relative to the volume integrals of the cytosine H5-H6 cross peak volumes. A pseudoatom was generated for the four methylene protons C(18)Ha, C(18)Hb, C(19)Ha, and C(19)Hb on the ligand molecule. Distance restraints from DNA protons to the pseudoatom were given a range of (1.8 to 7.0 Å). Ten intermolecular ligand-DNA restraints were used in modeling the AAAA complex, and 16 ligand-DNA restraints were used in modeling the IIIC complex.

2.3.5 Molecular Modeling. Models of the 1:1 distamycin complex with d(CGCAAAGGC):d(GCCTTTTGCG) and of the 2:1 distamycin complex with d(CGCIICCGGC):d(GCCIICCGCG) were built based on the NMR-derived distance restraints. Starting models of the two DNA duplexes were constructed using the Biopolymer module of Insight II (Biosym) from standard B-form DNA. Distamycin was assembled using the sketcher module of Insight II (Biosym) and assigned AMBER type potentials. The partial charges on the atoms were obtained by a MOPAC calculation on the ligand. The ligand model was manually docked into the minor groove using Insight II.

Energy minimizations were performed using the Discover module of Insight II (employing the AMBER forcefield). Force constants of 200 (kcal/mole)/Å² and 50

(kcal/mole)/Å² were used for the experimentally derived Watson-Crick hydrogen bond restraints and ligand-DNA NOE restraints, respectively. The cutoff distance for non-bonded interactions was set at 18 Å with a switching distance of 2 Å (thereby scaling all terms in the potential function smoothly to zero between 16 and 18 Å). A distance-dependent dielectric of the form $\epsilon = R$ was used to account for solvent effects.

The model refinements were performed in two steps. First, the coordinates of the DNA atoms were fixed. The energy of the complex was initially minimized using 100 steps of a steepest descents algorithm and 3000 steps of conjugate gradient minimization. The minimized starting model was then subjected to 9 psec of restrained molecular dynamics (rmd) at 300 K, followed by 3 psec of rmd at 200 K, 3 psec of rmd at 150 K, and finally 5 psec of rmd at 100 K. The model was then energy-minimized by conjugate gradient minimization to a final rms derivative of <0.05 (kcal/mole)/Å².

Second, the fix on the DNA atoms was released. The final model from the first step was then treated to the same preliminary energy minimization, dynamics, and final energy minimization protocols as shown above, with the exception that the complex underwent 9 psec of rmd at 200 K followed by 3 psec of rmd at 175 K, 3 psec of rmd at 150 K, 3 psec of rmd at 125 K, and finally 5 psec of rmd at 100 K.

2.3.6 Base Pair Opening Lifetime Measurements. The opening kinetics for the I•C base pairs in the oligomer [d(CGCGIICCCGCG)]₂ were measured using the methodology developed by Guéron and coworkers (for a complete description see Leroy et al., 1988). A sample of ~1.5 μmoles HPLC purified and desalted dodecamer was used. The NMR sample was ~3 mM duplex, 100 mM NaCl, 0.2 μM EDTA in 90% H₂O/10% D₂O in a total volume of 500 μl. The pH of the sample was adjusted to 8.8.

NH₃ was used as the imino proton exchange catalyst in these experiments. An NH₃/NH₄⁺ buffer solution was made by mixing 1 ml saturated NH₄Cl solution (30% w/v) with 60 μl concentrated aqueous ammonia solution (28.5% w/v), and the pH was adjusted to 8.8 using 0.1 M HCl. Using the pK_a of ammonia (9.2), the concentration of NH₃ in the

buffer was calculated to be 1.7 M. This buffer solution was titrated into the NMR sample and T_1 measurements of the inosine imino protons were taken after each buffer addition. T_1 measurements were done at concentrations of 0, 20, 25, 40, 50, 75, 100, 200, and 500 mM NH_3 in the NMR sample.

T_1 measurements were acquired on a 500 MHz GE Omega NMR spectrometer, using an inversion-recovery pulse sequence in which selective inversion of the imino protons is accomplished by a DANTE pulse train (Morris & Freeman, 1978), and the read pulse is a standard $1(x)3(-x)3(x)1(-x)$ sequence (Hore, 1983). Thirteen t_1 points were sampled after each addition of $\text{NH}_3/\text{NH}_4^+$ buffer, and each exponential decay was fit using the lsqfit routine on the NMR spectrometer. The T_1 data were converted into exchange lifetimes τ_{ex} using the equation (Leroy et al., 1988):

$$\frac{1}{\tau_{\text{ex}}} = \frac{1}{T_{1(\text{cat})}} - \frac{1}{T_{1(\text{no cat})}} \quad (\text{Equation 2-2})$$

where $T_{1(\text{no cat})}$ is the measured T_1 lifetime of an imino proton with no added catalyst and $T_{1(\text{cat})}$ is the measured T_1 lifetime at some point during the ammonia titration. The exchange lifetimes τ_{ex} were plotted against inverse NH_3 concentration. The opening lifetime of each base pair is determined by extrapolating the exchange lifetime τ_{ex} to infinite NH_3 concentration.

2.3.7 Job's Plots. The binding stoichiometry of distamycin to the symmetric DNA oligomer $[\text{d}(\text{CGCTTTAAAGCG})]_2$ was determined by UV-Vis absorption spectra taken using Job's method of continuous variations (Job, 1928). Absorption spectra were recorded with a Hewlett-Packard model 8452A diode array UV-Vis Spectrophotometer. All measurements were taken at ambient temperature. A 1 cm path length cell was used.

The following three stock solutions were made in 8 mL volumes: a blank solution of 0.1 M Tris HCl (pH 7.0) and 0.2 M NaCl, and DNA and distamycin stock solutions which contained 40 μM duplex and 40 μM distamycin, respectively, and were also 0.1 M in Tris HCl (pH 7.0) and 0.2 M in NaCl. The blank solution was used to blank the

spectrophotometer at the beginning of the experiment. The DNA and distamycin stock solutions were mixed in various ratios to yield test solutions, with the total volume of each test solution equaling 1 mL so that for each, [DNA duplex] + [distamycin] = 40 μ M. Thirteen ratios were used ranging from 100% DNA solution to 100% distamycin solution. Absorbance was monitored at 260, 302, 310, 314, and 320 nm.

The change in absorbance at a given wavelength of the ligand and the DNA due to complex formation was calculated according to the following equation:

$$\Delta A = A_{(\text{mixture})} - [\chi_{\text{dist}} * A_{(100\% \text{ dist solution})}] - [\chi_{\text{DNA}} * A_{(100\% \text{ DNA solution})}]$$

(Equation 2-3)

where $A_{(\text{mixture})}$ = measured absorbance of one of the mixtures of ligand and DNA at a given wavelength, $A_{(100\% \text{ dist solution})}$ = measured absorbance of the 40 μ M distamycin solution at the same given wavelength, $A_{(100\% \text{ DNA solution})}$ = measured absorbance of the 40 μ M DNA duplex solution at the same given wavelength, χ_{dist} = mole fraction of distamycin in the mixture, and χ_{DNA} = mole fraction of DNA duplex in the mixture. ΔA was plotted against the [distamycin]/[DNA duplex] ratio. The largest ΔA value at a given wavelength occurs at the ratio where all of the ligand and DNA is complexed, yielding the stoichiometry of the complex.

The experiment was repeated at 20 μ M concentration.

2.4 Distamycin and Netropsin Binding to Four Base Pair Sites

2.4.1 Distamycin Binding to AAAA:TTTT. The binding site AATT, which has been extensively studied by NMR and crystallography in complex with distamycin, is symmetric. Consequently, I performed an NMR study to examine the orientational preference of distamycin binding to the asymmetric A,T-rich sequence AAAA:TTTT. A 2D NOESY was acquired of a sample of the decamer d(CGCAAAGGC): d(GCCTTTTGCG) dissolved in H₂O solution. The base aromatic proton resonances and sugar H1', H2' and

Table 2-1. Chemical shift assignments of the d(CGCAAAGGC):d(GCCTTTGCG) duplex, free and in the 1:1 distamycin:DNA complex.^a

	H6/H8			H1'			H2/CH ₃		
	Free	Comp	$\Delta\delta$	Free	Comp	$\Delta\delta$	Free	Comp	$\Delta\delta$
Strand 1									
C ₁	7.62	7.56	-0.06	5.73	5.74	+0.01			
G ₂	7.93	7.96	+0.03	5.86	5.83	-0.03			
C ₃	7.31	7.20	-0.11	5.42	5.61	+0.19			
A ₄	8.19	8.36	+0.17	5.74	5.12	-0.62	7.17	7.47	+0.30
A ₅	8.08	8.05	-0.03	5.75	5.28	-0.47	7.00	7.69	+0.69
A ₆	7.99	7.86	-0.13	5.81	5.50	-0.31	7.03	7.67	+0.64
A ₇	7.87	7.73	-0.14	5.89	5.32	-0.57	7.43	7.63	+0.20
G ₈	7.39	7.42	+0.03	5.52	5.40	-0.12			
G ₉	7.56	7.62	+0.06	5.92	5.93	+0.01			
C ₁₀	7.36	7.36	0.00	6.15	6.19	+0.04			
Strand 2									
G ₁₁	7.95	7.95	0.00	5.97	5.99	+0.02			
C ₁₂	7.51	7.54	+0.03	6.08	6.13	+0.05			
C ₁₃	7.59	7.66	+0.07	6.00	6.02	+0.02			
T ₁₄	7.49	7.59	+0.10	6.09	5.90	-0.19	1.64	1.75	+0.11
T ₁₅	7.49	7.55	+0.06	6.14	6.13	-0.01	1.66	1.69	+0.03
T ₁₆	7.48	7.28	-0.20	6.10	5.49	-0.61	1.67	1.61	-0.06
T ₁₇	7.30	7.04	-0.26	5.81	5.08	-0.73	1.69	1.65	-0.04
G ₁₈	7.90	7.78	-0.12	5.79	5.57	-0.22			
G ₁₉	7.33	7.27	-0.06	5.75	5.70	-0.05			
C ₂₀	7.92	7.94	+0.02	6.12	6.17	+0.05			

^a Chemical shifts are given in ppm (± 0.01 ppm). The residual HOD resonance is referenced to 4.85 ppm (20 °C) for the free DNA and for the complexed DNA.

H2" resonances were assigned by standard methods (see Table 2-1). Distamycin was then titrated into the sample, with a 1D ^1H NMR spectrum acquired after each addition as shown in Figure 2-2. The ligand binds in slow exchange on the NMR time scale with no evidence of line broadening, indicating high affinity binding as observed for AATT. This result is consistent with a recent report in which the order of discrimination of distamycin for six different four A,T base pair sites was catalogued. Distamycin was found to bind the sites in order of decreasing affinity: AAAA=AATT > TAAT=ATTA > TTAA=TATA > ATAT, with the order varying slightly depending upon the flanking sequence (Abu-Daya et al., 1995). Additionally, only one dominant form of complex appears in the NMR titration of AAAA, as indicated by the appearance of only one set of complex resonances (Figure 2-2a).

2D NOESY spectra of the 1:1 complex of distamycin A with AAAA were acquired (Figure 2-3) to identify intermolecular contacts between the ligand and the DNA. Contacts were observed from the ligand pyrrole H3 and formyl resonances to adenosine H2 and sugar H1' resonances (Figure 2-4). These contacts establish that the ligand is bound with the formyl end of the drug in contact with the adenosine at the 5' end of the A-tract, and with the tripyrrole unit lying along the minor groove of the A-tract. With the sequence AAAT:ATTT, the predominant binding mode seems to be the same (Rajagopal, 1988). An NMR-restrained model of the distamycin:AAAA complex is shown in Figure 2-5. The pattern of NOEs between the ligand and the sugars changes along the length of the ligand, and is somewhat different from that seen in AATT, suggesting that the ligand is positioned differently relative to the DNA. This could arise from a difference in the shape of the two DNA sites, which is not unexpected since AATT is symmetric while AAAA is not.

At the lower right of Figure 2-3 a box highlights several weak cross peaks which have been determined by 2D ROESY (rotating Overhauser effect spectroscopy) experiments to arise from chemical exchange during the mixing time of the NOESY. These peaks align with the resonances from the ligand pyrrole H3 resonances of the major form

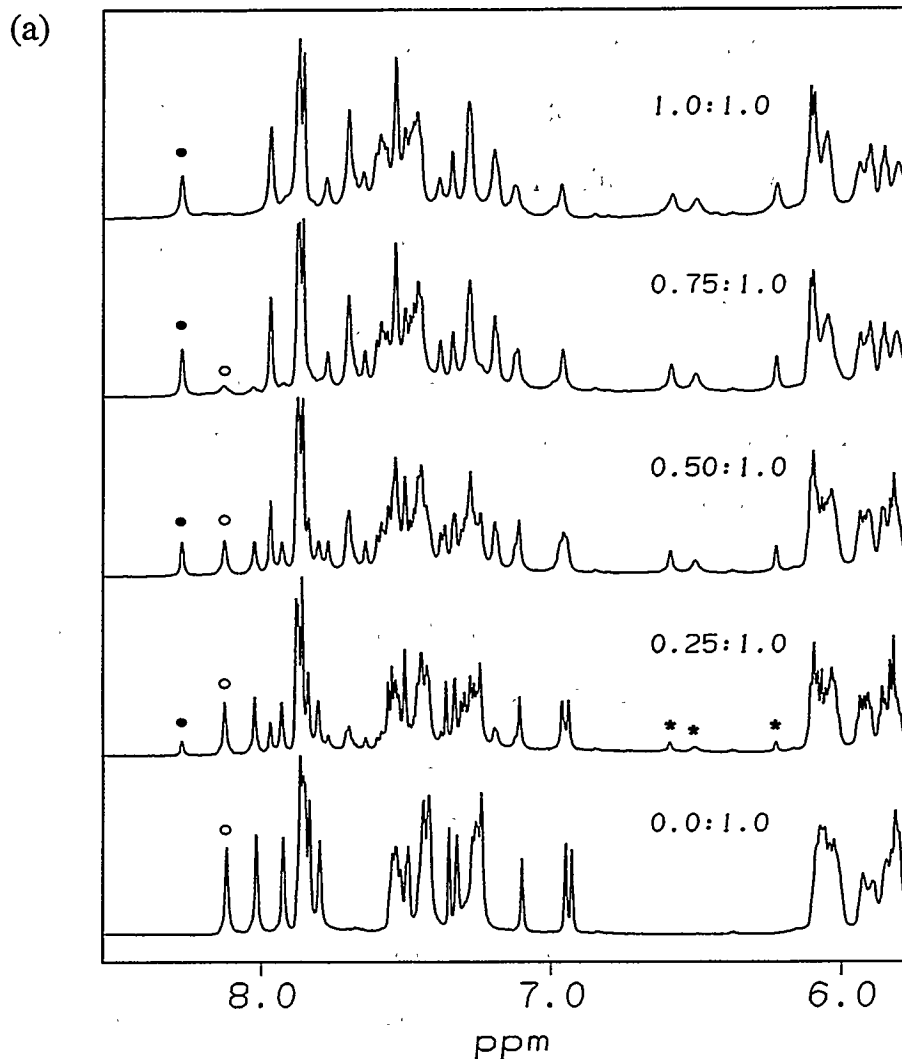
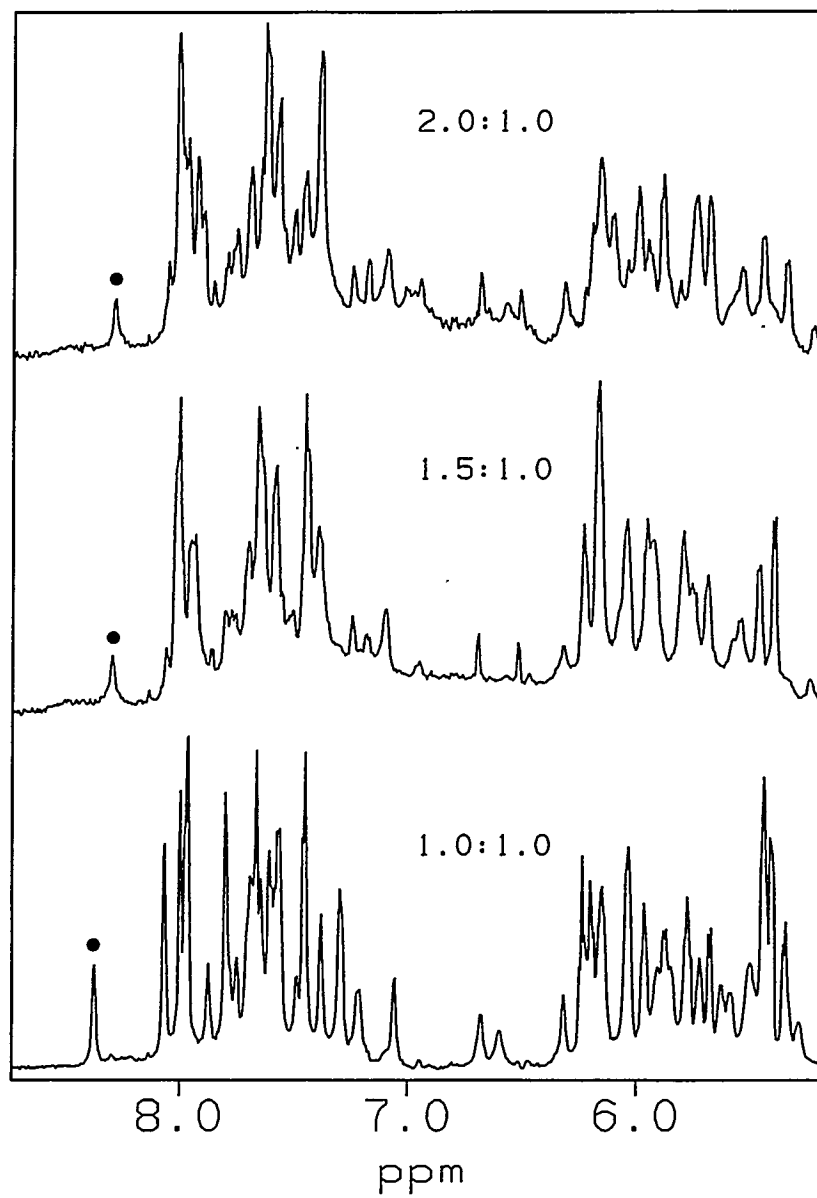


Figure 2-2. The downfield region of 1D ^1H NMR spectra acquired during the titration of d(CGCAAAGGC):d(GCCTTTTGCG) with distamycin (in D_2O , 20 $^\circ\text{C}$, 6024 Hz spectral width, 600 MHz, sample conditions 6.0 mM duplex DNA, 10 mM PO_4 , pH 7.0). The approximate distamycin:DNA stoichiometry is indicated for each spectrum. Open circles denote an adenosine H8 proton in the uncomplexed DNA, and filled circles denote an adenosine H8 proton in the 1:1 distamycin:DNA complex. Asterisks denote ligand pyrrole H3 protons in the complex. (a) titration to 1:1 ligand:DNA stoichiometry, (b) titration to 2:1 ligand:DNA stoichiometry.

(b)



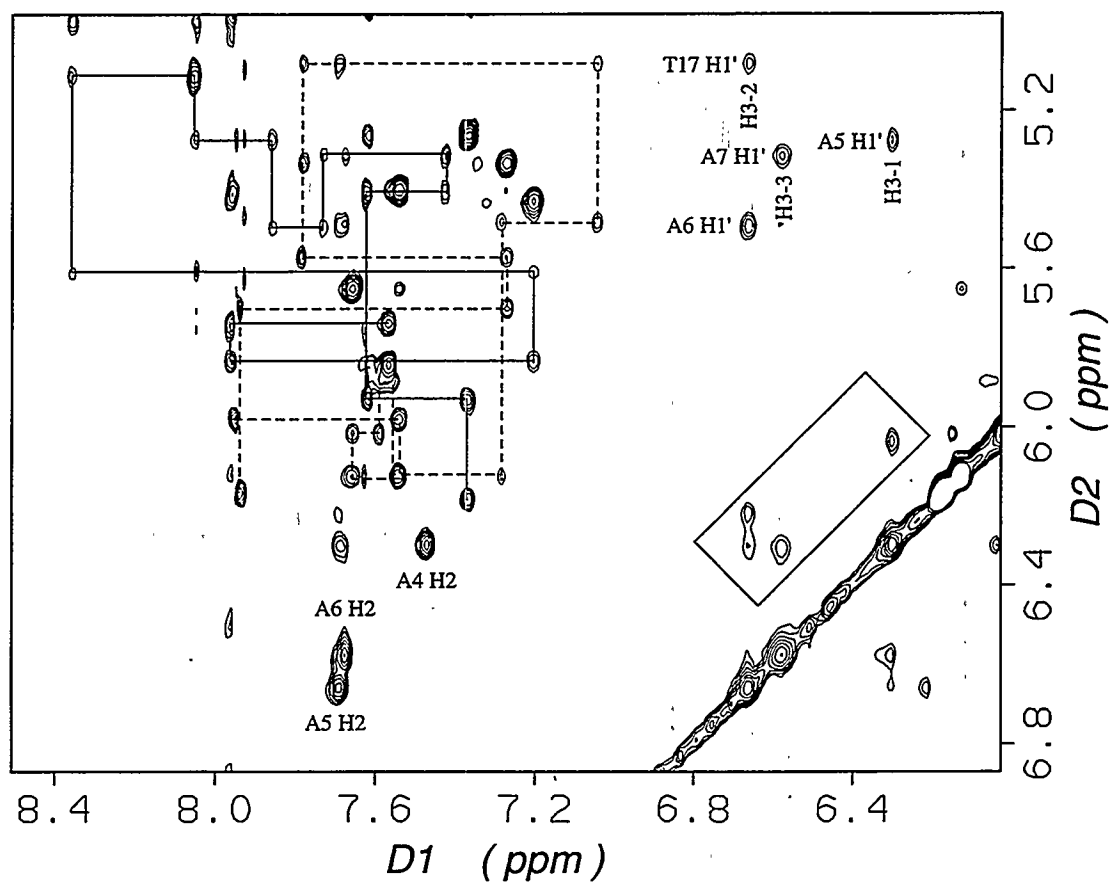


Figure 2-3. Expansion of the base aromatic and sugar H1' proton region of a 2D NOESY spectrum of the 1:1 complex of distamycin with d(CGCAAAGGC):d(GCCTTTTGCG) (in D₂O, 20 °C, 200 ms mixing time, 600 MHz). Sequential aromatic to H1' proton connectivities of the A-rich and the T-rich strands are denoted by solid and broken lines, respectively. Intermolecular ligand-DNA cross peaks are indicated. Labels above or below a cross peak denote the assignment along the D1 axis, and labels to the left or right of a cross peak denote the assignment along the D2 axis.

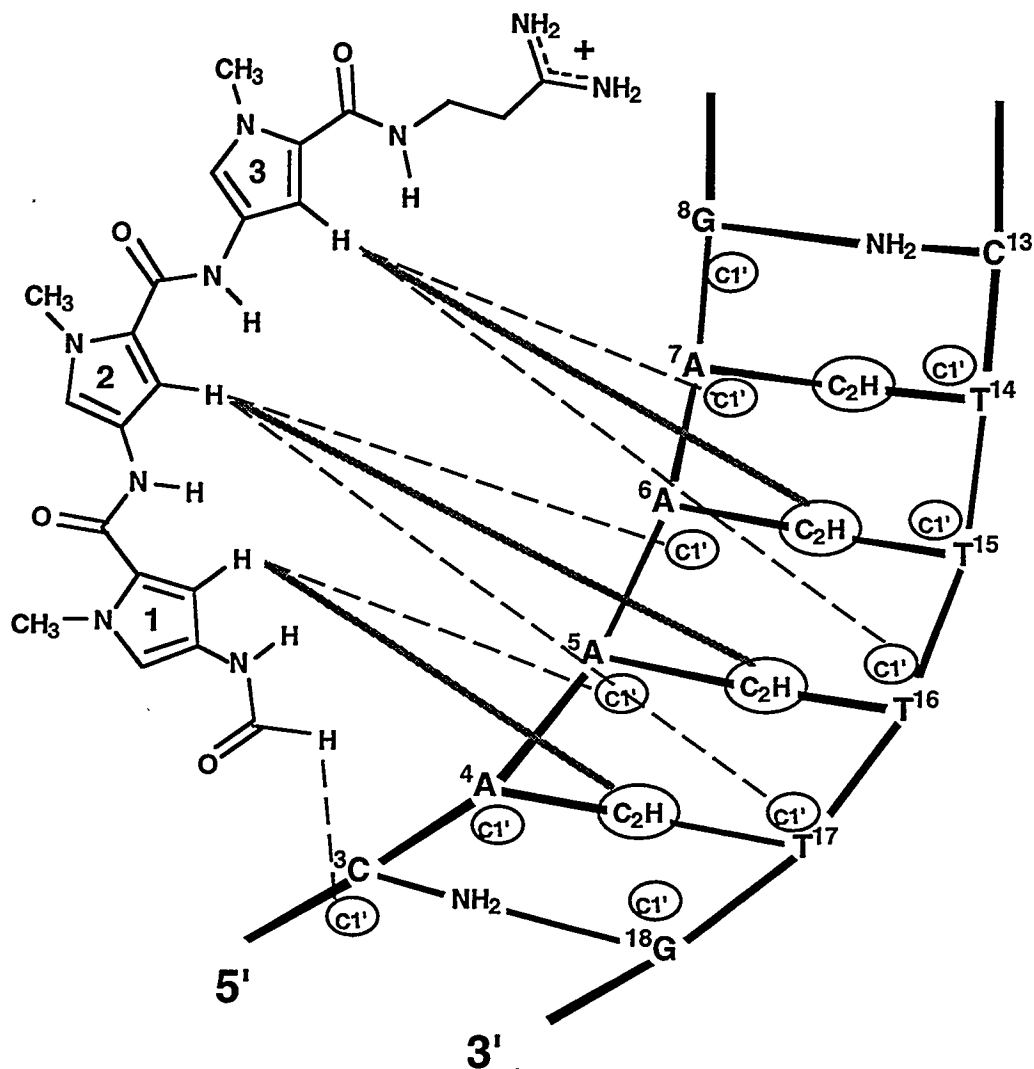


Figure 2-4. Summary of intermolecular ligand-DNA contacts observed in the 2D NOESY spectrum in D₂O of the 1:1 complex of distamycin with d(CGCAAAAGGC): d(GCCTTTTGCG).

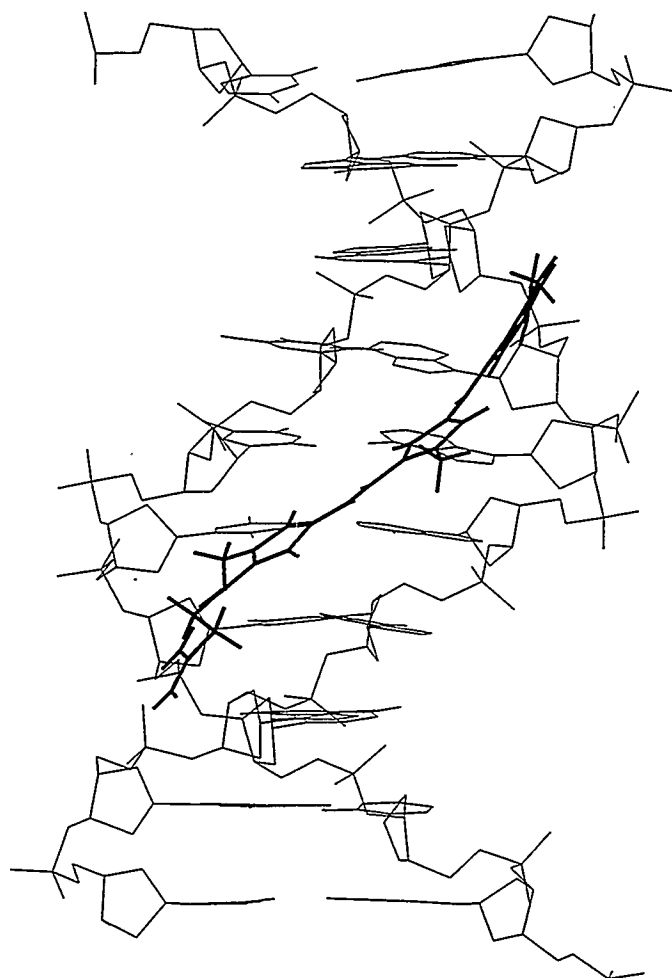


Figure 2-5. An example of a 1:1 distamycin:DNA complex. A molecular model of the 1:1 distamycin:DNA complex with d(CGCAAAAGGC): d(GCCTTTTGCG) obtained by energy minimization using semiquantitative distance restraints derived from NOESY data. For clarity, hydrogen atoms are omitted for the DNA but not for the ligand molecule.

of complex along on axis, and correspond to very weak peaks in the 1D spectrum along the other axis. These weak peaks probably arise from binding of the drug in the opposite orientation with a reduced affinity, not more than about 5% of the major form based on their relative intensities in the 1D spectrum. The equilibrium between the two complexes formed is shown in Figure 2-6. This orientational preference may be due to electrostatic effects which orient the charged end of the ligand, or in part due to van der Waals interactions since there is evidence that in A-tracts the minor groove (in uncomplexed DNA) is compressed from 5' to 3' along the A-rich strand (Burkhoff & Tullius, 1987; Nadeau & Crothers, 1989; Katahira et al., 1990).

Titration were performed which went to higher drug:DNA stoichiometries (Figure 2-2b). However, as in the case of AATT, no evidence was seen for 2:1 complex formation with this DNA sequence. The observed shifting of the DNA and ligand resonances is probably due to nonspecific binding which is in fast exchange on the NMR time scale. Because no 2:1 complex was seen to form with the related site TATA, it is likely that four A,T base pair sites are too short to accommodate a side-by-side 2:1 complex of the type observed for the five base pair site AAATT:AATTT.

2.4.2 Netropsin Binding to AAAA:TTTT. To compare the binding properties of netropsin (see Figure 2-1 for numbering scheme) with those of distamycin on an asymmetric DNA sequence, A. R. Leheny titrated netropsin into a sample of the same DNA sequence as above, d(CGCAAAGGC): d(GCCTTTTGCG). The NMR titration is shown in Figure 2-7. Similar to distamycin, netropsin binds to this sequence in a 1:1 stoichiometry and in slow exchange on the NMR time scale. However, with increasing concentrations of added netropsin, the disappearance of the free DNA resonances is concurrent with the appearance of *two* sets of complexed DNA resonances with approximately equal intensities. Refer to Figure 2-7, where there are four thymine imino resonances in the uncomplexed DNA (13.8 to 14.3 ppm) and eight in the complexed DNA

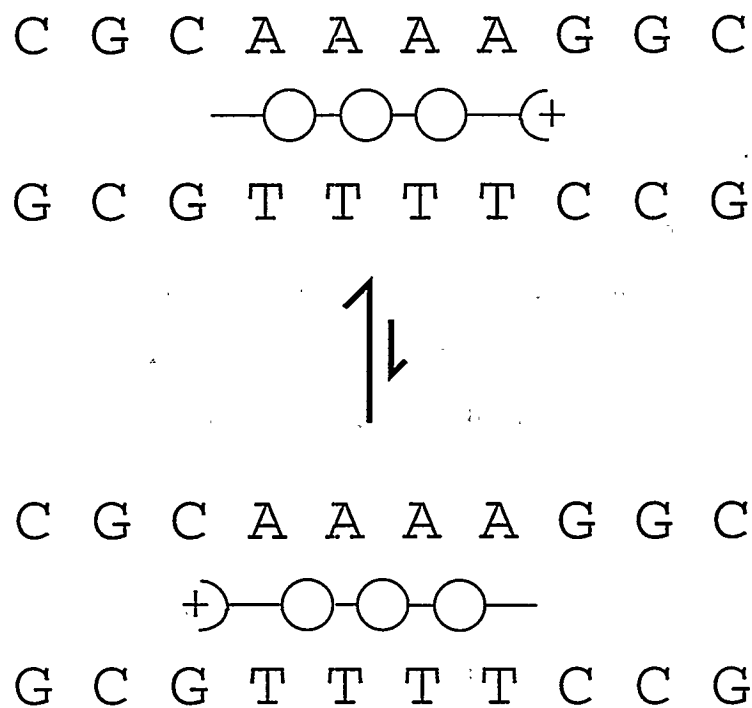


Figure 2-6. Schematic representation of the flip-flop exchange undergone by distamycin in complex with d(CGCAAAAGGC):d(GCCTTTTGGC).

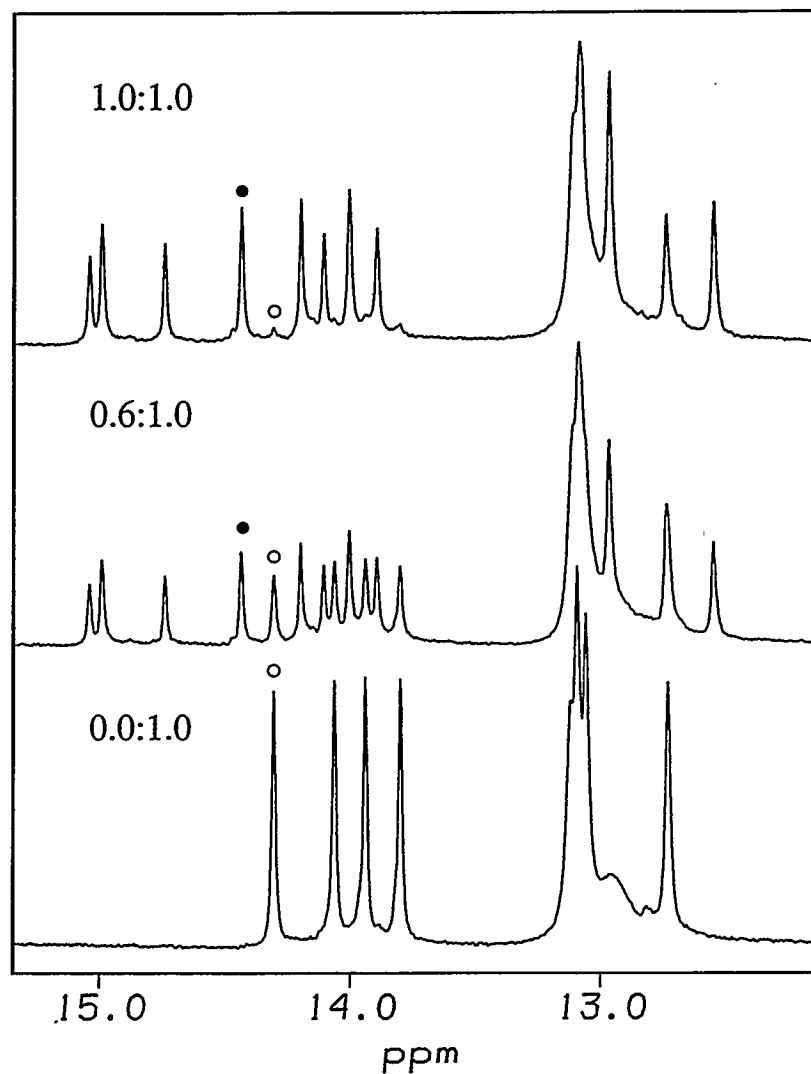


Figure 2-7. The imino proton region of 1D ^1H NMR spectra acquired during the titration of d(CGCAAAGGC):d(GCCTTTTGCG) with netropsin (in H_2O , 10°C , 13514 Hz spectral width, 600 MHz, sample conditions 2.0 mM duplex DNA, 10 mM PO_4 , pH 7). The approximate netropsin:DNA stoichiometry is indicated for each spectrum. Open circles denote a thymine H3 proton in the uncomplexed DNA, and filled circles denote a thymine H3 proton in one of the 1:1 netropsin:DNA complexes. Note that there is a small amount of uncomplexed DNA at 1:1 stoichiometry.

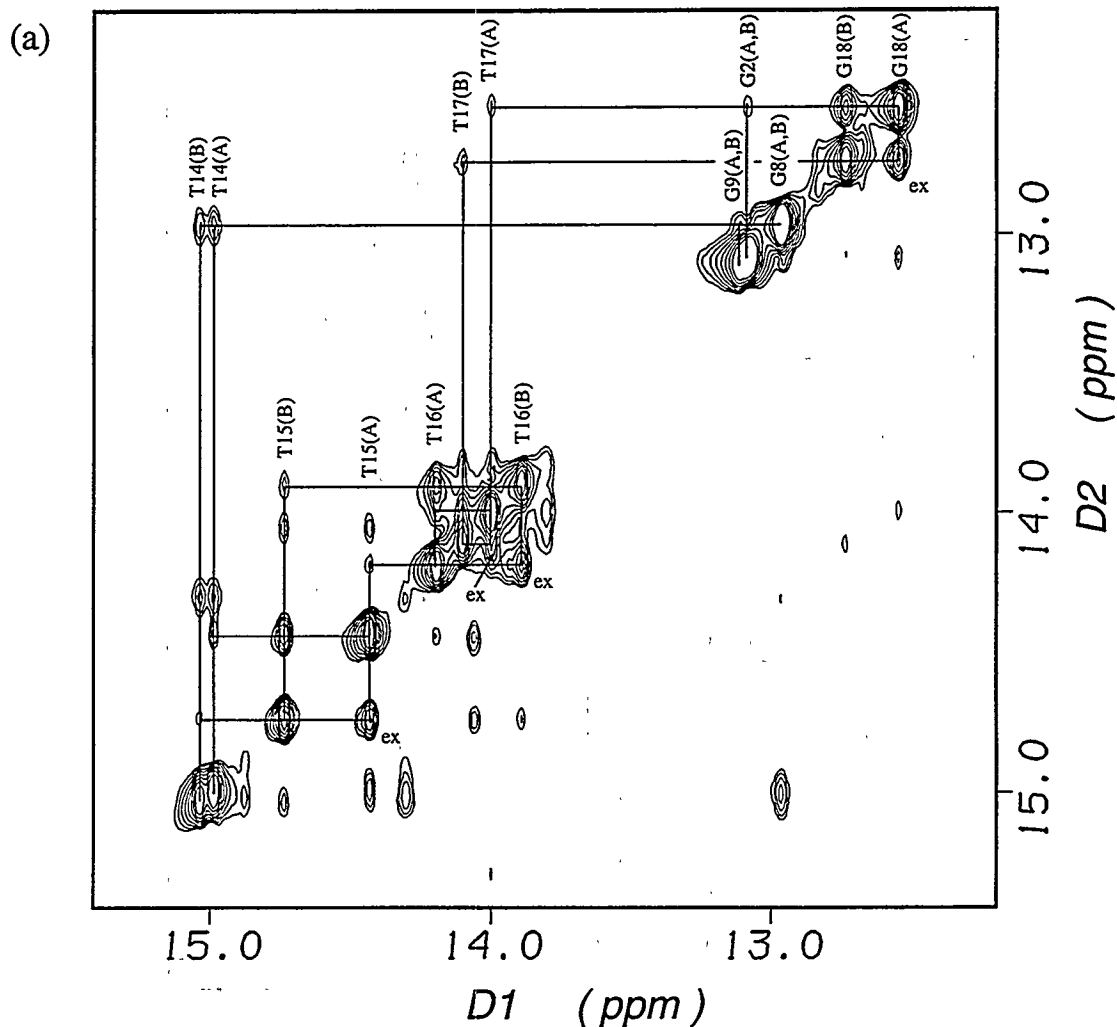
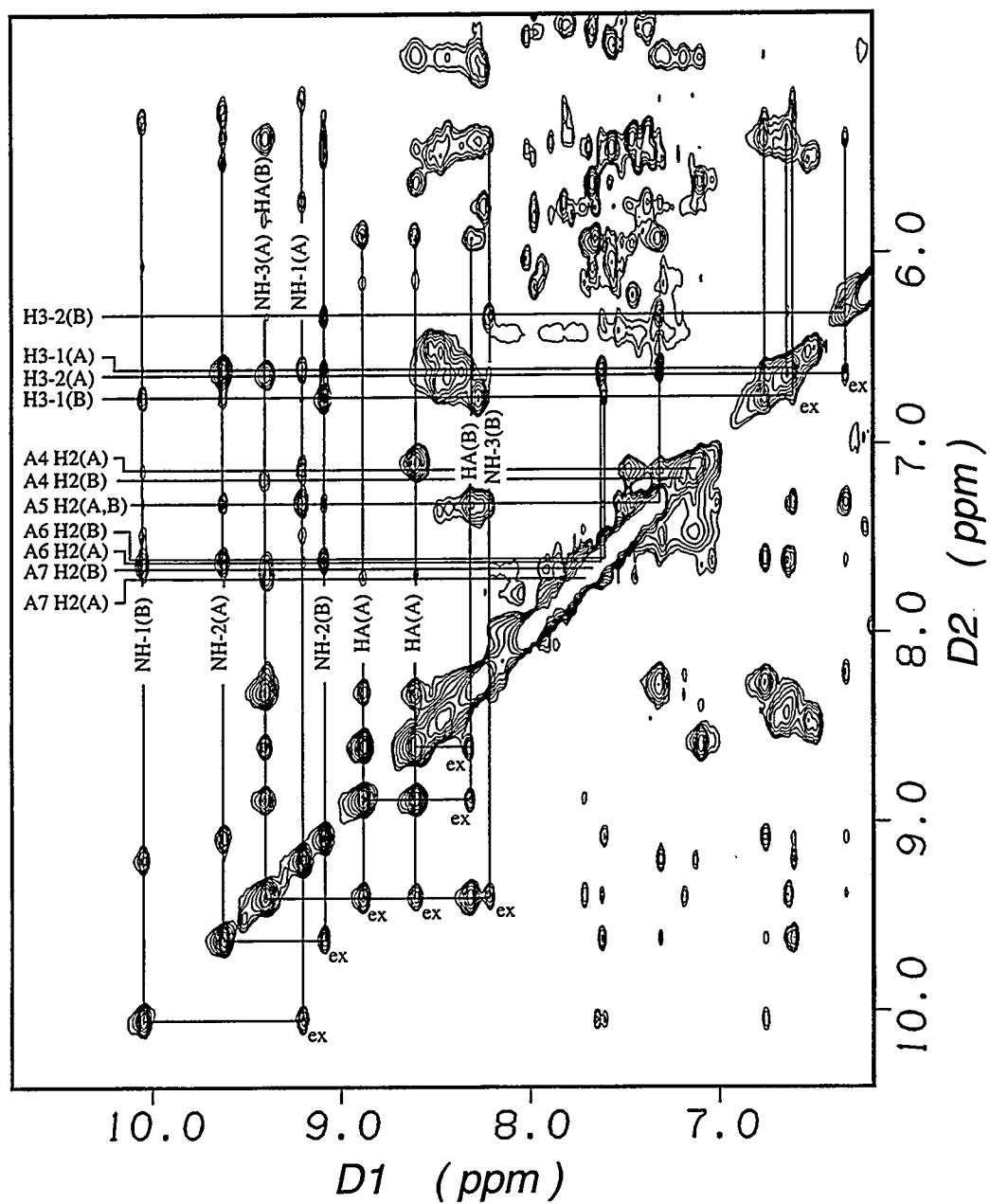


Figure 2-8. Two-dimensional NOESY spectrum of a mixture of two 1:1 netropsin:AAAA complexes (in H_2O solution, 10°C , $\tau_M = 200$ ms, 13514 Hz spectral width, 600 MHz, sample conditions 2.0 mM duplex DNA, 10 mM PO_4 , pH 7). The two complexes present are denoted by the labels A and B, where A is the slightly favored form. Resonance assignments are labeled above the diagonal. Exchange cross peaks are labeled below the diagonal. (a) Downfield imino-imino proton region. Extra weak cross peaks are due to the presence of a small amount of uncomplexed DNA. (b) Amide-aromatic proton region.

(b)



(13.8 to 15.1 ppm), indicating that netropsin forms two different 1:1 complexes with the AAAA site.

Based on 2D NOESY spectra of this mixture of complexes in D₂O and H₂O (Figure 2-8) taken at two different temperatures, I was able to determine the nature of the two complexes present. The spectra are quite complicated due to the presence of two sets of complexed DNA resonances, two sets of complexed ligand resonances, a set of exchange cross peaks, and many weaker cross peaks due to exchange-transferred NOE. The two sets of adenine H2 resonances were assigned, with the DNA H8, H6 and H1' protons only partially assigned with the help of comparison to the 1:1 distamycin:AAAA complex. Both sets of complexed ligand amide, pyrrole H3, and amidinium proton resonances were assigned. The ligand NH-4 and guanidinium protons were not assigned. Intermolecular contacts are observed between ligand amide proton and pyrrole H3 resonances and DNA sugar H1' and adenosine H2 resonances (Table 2-2).

Two distinct 1:1 ligand:DNA complexes were identified (denoted A and B) which are related by a flip-flop rearrangement of the ligand on the DNA (as in Figure 2-6). Other possible complexes, such as two complexes related by a sliding motion, were considered but not found. The sets of resonances from the two complexes have approximately equal intensities and linewidths, indicating the two complexes are of nearly equal affinity. However, based on integration of the thymine imino resonances in the two different complexes, complex A is very slightly favored over complex B by a factor between 1.0 and 1.5. In complex A netropsin binds in the orientation analogous to the major form of distamycin complexed with AAAA, with the guanidinium group at the 5' end of the A-tract and the amidinium group at the 3' end of the A-tract. The chemical shifts of the DNA protons are fairly similar in the two complexes. However, the chemical shifts of the ligand protons, particularly the amide protons, vary more between the two complexes (Figure 2-8). The differences in NH-1 (+0.85 ppm), NH-2 (-0.53 ppm), and NH-3 (-1.18 ppm)

Table 2-2. Ligand-DNA intermolecular contacts for two observed netropsin complexes with d(CGCAAAGGC): d(GCCTTTGCG).^a

Complex A		Complex B	
netropsin	DNA	netropsin	DNA
NH-1	A ₄ H2	NH-1	A ₇ H2
NH-1	A ₅ H2	NH-1	A ₆ H2
H3-1	A ₅ H2	H3-1	A ₆ H2
NH-2	A ₅ H2	NH-2	A ₆ H2
NH-2	A ₆ H2	NH-2	A ₅ H2
H3-2	A ₆ H2	H3-2	A ₅ H2
NH-3	A ₆ H2	NH-3	A ₅ H2
NH-3	A ₇ H2		

^a Observed in a NOESY spectrum acquired in H₂O solution at 10 °C, with $\tau_M = 200$ ms. Several contacts from ligand amide and pyrrole H3 protons to DNA H1' protons are not listed here due to difficulty in assigning the two sets of H1' resonances. See Table 2-1 for DNA numbering scheme.

chemical shifts indicate that netropsin may be positioned differently relative to the DNA from complex A to B, yielding essentially the same binding affinity in either case.

Therefore, it appears that netropsin exhibits almost no orientational preference with the AAAA site in contrast to distamycin, perhaps because of the presence of a second positive charge on netropsin. The result that netropsin does not distinguish well between the two binding orientations seems reasonable since netropsin is pseudosymmetric. Work is under way by M. Ban in the Wemmer lab to synthesize analogs of distamycin which have modified ends to further explore the contributions of the end groups to the orientational preference and affinity of distamycin binding. This work will have important consequences in understanding the determinants of ligand sequence selectivity. The question of orientational selectivity in ligand binding will be addressed further in Chapter 3, which describes studies of a modified netropsin binding to the AAAA site. This modified netropsin has only one positive charge like distamycin and displays some orientational binding preference (more than netropsin), although less than the 20-fold preference displayed by distamycin.

2.4.3 Distamycin Binding to IICC. I continued work begun by J. G. Pelton in examining the binding of distamycin to $[d(CGCGIICCCGCG)]_2$, where I represents deoxyinosine (2-desaminodeoxyguanosine). This oligomer is a minor groove analog of $[d(CGCGAATTCGCG)]_2$. An I•C base pair (Figure 2-9) has minor groove functional groups which are completely equivalent to an A•T pair, although the major groove functional groups and stacking interactions are more equivalent to G•C pairs. Since the functional groups available for minor groove recognition are equivalent in A•T and I•C base pairs, any difference in distamycin binding should arise from structural rather than functional differences. In particular, if the groove width is similar to that typical of G,C-rich regions (6-7 Å), rather than the narrow value seen in A,T-rich regions (3-4 Å), changes in drug affinity may be expected. The electrostatic potential of I•C pairs is

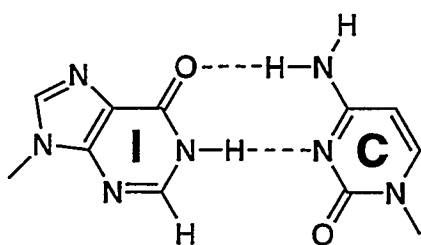
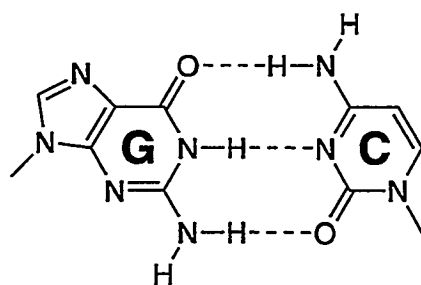
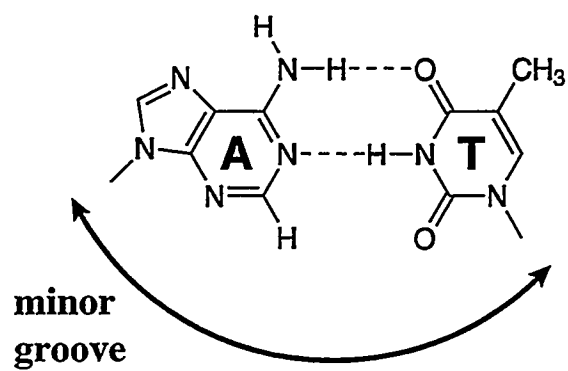


Figure 2-9. Adenine•thymine, guanine•cytosine, and inosine•cytosine base pairs, where inosine = 2-desamino-guanine.

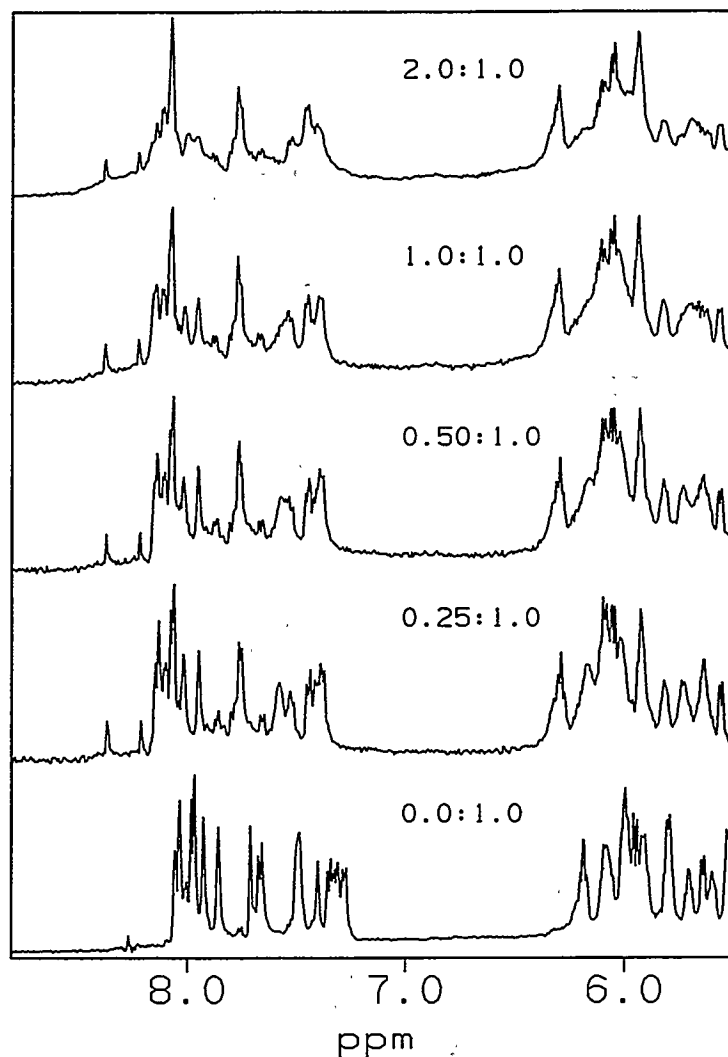


Figure 2-10. The downfield region of 1D ^1H NMR spectra acquired during the titration of $[\text{d}(\text{CGCGIICCCGCG})]_2$ with distamycin (in D_2O , 5000 Hz spectral width, 500 MHz, sample conditions 0.8 mM duplex DNA, 10 mM PO_4 , pH 7.0). The approximate distamycin:DNA stoichiometry is indicated for each spectrum. All spectra were acquired at 35 °C except for the first titration point (0.0:1.0), which was taken at 25 °C. The small set of sharp resonances which appear most notably downfield at approximately 8.2 and 8.3 ppm are due to hairpin which is present at higher temperatures.

probably much closer to that of G•C than A•T, which could contribute to differences in binding affinity.

1D NMR titrations of IICC show that distamycin binds to this site weakly and in intermediate exchange (Figure 2-10), similar to the behavior observed with TATA. This suggests that the free DNA has an intrinsically wider minor groove than AATT, in which a single ligand molecule would make contact with only one side, lowering its binding affinity. It also indicates that there is a substantial energetic cost to reducing the groove width in order to reestablish contact on both sides of the ligand. Although the groove may be wide enough to accommodate two ligands side-by-side, the absence of 2:1 complex formation strongly suggests that a four base pair site is too short to accommodate a 2:1 complex.

2.4.4 Base Pair Opening Kinetics for IICC. The results of the distamycin binding experiment to the oligomer [d(CGCGIICCCGCG)]₂ suggest that the structure of the IICC site is different than AATT. A-tracts in DNA (for a description of A-tracts, see Chapter 1), among their many distinct characteristics, have anomalously long base pair opening lifetimes (Guéron et al., 1990b). A•T and I•C pair opening lifetimes have been compared in the oligomers d(CGCAAAAAAGCG):d(CGCTTTTTTGCG) versus d(CGCHHHHIGCG): d(CGCCCCCCCCGCG). The I•C pairs exhibit 10-fold to 50-fold shorter opening lifetimes when compared to their corresponding A•T pairs, with the largest difference in lifetimes observed when comparing base pairs in the center of the tract. In comparing the oligomers d(CGCAAAAAGCG): d(CGCTTTTTTGCG) versus d(CGCAAIAAGCG): d(CGCTTCTTGCG) (Guéron et al., 1990b), Guéron and coworkers observed that the I•C pair has a lifetime half that of the corresponding A•T pair (Guéron et al., 1990b).

Because no comparison had been made of the base pair opening kinetics of this oligomer and its analog [d(CGCGAATTCGCG)]₂, I measured the opening lifetimes for the I•C base pairs in [d(CGCGIICCCGCG)]₂. Guéron and coworkers have developed a

relatively simple NMR technique to measure the exchange rate of DNA imino protons with solvent, and from these rates extrapolate individual base pair opening lifetimes (Leroy et al., 1988). I determined lifetimes of $7(\pm 5)$ ms for the $I_5 \cdot C_8$ pair and $9(\pm 5)$ ms for the central $I_6 \cdot C_7$ pair (Figure 2-11). The opening lifetimes have been measured for the A•T base pairs in the analog $[d(CGCGAATTCGCG)]_2$ under similar conditions (15 °C, pH 8.8, 150 mM NaCl, 0.5 mM EDTA, using NH_3 catalyst), and were found to be 8 ms for the $A_5 \cdot T_8$ pair and 26 ms for the $A_6 \cdot T_7$ pair (Leijon & Gräslund, 1992). The decreased relative lifetimes for the two central I•C pairs in the IICC site may be due in part to the different stacking propensities for A versus I. The stacking interactions of inosine are likely to be similar to guanine, and data suggests that guanine stacks more weakly than adenine (Shum & Crothers, 1983). In any case, it is clear that distamycin binds with high affinity to the sites AAAA and AATT which both exhibit anomalously long base pair lifetimes (Guéron et al., 1990a), and lower affinity to the sites IICC, which has more typical opening lifetimes, and TATA. Opening lifetimes of the site TATA have not been measured, but based on the observations of Guéron et al. on other sites with alternating A•T base pairs, it is expected that this site will have relatively short opening lifetimes as well.

These studies also have implications for observation of bound water in the minor groove of DNA by NMR (Kubinec & Wemmer, 1992; Liepinsh et al., 1992; Liepinsh et al., 1994). The observation of direct NOEs from minor groove DNA protons (adenine or inosine H2) to solvent water depends upon relatively long base pair opening lifetimes relative to the mixing times used (30 to 60 ms). This is necessary to minimize the exchange transferred NOE pathway which can occur via the imino proton (adenine/inosine H2 $\xleftrightarrow{\text{NOE}}$ thymine H3/inosine H1 $\xleftrightarrow{\text{exchange}}$ solvent H_2O), thereby maximizing the efficiency of magnetization transfer through the direct NOE (for a complete description see Otting et al., 1991). This implies that measurement of water bound in the minor groove of DNA by currently used techniques may be limited to sequences which have the desired

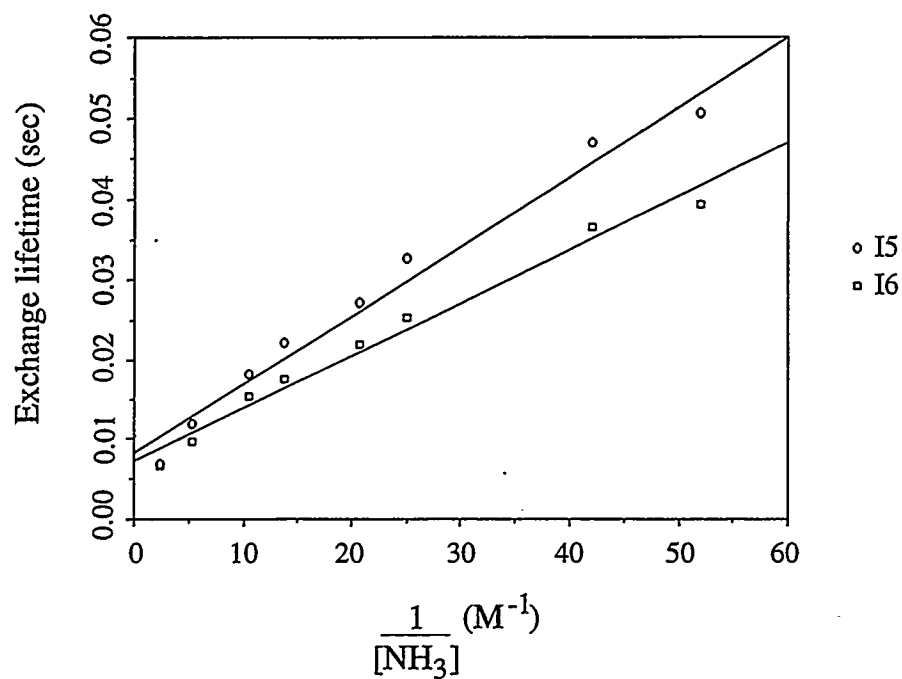


Figure 2-11. Fit of data for I₅ and I₆ imino exchange lifetimes in the duplex [d(CGCGIIICCGCG)]₂ at various ammonia concentrations (measured at 15 °C, in 100 mM NaCl, pH 8.8). By extrapolating to infinite ammonia concentration, the y-intercept yields the overall base pair opening lifetime.

base pair opening kinetics. Bound water has been observed in the minor groove of the AATT site in [d(CGCGAATTCGCG)]₂ (Kubinec & Wemmer, 1992; Liepinsh et al., 1992), and the absence of bound water has been observed in the TTAA site in [d(GTGGTTAACCAC)]₂ (Liepinsh et al., 1994). It was not possible to measure unambiguously the presence or absence of bound water in the minor groove of [d(CGCGIICCCGCG)]₂ because conditions could not be found which slowed the rapid exchange rate of the inosine imino protons.

2.5 Distamycin Binding to Five and Six Base Pair Sites

2.5.1 Distamycin Binding to IIICC:IICCC. J. G. Pelton studied the binding of distamycin to the sequence d(CGCAAATTGGC):d(GCCAATTTGCG) (Pelton & Wemmer, 1989; Pelton & Wemmer, 1990b). He observed that distamycin binds to this site both in a 1:1 ligand:DNA mode and in a 2:1 mode, as shown in equation 2-1. From calorimetry measurements, the binding constant of the second ligand (K_2) is 0.1 times that of the first ligand (K_1) (Rentzeperis et al., 1995).

To further investigate the determinants of distamycin binding and the structural features of I•C-containing DNA, I titrated distamycin into a sample of the analog d(CGCIICCGGC): d(GCCIICCCGCG). The NMR spectrum of free IIICC was assigned with standard 2D methods (Table 2-3). The titration of this sequence with distamycin is shown in Figure 2-12. There is no broadening upon addition of ligand, just growth of new resonances from complex with concomitant disappearance of the free DNA, up to a stoichiometry of 2:1.

A two-dimensional NOESY of the complex was acquired (Figure 2-13) and the spectrum was reassigned at this point, clearly indicating a side-by-side 2:1 complex (Figure 2-14, Table 2-3), completely analogous to the complex with AAATT. An NMR-restrained molecular model of the complex is shown in Figure 2-15. For AAATT several forms of 1:1 complex are observed at low levels of added ligand, with the 2:1 complex not appearing

Table 2-3. Chemical shift assignments of the d(CGCIICCGGC):d(GCCIIICCGCG) duplex, free and in the 2:1 distamycin:DNA complex.^a

	H6/H8			H1'			H2		
	Free	Comp	$\Delta\delta$	Free	Comp	$\Delta\delta$	Free	Comp	$\Delta\delta$
Strand 1									
C ₁	7.66	7.61	-0.05	5.78	5.73	-0.05			
G ₂	7.97	8.00	+0.03	5.91	5.89	-0.02			
C ₃	7.32	7.35	+0.03	5.48	5.84	+0.36			
I ₄	8.22	8.50	+0.28	5.90	5.58	-0.32	7.36	7.83	+0.47
I ₅	8.04	8.31	+0.27	5.90	5.77	-0.13	7.25	7.90	+0.65
I ₆	8.01	8.05	+0.04	6.06	5.30	-0.76	7.66	8.13	+0.47
C ₇	7.28	7.18	-0.10	5.97	5.54	-0.43			
C ₈	7.31	7.35	+0.04	5.78	5.24	-0.54			
G ₉	7.91	7.70	-0.21	5.54	5.59	+0.05			
G ₁₀	7.79	7.74	-0.05	6.00	5.90	-0.10			
C ₁₁	7.45	7.41	-0.04	6.19	6.18	-0.01			
Strand 2									
G ₁₂	7.99	7.96	-0.03	6.01	5.99	-0.02			
C ₁₃	7.50	7.53	+0.03	6.08	6.06	-0.02			
C ₁₄	7.47	7.46	-0.01	5.31	5.85	+0.54			
I ₁₅	8.27	8.51	+0.24	5.97	5.72	-0.25	7.44	7.80	+0.36
I ₁₆	8.11	8.41	+0.30	6.07	6.08	+0.01	7.65	8.22	+0.57
C ₁₇	7.32	7.17	-0.15	5.97	5.50	-0.47			
C ₁₈	7.49	7.39	-0.10	6.07	5.67	-0.40			
C ₁₉	7.39	7.44	+0.05	5.81	5.25	-0.56			
G ₂₀	7.95	7.78	-0.17	5.83	5.82	-0.01			
C ₂₁	7.36	7.35	-0.01	5.77	5.64	+0.13			
G ₂₂	7.96	7.93	-0.03	6.16	6.14	-0.02			

^a Chemical shifts are given in ppm (± 0.01 ppm). The residual HOD resonance is referenced to 4.85 ppm (20 °C) for the free DNA and to 4.91 ppm (15 °C) for the complexed DNA.

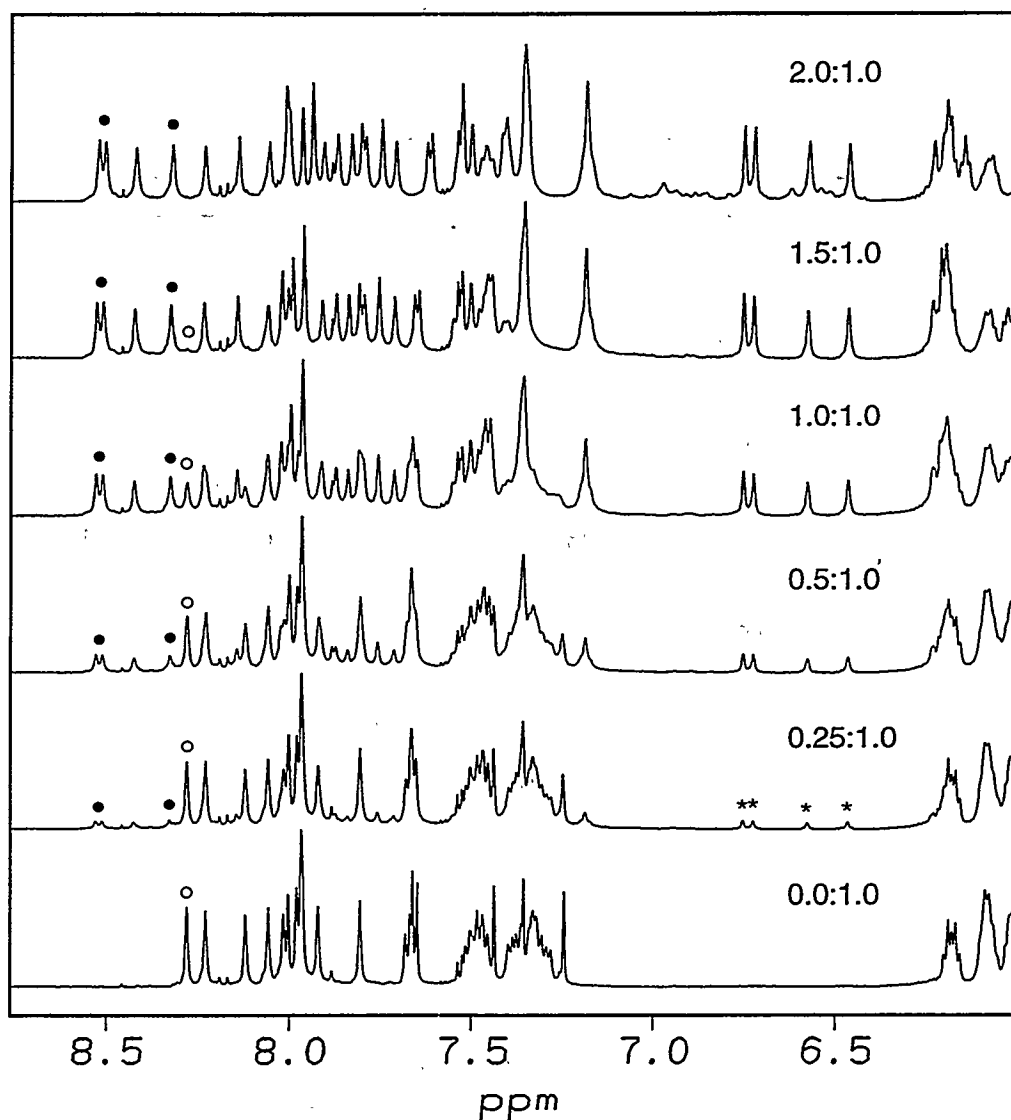
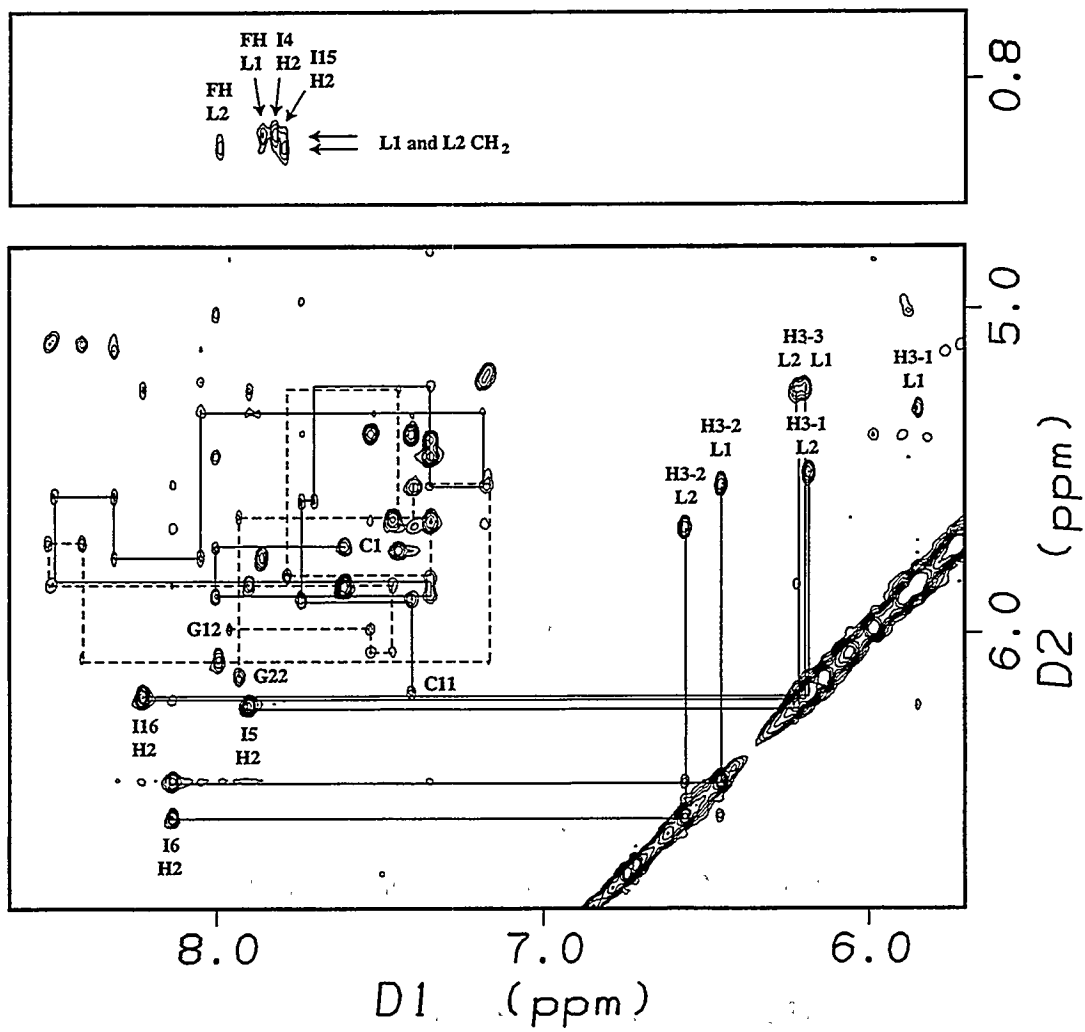


Figure 2-12. The downfield region of 1D ^1H NMR spectra acquired during the titration of d(CGCIICCGGC):d(GCCIIICCGCG) with distamycin (in D_2O , 15°C , 6024 Hz spectral width, 600 MHz, sample conditions 0.9 mM duplex DNA, 10 mM PO_4 , pH 7.0). The approximate distamycin:DNA stoichiometry is indicated for each spectrum. Open circles denote an inosine H8 proton in the uncomplexed DNA, and filled circles indicate downfield shifted inosine H8 resonances in the 2:1 complex. Asterisks indicate the ligand pyrrole resonances of the 2:1 complex. A small amount of free ligand can be seen in the 2:1 ratio spectrum.

Figure 2-13. Expansions of a 2D NOESY spectrum of the 2:1 distamycin:IIIC complex (in D₂O, 15 °C, τ_M = 200 ms, 600 MHz). Lower panel: The base aromatic and sugar H1' proton region. The sequential connectivities of the I-rich and the C-rich strands are denoted by solid and broken lines, respectively. Labels above or below a cross peak denote the assignment along the D1 axis, and labels to the left or right of a cross peak denote the assignment along the D2 axis. Ligand 1 (L1) and ligand 2 (L2) lie against the I-rich DNA strand and the C-rich DNA strand, respectively. Intermolecular ligand-DNA cross peaks are indicated between ligand pyrrole H3 protons and sugar H1' protons (right portion of the figure), between ligand pyrrole H3 protons and inosine H2 protons (lower left). Upper panel: The base aromatic and methylene proton region. Intermolecular ligand-DNA cross peaks shown are between pyrrole methylene protons and inosine H2 protons. Intermolecular ligand-ligand cross peaks shown are between the formyl proton each ligand and the methylene protons on the opposite ligand. Two other ligand-ligand cross peaks can be observed in frame (a) near the diagonal, one between L2 H3-2 and L1 H3-2, and the other between L2 H3-3 and L1 H3-1.



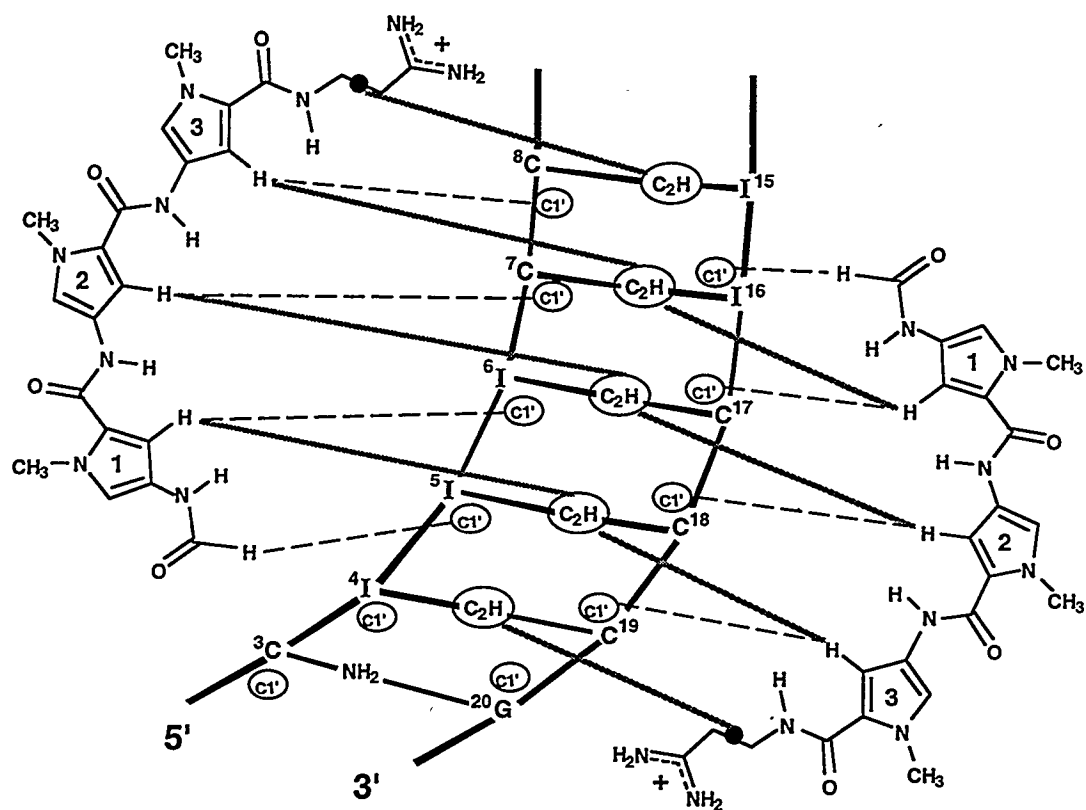


Figure 2-14. Summary of all intermolecular ligand-DNA contacts observed in the 2D NOESY spectrum in D₂O of the 2:1 complex of distamycin with d(CGCHCCGGC):d(GCCCHCCGCG). Ligand-ligand contacts are not shown for clarity.

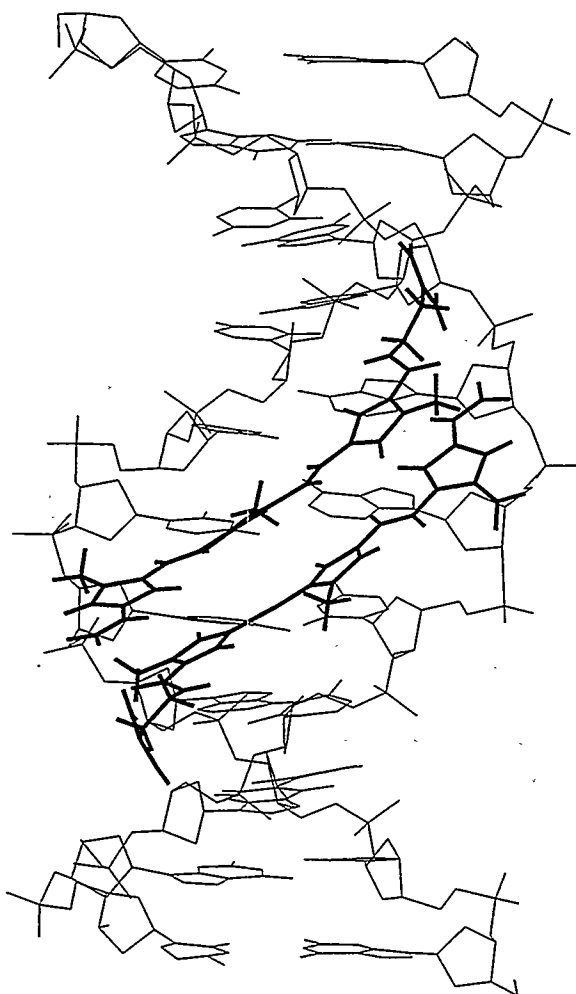


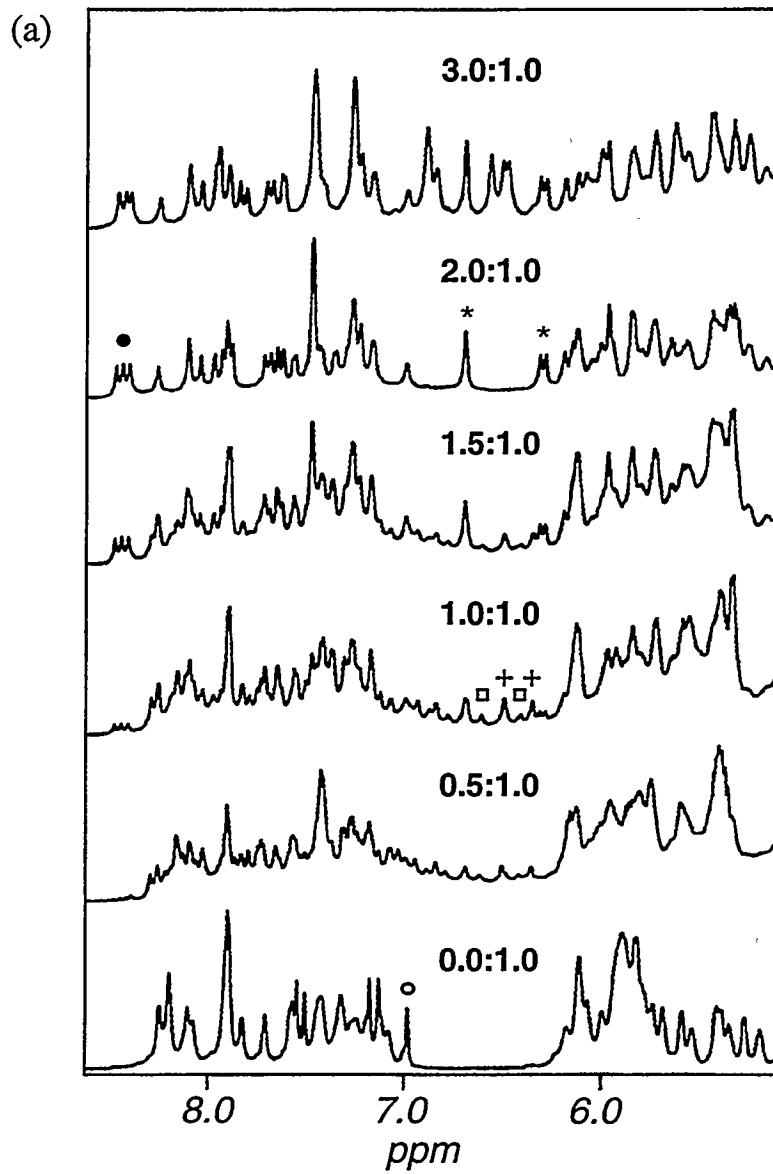
Figure 2-15. An example of a 2:1 distamycin:DNA complex. A molecular model of the 2:1 distamycin:DNA complex with d(CGCIHCCGGC): d(GCCIHCCCGCG) obtained by energy minimization using semiquantitative distance restraints derived from NOESY data. For clarity, hydrogen atoms are omitted for the DNA but not for the ligand molecules.

until nearly one equivalent of ligand has been added. Strikingly, for IIIIC no peaks are seen for the 1:1 complex, even at low stoichiometries. Only resonances for the 2:1 complex are seen, indicating a very high degree of cooperativity ($K_2/K_1 > 100$) for distamycin binding to this sequence.

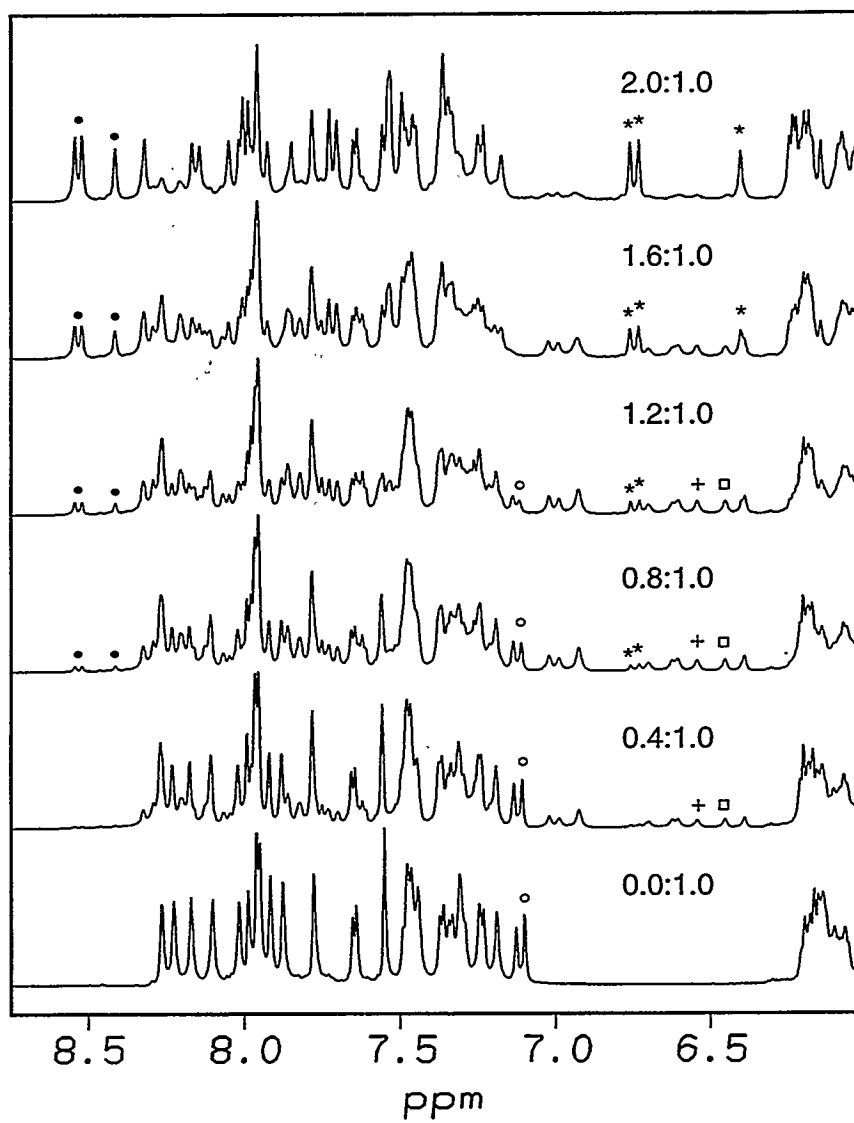
This suggests that the uncomplexed DNA has an intrinsically wider minor groove than AAATT, in which a single ligand would make contact with only one side, lowering its binding affinity. This suggestion is supported by evidence that distamycin binds exclusively in the 2:1 mode to G,C-containing sequences (Geierstanger et al., 1993; Geierstanger et al., 1994a), which in crystal structures have wider minor grooves (Yoon et al., 1988; Heinemann & Alings, 1989; Goodsell et al., 1993), perhaps due to the presence of the guanine 2-amino group. It also supports the idea that there is a substantial energetic cost to reducing the groove width in order to reestablish contact on both sides of the ligand. In the 2:1 complex, however, the groove is filled, and the two ligand molecules make close contacts with both sides of the groove. The lack of exchange broadening at intermediate ratios of added ligand and evidence from calorimetric and spectroscopic measurements (L. A. Marky, unpublished results) suggest that the overall binding affinity of the 2:1 IIIIC complex is at least as high as that of the 2:1 AAATT complex.

2.5.2 Distamycin Binding to Other I,C-Containing Sites. Because of the difference observed in the cooperativity of binding to AAATT:AATTT compared to IIIIC:IIICC, I went on to examine the cooperativity of distamycin binding to analogs of the AAATT:AATTT site which were only partially substituted by I•C base pairs. I examined the sites AAITT:AACTT, AAICT:AICTT, and IAATC:IATTC. The 1D NMR titrations are shown in Figure 2-16. At low ligand:DNA ratios, three sets of ligand pyrrole H3 protons are observed. At higher ligand:DNA ratios, one set of ligand pyrrole H3 protons increases in intensity, while the other two sets decrease in intensity and disappear at 2:1 ligand:DNA stoichiometry. 2D NOESY experiments run at this point confirmed that

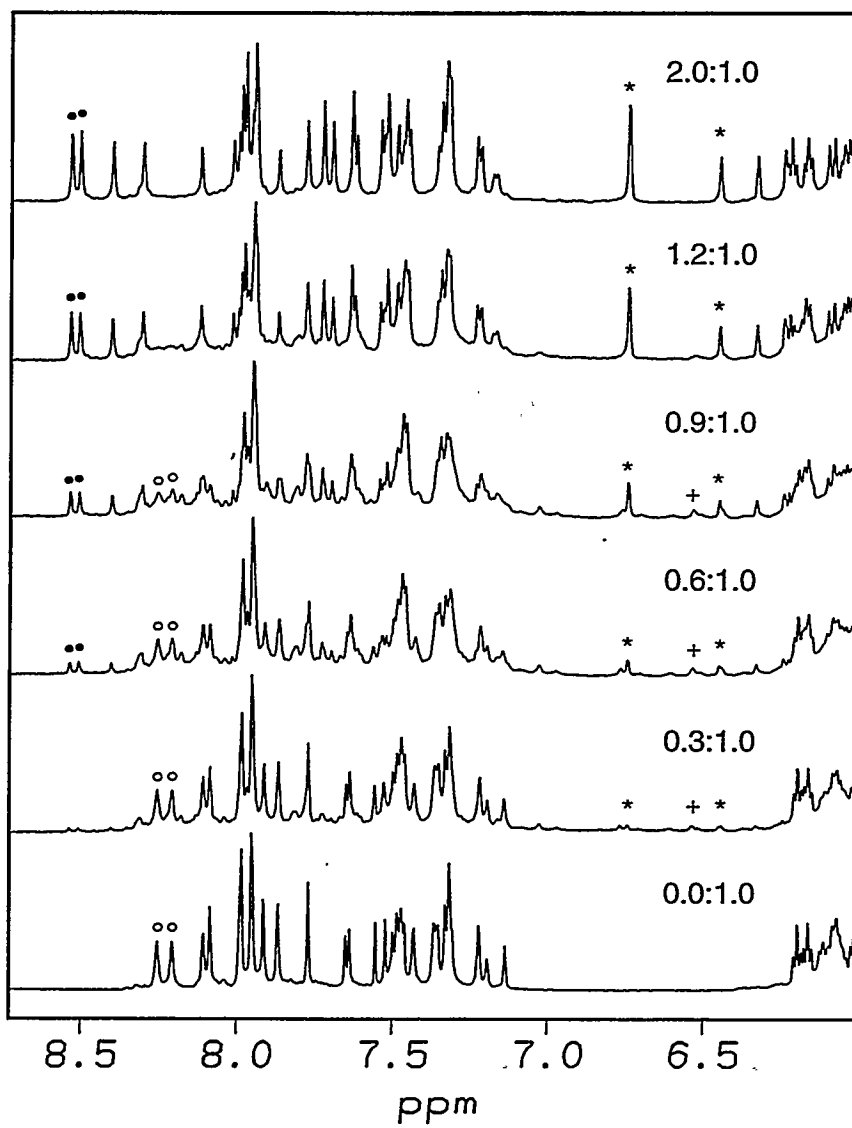
Figure 2-16. The downfield region of 1D ^1H NMR spectra acquired in D_2O solution during the distamycin titration of (a) $\text{d}(\text{CGCAAATTGGC}):\text{d}(\text{GCCAATTTGCG})$ (35°C , 500 MHz) (Pelton & Wemmer, 1990b), (b) $\text{d}(\text{CGCAAITTGGC}):\text{d}(\text{GCCAACTTGCG})$ (20°C , 5435 Hz spectral width, 600 MHz, 2.9 mM duplex DNA, 5 mM PO_4 , pH 7.0), (c) $\text{d}(\text{CGCAAICTGGC}):\text{d}(\text{GCCAICTTGCG})$ (25°C , 5556 Hz spectral width, 600 MHz, 2.0 mM duplex DNA, 10 mM PO_4 , pH 7.0), and (d) $\text{d}(\text{CGCIAATCGGC}):\text{d}(\text{GCCIAATTCGCG})$ (25°C , 4608 Hz spectral width, 500 MHz, 2.0 mM duplex DNA, 10 mM PO_4 , pH 7.0). The approximate distamycin:DNA stoichiometry is indicated for each spectrum. The open circles denote free DNA resonances and the filled circles indicate downfield shifted adenine/inosine H8 resonances in the 2:1 complex. Asterisks indicate the ligand pyrrole resonances of the 2:1 complex, and plus signs denote ligand pyrrole resonances in a 1:1 complex. In sections (a) and (b), plus signs denote ligand pyrrole resonances in one 1:1 complex, while open squares denote ligand pyrrole resonances in a second 1:1 complex. The two 1:1 complexes were not characterized in detail, although it is likely that they are related by a flip-flop symmetry. Also in section (b), the actual distamycin:DNA stoichiometry is slightly less than the ratio shown, as evidenced by the small amount of 1:1 complex still visible at the last titration point.



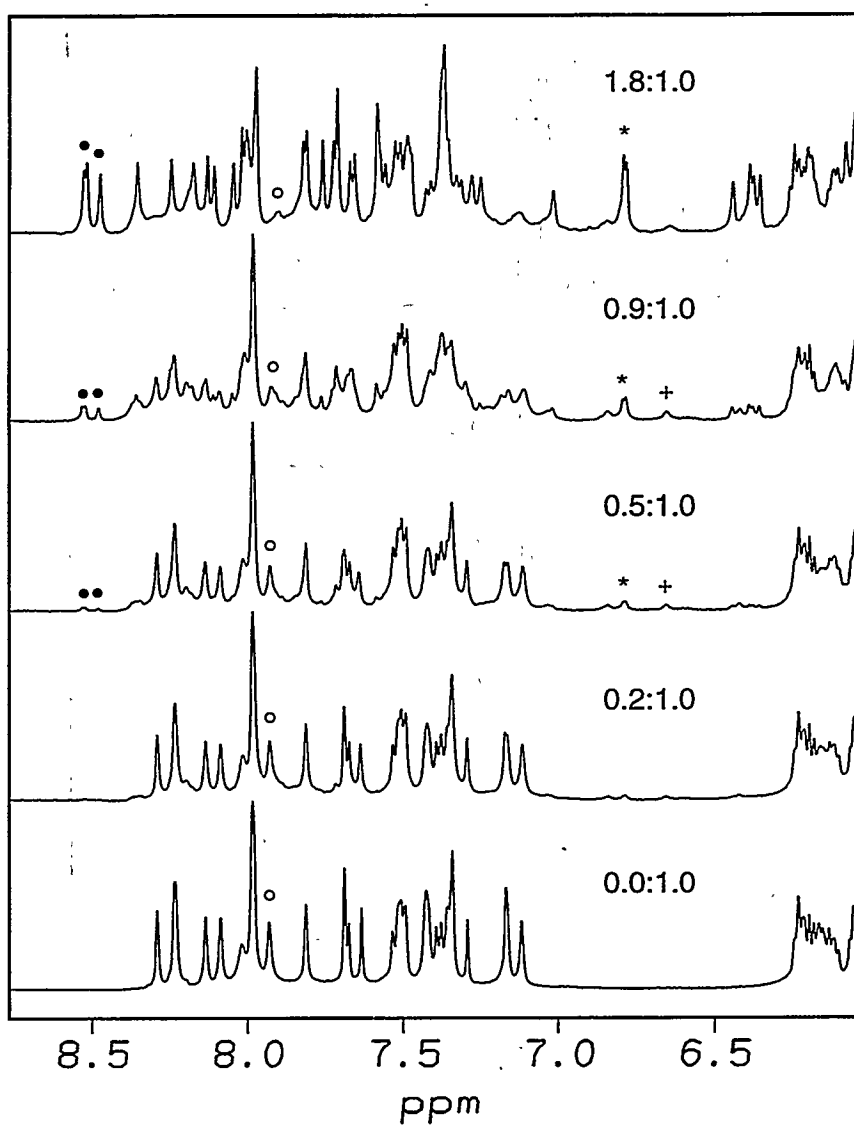
(b)



(c)



(d)



in all cases the complex formed is indeed a 2:1 complex analogous to that determined for the AAATT:AATTT and IIICC:IIICC sites.

A quantitative measure of the cooperativity can be made by integrating peaks corresponding to uncomplexed DNA, all 1:1 complexes formed, and the 2:1 complex and fitting the data to theoretical curves based on equation 2-1. For the site AAITT:AACTT, the ratio K_2/K_1 was found to be between 0.1 and 0.15 (Figure 2-17). The two sites studied which contain two I•C substitutions are approximately noncooperative ($K_2/K_1 \approx 1$). When compared with the cooperativities of the sites AAATT:AATTT and IIICC:IIICC, clearly distamycin binding in the 2:1 mode becomes increasingly cooperative upon successive replacement of A•T base pairs with I•C base pairs (Figure 2-18). Together, this information strongly suggests that the minor groove is wider in I,C regions of DNA compared to A,T rich regions, and is consistent with the hypothesis that the ability of A•T base pairs to form bifurcated hydrogen bonds in the major groove results in the high propeller twists and unusually narrow minor groove of A-tracts (Coll et al., 1987; Nelson et al., 1987; Yoon et al., 1988), since I•C base pairs are unable to form such hydrogen bonds.

My NMR results also correlate well with gel electrophoresis studies which measured the extent of DNA bending in oligomers containing the site AAAAA:TTTTT, where variants were made containing one or more I•C substitutions (Diekmann et al., 1987; Koo & Crothers, 1987). The degree of bending decreases with the number of I•C base pairs in the tract, and complete substitution in the site IIIII:CCCCC results in only a small amount of bending. Hagerman studied the effects of methylation at the pyrimidine 5 position on the curvature of A-tracts and I-tracts by gel electrophoretic mobility. He observed that in general, substitution of dU for dT in A-tracts induces bending and substitution of d(5^{me}-C) for dC in I-tracts reduces bending (Hagerman, 1990a). In my studies, the sites AAICT:AICTT and AAI(5^{me}-C)T:AI(5^{me}-C)TT appear to bind distamycin

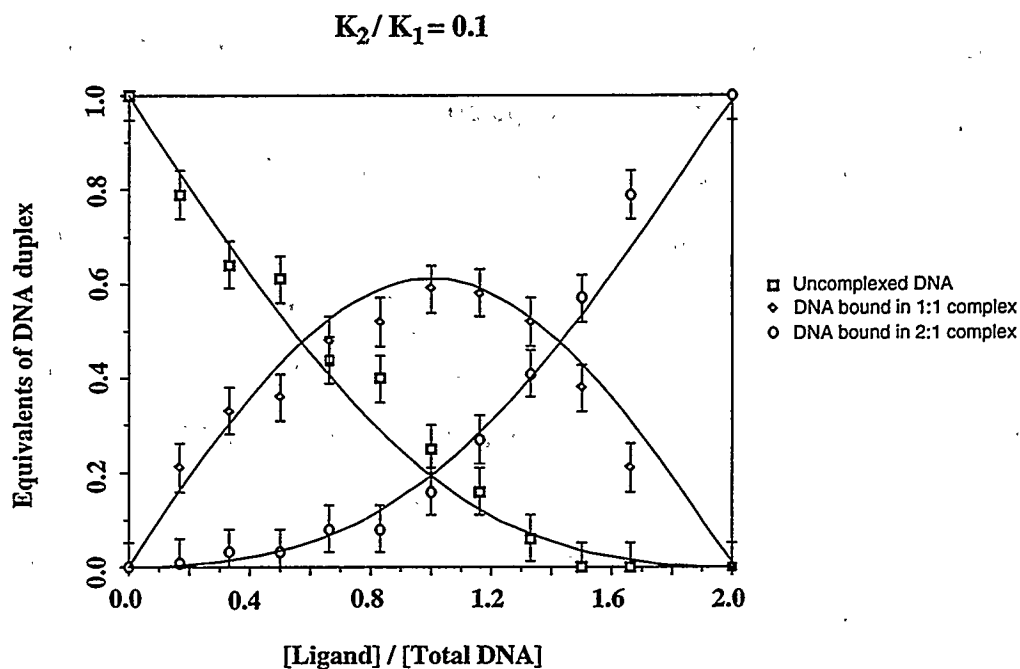


Figure 2-17. Integrated intensities of resonances corresponding to free DNA, 1:1 complex, and 2:1 complex during the 1D NMR titration of d(CGCAAITTGGC): d(GCCAACTTGCG). The intensities shown for the 1:1 complex points constitute the sum of the intensities for the two distinct 1:1 complexes observed. The theoretical curves shown (solid lines) are for the case where $K_2/K_1 = 0.1$.

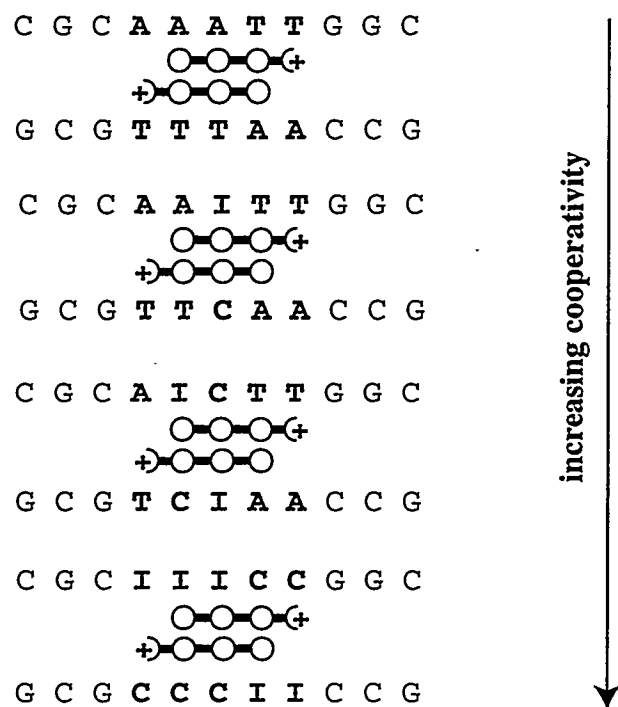


Figure 2-18. Schematic representation of 2:1 distamycin complexes with I,C-containing sites. The DNA sequences are displayed according to increasing cooperativity of ligand binding.

with similar cooperativity, suggesting that minor groove geometry is not sensitive to the presence of major groove methyl groups.

J. G. Pelton found that in solutions of distamycin and the oligomer [d(CGCAAATTTGCG)]₂ at crystallization conditions, the predominant form of complex present was the 2:1 complex (Pelton, 1990). Yet, the form which selectively crystallized was the 1:1 complex which was present in much smaller amounts (Coll et al., 1987). Interestingly, the only 2:1 distamycin:DNA complex yet crystallized is with the oligomer [d(ICICICIC)]₂ (Chen et al., 1994). In this structure, the two ligands bind in an "open" complex which is proposed for six base pair sites, but is somewhat different than the typical complexes observed on five base pair sites. In any case, it is likely that the high cooperativity of distamycin binding to I,C-containing oligomers contributed to the ability of this 2:1 complex to be crystallized.

2.5.3 Distamycin Binding to TTAAA. J. G. Pelton studied the binding of distamycin to the sequence d(CGCAAATTTGCG)₂. From integration of the 1D NMR titration data, the binding constant of the second ligand (K_2) is approximately 1.7 times that of the first ligand (K_1). I chose to study the related site d(CGCTTTAAAGCG)₂. The main difference between these two sites is that the former contains one 5'-AT-3' step, whereas the latter contains one 5'-TA-3' step, which is likely to disrupt the A-tract structure. The hypothesis is that if TTAAA has a wider minor groove, then distamycin may bind more cooperatively.

The 1D NMR titration of TTAAA is shown in Figure 2-19. No ligand resonances appear in the one-dimensional spectrum as ligand is added. It is likely that the ligand resonances are broadened due to exchange processes on an intermediate time scale. With increasing concentrations of added ligand the DNA resonances broaden and shift. In an attempt to sharpen the resonances, a D₂O NOESY of the 2:1 stoichiometry was acquired at higher temperature (45 °C). The standard pattern of cross peaks which arise between sugar H1' and base aromatic protons in B-form DNA can be seen only for the ends of the DNA

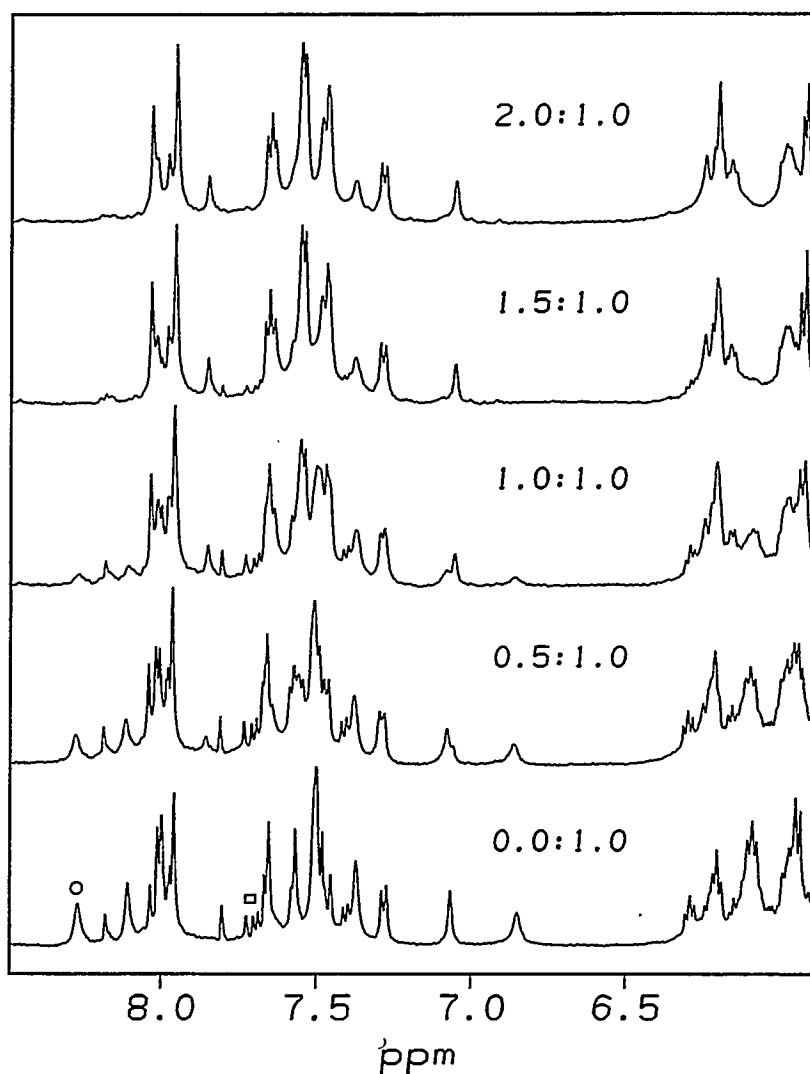


Figure 2-19. The downfield region of 1D ^1H NMR spectra acquired during the titration of $[\text{d}(\text{CGCTTTAAAGCG})]_2$ with distamycin (in D_2O , 25°C , 500 MHz, 4608 Hz spectral width, sample conditions 1.0 mM duplex DNA concentration, 10 mM PO_4 , pH 7.0). The approximate distamycin:DNA stoichiometry is indicated for each spectrum. The ligand resonances are too broad to be observed. The open circle denotes an adenosine H8 proton resonance in the uncomplexed DNA which broadens with added ligand. The open squares denote sharp hairpin resonances which are present at low concentrations, which decrease with increasing ligand concentration and disappear at a 2:1 ratio.

Distamycin binding to d(CGCTTTAAAGCG)₂

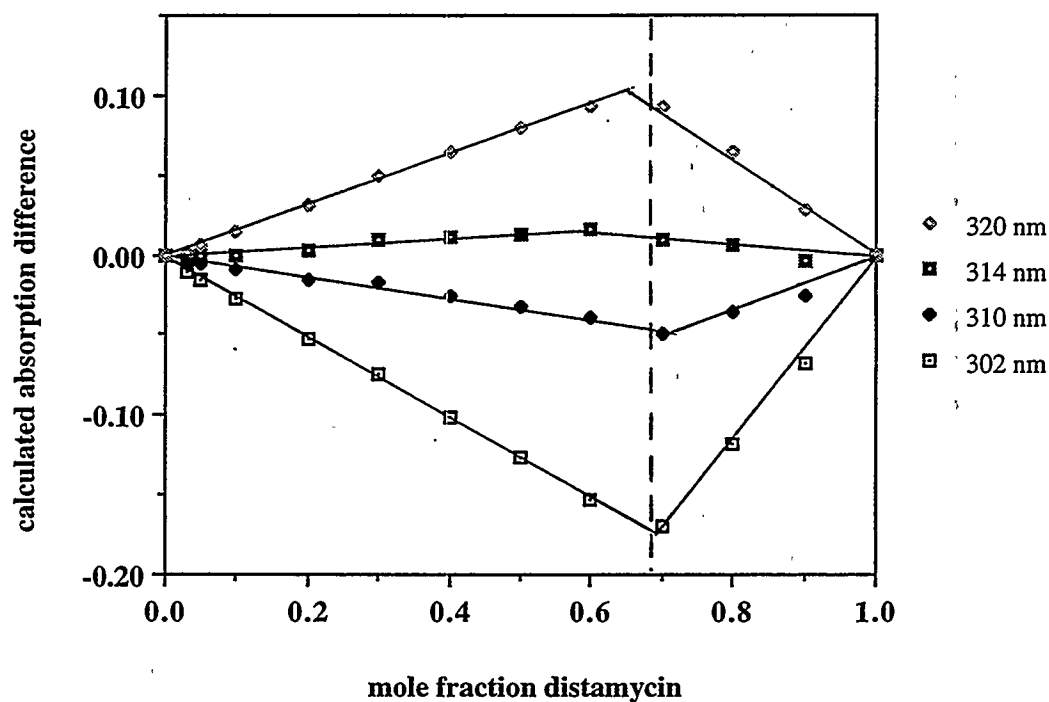


Figure 2-20. Job's plot of distamycin binding to [d(CGCTTTAAAGCG)]₂. The maximum ΔA is indicated, yielding the stoichiometry of this complex.

duplex, from bases C₁ through C₃, and again from G₁₀ through G₁₂. Resonances from the central TTTAAA tract are broadened as are the ligand resonances, indicating that this system is in intermediate exchange on the NMR time scale. This exchange behavior suggests that distamycin binds with lower affinity to TTTAAA relative to AAATTT. However, it is not possible to draw any conclusions about the cooperativity of distamycin binding to this site.

To verify the stoichiometry of the complex formed, I performed UV studies to obtain a Job's plot (Figure 2-20). The maximum ΔA at a given wavelength indicates the stoichiometry of the complex, and clearly shows that two ligands bind to this site. This experiment supported earlier work by Tondelli and coworkers, who studied the related sequence [d(CGTTTAAACG)]₂ using Job's method and also found two distamycin molecules bound (Capobianco et al., 1991).

2.5.4 Distamycin Binding to ATATAT. A crystal structure of the DNA sequence d(CGCATATATGCG)₂ has been determined in the absence of distamycin (Yoon et al., 1988). In the A,T region the groove is somewhat wider than is seen in crystal structures of oligomers with central sequences AATT and AAAAAA (Wing et al., 1980; Coll et al., 1987; Nelson et al., 1987; Yoon et al., 1988; Larsen et al., 1991). I have examined distamycin binding to this sequence.

Upon addition of distamycin, there is broadening and appearance of new resonances, indicating that the rate of ligand dissociation from the DNA is slow on the NMR time scale (Figure 2-21). A NOESY spectrum was acquired at 35 °C after the addition of 2 moles of distamycin/mole of duplex oligomer, and inspection of the aromatic base to sugar H1' proton region revealed that some critical cross peaks were weak or missing, mostly in the ATATAT tract. The A₆ H8 to A₆ H1' and the T₇ H6 to A₆ H1' cross peaks were not observable. At 50 °C the resonances in the A,T-rich region of the ATATAT complex sharpen slightly, enabling more cross peaks to be identified. The NOESY spectrum at 50 °C clearly indicates a 2:1 complex, with two ligands bound side by

side, based upon ligand-DNA and ligand-ligand cross peaks (Figure 2-22). The complex appears to be symmetric since only one set of ligand resonances appears in the spectrum, and the two DNA strands are degenerate in spectra of the complex.

However, the NOE contacts observed are not consistent with any single symmetric 2:1 complex. Analysis of line broadening as a function of temperature in the 2:1 distamycin:[d(CGCAAATTTGCG)]₂ complex indicates a dynamic process which does not involve the dissociation of the ligands (Pelton & Wemmer, 1990a). In that site as well as in the ATATAT site, the NOE patterns and the increased broadening of resonances in the binding sites suggests that there are sliding motions of the two ligands relative to one another and relative to the DNA (Figure 2-23). This sliding is fast on the NMR time scale, and results in an averaged set of resonances for the ligand and the DNA. Because the set of available complexes are averaged, the symmetry of the two DNA strands is maintained, resulting in the single NOE connectivity pathway shown in Figure 2-22.

It is important to note that peaks from the 2:1 complex appear well below stoichiometric addition of ligand, indicating a positive cooperativity in binding. Since no 1:1 complex is observed, the degree of cooperativity is somewhat higher than for AAATTT, and the binding constant of the second ligand is estimated to be more than 10 times the binding constant of the first ligand. This suggests that the wider groove in free ATATAT leads to weaker binding in the 1:1 mode, but still allows facile binding in the 2:1 mode. Relative to the minor groove width seen in the crystal structure, it is still necessary to widen the groove to accommodate two drugs binding side by side. These results, together with studies of distamycin binding to the sites AAAAA and ATATA (Chang, 1993; Wemmer et al., 1994) establish a clear trend in binding behavior, with the lowest cooperativity in runs of consecutive adenines, and the highest in the alternating sequences (Figure 2-24).

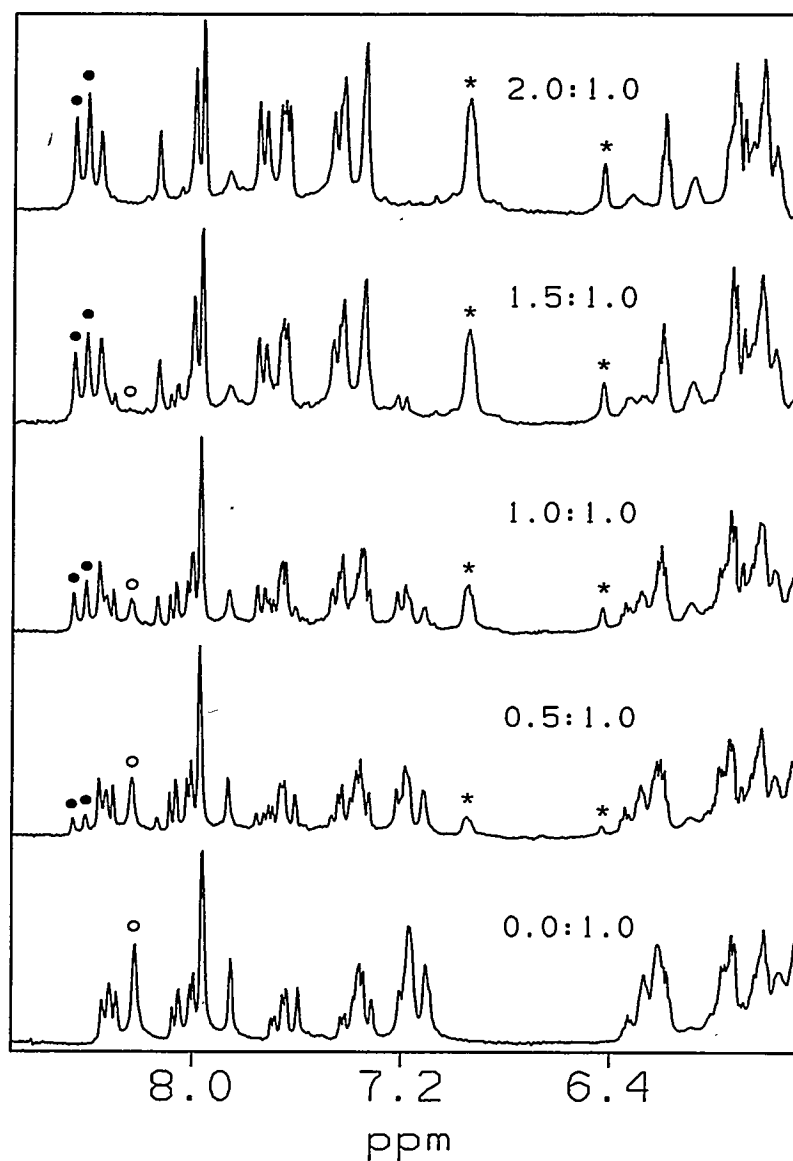


Figure 2-21. The downfield region of 1D ^1H NMR spectra acquired during the titration of $[\text{d}(\text{CGCATATATGCG})]_2$ with distamycin (in D_2O , 25°C , 4608 Hz spectral width, 500 MHz, sample conditions 0.7 mM duplex DNA concentration, 10 mM PO_4 , pH 7.0). The approximate distamycin:DNA stoichiometry is indicated for each spectrum. Open circles denote a proton in the uncomplexed DNA, and filled circles denote a proton in the 2:1 distamycin:DNA complex. Asterisks denote ligand pyrrole protons in the complex.

Figure 2-22. Expansions of a 2D NOESY spectrum of the 2:1 distamycin: [d(CGCA⁺TATATGCG)]₂ complex (in D₂O, 50 °C, 200 ms mixing time, 6024 Hz spectral width, 600 MHz, sample conditions 0.7 mM duplex DNA concentration, 10 mM PO₄, pH 7.0). Lower panel: The base aromatic and sugar H1' proton region. The sequential connectivities of the symmetric DNA duplex are denoted by solid lines. Labels above or below a cross peak denote the assignment along the D1 axis, and labels to the left or right of a cross peak denote the assignment along the D2 axis. Intermolecular ligand-DNA cross peaks are indicated between ligand pyrrole H3 protons and sugar H1' protons (right portion of the box) and between ligand pyrrole H3 protons and adenine H2 protons (lower left). Upper panel: The base aromatic and methylene proton region. Intermolecular ligand-ligand cross peaks shown are between the ligand formyl proton and the methylene protons on the opposite ligand.

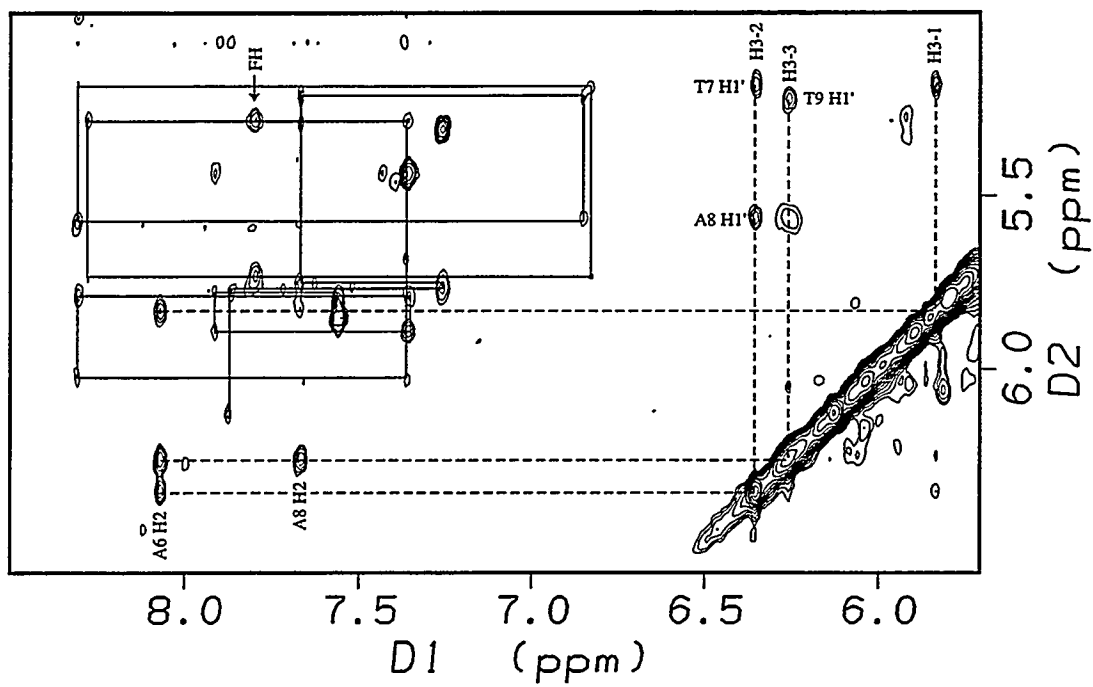
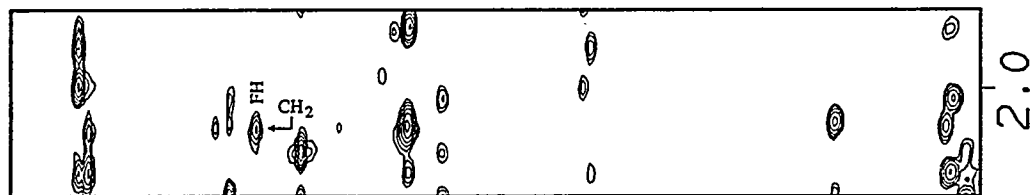
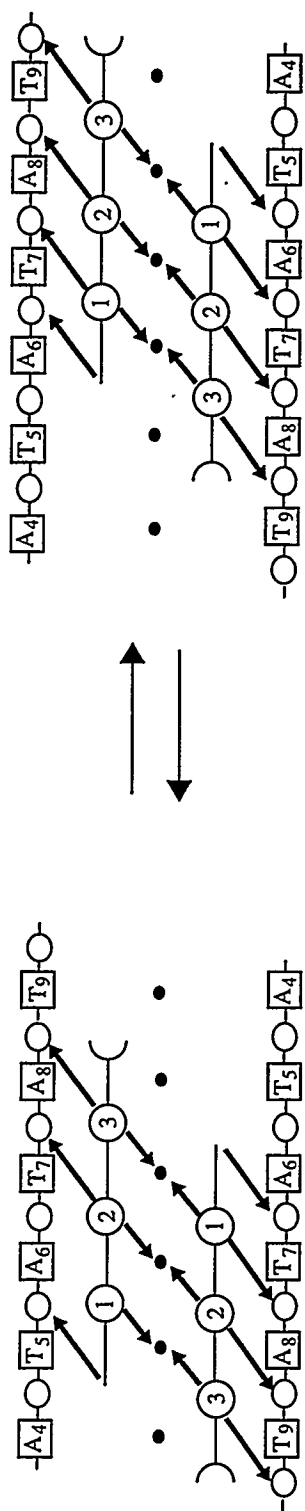
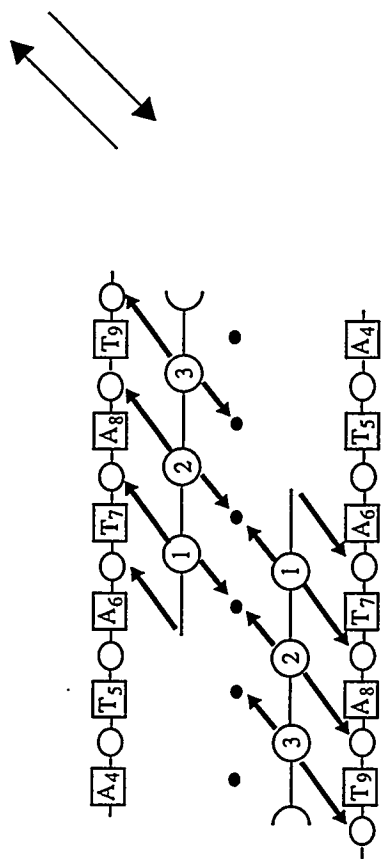


Figure 2-23. Schematic representation of the sliding model of the 2:1 complexes formed with the sequence $[d(\text{CGCATATATGCG})]_2$. The equilibria shown indicate sliding between three different 2:1 complexes which is fast on the NMR time scale. The DNA is represented by open squares (bases), open circles (sugar rings), and filled circles in the center of the groove (adenine H2 protons). Intermolecular NOEs observed from ligand pyrrole H3 and formyl protons to DNA sugar H1' and adenine H2 protons are denoted by thick arrows. Note that the H3-1 to A₆ H1' cross peak was not identified in the 2D NOESY spectrum, probably because the A₆ H1' resonance is broad and these two resonances are very similar in chemical shift, making it difficult to observe this cross peak.



ATATA subsite

TATAT subsite



ATATAT subsite

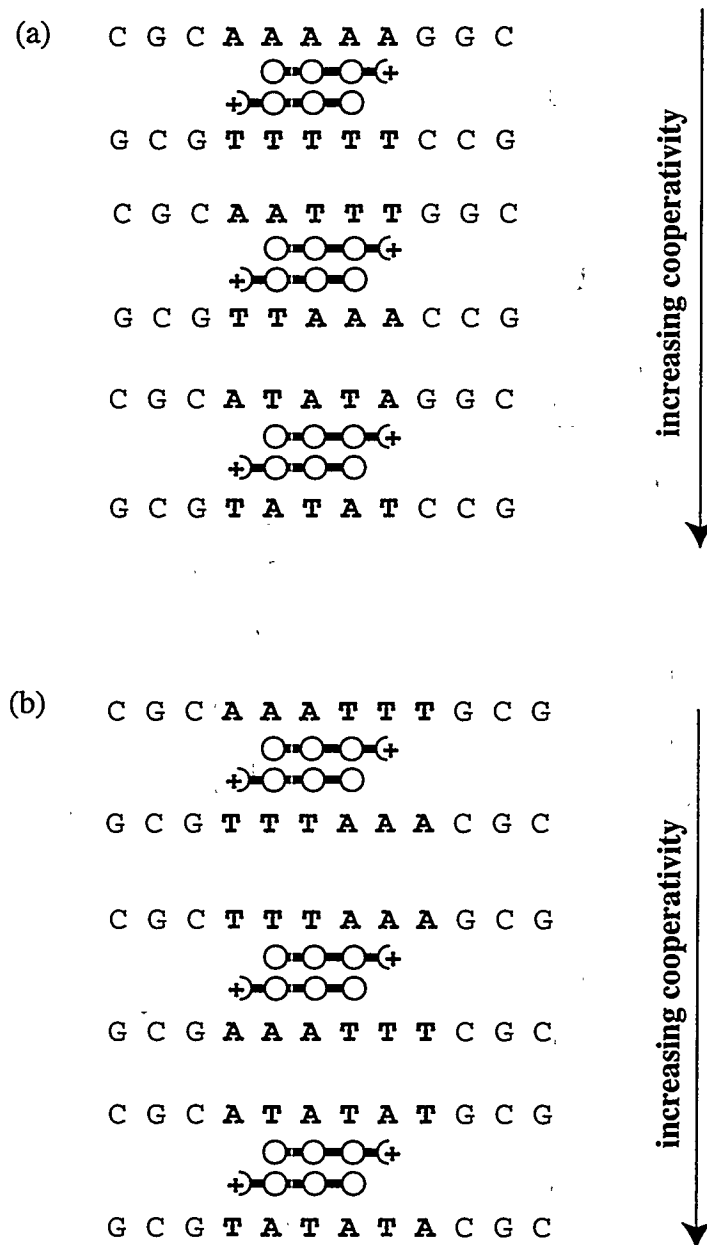


Figure 2-24. Schematic representation of 2:1 distamycin complexes with A,T-rich sites. The DNA sequences are displayed according to increasing cooperativity of ligand binding. (a) Five base pair sites. (b) Six base pair sites. Note that the six base pair sites undergo rapid exchange between different 2:1 complexes as shown in Figure 2-23.

2.6 Conclusions

Distamycin is a sensitive probe of sequence-dependent DNA structure. Once the multiple modes of distamycin binding were established, it became clear that more studies were needed to understand the determinants of the binding behavior of distamycin. Distamycin binds the asymmetric site AAAA with high orientational preference, but netropsin binds equally well in either orientation. This result highlights the importance of end groups in determining ligand binding affinity and sequence selectivity. Focus is increasingly turning toward understanding the effects of end groups and linkers as research in the Wemmer lab advances with novel linked and cyclized two-drug complexes based on distamycin and various lexitropsins. Further insight into these questions will undoubtedly contribute to efforts in rational design of sequence specific ligands.

The studies described early in this chapter indicate that ligand affinity at four base pair sites is affected by the width of the minor groove. Four base pair sites are too short to accommodate a second ligand in the 2:1 complex, and it appears to be energetically unfavorable to narrow a wide groove even by a small amount. The results of studies on various five and six A,T base pair sites indicate that minor groove width increases with the presence of 5'-TA-3' steps, and is highest in alternating sequences. The 2:1 complex is favored in sequences with wide minor grooves so that close contacts can be made between the ligands and the walls of the minor groove. Distamycin binding behavior seems to correlate well with the findings that A-tract DNA has distinct structural characteristics, since A-tracts seem to favor the 1:1 distamycin:DNA mode of binding.

The results from the I,C-containing oligomers indicate that I-C regions of DNA have properties somewhat unlike those of A-tracts due to differences in base pair stacking, propeller twist, and other structural parameters. The I,C regions are likely to have a wider minor groove, leading to higher cooperativity of distamycin binding despite the fact that the functional groups in the minor groove are identical in these two sequences. The difference in the shape of the groove is not unexpected, since the stacking of I•C pairs should be more

like G•C than A•T, leading to a groove more typical of a G,C-rich segment. This factor should be considered by researchers who utilize I•C base pairs as probes for minor groove interactions, and suggests that substitution of I•C for A•T pairs may provide a mechanism for characterization of sequence-dependent backbone interactions, such as those suggested in some repressor-operator complexes.

Chapter 3

The Binding Modes of a Rationally Designed Photoactivated DNA

Nuclease: Netropsin-Diazene Bound to

d(CGCAAAAGGC):d(GCCTTTTGCG)²

3.1 Summary

The complex between the rationally designed synthetic DNA cleaving agent netropsin-diazene and the double stranded DNA oligomer d(CGCAAAAGGC):d(GCCTTTTGCG) was characterized by two-dimensional NMR spectroscopy in solution. Netropsin-diazene is a conjugated ligand containing a derivative of the A,T-specific minor groove binding ligand netropsin. The π -diyl trimethylenemethane based compounds are a new class of DNA nucleases. Photolysis of netropsin-diazene bound to DNA generates a trimethylene-methane diradical intermediate that induces single strand breaks in the DNA in a sequence-specific manner. This study was undertaken to test the following design criteria: 1) binding of the diazene and subsequent reactive diyl to the DNA, 2) achievement of sequence selectivity in the ligand binding, and 3) prevention of diyl dimerization side reactions by physical separation from other diyls. A total of 16 intermolecular ligand-DNA distance restraints derived from NOE data were used to obtain the energy-minimized model of the netropsin-diazene:DNA complex. In the predominant conformation of the netropsin-diazene:DNA complex, the ligand is bound to the minor groove of the oligomer with the diazene at the 5' end of the A-tract. This form of the complex exchanges with a minor conformation in which the ligand is in the opposite orientation. The netropsin-diazene molecule appears to have fulfilled all of the design criteria, binding to the DNA duplex

² Reprinted in part with permission from Spielmann, H.P., Fagan, P.A., Bregant, T.M., Little, R.D. & Wemmer, D.E. (1995) *Nucleic Acids Res.* 23(9), 1576-1583. Copyright 1995 Oxford University Press.

studied in the minor groove of the central AAAA tract in a 1:1 mode, preventing diyl dimerization and presumably preventing other side reactions from occurring.

3.2 Introduction

The rational design of new ligands with novel biological activities and enhanced affinities for defined DNA sequences requires detailed structural knowledge of the interactions responsible for complex formation. The development of artificial sequence-specific DNA cleaving agents requires the simultaneous solution of the problem of DNA sequence recognition and the introduction of a functionality that effects cleavage.

The enediyne (1,5-diyn-3-ene) family of antitumor antibiotics is a class of molecules that upon activation, form diradical intermediates that cleave DNA and cause cell death. This family includes the natural products esperamicin, the calicheamicins, neocarzinostatin chromophore, and dynemicin (Figure 3-1) (for a review see Nicolaou & Dai, 1991 and Nicolaou et al., 1993). These molecules bind to the minor groove of short sequences of B-form DNA with high sequence specificity (up to four base pairs). The complex between calicheamicin and DNA has been characterized by NMR spectroscopy (Paloma et al., 1994; Ikemoto et al., 1995).

The mechanism of DNA cleavage by enediynes involves abstraction of a hydrogen atom from the minor groove via a 1,4-benzenoid diradical (Figure 3-2a). However, although they are highly potent antitumor agents, the natural products have relatively high toxicity. Therefore, there is much interest in designing synthetic analogs which have lowered toxicity and high activity against tumor cell lines. One way of achieving this goal is to synthesize less potent cleaving agents with higher sequence selectivity. This has been attempted by efforts to synthesize less active enediyne analogs. These efforts have met with mixed success.

Delocalized trimethylenemethane (TMM) π -diyls (Figure 3-2b) are expected to be less reactive towards hydrogen atom abstraction than are the enediyne-derived 1,4 σ -diyls

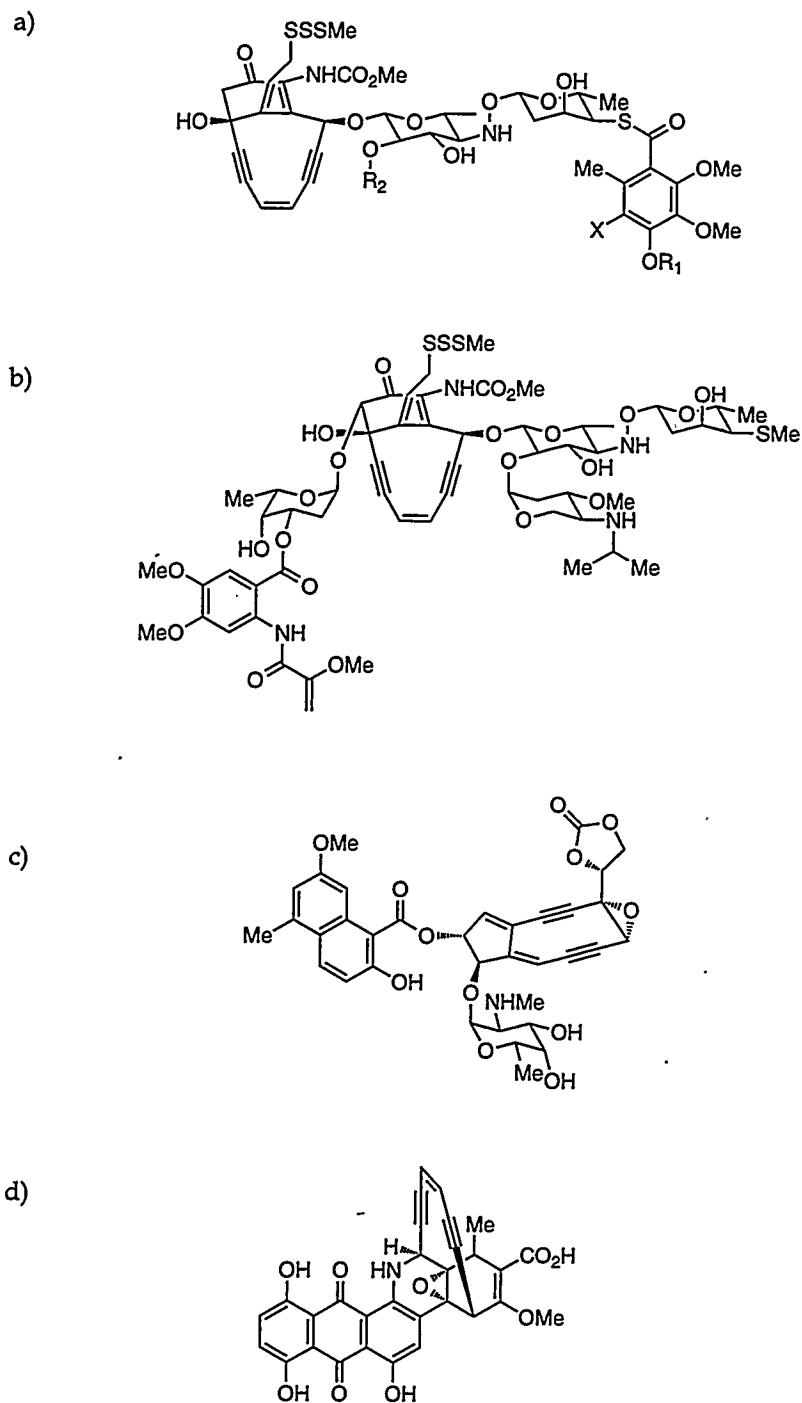


Figure 3-1. Representatives of the enediyne family of antitumor antibiotics: (a) calicheamicin γ_1 , (b) esperamicin A₁, (c) neocarzinostatin chromophore, and (d) dynemicin A.

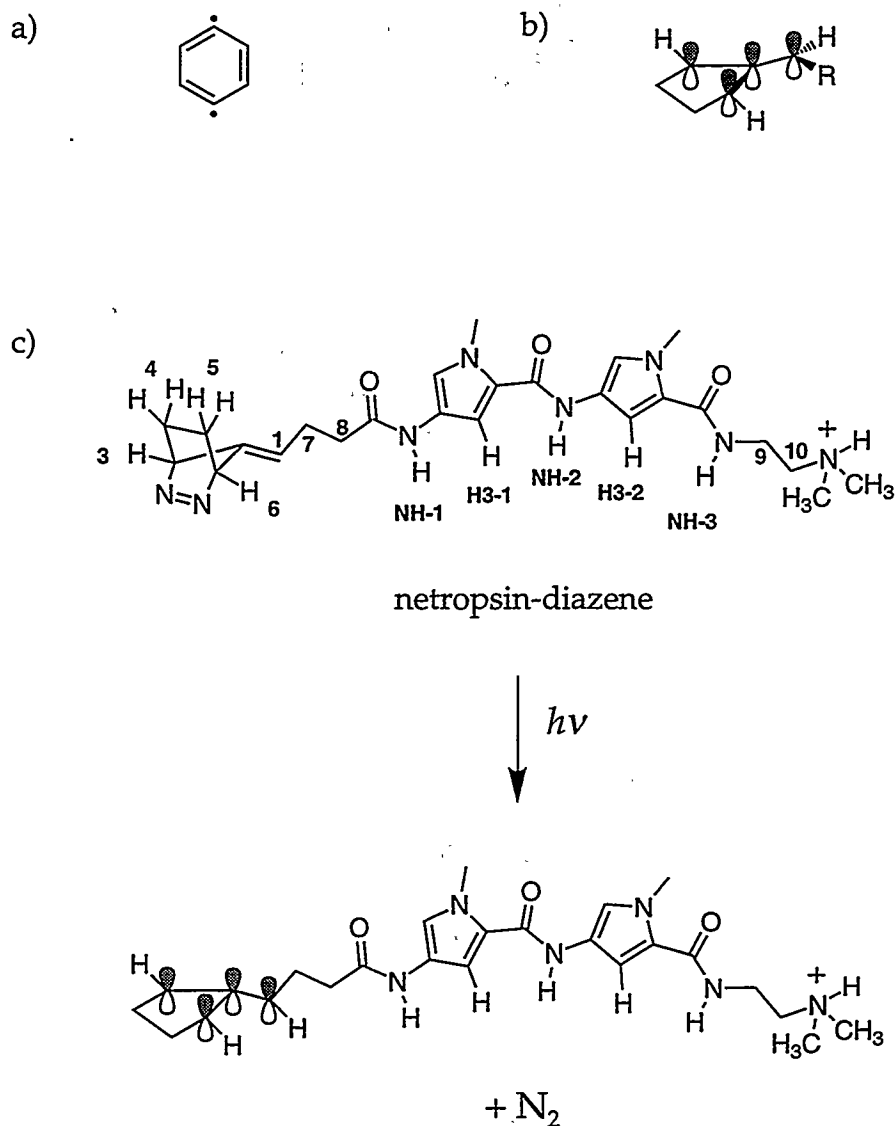


Figure 3-2. (a) A 1,4-benzenoid diradical, (b) a delocalized TMM diradical, and (c) diyl formation by photolysis of netropsin-diazene. The diazene chromophore is specifically excited by irradiation with light from a 450 W Hanovia lamp, transmitted through Pyrex and a 3.9% naphthalene/ethanol liquid filter. The numbering scheme for netropsin-diazene is shown.

(Berson, 1982; Little et al., 1992). In principle the reduced reactivity of the delocalized π -diyl may result in decreased overall toxicity (Nicolaou & Dai, 1991). TMM diyls have been shown to cleave DNA (Bregant et al., 1994), and the mechanism of cleavage is currently being studied (R. D. Little, unpublished results). Additionally, TMM diyls can be generated by the thermolysis or photolysis of diazenes (Little et al., 1992; Billera & Little, 1994; Bregant et al., 1994). Photoactivation has the added advantage of allowing precise control of the initiation and duration of the cleaving reaction. Therefore, we have focused this study on a DNA cleaving agent based on diazene, a TMM radical precursor.

Another way of reducing toxicity is to increase the sequence specificity of the cleaving agent to limit unwanted side reactions. As described in detail in Chapters 1 and 2, specific sequences of DNA can be targeted by ligands based on the antibiotics distamycin and netropsin. Various lexitropsins have previously been conjugated to DNA-cleaving agents (Figure 3-3). The earliest case was Dervan and coworkers, who used a tethered Fe•EDTA DNA cleaving functionality to make affinity cleaving reagents via local generation of the diffusible hydroxyl radical (Schultz et al., 1982; Schultz & Dervan, 1984). Most recently, lexitropsin conjugates have been synthesized which utilize flavin (Bouziane et al., 1995), bleomycin (Huang et al., 1995) and various enediynes (Nicolaou et al., 1993; Semmelhack et al., 1994) to cause DNA cleavage. The sequence specificity of these molecules in cleaving DNA has in many cases been lower than expected. It has been suggested that precise docking of the ligand into the DNA site may be required (Nicolaou et al., 1993), prompting the need for structural studies of these complexes.

A DNA cleaving agent based on diazene, a TMM diradical precursor, and netropsin was designed and synthesized in the laboratory of Professor R. D. Little (Figure 3-2c) (Bregant et al., 1994). The use of a lexitropsin is a particularly attractive choice because the binding behavior and DNA sequence selectivity of lexitropsins are well-documented, and the minimum length of a lexitropsin binding site is four base pairs. This site is longer than sites recognized by most enediyne molecules, and longer binding sites afford higher

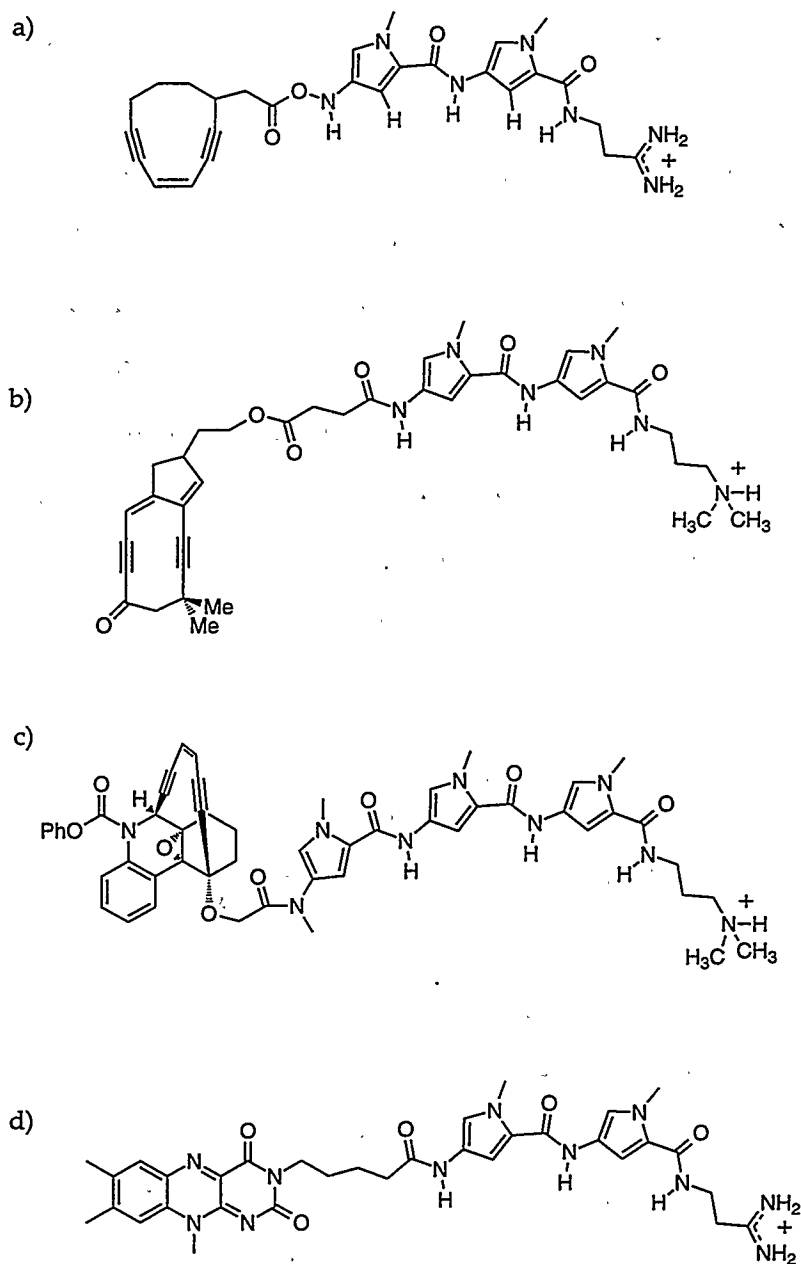


Figure 3-3. Synthetic conjugated ligands consisting of a DNA cleaving agent and a lexitropsin: (a) netropsin-(cyclodeca-4-ene-2,6-diyn-1-ol), (b) netropsin-neocarzinostatin, (c) distamycin-dynemicin, and (d) netropsin-flavin.

overall specificity, which could result in lower overall toxicity. Additionally, since several different lexitropsins may be utilized, there is a selection of possible DNA targets available. The design criteria were to: 1) bind the diazene and subsequent reactive diyl to the DNA, 2) achieve sequence selectivity in the ligand binding, and 3) prevent diyl dimerization side reactions by physical separation from other diyls. Although lexitropsins often bind DNA in a side-by-side 2:1 ligand:DNA complex, the two ligands are antiparallel, so that even in this case the diyls would be separated.

I, together with Dr. H. P. Spielmann, used NMR spectroscopy to determine the binding modes of netropsin-diazene to d(CGCAAAGGC):d(GCCTTTTGCG). Intermolecular NOE contacts between the ligand and the minor groove of the DNA molecule unambiguously confirm the formation of a 1:1 ligand:DNA complex at the AAAA:TTTT binding site. The ligand binds in fast exchange on the NMR time scale, suggesting that it binds with lower affinity than the parent compound netropsin. Two complexes are formed related by a flip-flop rearrangement of the ligand on the DNA.

3.3 Materials and Methods

3.3.1 Synthesis of Ligands and Oligonucleotides. The netropsin-diazene was synthesized as described previously (Bregant et al., 1994). The oligomers d(CGCAAAGGC) and d(GCCTTTTGCG) were synthesized and purified as described in Chapter 2. The numbering scheme for the DNA is given in Table 2-1.

3.3.2 NMR Sample Preparation. NMR samples were prepared by dissolving the decamer oligonucleotide duplexes in 0.5 mL of 25 mM sodium phosphate buffer (pH 7.0) and then lyophilizing to dryness. For experiments carried out in D₂O the solid was redissolved in 0.5 mL 99.96% D₂O (Cambridge Isotope Laboratories), while for experiments in H₂O a 90% H₂O/10% D₂O mixture (0.5 mL) was used. Solid netropsin-diazene was suspended in H₂O, and titrated with HCl until a clear yellow solution was obtained, yielding stock solution of 40 mM as determined by UV absorbance. An extinction coefficient $\epsilon_{296} = 21,500 \text{ M}^{-1} \text{ cm}^{-1}$ was used for the netropsin-diazene. The

ligand stock solution was stored at -70°C . Extinction coefficients at 260 nm for d(CGCAAAGGC) and d(GCCTTTTGCG) were calculated to be $1.02 \times 10^5 \text{ M}^{-1}\text{cm}^{-1}$, and $8.72 \times 10^4 \text{ M}^{-1} \text{ cm}^{-1}$ respectively (Warshaw & Cantor, 1970). The concentration of the double-stranded DNA sample was 2.3 mM.

3.3.3 NMR Experiments. A more general description of the NMR experiments and data analysis can be found in Chapter 2. The parameters of the NMR experiments performed are given here. Netropsin-diazene was titrated into the NMR sample containing duplex DNA in approximately 0.2 mole equivalents per addition. One-dimensional spectra in H_2O were acquired on a 500 MHz spectrometer with 8192 complex points averaged over 128 scans with a spectral width of 12048 Hz.

NOESY spectra in D_2O were collected on a 600 MHz spectrometer at 25°C , with 1024 complex points in t_2 using a spectral width of 6024 Hz and a mixing time of 200 ms. 512 t_1 experiments were recorded and zero-filled to 1024 points. For each t_1 value 64 scans were signal averaged using a recycle delay of 2 s. NOESY spectra in H_2O were acquired at 25°C , with 2048 complex points in t_2 using a spectral width of 13,514 Hz and a mixing time of 200 msec, with 1:1 water suppression as described in Chapter 2. 480 t_1 experiments with 64 scans were recorded and zero-filled to 2048 points. TOCSY spectra in D_2O were acquired at 25°C , with 1024 complex points in t_1 with a spectral width of 4800 Hz and a mixing time of 60 ms.

The DNA resonances were assigned as described in Chapter 2, and the ligand resonances were assigned as described below (Pelton & Wemmer, 1989; Pelton & Wemmer, 1990b; Dwyer et al., 1992; Mrksich et al., 1992).

3.3.4 Distance Restraints. The general procedure for extracting distance restraints from NMR data is given in Chapter 2. For the netropsin-diazene complex, intermolecular distance restraints were generated from the volume integrals of the cross peaks in the H_2O NOESY spectra acquired at a mixing time of 200 msec. A pseudoatom was generated for each of the following sets of ligand protons: 1) the six $\text{N}-(\text{CH}_3)_2$ protons, 2) the four

Table 3-1. Ligand-DNA restraints used in modeling the netropsin-diazene complex with d(CGCAAAAGGC):d(GCCTTTTGCG).

netropsin-diazene	DNA	NOE classification ^a
NH-2	A ₅ H2	w
H3-1	A ₅ H2	s
NH-1	A ₅ H2	s
H3-2	A ₆ H2	s
NH-2	A ₆ H2	s
H3-1	A ₆ H1'	m
NH-2	A ₆ H1'	w
H3-2	A ₇ H1'	m
H3-2	T ₁₆ H1'	m
H3-1	T ₁₇ H1'	m
N-(CH ₃) ₂	A ₇ H2	p
H7a, H7b, H8a, H8b	A ₄ H2	p
H9a, H9b, H10a, H10b	G ₈ H1'	p
H9a, H9b, H10a, H10b	G ₉ H1'	p
H9a, H9b, H10a, H10b	T ₁₄ H1'	p
H9a, H9b, H10a, H10b	T ₁₅ H1'	p

^a NOE classification: s = strong (1.8 to 2.8 Å), m = medium (1.8 to 3.5 Å), w = weak (1.8 to 5.0 Å), p = pseudo atom (1.8 to 7.0 Å).

methylene protons 7a, 7b, 8a, and 8b, and 3) the four methylene protons 9a, 9b, 10a, 10b. Distance restraints from DNA protons to these pseudoatoms were given a range of (1.8 to 7.0 Å). The 16 intermolecular ligand-DNA restraints used in the modeling are listed in Table 3-1.

3.3.5 Model Refinement. The starting model was generated and energy minimizations were performed using the methods described in Chapter 2. The energy of the complexes was initially minimized using 100 steps of a steepest descents algorithm and 3000 steps of conjugate gradient minimization. The final model was created by subjecting the minimized starting model to 13 psec of restrained molecular dynamics (rmd) at 200 K, which was subsequently cooled to 100 K over 10 psec, followed by 4 psec of rmd at 100 K. The model was then energy-minimized by conjugate gradient minimization to a final rms derivative of <0.05 (kcal/mole)/Å².

3.4 Results

3.4.1 Titration of the AAAA Site with Netropsin-Diazene. NMR spectra recorded at 25 °C during the titration of the duplex d(CGCAAAGGC): d(GCCTTTTGCG) with netropsin-diazene are shown in Figure 3-4. The ligand is complexed to the DNA in a 1:1 ratio at all stoichiometries. The DNA resonances broaden and shift upon addition of ligand, and only one set of DNA resonances exists throughout the titration, indicating that the ligand is binding to the DNA in fast exchange on the NMR time scale. In contrast, when the same DNA duplex is titrated with the parent compound netropsin, distinct sets of resonances for the free and ligand bound DNA exist, indicating that netropsin binds in slow exchange on the NMR time scale (Chapter 2).

3.4.2 NMR Resonance Assignments. Part of a NOESY spectrum of the netropsin-diazene:DNA complex in D₂O is shown in Figure 3-5. The chemical shifts of the deoxyribose H1' and DNA base H6, H8 and adenine H2 resonances, and those of the ligand resonances are listed in Tables 3-2 and 3-3, respectively. Assignments of the

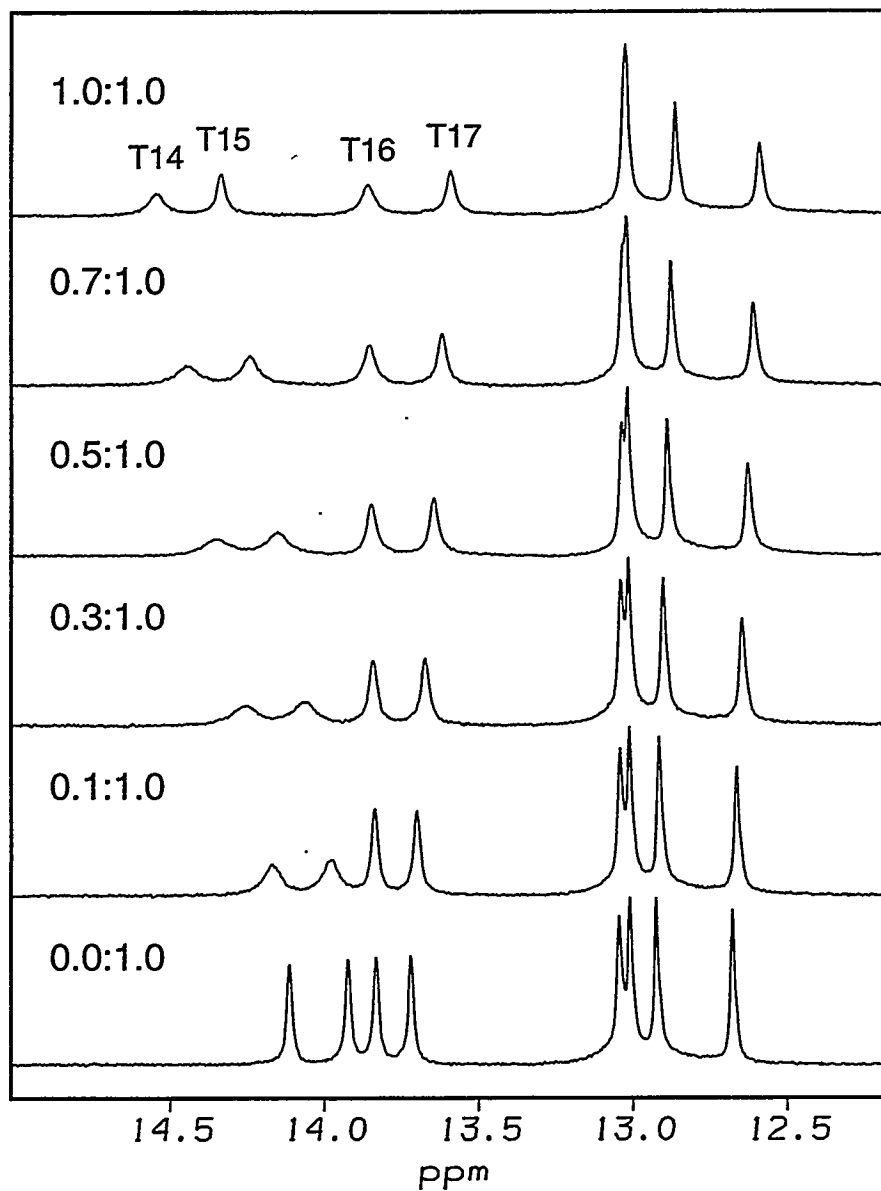


Figure 3-4. Imino region of ^1H NMR spectra acquired at several points in a titration of $\text{d(CGCAAAGGC):\text{d(GCCTTTTGCG)}$ with netropsin-diazene at 25 °C (in H_2O). The molar netropsin-diazene:DNA ratios are indicated for each spectrum.

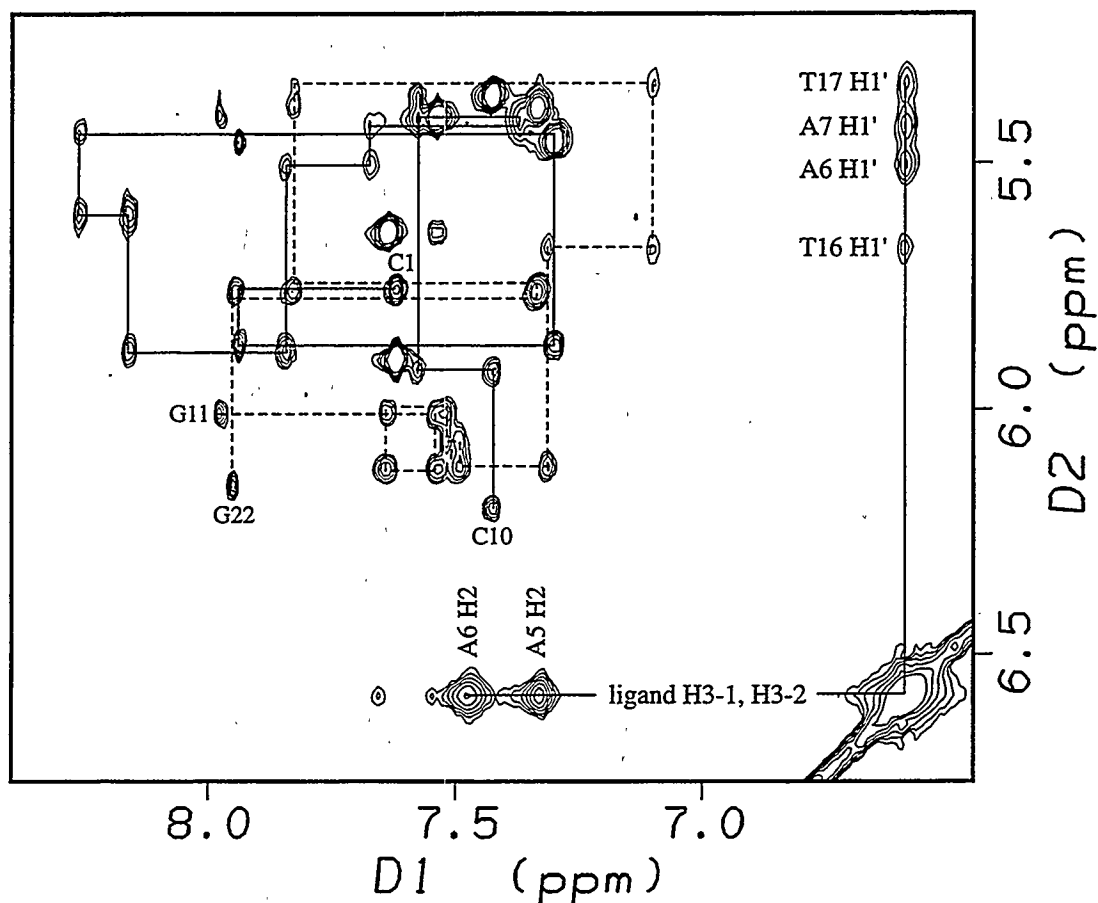


Figure 3-5. Expansion of the aromatic to sugar H1' region of a NOESY spectrum of the 1:1 complex of netropsin-diazene with d(CGCAAAGGC):d(GCCTTTTGCG) (in D₂O, 25 °C, τ_M = 200 ms). Sequential aromatic to H1' connectivities for the 5'-AAAA-3' strand are shown as a solid line; those for the 5'-TTTT-3' strand are shown as a broken line. NOE cross peaks from ligand pyrrole protons to various DNA protons are labeled. The ligand numbering scheme is as according to Figure 3-2c.

Table 3-2. Chemical shift assignments of the d(CGCAAAGGC):d(GCCTTTTGCG) duplex, free and in the netropsin-diazene:DNA complex.^a

	H6/H8			H1'			H2		
	Free	Comp	$\Delta\delta$	Free	Comp	$\Delta\delta$	Free	Comp	$\Delta\delta$
Strand 1									
C ₁	7.62	7.64	+0.02	5.73	5.76	+0.03			
G ₂	7.93	7.96	+0.03	5.86	5.87	+0.01			
C ₃	7.31	7.32	+0.01	5.42	5.44	+0.02			
A ₄	8.19	8.28	+0.09	5.74	5.60	-0.14	7.17	7.15	-0.02
A ₅	8.08	8.19	+0.11	5.75	5.88	+0.13	7.00	7.35	+0.35
A ₆	7.99	7.86	-0.13	5.81	5.51	-0.30	7.03	7.50	+0.47
A ₇	7.87	7.69	-0.18	5.89	5.42	-0.47	7.43	7.67	+0.24
G ₈	7.39	7.39	0.00	5.52	5.41	-0.11			
G ₉	7.56	7.60	+0.04	5.92	5.92	0.00			
C ₁₀	7.36	7.44	+0.08	6.15	6.20	+0.05			
Strand 2									
G ₁₁	7.95	8.00	+0.05	5.97	6.01	+0.04			
C ₁₂	7.51	7.56	+0.05	6.08	6.12	+0.04			
C ₁₃	7.59	7.67	+0.08	6.00	6.01	+0.01			
T ₁₄	7.49	7.54	+0.05	6.09	6.07	-0.02			
T ₁₅	7.49	7.51	+0.02	6.14	6.11	-0.03			
T ₁₆	7.48	7.33	-0.15	6.10	5.67	-0.43			
T ₁₇	7.30	7.12	-0.18	5.81	5.34	-0.47			
G ₁₈	7.90	7.85	-0.05	5.79	5.76	-0.03			
G ₁₉	7.33	7.35	+0.02	5.75	5.77	+0.02			
C ₂₀	7.92	7.97	+0.05	6.12	6.15	+0.03			

^a Chemical shifts are given in ppm (± 0.01 ppm). The residual HOD resonance is referenced to 4.84 ppm (20 °C) for the free DNA, and to 4.78 ppm (25 °C) for the complexed DNA.

Table 3-3. Chemical shift assignments of netropsin-diazene complexed with d(CGCAAAAGGC):d(GCCTTTTGCG).^{a,b}

Proton	Chemical shift
NH-1	9.45
H3-1	6.59
NH-2	9.23
H3-2	6.59
NH-3	8.23
C4(a,b)H ^c , C5(a,b)H ^c	1.10, 1.14, 1.81, 1.87
C7(a,b)H ^c , C8(a,b)H ^c	2.2 to 2.5
C9(a,b)H ^c , C10(a,b)H ^c	3.31, 3.40, 3.57, 3.60
C1H ^{c,d} , C3H ^{c,d} , C6H ^{c,d}	5.34, 5.39, 5.73, 5.74
N-(CH ₃) ₂	3.06

^a Chemical shifts are given in ppm (± 0.01 ppm), with the residual HOD resonance referenced to 4.78 ppm (25 °C). ^b Refer to Figure 3-2c for numbering system used.

^c Resonances assigned as a group. ^d More than one resonance is observed for some of the diazene protons because the ligand binds the DNA diastereotopically.

deoxyribose H1', H2', H2'', H3', thymine methyl, and base H5, H6, H8, and H2 resonances were made from the D₂O NOESY spectra, using standard methods (Hare et al., 1983). Assignments were extended to some of the deoxyribose H4' resonances as well. The correct assignment of the adenosine base H2 resonances established the orientation of the ligand in the complex. These assignments were made based on cross peaks from the H2 protons to intrastrand and interstrand deoxyribose H1' protons in the D₂O NOESY. Cross peaks from the H2 protons to thymidine imino protons in the H₂O NOESY confirm these assignments. Intraresidue NOE cross peak patterns arising from deoxyribose protons as well as interresidue NOE cross peak patterns indicate that the DNA is B-form in the complex, and that there is no obvious distortion of the DNA helix as a result of ligand binding.

The ligand methylene 9a, 9b, 10a, and 10b (see Figure 3-2c for numbering scheme) proton resonances and N-(CH₃)₂ resonances were assigned using the D₂O NOESY and TOCSY spectra (Figure 3-6). The resonance at 3.06 ppm integrates to six protons in the spectrum, and gives strong NOESY cross peaks to four other ligand resonances. These four resonances give rise to strong TOCSY and NOESY cross peaks to each other, indicating that they are J-coupled and are close in space. However, the resonance at 3.06 ppm does not give rise to TOCSY cross peaks to these four resonances. This information is consistent with our assignments of the resonance at 3.06 ppm to the N-(CH₃)₂ protons and the other four resonances to the adjacent methylene protons 9a, 9b, 10a, and 10b (see Table 3-3). Among these four resonances, it was impossible to determine which resonance corresponded with which of the four methylene protons, so that specific assignments could not be made.

Two of the methylene resonances 9a, 9b, 10a, and 10b give NOE cross peaks to a ligand amide resonance. This amide resonance at 8.23 ppm is assigned to NH-3. Several broad resonances in the region 2.2-2.5 ppm give rise to NOESY cross peaks to another ligand amide resonance at 9.45 ppm. This amide must be NH-1, and the broad upfield

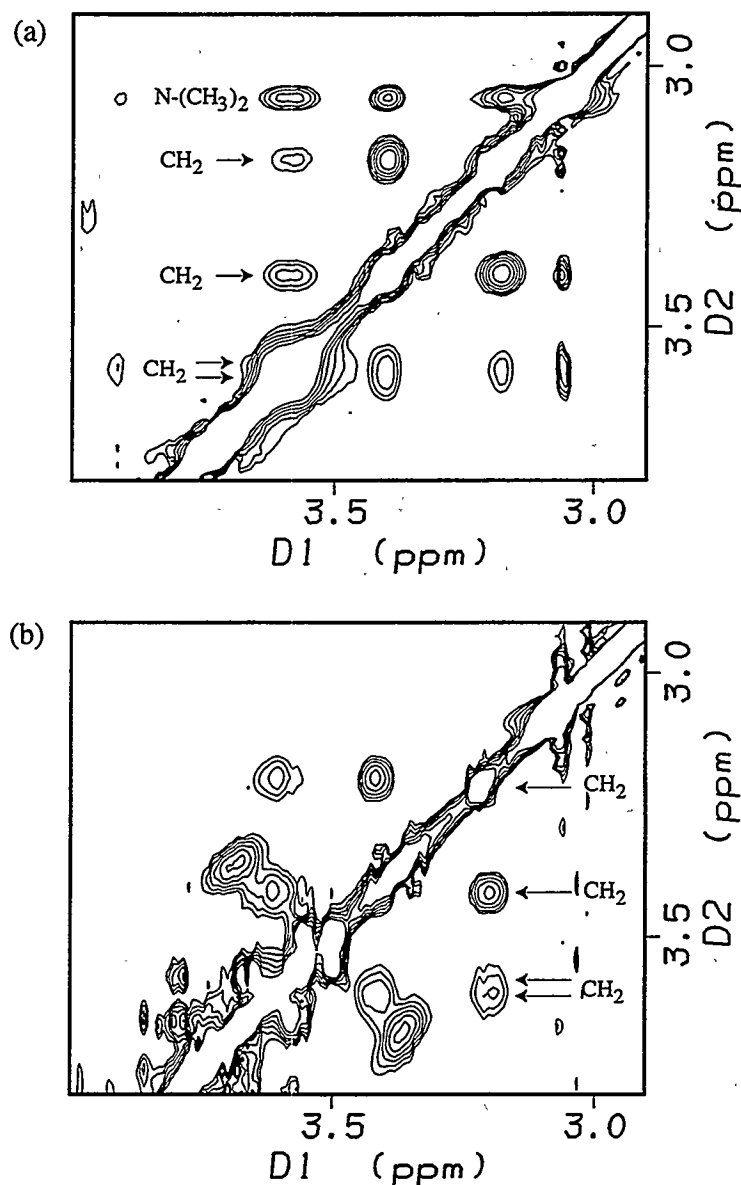


Figure 3-6. (a) Expansion of the upfield diagonal region of a NOESY spectrum of the 1:1 complex of netropsin-diazene with d(CGCAAAGGC):d(GCCTTTTGCG) (in D_2O , 25 °C, $\tau_M = 200$ ms). Labels indicate the assignments of the nonexchangeable protons on the dimethylammonium end of the molecule. All five resonances are connected by NOEs to each other. (b) Expansion of the corresponding region of a TOCSY spectrum of the same complex (in D_2O , 25 °C, $\tau_M = 60$ ms). The resonance at 3.06 ppm does not give rise to TOCSY cross peaks to any of the methylene resonances as labeled.

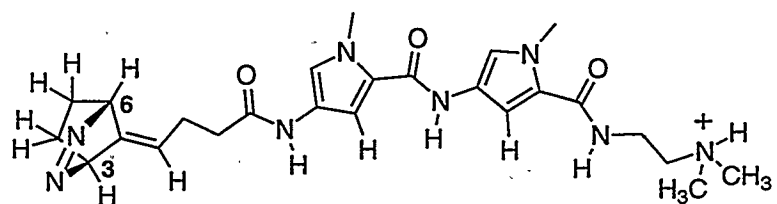
resonances are assigned to the ligand methylene protons 7a, 7b, 8a, and 8b. These methylene protons have additional NOESY cross peaks to DNA A7 and A4 H₂ protons and to an unassigned proton of the bicyclic diazene framework. The third amide resonance at 9.23 ppm is assigned to NH-2. The two netropsin-diazene pyrrole H3-1 and H3-2 resonances are degenerate in chemical shift, and were assigned based on their chemical shift and based on NOESY cross peaks to the ligand NH-1, NH-2 and NH-3 resonances and to various DNA resonances.

The bicyclic diazene moiety contains seven protons, labeled 1, 3, 4a, 4b, 5a, 5b, and 6 in Figure 3-2c. We observe intramolecular NOEs within the bicyclic diazene moiety, and from methylene protons 7a, 7b, 8a, and 8b to the bicyclic diazene framework. However, using both NOESY and TOCSY spectra we were unable to unambiguously assign the diazene resonances. No intermolecular NOESY cross peaks from any of the seven bicyclic diazene resonances to the DNA were observed.

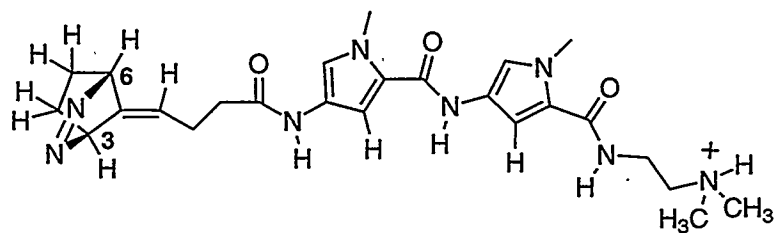
Two separate sets of ligand resonances are observed for the diazene proton 1-6 resonances, which were attributed to two enantiomers. The enantiomers result from the 3*S*-*E* and 3*S*-*Z* (Figure 3-7) configurations of the bicyclic diazene fragment of the ligand. The enantiomeric netropsin-diazene binds the DNA diastereotopically. There is not enough evidence to determine if there is a preference for one enantiomer binding over the other.

3.4.3 Intermolecular Contacts. Cross peaks from DNA protons to ligand pyrrole protons H3-1 and H3-2 and amide protons NH-1 and NH-2 indicate that the ligand pyrrole-pyrrole ring system spans the 5'-A₄-A₅-A₆-A₇-3' sequence (Table 3-1 and Figure 3-8). This is further supported by cross peaks from DNA protons to the ligand methylene protons at either end of the molecule.

A set of NOEs connects the dimethylamino protons to the H₂ proton of A₇, although a weaker set to A₄ H₂ also occurs which cannot be explained by a single model of the ligand-DNA complex (Figure 3-9). The presence of two mutually exclusive sets of NOESY cross peaks clearly indicates that the ligand is in one predominant orientation with



3S-E, 6R-Z
netropsin-diazene



3S-Z, 6R-E
netropsin-diazene

Figure 3-7. The two enantiomers of netropsin-diazene.

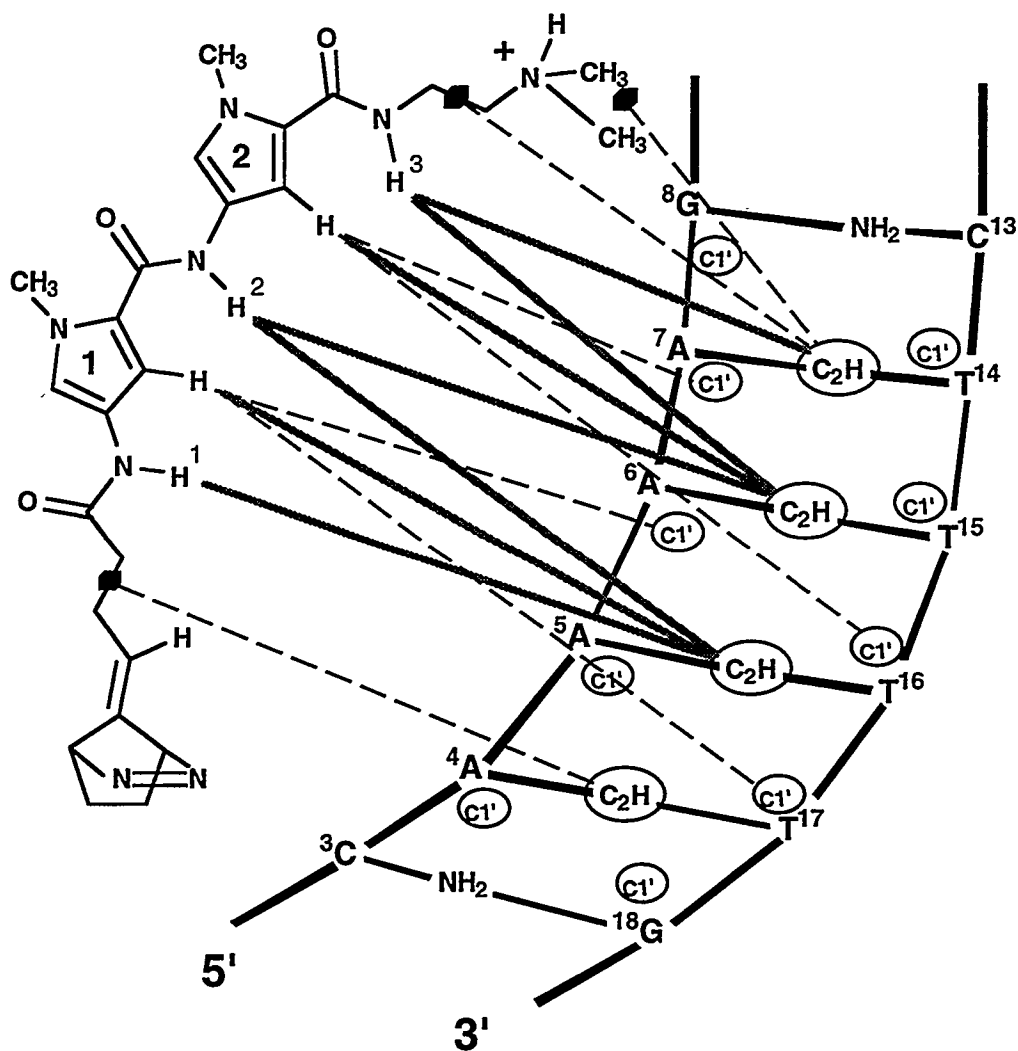


Figure 3-8. Schematic of selected intermolecular NOEs between the netropsin-diazene and d(CGCAAAGGC):d(GCCTTTTGCG) in the major form of the complex. The black diamonds are pseudoatoms.

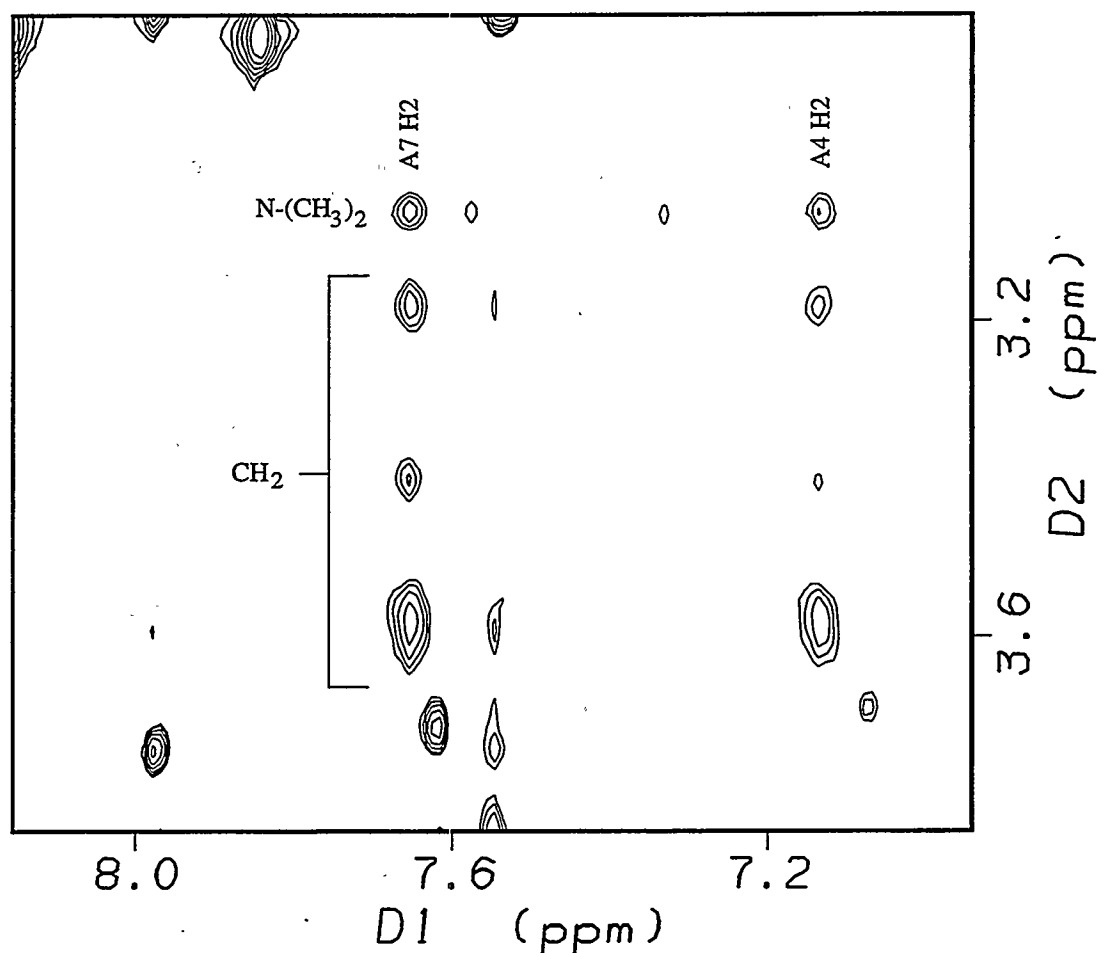


Figure 3-9. Expansion of the aromatic to upfield region of a NOESY spectrum of the 1:1 complex of netropsin-diazene with d(CGCAAAGGC):d(GCCTTTTGCG) (in D₂O, 25 °C, τ_M = 200 ms). Resonances assigned to nonexchangeable ligand protons and to the A₄ H₂ and A₇ H₂ protons are labeled. The presence of these two sets of cross peaks between the ligand and DNA protons at either end of the A-tract clearly indicates that the ligand is exchanging between two conformations.

a minor alternate conformation, and is consistent with the occurrence of flip-flop exchange between the two conformations (Figure 2-6) (Klevit et al., 1986). The set of NOEs which defines the major form contains cross peaks from the methylene protons 9a, 9b, 10a, and 10b to the A₇ H₂ proton (Figure 3-9) and cross peaks from the N-(CH₃)₂ protons to the A₇ H₂ proton as well as to G₈, G₉, and T₁₄ deoxyribose H1' protons. The minor form is defined by cross peaks from the same ligand methylene protons to the A₄ H₂ proton (Figure 3-9), and from the N-(CH₃)₂ protons to the A₄ H₂ proton and to the G₁₈ and C₁₉ deoxyribose protons. NOEs are observed between NH-1 of the ligand and A₅, A₆, and A₇ H₂ protons.

3.4.4 Molecular Modeling. A total of 16 intermolecular ligand-DNA distance restraints derived from NOE data were used to obtain the energy-minimized model of the netropsin-diazene:DNA complex with AAAA:TTTT (Figure 3-10). In the major form of the complex the ligand fills the minor groove, with the positively charged tail group pointing toward the 3' end of the A tract. Although the dimethylammonium proton is likely to be involved in hydrogen bonding interactions with the DNA, specific acceptors cannot be resolved by the current data and model. The minor form of the complex was not modeled, but is undoubtedly very similar.

3.5 Discussion

3.5.1 DNA Cleavage. The most important feature of the netropsin-diazene:DNA complex is the placement of the bicyclic diazene unit in the minor groove near the H1', H4' and H5' protons (the other sugar protons are in the major groove) of bases C₃, A₄, G₁₈, and C₁₉ in the major form of complex. These protons are the likely targets for radical initiated cleavage reactions (Hatayama & Goldberg, 1978; Hatayama & Goldberg, 1979; Nicolaou & Dai, 1991; Goldberg, 1993). The diazene is localized to a very short segment of the DNA by the binding properties of the netropsin section of this molecule. However, from this study there are insufficient data to characterize the precise orientation of the

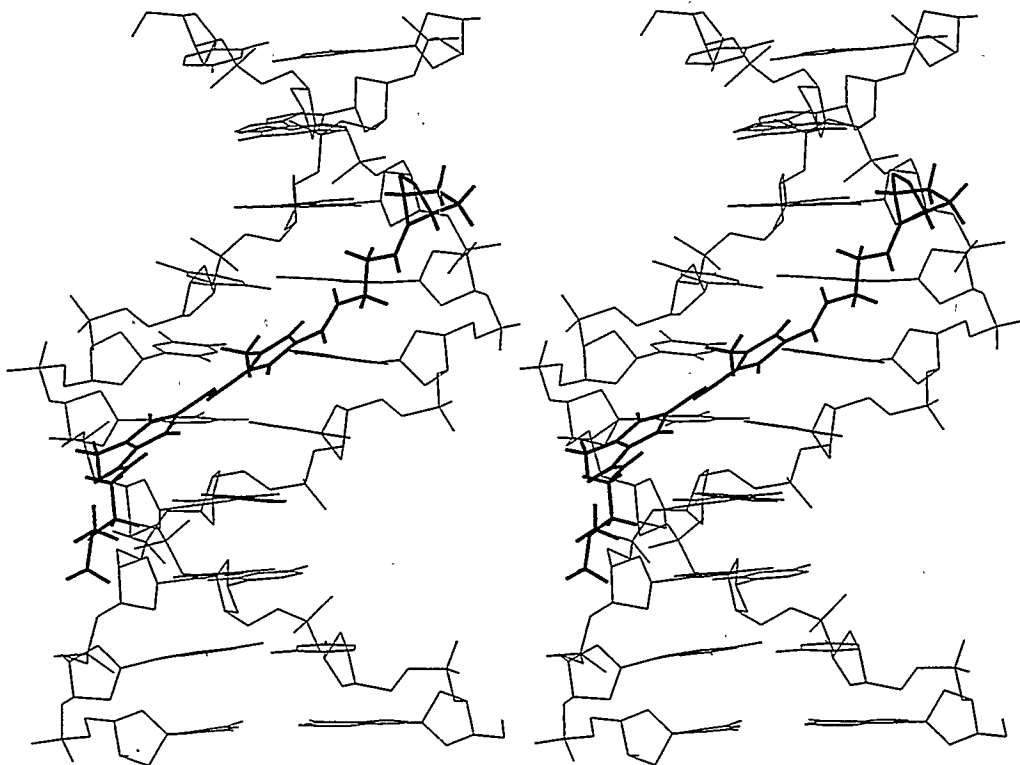


Figure 3-10. Stereo view of a molecular model of the 1:1 netropsin-diazene:DNA complex with d(CGCAAAAGGC): d(GCCTTTTGCG) obtained by energy minimization using semiquantitative distance restraints derived from NOESY data. For clarity, hydrogen atoms are omitted for the DNA but not for the ligand molecule.

diazene moiety with respect to the DNA. The diazene moiety may not have a preferred orientation in the complex with respect to the DNA, and is likely to be in equilibrium among several bound conformations. This is supported by the complete lack of intermolecular NOESY cross peaks from the diazene moiety to the DNA in the presence of intramolecular NOESY cross peaks from the diazene to the netropsin moiety.

Based on the NMR derived model, the delocalized π -TMM diyl that is formed upon photolysis of the diazene is localized in the area near the backbone of the DNA. One possible cleavage mechanism involves the abstraction of a hydrogen atom from the sugar-phosphate backbone to give a DNA centered radical that can react further, ultimately leading to strand scission (Billera & Little, 1994). Other diyl dependent cleavage mechanisms are also accommodated by the minor groove binding and localization of the diazene. Thus far, it is known that DNA strand cleavage is diyl-dependent, but it is unclear by what mechanism the diradical generated from the netropsin-diazene effects DNA strand cleavage. Work is in progress to characterize the cleavage chemistry (Allan, A. and Little, R. D., unpublished results).

Recent studies have shown that photolysis of netropsin-diazene in the presence of DNA leads to the formation of single stranded breaks in the phosphodiester backbone of DNA (Bregant et al., 1994), and characterization of the cleavage products to determine the exact site of cleavage is under way. It appears that double strand breaks are more important than single strand lesions with respect to mutagenicity and cytotoxicity. Single stranded breaks are rapidly repaired in mammalian cells, but persistent double stranded breaks are associated with cell-killing.

3.5.2 Binding Site. Intermolecular ligand-DNA NOE contacts unambiguously confirm the formation of the netropsin-diazene:DNA complex to a four base pair binding site in the minor groove of DNA. The ligand spans the AAAA:TTTT binding site. In the major form of the complex, the bicyclic diazene moiety of the ligand points toward the 5'-end of the A-rich DNA strand. The ligand arrangement is determined by specific hydrogen

bonds between the amide protons of the ligand and the DNA. Evidence for these hydrogen bonds is provided indirectly by NOE cross peaks between the ligand amide protons and DNA A₄, A₅, A₆, and A₇ base H₂ protons in the H₂O NOESY spectrum. Additionally, NOESY contacts to the DNA from the ligand H₃ pyrrole protons, methylene protons on both ends of the ligand and the N-(CH₃)₂ protons support this model.

These results are in agreement with the observed cleavage pattern on a 517 base pair fragment of DNA in which the highest frequency of cleavage occurred at the 5' end of netropsin binding sites (Bregant et al., 1994). Although the present study demonstrates a specificity for the AAAA site, a complete characterization of the sequence specificity of netropsin-diazene for A,T containing sites would require independent determination of the binding affinity of netropsin-diazene for all possible four base pair DNA sequences.

3.5.3 Complex Stoichiometry. Exclusively 1:1 ligand:DNA complexes were observed during the titration (Figure 3-4). At approximately 2:1 stoichiometry, some additional broadening of lines occurs, but no significant changes in chemical shifts or new sets of resonances are observed. Diyl dimerization side reactions are prevented by the physical separation of netropsin-diazene molecules from each other. Two diazenes could be placed in close proximity to each other in either a parallel side-by-side 2:1 ligand:DNA complex, or in an antiparallel end-to-end 2:1 ligand:DNA complex. No evidence was observed to suggest that either of these two cases exists in this system. There is no evidence for a 2:1 binding mode of netropsin on any DNA sequence.

This result agrees with the previously characterized monomeric complexes of distamycin with AAAA:TTTT (Wemmer et al., 1990) and AATT:AATT (Pelton & Wemmer, 1988), and the complex of netropsin with AAAA:TTTT (Chapter 2). The relative positions and geometries of the ligands bound in the minor groove are very similar in all of these complexes.

3.5.4 Ligand Orientation. Two ligand orientations in the complex were observed. The presence of NOEs between nonexchangeable protons on the dimethylammonium end

of the ligand and A H2 protons at either end of the A-tract (Figure 3-10) cannot be explained by one ligand conformation. These data are consistent with the occurrence of flip-flop exchange between two conformations (Klevit et al., 1986). In the major orientation, the bicyclic diazene is located near the C₃•G₁₈ and A₄•T₁₇ base pairs, and in the minor orientation it is located near the A₇•T₁₄ and G₈•C₁₃ base pairs. Based on relative intensities of ligand-DNA NOE cross peaks in the two ligand conformations, the major form is preferred over the minor form in approximately a 3:1 ratio. The directionality of netropsin-diazene binding is intermediate between the complexes of distamycin (> 20-fold) and netropsin (almost none) with AAAA:TTTT (Chapter 2). The preferred orientation is consistent among all three ligands.

3.5.5 Binding Affinity. Netropsin-diazene binds to AAAA in fast exchange on the NMR time scale (Figure 3-4). The parent compounds distamycin A and netropsin have been shown to bind DNA with submicromolar affinity, and in slow exchange on the NMR time scale (Marky et al., 1985; Breslauer et al., 1987; Wemmer et al., 1990). The increased exchange rate of netropsin-diazene suggests that its binding affinity is reduced relative to that of distamycin and netropsin.

Differences in shape, charge distribution or hydrogen bonding may be responsible for the lowered DNA binding affinity of netropsin-diazene. Netropsin is a di-cation while netropsin-diazene is a monocation at physiological pH. The positively charged tail groups in these molecules contribute to the electrostatic attraction to the polyanionic DNA. The more positive charge on a ligand, the stronger the binding to DNA. The hydrophobic bicyclic diazene moiety has a pK_a of -1.4, and therefore remains uncharged at physiological pH (Nelsen et al., 1986). The bicyclic diazene moiety is hydrophobic which may prevent favorable interactions with the functional groups in the minor groove. Additionally, the bulky bicyclic diazene is positioned in the minor groove in the area of the exocyclic 2-amino group of the C₃•G₁₈ base pair, which may prevent the diazene from

binding deeply into the groove. The substitution of an A,T base pair at that position in the duplex may be more sterically favorable for binding.

Netropsin-diazene differs further from netropsin and distamycin in that the cationic tail consists of a dimethylammonium group instead of an amidinium group (see Figure 3-2c). P3 is a distamycin derivative in which an acetyl moiety is substituted for the formyl group, and a dimethylammonium group is substituted for the amidinium group. P3 binds calf thymus DNA with a 28-fold lower affinity than distamycin (Bruice et al., 1992). It is likely that the lowered affinity of P3 for DNA is at least partially due to the dimethylammonium tail, and may also account for part of the difference in the affinities of netropsin-diazene and netropsin for DNA.

3.5.6 Molecular Design. A great deal of progress has been made towards understanding the sequence specific recognition of DNA sequences by minor groove binding drugs. X-ray and NMR studies of complexes of netropsin and distamycin with various DNA oligomers have proven useful in designing oligo(pyrrolicarboxamides) for recognition of longer A,T tracts of DNA (Geierstanger et al., unpublished). The synthesis of distamycin and netropsin analogs that bind to G•C base pairs has made it possible to target many different four or five base pair sequences in DNA for specific binding (Geierstanger et al., 1994a; Geierstanger et al., 1994b).

Distamycin analogs and lexitropsins can be modified both to carry out additional tasks, such as alkylation and cleavage (Schultz & Dervan, 1983; Schultz & Dervan, 1984; Baker & Dervan, 1985; Baker & Dervan, 1989; Church et al., 1990; Chen et al., 1993; Lee et al., 1993b; Sigurdsson et al., 1993; Zhang et al., 1993; Lee et al., 1994). The TMM based DNA cleaving moiety adds new versatility to these molecules. The netropsin-diazene molecule demonstrates that TMM can be used to cleave DNA, and the binding and placement of the diazene could be optimized by building more sophisticated systems. New sequence selective reagents based on the TMM could be generated by increasing the number of pyrrolicarboxamide and/or introducing imidazolecarboxamide units into the

ligand, or by linking molecules with selective binding properties together (Youngquist & Dervan, 1985).

3.6 Conclusions

The DNA binding modes of a new rationally designed DNA cleaving agent based on a TMM diyl have been characterized using the powerful tools of two-dimensional NMR spectroscopy. The molecule appears to have fulfilled all of the criteria used in the design. Netropsin-diazene binds to the DNA duplex studied in the minor groove of the central AAAA tract in a 1:1 mode, preventing diyl dimerization and presumably preventing other side reactions from occurring. The data do not, however, allow us to describe the precise orientation of the TMM diyl in the minor groove. Further studies on the binding properties and cleavage mechanism of netropsin-diazene and related molecules will prove useful in the further development of this new class of DNA nucleases.

Chapter 4

An NMR Study of [d(CGCGAATTCGCG)]₂ containing an Interstrand Cross-Link Derived from a Distamycin-Pyrrole Conjugate

4.1 Summary

Minor groove-binding compounds related to distamycin A bind DNA with high sequence selectivity, recognizing sites which contain various combinations of A•T and G•C base pairs. These molecules have the potential to deliver crosslinking agents to the minor groove of a target DNA sequence. I have studied the covalent, DNA-DNA crosslinked complex of 2,3-bis-(hydroxymethyl)pyrrole-distamycin and the duplex [d(CGCGAATTCGCG)]₂. The alkylating pyrrole design is based on the pharmacophore of mitomycin C, and is similar in substructure to another important class of natural products, the oxidatively activated pyrrolizidine alkaloids. Ligand-DNA NOEs confirm that the tri(pyrrolocarboxamide) unit of the ligand is bound in the minor groove of the central A•T tract. However, it is shifted by one base pair with respect to the distamycin A binding site on this DNA sequence. The crosslink bridges the 2-amino position of two guanine residues, G₄ and G₂₂. The C₃•G₂₂ and G₄•C₂₁ base pairs exhibit Watson-Crick base pairing, with some local distortion as evidenced by unusual intensities observed for DNA-DNA NOE cross peaks. The model is compared to a related structure of a crosslinked mitomycin C:DNA complex.

4.2 Introduction

Several effective antitumor substances function by alkylating DNA (Iyer & Szybalski, 1963; Pinto & Lippard, 1985). Although the DNA sequence selectivity of these agents can be high, the length of the sequence recognized is invariably quite small. For example, mitomycin C, which has been used clinically to treat a variety of cancers,

selectively crosslinks guanines in the sequence CG to form a crosslink in the minor groove (Teng et al., 1989; Borowy-Borowski et al., 1990; Millard et al., 1990; Norman et al., 1990). It is possible that a molecule which could target a very restricted set of binding sites might be therapeutically more valuable. The crosslinking agent described herein was designed as an initial step toward the goal of achieving selectivity for longer DNA sequences.

Structural studies of complexes of netropsin and distamycin with various DNA oligomers has led to the design of analogs which have been successful in recognizing specific mixed A,T- and G,C-containing sequences (Lown et al., 1986; Wade & Dervan, 1987; Dwyer et al., 1992; Mrksich et al., 1992; Geierstanger et al., 1993; Lee et al., 1993a; Geierstanger et al., 1994a). Most recently an analog has been developed which binds an entirely G,C core site four base pairs long (Geierstanger et al., 1994b). This well-characterized class of molecules has the potential to target a variety of DNA sequences with predictable binding behavior, which could deliver a crosslinking agent to the minor groove of the target DNA sequence.

2,3-bis-(hydroxymethyl)pyrrole, an inefficient DNA crosslinking agent alone, has been synthetically tethered to a distamycin analog to form a 2,3-bis-(hydroxymethyl)pyrrole-distamycin conjugate (XL-Dst) (Figure 4-1) (Sigurdsson et al., 1993). The bis-(hydroxymethyl)pyrrole function mimics in part the functionality present in reductively activated mitomycins or oxidatively activated pyrrolizidine alkaloids (Figure 4-2) (Woo et al., 1993). This sub-structure crosslinks the minor groove amino group of guanines of the sequence d(CG), albeit slowly. The conjugation of this agent to distamycin results in a molecule shown to have a high efficiency of interstrand crosslinking in both a linearized plasmid and synthetic DNA oligomers (Sigurdsson et al., 1993), and specificity for sites bearing the distamycin binding sequence adjacent to a pyrrole crosslinking sequence. The distamycin portion of the ligand is important: the conjugate is 1000-fold more active than 2,3-bis-(hydroxymethyl)-1-methylpyrrole alone (Sigurdsson et al., 1993).

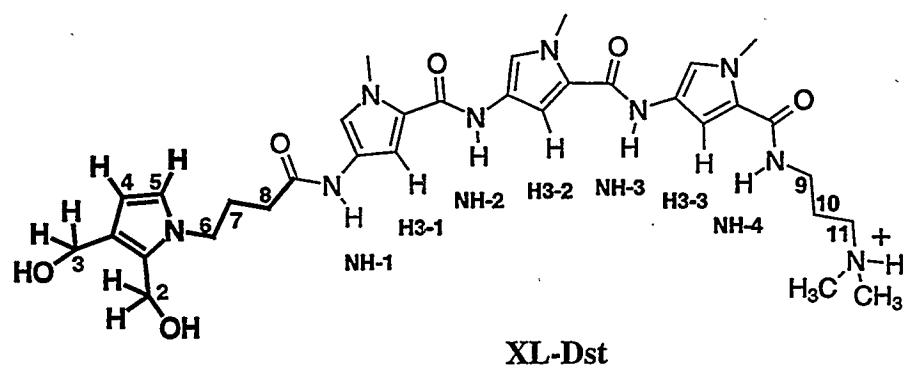
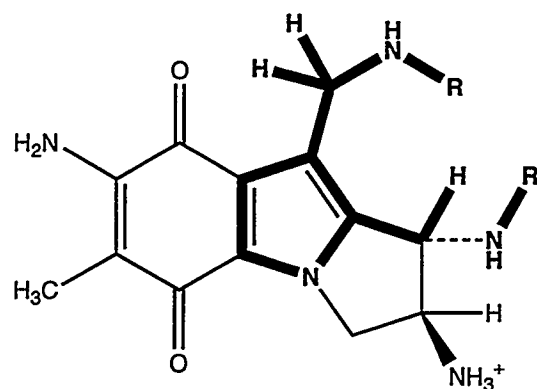
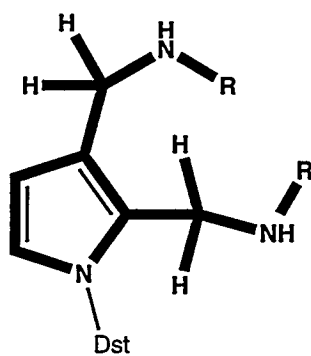


Figure 4-1. Structure of the 2,3-bis-(hydroxymethyl)pyrrole-distamycin conjugate (XL-Dst).



Mitomycin C adduct



XL-Dst adduct

Figure 4-2. Comparison of the structures of the crosslinking agents mitomycin C and 2,3-bis-(hydroxymethyl)pyrrole.

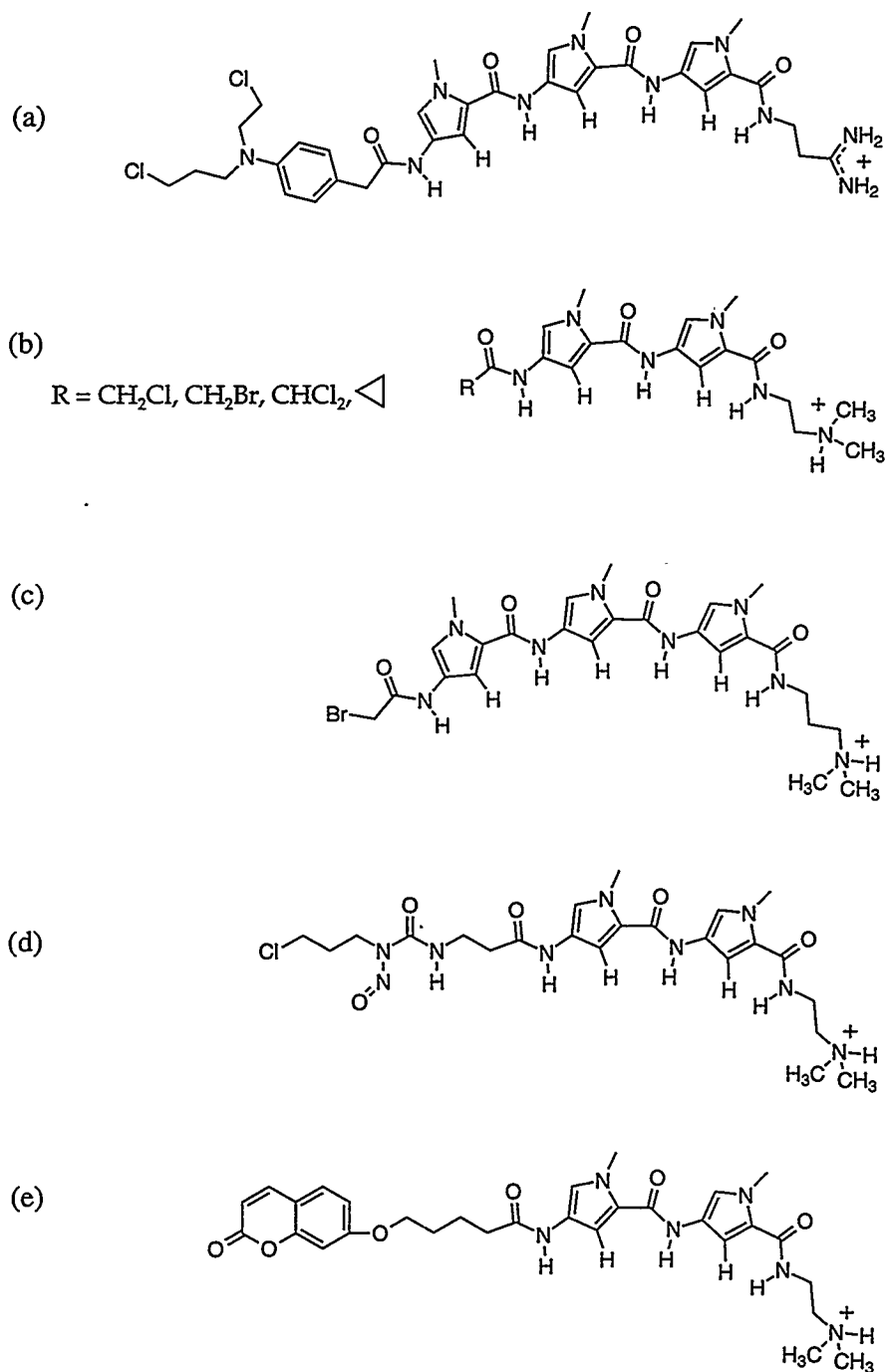


Figure 4-3. Structures of several designed ligands that have been synthesized which contain an alkylating group tethered to a distamycin analog. a) FCE-24517; b) various acetamido-lexitropsins; c) N-bromoacetyl-distamycin; d) chloroethylnitrosourea-lexitropsin; e) coumarin-lexitropsin conjugate.

Other DNA alkylating reagents have been conjugated to members of the distamycin/lexitropsin class of molecules (Figure 4-3). One such agent, FCE-24517, a benzoyl mustard derivative of distamycin, has been shown to be cytotoxic and is currently in phase I clinical trials (Sessa et al., 1994). When tested alone, many of the alkylating agents have no sequence preference and alkylate in the DNA major groove, but in the conjugated molecule alkylate sequence specifically and in the minor groove. Most form DNA monoadducts (Baker & Dervan, 1985; Baker & Dervan, 1989; Church et al., 1990; Broggini et al., 1991; Chen et al., 1993; Zhang et al., 1993). There has been more limited success in designing conjugates which form the generally more cytotoxic interstrand crosslinks (Lee et al., 1993b; Lee et al., 1994). To my knowledge, no detailed structural studies have been done on either covalent or noncovalent complexes of any of these conjugated compounds with DNA.

I have used NMR spectroscopy to determine the binding mode of a 2,3-bis(hydroxymethyl)pyrrole-distamycin conjugate which is covalently crosslinked to both strands of the duplex [d(CGCGAATTCGCG)]₂. I present a semi-quantitative model of the crosslinked DNA, and discuss the ligand binding site. The model is compared to a related structure of a crosslinked mitomycin C:DNA complex.

4.3 Materials and Methods

4.3.1 Synthesis and Purification of Crosslinked Oligonucleotides. XL-Dst was synthesized as described previously (Sigurdsson & Hopkins, 1994). The oligonucleotide d(CGCGAATTCGCG) was synthesized on an automated DNA synthesizer and purified by standard methods (see Chapter 2).

Two crosslinked oligonucleotide samples were synthesized. The first sample was synthesized in the P.B. Hopkins laboratory at the University of Washington, Seattle, according to procedures detailed elsewhere (Fagan et al., in press). A summary will be presented here. The purified oligomer d(CGCGAATTCGCG) (2.3 μ mol duplex) was

dissolved in H₂O solution; and the sample was annealed. The solution was then treated with XL-Dst (7.5 μ moles total) in CH₃OH. The mixture was allowed to stand for at ambient temperature for 4.5 days. The DNA-containing product was recovered by ethanol precipitation. The crosslinked DNA was separated from native DNA and alkylated single strands on a 20% denaturing polyacrylamide gel. The interstrand crosslinked DNA was eluted from the gel and purified through a Waters Sep-Pak C18 cartridge, to yield 137 OD₂₆₀ of interstrand crosslinked DNA.

I synthesized the second crosslink sample in the Wemmer lab. One equivalent of XL-Dst dissolved in 50% MeOH/50% H₂O solution was titrated into a solution of 1.7 mM duplex [d(CGCGAATTCGCG)]₂ (10 mM potassium phosphate buffer, pH 7.0) in an NMR tube at 25 °C, with the goal of studying the noncovalent XL-Dst:DNA complex. In an attempt to lower the binding affinity and sharpen the NMR resonances, the solution was made 500 mM in NaCl. The sample then sat for 9 days at 4 °C. The sample was then checked again by NMR, and the resonances were found to be sharper. An overnight experiment was run at 45 °C, and these data confirmed that the ligand had crosslinked the DNA, generating the same complex as the first sample. This sample provided most of the useful data described in this chapter.

The numbering scheme for the crosslinked oligomer is:

5'- C₁ G₂ C₃ **G₄** A₅ A₆ T₇ T₈ C₉ G₁₀ C₁₁ G₁₂ -3' (strand 1)

3'- G₂₄ C₂₃ **G₂₂** C₂₁ T₂₀ T₁₉ A₁₈ A₁₇ G₁₆ C₁₅ G₁₄ C₁₃ -5' (strand 2)

Bold lettering indicates the sites of crosslinking.

4.3.2 NMR Sample Preparation. NMR samples were prepared by dissolving the dried, crosslinked oligonucleotide duplex in 0.5 mL of 10 mM potassium phosphate buffer, 50-500 mM NaCl (pH 7.0), and then lyophilizing to dryness. For experiments carried out in D₂O, the solid was redissolved in 0.5 mL of 99.96% D₂O (Cambridge Isotope Laboratories), and for experiments in H₂O a 90% H₂O/10% D₂O solution was used. The concentration of the NMR sample was 1.0 mM in duplex.

4.3.3 NMR Experiments and Resonance Assignments. A general description of the NMR experiments and data analysis can be found in Chapter 2. The parameters of the NMR experiments performed are given here. 2D NOESY spectra in H₂O were acquired at 30 °C with a spectral width of 13514 Hz, 64 scans per increment and mixing time $\tau_M = 200$ ms, and at 45 °C with a spectral width of 13889 Hz, 64 scans per increment and $\tau_M = 350$ ms. 2D NOESY spectra in D₂O were acquired at 30 °C and 35 °C, with a spectral width of 5000 Hz, 64 scans per increment and $\tau_M = 200$ ms. A 2D TOCSY in D₂O was acquired at 35 °C, with a spectral width of 5000 Hz and $\tau_M = 50$ ms. All spectra were acquired in TPPI mode (Drobny et al., 1978; Marion & Wüthrich, 1983).

4.3.4 Distance Restraints and Model Refinement. A model of XL-Dst crosslinked to [d(CGCGAATTCGCG)]₂ was built based on the NMR data. Ligand-DNA and DNA-DNA distance restraints were generated as described in Chapter 2. Model refinement was performed with and without intramolecular DNA restraints in the region of the crosslink. Some long distance restraints of 4.0 Å to 7.0 Å derived from several missing base aromatic-sugar H1' NOESY cross peaks were used.

The model of the [d(CGCGAATTCGCG)]₂ duplex was constructed using the Biopolymer module of Insight II (Biosym) from standard B-form DNA. A new potential was added to the AMBER forcefield for the adducted guanine N2 atoms. The starting model was generated as described in Chapter 2.

The model refinements were performed in two steps, following the general procedure given in Chapter 2. In both steps, the energy of the complex was initially minimized using 100 steps of a steepest descents algorithm and 3000 steps of conjugate gradient minimization. The final model was created by subjecting the minimized starting model to 12 psec of restrained molecular dynamics (rmd) at 300 K, which was subsequently cooled to 100 K over 6 psec, followed by 5 psec of rmd at 100 K. The model was then energy-minimized by conjugate gradient minimization to a final rms

derivative of <0.05 (kcal/mole)/Å². Analysis of the helical parameters of the final model was performed using CURVES.

4.4 Results

4.4.1 Crosslinked Sample Preparation. The cross-linked dodecamer was prepared by direct admixture of two molar equivalents of XL-Dst with the synthetic dodecamer. Interstrand cross-linked DNA was separated from residual monoadducts and unreacted DNA using preparative denaturing PAGE. The cross-linked material had an electrophoretic mobility roughly half that of the native single strands. The covalent connectivity of a closely related interstrand cross-link at the sequence CGAATT has previously been relatively well characterized (Sigurdsson et al., 1993; Sigurdsson & Hopkins, 1994), and that connectivity is assumed in the present case.

4.4.2 Titration of the AATT Site with XL-Dst. NMR spectra were recorded at 25 °C at a number of points in the titration of the duplex [d(CGCGAATTCGCG)]₂ with XL-Dst. The DNA resonances broaden significantly and shift upon addition of ligand, indicating that the complex is on the fast side of intermediate exchange on the NMR time scale. Upon addition of 500 mM NaCl and increasing the temperature to 45 °C, the resonances sharpen slightly. However, due to the large line widths we were unable to further characterize the noncovalent complex of XL-Dst with [d(CGCGAATTCGCG)]₂ in any detail.

In contrast, when the same DNA duplex is titrated with the parent compound distamycin, distinct sets of resonances for the free and ligand bound DNA exist (Klevit et al., 1986), indicating a much slower dissociation rate for the unmodified ligand. This difference in rate probably reflects weaker binding of the modified ligand in the noncovalent DNA complex.

4.4.3 One-Dimensional NMR of Crosslinked DNA. I performed preliminary NMR studies on three versions of XL-Dst covalently crosslinked to

d(CGCGAATTCGCG)₂. The compounds differed in the length of the (CH₂)_n tether connecting the crosslinking pyrrole with the distamycin unit, where n = 2, 3, or 4. The 1D NMR spectra of the n = 2 and n = 4 crosslinked complexes show little dispersion and very broad imino proton resonances. The NMR spectrum of the n = 3 crosslinked complex was chosen for further study because the resonances are relatively narrow with better chemical shift dispersion.

4.4.4 NMR Resonance Assignments of DNA. Upon formation of the covalent complex, the symmetry of the DNA oligomer [d(CGCGAATTCGCG)]₂ is broken, and the two DNA strands have nondegenerate resonances which can be assigned independently. Chemical shift assignments of the DNA base H6, H8, adenine H2, and thymine methyl proton and sugar H1', H2', and H2'' proton resonances in the free and the crosslinked DNA were made from the D₂O NOESY spectra (see Tables 4-1 and 4-2).

The adenosine H2 assignments were made based on cross peaks from the H2 protons to intrastrand and interstrand deoxyribose H1' protons in the D₂O NOESY. Cross peaks from the H2 protons to thymidine imino protons in the H₂O NOESY confirm these assignments.

Several pieces of evidence define the location of the crosslink. The 2-amino protons on the two adducted guanines resonate at 8.89 ppm and 8.08 ppm (at 45 °C), and give rise to a weak NOESY cross peak. NOE cross peaks are also observed from the two adducted amino protons to their respective imino protons. The resonance at 8.08 ppm also has a contact to the A₅ H2 and therefore is assigned to the G₄ 2-amino proton, so that the resonance at 8.89 ppm must then be assigned to the G₂₂ 2-amino proton.

4.4.5 NMR Resonance Assignments of the Ligand. Part of a NOESY spectrum of the XL-Dst:DNA complex in H₂O is presented in Figure 4-4 showing several ligand-DNA cross peaks. Intense NOE cross peaks are observed from DNA adenine H2 protons to XL-Dst pyrrole H3 and amide NH protons. In identifying the location of the crosslink, the

Table 4-1. Chemical shift assignments of the d(CGCGAATTCGCG)₂ duplex, free and in the distamycin crosslink:DNA complex.^a

	H6/H8			H1'			A H2/T CH ₃		
	Free	Comp	$\Delta\delta$	Free	Comp	$\Delta\delta$	Free	Comp	$\Delta\delta$
Strand 1									
C1	7.61	7.66	+0.05	5.79	5.77	-0.02			
G2	7.93	7.95	+0.02	5.88	5.95	+0.07			
C3	7.25	7.08	-0.17	5.63	5.84	+0.21			
G4	7.82	7.87	+0.05	5.45	6.00	+0.55			
A5	8.08	8.27	+0.19	5.98	6.22	+0.24	7.27	7.55	+0.28
A6	8.07	8.36	+0.29	6.14	6.20	+0.06	7.63	8.17	+0.54
T7	7.07	7.01	-0.06	5.88	5.63	-0.25	1.24	1.34	+0.10
T8	7.36	7.12	-0.24	6.08	5.70	-0.38	1.51	1.48	-0.03
C9	7.45	7.29	-0.16	5.66	5.30	-0.36			
G10	7.89	7.85	-0.04	5.85	5.94	+0.09			
C11	7.31	7.34	+0.03	5.78	5.79	+0.01			
G12	7.93	7.97	+0.04	6.15	6.19	+0.04			
Strand 2									
C13	7.61	7.64	+0.03	5.79	5.80	+0.01			
G14	7.93	7.98	+0.05	5.88	5.88	0.00			
C15	7.25	7.29	+0.04	5.63	5.74	+0.11			
G16	7.82	7.91	+0.09	5.45	5.43	-0.02			
A17	8.08	8.19	+0.11	5.98	6.05	+0.07	7.27	7.40	+0.13
A18	8.07	8.11	+0.04	6.14	5.77	-0.37	7.63	8.23	+0.60
T19	7.07	6.90	-0.17	5.88	5.54	-0.34	1.24	1.14	-0.10
T20	7.36	6.99	-0.37	6.08	5.35	-0.73	1.51	1.37	-0.14
C21	7.45	7.02	-0.43	5.66	5.41	-0.25			
G22	7.89	7.87	-0.02	5.85	5.97	+0.12			
C23	7.31	7.51	+0.20	5.78	6.10	+0.32			
G24	7.93	8.00	+0.07	6.15	6.17	+0.02			

^a Chemical shifts are given in ppm (± 0.01 ppm). The residual HOD resonance is referenced to 4.65 ppm (35 °C) for both the free and the complexed DNA. Bold lettering indicates the sites of crosslinking.

Table 4-2. Chemical shift assignments of the d(CGCGAATTCGCG)₂ duplex, free and in the distamycin crosslink:DNA complex.^a

	H2'			H2''		
	Free	XL	$\Delta\delta$	Free	XL	$\Delta\delta$
Strand 1						
C1	1.93	1.98	+0.05	2.39	2.41	+0.02
G2	2.64	2.68	+0.04	2.71	2.75	+0.04
C3	1.83	1.03	-0.80	2.26	2.00	-0.26
G4	2.64	2.67	+0.03	2.73	2.76	+0.03
A5	2.67	2.82	+0.15	2.90	3.07	+0.17
A6	2.59	2.66	+0.07	2.88	2.85	-0.03
T7	1.95	1.74	-0.21	2.54	2.45	-0.09
T8	2.15	1.82	-0.33	2.53	2.49	-0.04
C9	2.04	1.81	-0.23	2.39	2.12	-0.27
G10	2.62	2.66	+0.04	2.69	2.77	+0.08
C11	1.88	1.91	+0.03	2.32	2.34	+0.02
G12	2.37	2.44	+0.07	2.59	2.68	+0.09
Strand 2						
C13	1.93	1.92	-0.01	2.39	2.40	+0.01
G14	2.64	2.70	+0.06	2.71	2.75	+0.04
C15	1.83	1.82	-0.01	2.26	2.32	+0.06
G16	2.64	2.73	+0.09	2.73	2.77	+0.04
A17	2.67	2.77	+0.10	2.90	2.87	-0.03
A18	2.59	2.30	-0.29	2.88	2.84	-0.04
T19	1.95	1.69	-0.26	2.54	2.42	-0.12
T20	2.15	1.66	-0.49	2.53	2.30	-0.23
C21	2.04	1.05	-0.99	2.39	1.78	-0.61
G22	2.62	2.67	+0.05	2.69	2.69	0.00
C23	1.88	1.91	+0.03	2.32	2.44	+0.12
G24	2.37	2.42	+0.05	2.59	2.68	+0.09

^a Chemical shifts are given in ppm (± 0.01 ppm). The residual HOD resonance is referenced to 4.65 ppm (35 °C) for the free and the crosslinked DNA.

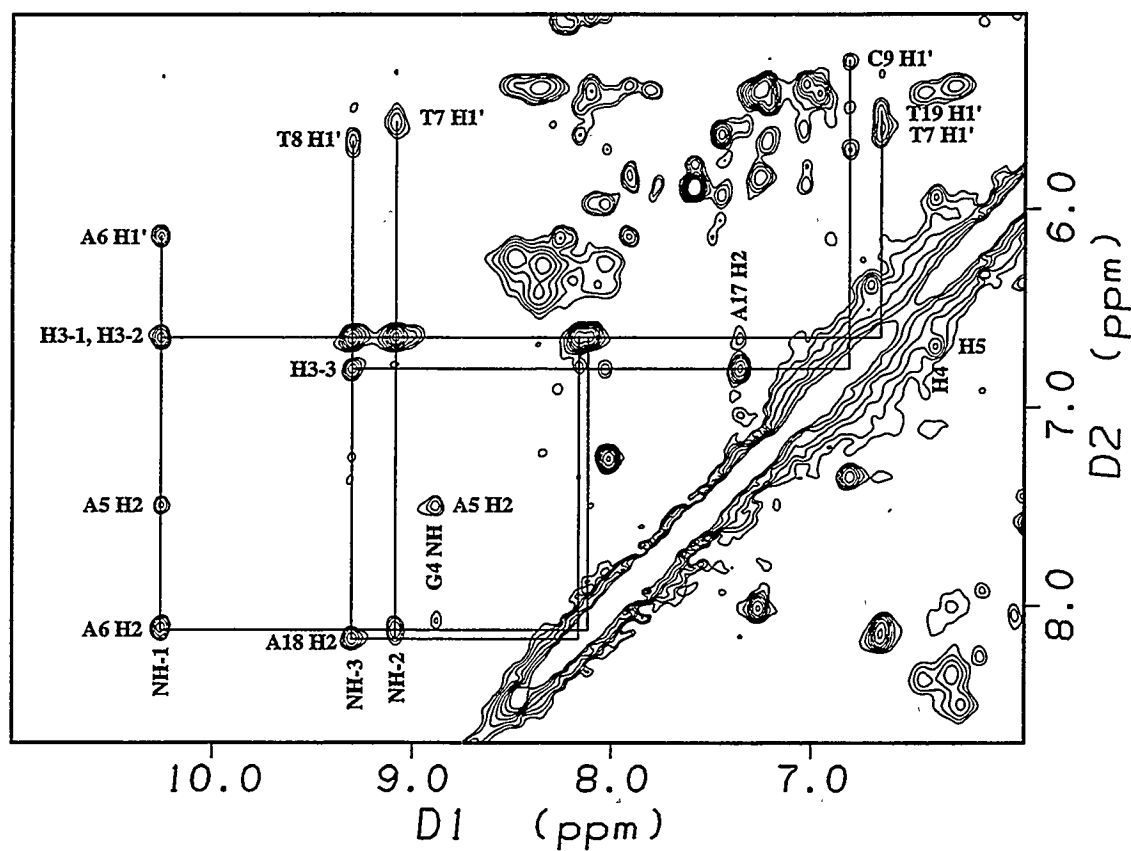


Figure 4-4. Expansion of the amide and aromatic to deoxyribose H1' region of a NOESY spectrum of the XL-Dst:[d(CGCGAATTCGCG)]₂ complex (in H₂O, 45 °C, τ_M = 350 ms). Various ligand-DNA and ligand-ligand NOE cross peaks are labeled.

orientation of the ligand with respect to the DNA is established, thereby facilitating the unambiguous assignment of ligand protons. Chemical shift assignments for the ligand are listed in Table 4-3. It was not possible to stereospecifically assign the methylene protons in the tether or in the $N(\text{CH}_3)_2$ tail.

The H4 and H5 protons on the crosslinking pyrrole are assigned to 6.39 ppm and 6.70 ppm (at 45 °C), respectively. An NOE cross peak was observed between the H4 resonance and the G₂₂ C1'H resonance. The methylene protons on the crosslinking pyrrole (H2a, H2b, H3a, and H3b) were assigned, although not stereospecifically, based on NOEs to the crosslinked guanine amino groups. NOEs are observed from the H3 protons to the ligand H4 and the DNA G₂₂ N2H protons, and from the ligand H2 protons to the DNA G₄ N2H proton.

Several methylene proton resonances were assigned to the ligand protons in the tether, H6(a,b), H7(a,b), and H8(a,b), because they exhibit cross peaks to NH-1 (the first amide of the distamycin fragment) and the crosslinking pyrrole H5. Some of the protons in this group have NOEs to the G₄ N2H and A₅ H2 protons, consistent with the location of the crosslink at G₄ and G₂₂.

4.4.5 Ligand-DNA Contacts. NOE cross peaks localize the ligand in the minor groove of the central AATT:AATT tract of the duplex (Table 4-4). Cross peaks from DNA H2 and H1' protons to ligand pyrrole H3 and amide NH protons indicate that the ligand pyrrole ring system spans the 5'-A₅-A₆-T₇-T₈-3' sequence (Figure 4-5), in the minor groove. Several additional contacts are made between the XL-Dst protons and the DNA. These include NOEs from the crosslinking pyrrole H4 to G₂₂ C1'H, from pyrrole H2 and H3 protons to G₄ N2H and G₂₂ N2H, respectively, and from the ligand tether methylene protons to G₄ N2H.

4.4.7 Molecular Modeling of the Crosslink. A total of 16 ligand-DNA distance restraints derived from NOE data (see Table 4-4) and the two covalent ligand-DNA bonds were used to obtain the energy-minimized model of the XL-Dst complex with

Table 4-3. Chemical shift assignments of XL-Dst in the interstrand crosslinked complex with [d(CGCGAATTCGCG)]₂.^{a, b}

Proton	Chemical shift
H2(a,b) ^c	3.75, 4.07
H3(a,b) ^c	3.80, 4.18
H4	6.37
H5	6.70
H6(a,b), H7(a,b), H8(a,b) ^{c,d}	2.26, 2.27, 2.45, 2.70, 3.73
NH-1	10.25
H3-1	6.65
NH-2	9.09
H3-2	6.66
NH-3	9.30
H3-3	6.80
NH-4	not assigned
H9(a,b), H10(a,b), H11(a,b) ^c	not assigned
N-(CH ₃) ₂	not assigned

^aChemical shifts are given in ppm (± 0.01 ppm), with the residual HOD resonance referenced to 4.52 ppm (45 °C). ^bRefer to Figure 1 for numbering system used.

^cResonances assigned as a group. ^dOne resonance in this group not assigned.

Table 4-4. Ligand-DNA contacts for the XL-Dst:[d(CGCGAATTCGCG)]₂ interstrand crosslinked complex.

XL-Dst	DNA	NOE classification ^a
NH-1	A ₅ H2	w
NH-1	A ₆ H1'	m
H3-1	A ₆ H2	s
H3-1	T ₂₀ H1'	m
H3-1	T ₇ H1'	m
NH-2	A ₆ H2	w
NH-2	T ₇ H1'	m
H3-2	A ₁₈ H2	s
H3-2	T ₁₉ H1'	w
H3-2	T ₈ H1'	w
NH-3	A ₁₈ H2	m
NH-3	T ₈ H1'	w
H3-3	A ₁₇ H2	s
H3-3	A ₁₈ H2	w
H3-3	C ₉ H1'	m
crosslinking pyrrole H4	G ₂₂ H1'	m

^aNOE classification: s = strong (1.8 to 2.8 Å), m = medium (1.8 to 3.5 Å), w = weak (1.8 to 5.0 Å).

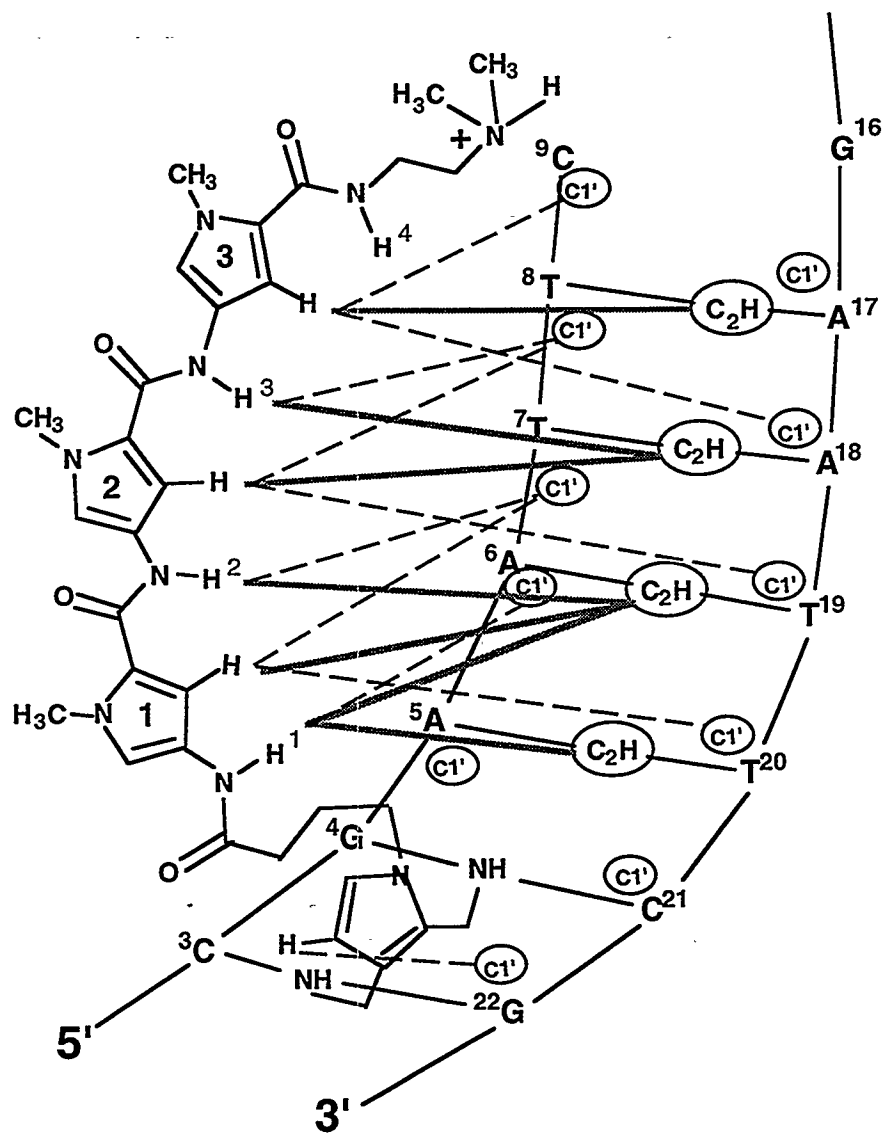


Figure 4-5. Schematic of the ligand-DNA NOE contacts used in modeling the interstrand crosslinked XL-Dst:DNA complex.

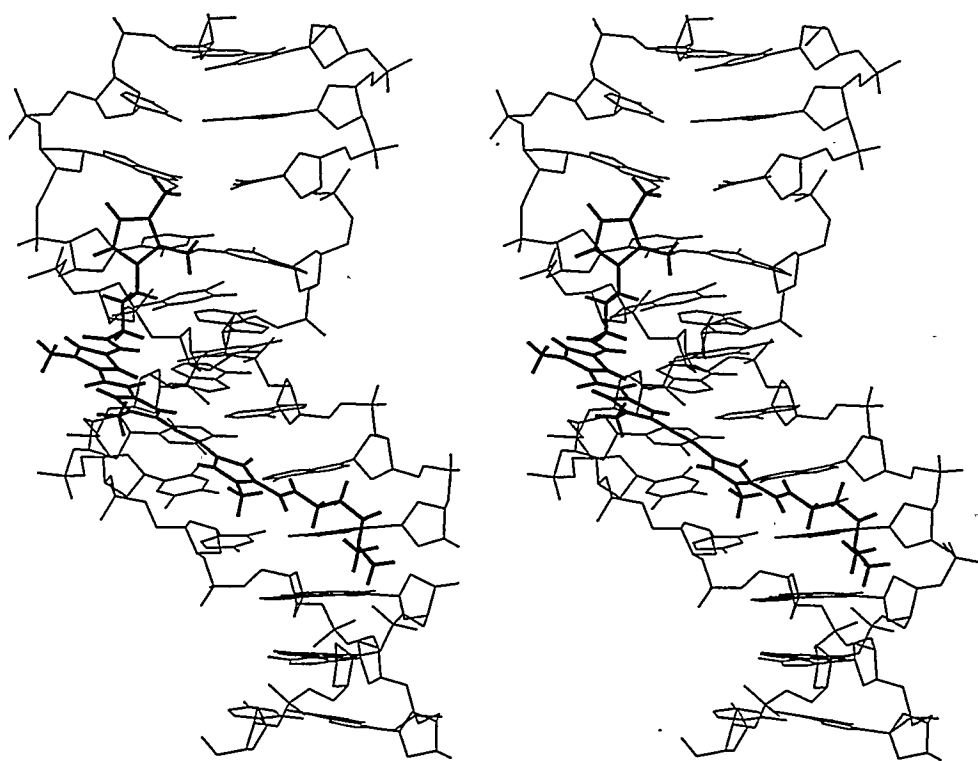


Figure 4-6. Stereo view of a molecular model of the XL-Dst:[d(CGCGAATTCGCG)]₂ complex obtained by energy minimization using semiquantitative distance restraints derived from NOESY data. For clarity, hydrogen atoms are omitted for the DNA but not for the ligand molecule.

CGAATT:AATTCG (Figure 4-6). There is some local distortion of the DNA helix at the crosslinking site which does not extend beyond the base pairs on either side of the crosslink. Although the dimethylammonium proton is likely to be involved in hydrogen bonding interactions with the DNA, specific acceptors cannot be resolved by the current data and model.

4.5 Discussion

4.5.1 Site of Crosslink. XL-Dst adducts guanine bases (Sigurdsson & Hopkins, 1994) and both N2 amino groups of deoxyguanosine residues at the sequence CG are necessary for efficient DNA-DNA interstrand crosslinking (Sigurdsson et al., 1993). The C₃•G₂₂ and G₄•C₂₁ base pairs are stabilized by the interstrand crosslink. The G₄ and G₂₂ imino proton resonances exhibit slow exchange with solvent water even at 45 °C, where other imino proton resonances are broadened due to increased exchange with solvent. In the interstrand crosslinked complex of mitomycin C with [d(TACGTA)]₂, the imino proton resonances on the two crosslinked guanines are similarly unaffected by increasing temperature (Norman et al., 1990).

The NMR data confirm that the lesion occurs at the guanine N2 position. In unadducted DNA oligomers, the guanine 2-amino proton resonances are broad due to the rotation of the amino group about the C2-N2 bond on a millisecond time scale. Covalent attachment of a ligand to N2 prevents this rotation, resulting in one narrow NMR resonance for the remaining amino proton. The two substituted amino proton resonances on G₄ and G₂₂ have contacts to nearby ligand and DNA protons and intense NOE cross peaks to the imino proton on the same guanine base. Similar patterns are observed in the crosslinked complex of mitomycin C with [d(TACGTA)]₂ (Norman et al., 1990), in which mitomycin C forms interstrand crosslinks in the minor groove at N2 of guanines in the sequence CG. The chemical shifts of the substituted G₄ and G₂₂ amino (8.89 and 8.08 ppm) and imino (13.17 and 12.45 ppm) protons in the XL-Dst:DNA crosslinked complex are similar to

those observed in the mitomycin C complex (aminos at 9.36 and 8.87 ppm, and iminos at 13.12 and 12.60 ppm).

Additional evidence that the crosslink is in the minor groove arises from NOEs from hydroxymethyl pyrrole protons and ligand tether protons to DNA nucleotides G₄, A₅ and G₂₂. I was unable to confirm scalar coupling between the ligand hydroxymethyl pyrrole H2 protons and the G₄ amino proton, and between the hydroxymethyl pyrrole H3 protons and the G₂₂ amino proton.

4.5.2 Distortion of the DNA Due to Crosslink. The model of the XL-Dst:[d(CGCGAATTCGCG)]₂ complex is shown in Figure 4-6. Weak intraresidue NOE cross peaks from base protons to H1' protons indicate that the bases are all in *anti* geometry. Several extremely weak or missing cross peaks are indicative of distortion in the two crosslinked base pairs which extends to the base pairs on either side of the crosslink. In the refined model we observe a large negative buckle for the C₃•G₂₂ base pair and a large positive buckle for the G₄•G₂₁ base pair. Additionally, the model shows a decrease in helical twist between these two base pairs, indicating unwinding at the crosslink site.

The slide between the two base pairs at the crosslink site is unusual compared to standard B-DNA, so that C₃ and C₂₁ are moved from the edges of the groove toward the center of the helix, and G₂₂ and G₄ are pushed away from the center of the helix (Figure 4-7). As a result, the crosslinked guanine bases are positioned over the deoxyribose H2' and H2'' protons of C₃ and C₂₁. The NMR data of the crosslinked complex agree with this model. The H2' resonances of C₃ and C₂₁ are dramatically upfield shifted, by 0.80 ppm and 0.99 ppm respectively, relative to the uncomplexed DNA. An upfield shift of this magnitude has not been observed in any complexes of distamycin and related analogs with DNA in the Wemmer laboratory. I conclude that these unusual shifts are due to distortion of the DNA structure by the crosslink. The cytosine H2' and H2'' resonances undergo an upfield ring current shift due to stacking over the aromatic guanine base as seen in the model, similar to ring current shifts observed for protons which are stacked over aromatic

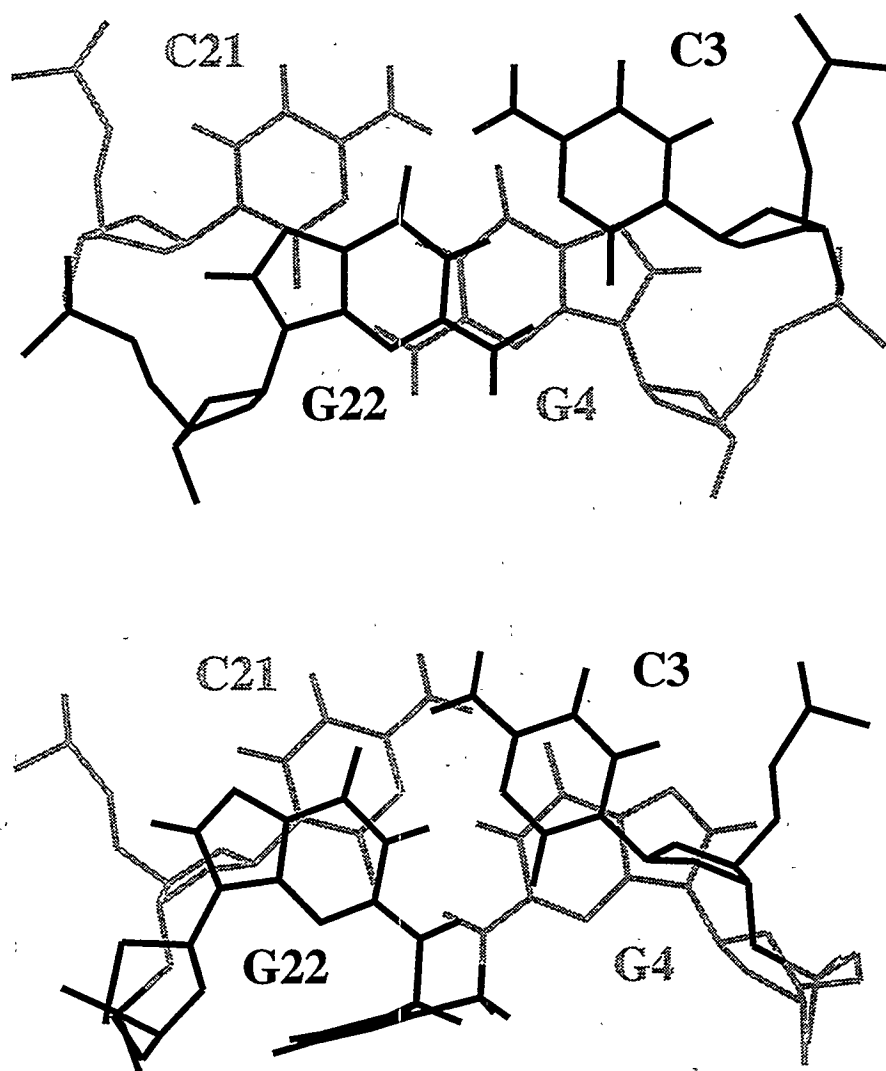


Figure 4-7. View down the helical axis of (a) a standard B-form DNA duplex of the sequence CG (built using Insight II software), and (b) the base pairs C₃•G₂₂ and G₄•C₂₁ in the XL-Dst:AATT crosslinked complex. Explicit hydrogen atoms are shown only for the aromatic bases and the crosslinking pyrrole unit.

amino acid sidechains in protein NMR studies.

4.5.3 Overall Binding Site. An important feature of XL-Dst is that it requires both a distamycin binding sequence and an adjacent pyrrole crosslinking sequence to crosslink DNA efficiently, resulting in a six base pair binding site CGAATT. Ligand-DNA NOEs unambiguously confirm the formation of the XL-Dst:DNA complex to this site in the minor groove. The tether is fully extended, bridging the A₁₇•T₈ base pair. It is not clear whether the tether can tolerate a G•C base pair at this site. The amino group at the 2 position of the guanine may sterically interfere with the ligand binding and/or reactivity at the crosslinking site. The N-(CH₃)₂ tail extends across the G₉•C₁₆ base pair, but no specific contacts are observed.

From the model, hydrogen bonds can be inferred between ligand amide protons and adenine N3 and thymine O2 as seen in other models of distamycin complexed to DNA (Klevit et al., 1986; Pelton & Wemmer, 1988; Pelton & Wemmer, 1989). The tripyrrole unit contacts three adenine H2 protons in the sequence ATT, with the "head" of the ligand (pyrrole ring number 1 in Figure 4-1) nearest the adenine, pointing toward the 5' direction of the strand. The noncovalent complex of distamycin A with [d(CGCGAATTCGCG)]₂ is depicted in Figure 4-8. It appears that the covalent linkage in the crosslinked complex forces the tripyrrole unit to shift by one base pair toward the 3' end of the first strand, relative to the distamycin binding site. However, it is difficult to predict the preferred subsite, AAT or ATT, for the noncovalent binding of XL-Dst to d(CGCGAATTCGCG)₂ by analogy to these complexes.

4.5.4 Crosslinking Mechanism. XL-Dst is a 1000-fold more efficient crosslinker than 2,3-bis(hydroxymethyl)pyrrole alone (Sigurdsson & Hopkins, 1994). Studies of efficiency of XL-Dst crosslinking at different DNA sequences revealed that both a distamycin binding site and a CG crosslinking site are necessary for efficient DNA-DNA interstrand crosslinking. The yield of crosslinked DNA is significantly higher for the site AATTCG, as compared to the site CGCGCG, suggesting that noncovalent binding is

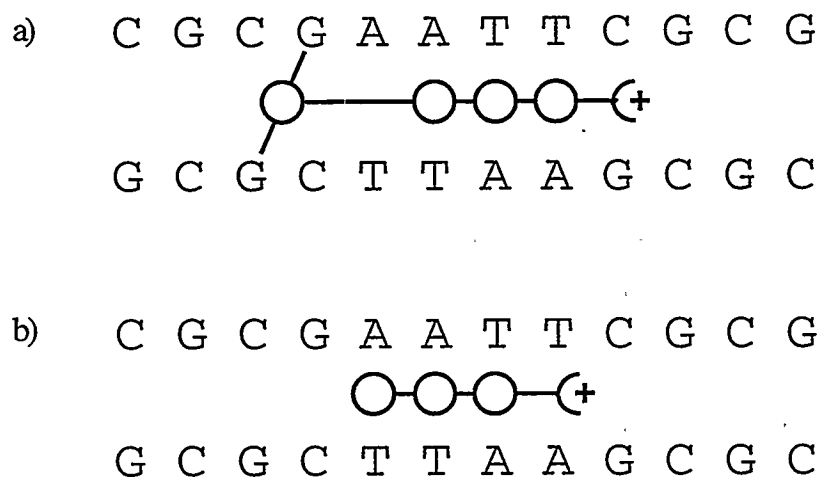


Figure 4-8. Schematic of (a) the XL-Dst:CGAATT crosslinked complex, and (b) the distamycin:AATT complex.

associated with the crosslinking event (Sigurdsson & Hopkins, 1994). Presumably, noncovalent binding orients the bis(hydroxymethyl)pyrrole unit so that efficient crosslinking can occur. Therefore, the noncovalent complex of XL-Dst with DNA is of some interest, and would be useful in understanding the mechanism of crosslinking. In this study, I have examined only the crosslinked XL-Dst:AATT complex, since the resonances of the noncovalent complex are broad due to intermediate exchange on the NMR time scale.

To my knowledge, there is no system in which both the structures of the noncovalent ligand-DNA complex and the crosslinked DNA are known. Systematic measurements of crosslinking kinetics by XL-Dst with various DNA sequences should prove useful in characterizing the reactive noncovalent complex between XL-Dst and AATT.

4.5.5 Comparison to Mitomycin C. The design of the bis(hydroxymethyl)pyrrole unit was modeled from the pharmacophore in mitomycin C (Figure 4-2). There are many similarities between the results observed in this study and those observed in the high resolution structure of the crosslinked mitomycin C:DNA complex (Norman et al., 1990).

In the XL-Dst crosslink, I observe NOE cross peaks indicating base stacking throughout the helix, but observe that several NOE cross peaks between DNA base protons and sugar H1' protons near the C₃•G₂₂ and G₄•C₂₁ base pairs are significantly weaker than in normal B-form DNA. Patel and coworkers observed similar indications of base stacking and unusually weak NOE cross peaks at the mitomycin C crosslink site (Norman et al., 1990).

The overall geometry of the modeled XL-Dst:DNA complex in this study is similar to the calculated structure of the crosslinked mitomycin C:DNA complex. In our model, the crosslinking pyrrole ring interacts with the bases C₂₁ and G₂₂ on one strand of DNA rather than being positioned in the center of the minor groove. This geometry was proposed by Weidner et al. based on molecular modeling (Weidner et al., 1990). This structure

appears to be a consequence of the covalent geometry of the crosslink, since energy minimizations arrive at this structure from different starting structures and without any ligand-DNA restraints. This structure is consistent with an observed NOESY cross peak between the H4 proton on the pyrrole ring and the H1' proton of G₂₂. Patel and coworkers came to the same conclusion based on their NMR-restrained structure calculations and on unusual chemical shifts of the sugar H1' and H2'' protons on the analogous strand in the region of the mitomycin C:DNA crosslink (Norman et al., 1990).

The minor groove width of GC regions has typically been measured to be about 5-7 Å (Saenger, 1983; Heinemann & Alings, 1989). The nonplanar five-membered ring in mitomycin C may widen the minor groove to 9.2 Å or 10.8 Å, depending on its pucker (Norman et al., 1990). The bis(methoxy)pyrrole unit is planar. The minor groove width in the region of the XL-Dst crosslink (10-11 Å) is similar to that in the mitomycin C adduct.

4.5.6 Molecular Design. Designed minor groove binding peptides have been successful in recognizing specific mixed A,T- and G,C-containing sequences of DNA (Dwyer et al., 1992; Mrksich et al., 1992; Geierstanger et al., 1994a). These ligands derive their high binding affinity and sequence selectivity from a unique antiparallel side-by-side two-drug binding mode (Pelton & Wemmer, 1989). Future directions of this study include generating a two-drug complex which contains a 2,3-bis-(hydroxymethyl)pyrrole unit. The increased selectivity of the side-by-side two-drug binding mode combined with the site selective reactivity of the hydroxymethylpyrrole may decrease the overall frequency of crosslinking, reducing unwanted crosslinks and toxicity.

The properties and dimensions of tethers are becoming increasingly important in ligand design. Tethers have been used to link a variety of reactive groups to lexitropsins (Baker & Dervan, 1985; Baker & Dervan, 1989; Church et al., 1990; Broggini et al., 1991; Chen et al., 1993; Lee et al., 1993b; Zhang et al., 1993; Lee et al., 1994). Two or more lexitropsins have been tethered together end-to-end (Dervan, 1986; Mrksich et al., 1994),

substantially increasing the length of the ligand binding site to ten or more base pairs. Tethering with the use of a flexible linker is necessary to maintain proper phasing of the ligand with respect to the rise of the DNA helix, since lexitropsins consisting of more than four N-methylpyrrole or N-methylimidazole units lose proper phasing with respect to the rise of the DNA helix. Tethers have also been used to link two lexitropsins through the two central N-methylpyrroles, generating an H-shaped molecule which binds with the two lexitropsin units antiparallel and side-by-side (Dwyer et al., 1993; Chen & Lown, 1994). Through this design it is possible to prevent unwanted binding modes from occurring, and to enhance binding affinity over the binding by two monomeric ligands. The finding in this study that the XL-Dst tether bridges one base pair, with the binding site for the tripyrrole unit shifted by one base pair relative to the distamycin binding site, may be due in part to the properties of the tether. This emphasizes the need for further investigation and systematic characterization of different linkers and their minor groove binding properties in 1:1 and 2:1 ligand:DNA complexes.

A critical issue in the design of sequence-specific DNA ligands which contain a reactive group is the effect of the reactive group on the affinity and sequence specificity of the ligand. In the cases of XL-Dst and netropsin-diazene (Spielmann et al., 1995), each molecule contains a bulky hydrophobic group appended to the "head" of the molecule, and the NMR exchange behavior indicates lowered noncovalent binding affinity relative to the parent molecules distamycin or netropsin. Since the noncovalent binding event is critical to the delivery of the reactive group to the proper target site, this highlights the need for studies examining how the presence of a reactive group changes the binding selectivity of such conjugated ligands.

4.6 Conclusions

I have characterized the DNA binding mode of a new rationally designed DNA crosslinking agent based on distamycin using two-dimensional NMR spectroscopy. In this study, we

have chosen a distamycin derivative as a starting point for the rational design of a sequence-selective crosslinker. This system holds excellent promise for the eventual design of a compound which forms interstrand DNA crosslinks with high sequence selectivity for both A,T and G,C containing sequences. Further studies on the binding properties and mechanism crosslinking of this distamycin conjugate and related molecules will prove useful in the development of effective antitumor and antiviral agents.

Chapter 5

An NMR Study of the Conformation of the 2-Aminopurine•Cytosine Mismatch in DNA

5.1 Summary

DNA polymerase makes errors by misincorporating natural DNA bases and base analogs, or by copying an unnatural template base. Because of the wide variety of possible mismatches and the varying efficiency with which they are repaired, structural studies are necessary to understand in detail how these mispairs differ and can be distinguished from standard Watson-Crick base pairs. 2-Aminopurine (AP) is a highly mutagenic base analog which, due to its highly fluorescent properties, is widely used to measure rates of misinsertion and proofreading by DNA polymerase. The objective of this study was to determine the geometry of the AP•C mispair in DNA at neutral pH. Although several studies have focused on the AP•C mispair in DNA, there is not as of yet consensus on its structure. At least four models have been proposed for this mispair. Through the use of NMR spectroscopy with selective ^{15}N -labeling of exocyclic amino nitrogens on bases of interest, it is possible to resolve ambiguities in previous studies. I show here that, in two different DNA sequences, the AP•C mispair at neutral and high pH is in a wobble geometry. The structure and stability of this base mispair is dependent upon the local base sequence. Studies which utilize 2-aminopurine to measure rates of DNA base insertion, proofreading, or repair must now consider the AP•C wobble base pairing. The dependence of the stability of the AP•C mispair on the local sequence context highlights the need for consideration of this effect in such studies.

5.2 Introduction

Fidelity in replication is of the utmost importance in maintaining an error-free genome. Overall mutation rates are typically only 10^{-9} to 10^{-11} (Drake et al., 1969; Saenger, 1983) per base replicated. DNA polymerase makes errors by misincorporating natural DNA bases and base analogs, or by copying an unnatural template base. These misincorporated bases can form stable mispairs in Watson-Crick-like and various non-Watson-Crick geometries with the base on the opposite strand. Several DNA base mismatches have been characterized structurally (reviewed in Hunter, 1992) (Figure 5-1), including the G•T wobble-mispair (Patel et al., 1982; Hare et al., 1986), A•A and T•T mispairs (Gervais et al., 1995), the C•A mispair (Boulard et al., 1992), and the well-studied G•A mispair which adopts several different conformations depending on local sequence context and pH (Privé et al., 1987; Gao & Patel, 1988; Nikonowicz et al., 1991; Cheng et al., 1992; Chou et al., 1992; Greene et al., 1994). The surprising variety of base pairing geometries which have been characterized emphasizes the importance of understanding non-standard base pairing interactions and their biological relevance. Mismatched unnatural DNA bases, including 6,8-diketoguanine (McAuley-Hecht et al., 1994), N4-methoxy-cytosine (Stone et al., 1991; Nedderman et al., 1993) and 2-aminopurine (Sowers et al., 1986) have also been studied by structural methods in attempts to explain their mutagenicities (Figure 5-2).

Polymerase proofreading and mismatch repair machinery are needed to recognize and repair mismatched base pairs. It is not clear by what mechanism the recognition of non-Watson-Crick base pairs occurs, but it has been suggested that DNA polymerase screens for the base-pairing free energy of the pair and the precise geometry of the Watson-Crick base pair (Kornberg, 1980; Echols & Goodman, 1991). Because of the wide variety of possible mismatches and the varying efficiency with which they are repaired, structural studies can contribute to understanding in detail how these mispairs differ and can be distinguished from standard Watson-Crick base pairs.

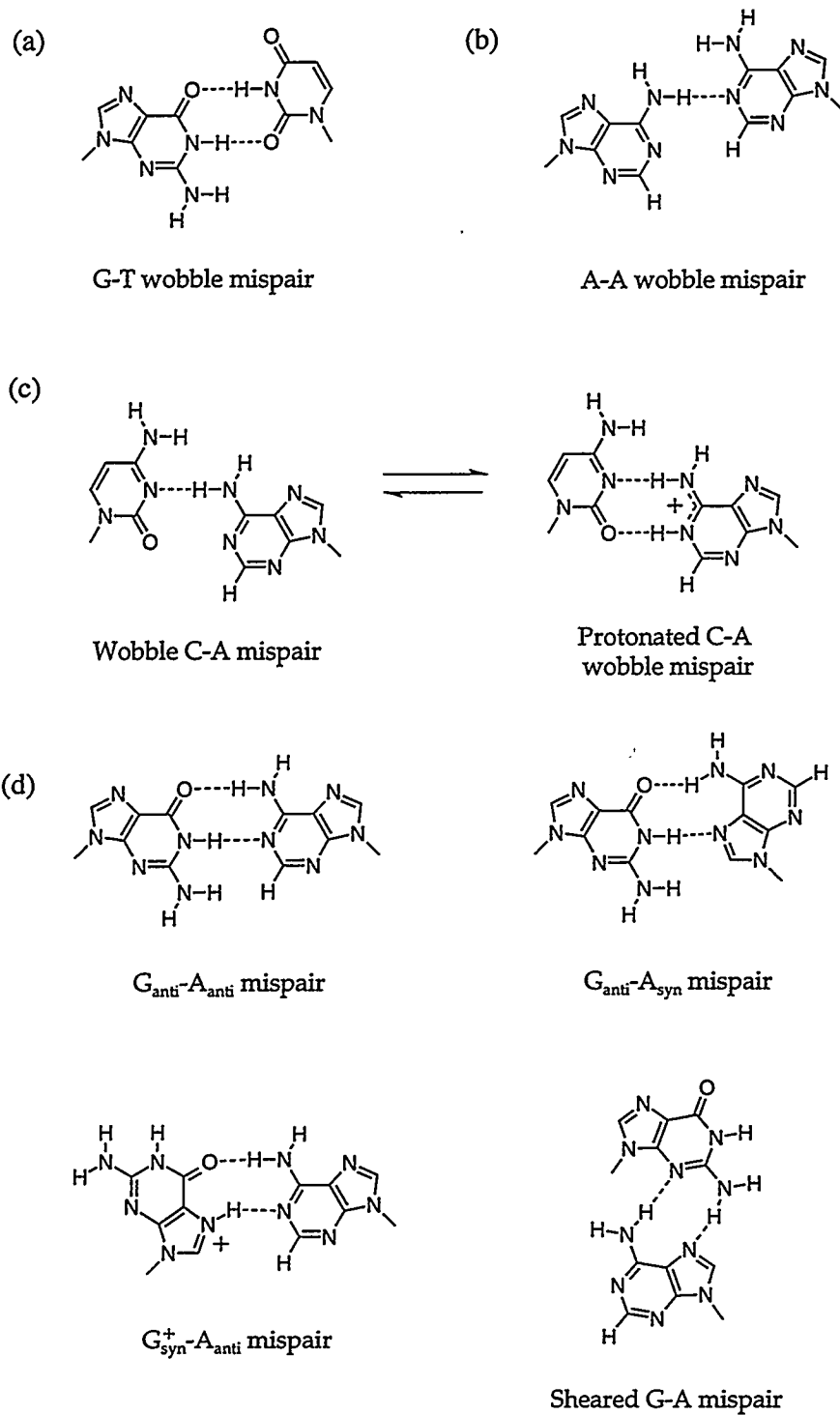


Figure 5-1. Several mismatches between natural DNA bases which have been characterized by structural methods.

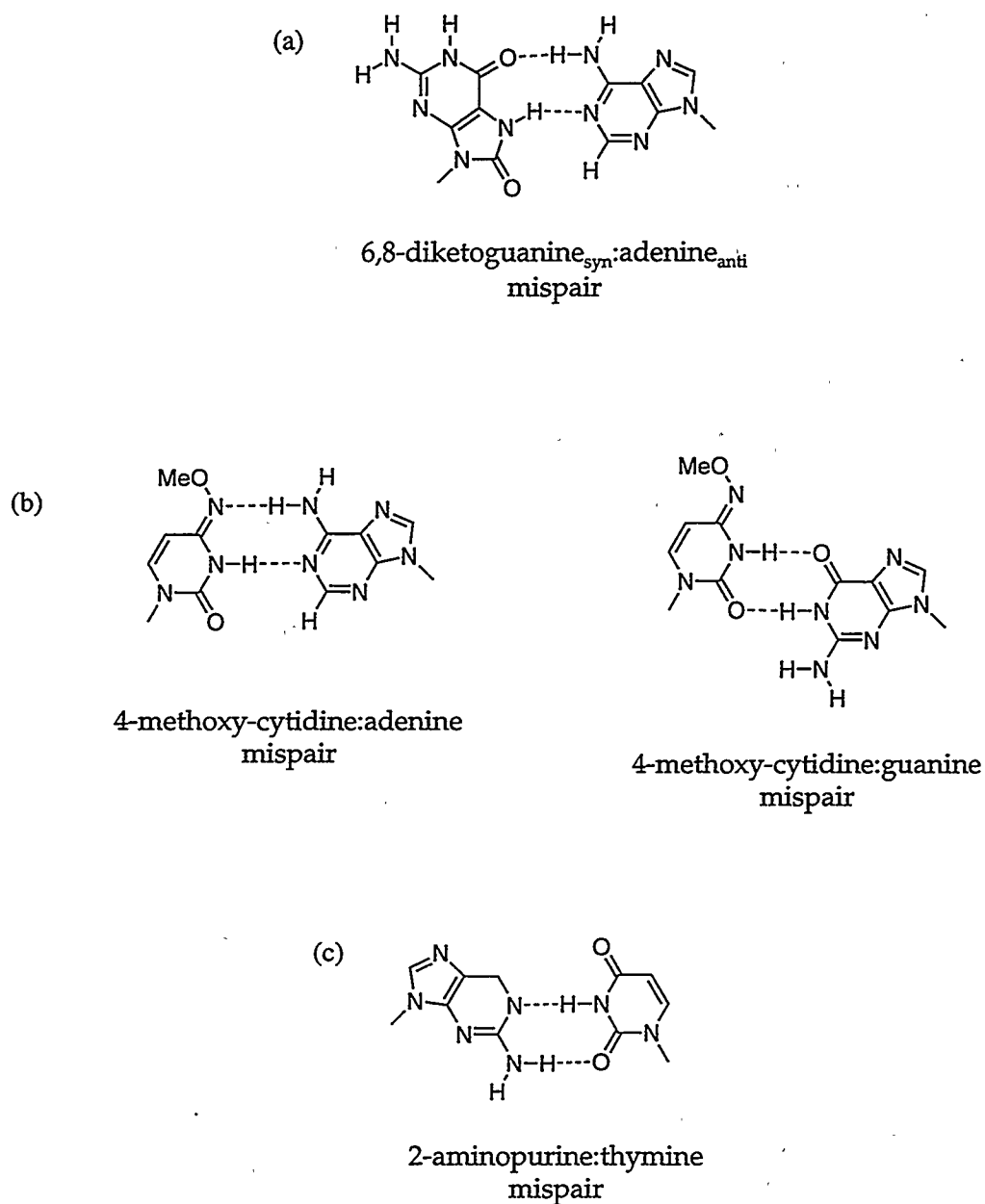


Figure 5-2. Examples of mutagenic base analogs which have been shown to mispair in DNA oligomers.

Normal replication is impaired when base analogs are incorporated into DNA, allowing mutations to arise through subsequent replication steps or repair processes. 2-Aminopurine (AP), a highly mutagenic base analog, has been studied for decades (for a review, see Ronen, 1979). AP is known to cause increased frequency of G•C → A•T and A•T → G•C transitions. The proposed pathway for such transitions requires that AP mispair with both thymine and cytosine (Freese, 1959; Ronen, 1979). As a result, much discussion about the cause of AP mutagenicity has focused on its base pairing interactions with pyrimidines (Watanabe & Goodman, 1982).

AP preferentially mispairs with thymine during replication (Bessman et al., 1974; Watanabe & Goodman, 1981). AP has been shown by NMR spectroscopy to form a stable mispair with thymine in a DNA oligomer, stabilized by two hydrogen bonds in a Watson-Crick geometry (Sowers et al., 1986) (Figure 5-2c). AP incorporation opposite thymine is favored by approximately 25-fold over cytosine (Watanabe & Goodman, 1981; Watanabe & Goodman, 1982). Although several studies have focused on the AP•C mispair in DNA, there is not as of yet consensus on its structure. At least four models have been proposed for this interaction (Figure 5-3): 1) a Watson-Crick geometry involving one hydrogen bond (Ronen, 1979), 2) a protonated cytosine resulting in Watson-Crick geometry, involving two hydrogen bonds (at neutral pH) (Sowers et al., 1986), 3) a wobble geometry, (Sowers et al., 1989) and 4) the imino tautomer of either C or AP, paired with the amino tautomer of the other base (Janion & Shugar, 1973; Goodman & Ratliff, 1983).

The rare imino tautomer form of bases has long been used to explain spontaneous and base analog-induced mutation (Watson & Crick, 1953a; Watson & Crick, 1953b; Freese, 1959; Topal & Fresco, 1976; Saenger, 1983). However, calculated equilibrium constants of imino/amino tautomeric equilibria do not explain observed mutational rates (Ronen, 1979), and structural studies to date have found no evidence of the imino tautomer form of cytosine or 2-aminopurine in DNA oligomers. NMR evidence has been presented

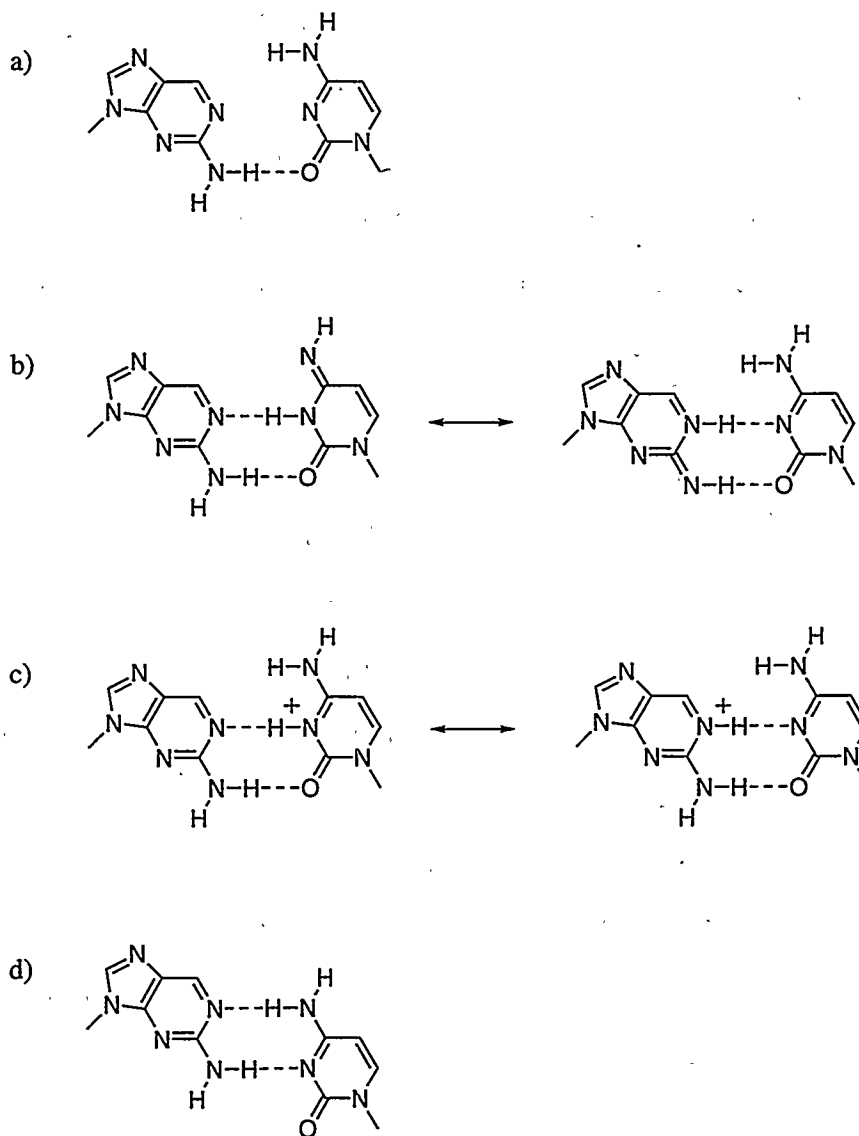


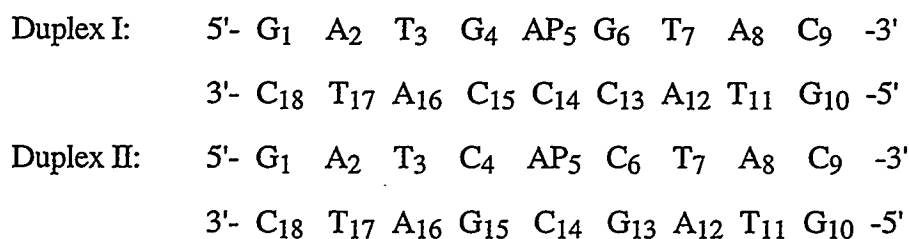
Figure 5-3. Possible structures for the 2-aminopurine•cytosine mispair: (a) Watson-Crick, (b) imino tautomer of either AP or C paired with the amino tautomer of the other base to form a Watson-Crick geometry, (c) protonated Watson-Crick, and (d) wobble.

for an AP•C mispair which is protonated at the cytosine N3 position at neutral pH (Sowers et al., 1986). At high pH, ^{15}N multiplicity and chemical shift data show normal amino tautomers in both bases and suggest that the AP•C mispair is in a wobble geometry (Sowers et al., 1989).

Owing to the highly fluorescent nature of AP (Ward et al., 1969), there has been a recent resurgence of interest in utilizing this analog as a "real-time" probe to study relaxation dynamics of DNA (Guest et al., 1991) and the presteady state kinetics of DNA synthesis (Bloom et al., 1993; Frey et al., 1995) and proofreading (Bloom et al., 1994). In this study I address the question of the geometry of the AP•C mispair in DNA over a wide range of pH values, and in the context of nearest neighboring base pairs. Through the use of selective ^{15}N -labeling of exocyclic amino nitrogens on bases of interest, it is possible to resolve ambiguities in previous studies. The results here show that, in two different DNA sequences, the AP•C mispair at neutral and high pH is in a wobble geometry. The structure and stability of this base mispair is dependent upon the local base sequence.

5.3 Materials and Methods

5.3.1 Synthesis and Purification of Oligonucleotides. The sequences of the two DNA duplexes in this study are:



where AP = 2-aminopurine.

In samples of duplexes I and II, C₁₄ was ^{15}N -labeled at the exocyclic N4 position. A sample of duplex II was also synthesized which contained AP₅ ^{15}N -labeled at the exocyclic N2 position and C₁₄ ^{15}N -labeled at the exocyclic N4 position, by methods as

published elsewhere (Acedo et al., 1994). The oligonucleotides were purified by standard reverse-phase HPLC methods.

5.3.2 NMR Sample Preparation. The concentrations of the single-stranded oligonucleotides were determined by UV absorbance. The extinction coefficients at 260 nm were calculated to be $0.92 \times 10^5 \text{ M}^{-1}\text{cm}^{-1}$ and $0.84 \times 10^5 \text{ M}^{-1}\text{cm}^{-1}$ for d(G-A-T-G-AP-G-T-A-C), assuming AP = G, and d(G-T-A-C-C-C-A-T-C), respectively. For duplex II, calculated extinction coefficients were $\epsilon_{260} = 0.85 \times 10^5 \text{ M}^{-1}\text{cm}^{-1}$ for d(G-A-T-C-AP-C-T-A-C) assuming AP = G, and $\epsilon_{260} = 0.90 \times 10^5 \text{ M}^{-1}\text{cm}^{-1}$ for d(G-T-A-G-C-G-A-T-C). Equimolar amounts of the complementary strands were mixed. Duplexed oligonucleotides were dialyzed using 1 kDa molecular weight cutoff cellulose tubing, against 300 mM NaCl, 30 mM NaCl, and then deionized water for 24 hours each.

NMR samples were prepared by dissolving the dried nonamer oligonucleotide duplexes in 0.5 mL of 10 mM potassium phosphate buffer, 100 mM NaCl (pH 7.0), 0.1 mM EDTA, and then lyophilizing to dryness. For experiments carried out in D₂O the solid was redissolved in 0.5 mL 99.96% D₂O (Cambridge Isotope Laboratories), and for experiments in H₂O a 90% H₂O/10% D₂O solution was used. In pH titrations, the pH of the samples was adjusted using 0.1 M or 0.01 M HCl and 0.1 M or 0.01 M NaOH. The concentration of the duplex I NMR sample was 1.0 mM, and that of the 4-¹⁵NH₂-C₁₄ containing duplex II NMR sample was 1.8 mM. A sample of duplex II which contained 4-¹⁵NH₂-C₁₄ and 2-¹⁵NH₂-AP₅ was 1.0 mM in concentration.

5.3.3 NMR Experiments and Resonance Assignments. NMR experiments were performed on a 500 MHz (proton frequency) GE Omega spectrometer or on a Bruker AMX-600 spectrometer. 1-Dimensional ¹⁵N spectra were acquired at 60 MHz and 25 °C using a broad band probe, with a spectral width of 200 ppm and 15,000-23,000 scans, using a relaxation delay of 2s. ¹⁵N spectra were indirectly referenced downfield of neat ¹⁵NH₃ (Live et al., 1984). 1-Dimensional ¹H spectra in H₂O were acquired with a spectral width of 20 ppm and 256 scans, using a relaxation delay of 2s. 1-Dimensional ¹⁵N-filtered

^1H spectra in H_2O were acquired with or without ^{15}N decoupling during acquisition (Sklenar et al., 1987). 1D NOE experiments were performed on the AMX-600, using duplex II in H_2O at 5 °C and pH 8.7, with a spectral width of 22 ppm and 1024-2048 scans each for the reference and saturated spectra. pH titrations were performed on samples of duplexes I and II in 150 mM NaCl, and monitored by 1D ^1H NMR at 15 °C. Typically, the intense solvent resonance was suppressed using a 1:1 read pulse (Kime & Moore, 1983).

2D NOESY spectra of duplex I in H_2O were acquired at 20 °C (pH 6.5) and 26 °C (pH 7.0), with the mixing time $\tau_M = 200$ ms. A 2D NOESY spectrum in D_2O was acquired at 20 °C (pH 6.5) with $\tau_M = 200$ ms. 2D NOESY spectra of duplex II in H_2O were acquired at 15 °C and 20 °C (pH 7.7). A ^{15}N -decoupled 2D NOESY of duplex II in H_2O was acquired at 10 °C with $\tau_M = 150$ ms. Spectral widths were typically 14,000 Hz in H_2O , and 6,000 Hz in D_2O solution. Between 450 and 600 t_1 increments were acquired in the NOESY experiments. Data processing and analysis were performed as described in Chapter 2.

5.4 Results

5.4.1 Nonexchangeable Proton Resonance Assignments. Part of a NOESY spectrum of duplex II in H_2O is shown in Figure 5-4. The chemical shifts of the DNA base H6/H8/H2, thymine methyl and deoxyribose H1' resonances are listed in Table 5-1. Intraresidue deoxyribose NOE cross peak patterns and intensities and interresidue NOE cross peak patterns indicate that most of the DNA is near the standard B-form, with some minor distortion of the DNA helix in the region of the mismatch.

5.4.2 Resonance Assignment of Mismatch Protons. The ^{15}N -bound protons could be identified in the 1D ^1H spectrum, because they appear as doublets split by 92 Hz, the ^{15}N - ^1H one bond scalar coupling constant. These proton resonances could be more clearly

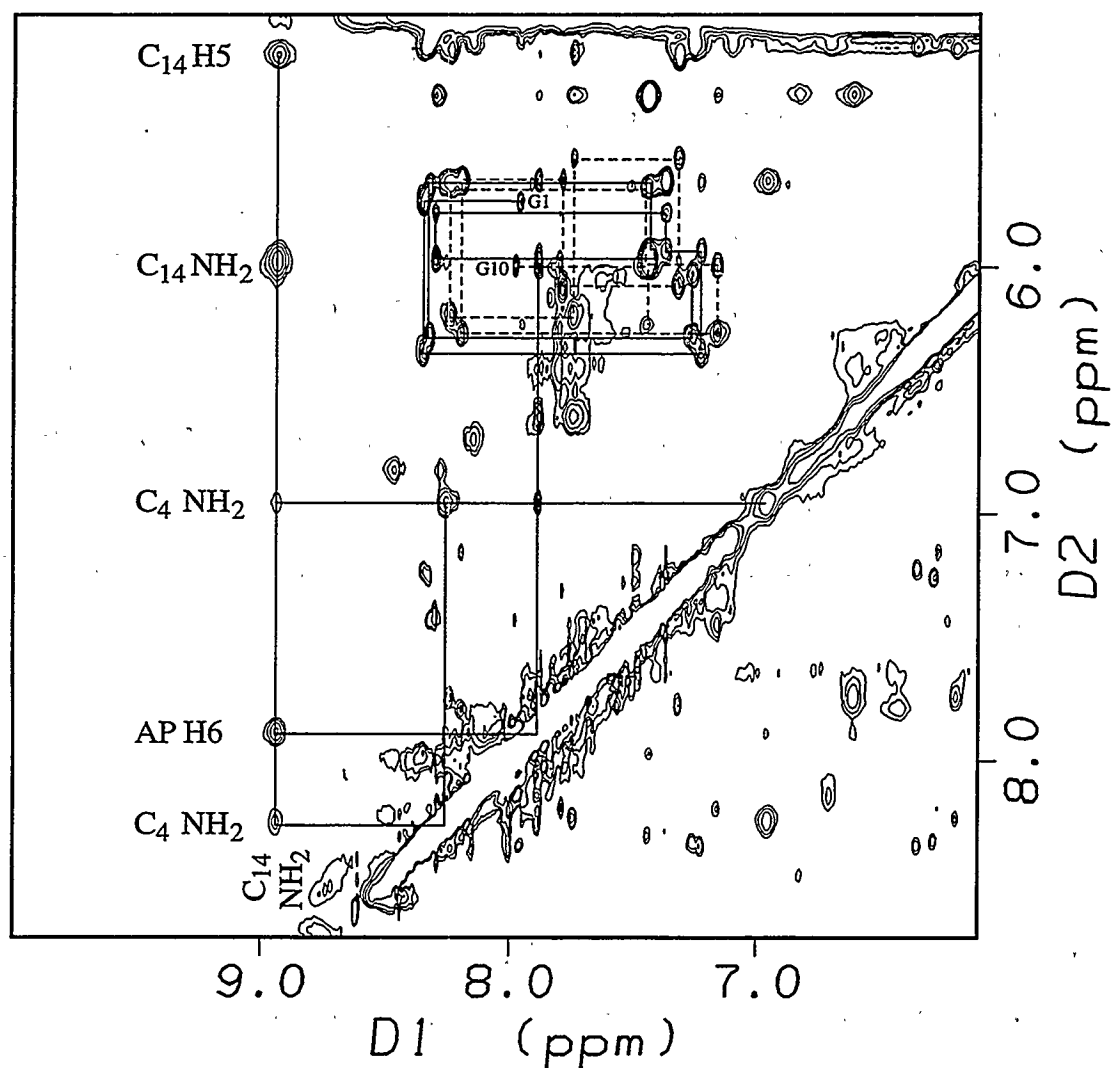


Figure 5-4. Expansion of the amino and aromatic to deoxyribose H1' region of a ^{15}N -decoupled ^1H - ^1H NOESY spectrum of $^{15}\text{NH}_2\text{-C}_{14}$ labeled duplex II (in H_2O solution, 15°C , $\text{pH} = 7.7$, 175 mM NaCl , $\tau_{\text{M}} = 200\text{ ms}$). Various NOE cross peaks to the mispaired C_{14} amino group are labeled.

Table 5-1. Nonexchangeable proton chemical shift assignments of d(GATGPGTAC):d(GTACCCATC), duplex I, and d(GATCPCTAC):d(GTAGCGATC), duplex II.^a

	H6/H8			H1'			A H2/ P H6/ T CH ₃		
	I	II	$\Delta\delta$	I	II	$\Delta\delta$	I	II	$\Delta\delta$
Strand 1									
G ₁	7.95	8.01	+0.06	5.71	5.76	+0.05			
A ₂	8.34	8.43	+0.09	6.32	6.41	+0.09	7.92	8.06	+0.14
T ₃	7.15	7.27	+0.12	5.73	5.98	+0.25	1.42	1.43	+0.01
G ₄ /C ₄	7.72	7.42	-0.30	5.48	5.82	+0.34			
AP ₅	8.14	8.34	-0.20	5.78	6.00	+0.22	7.45	7.94	+0.49
G ₆ /C ₆	7.55	7.48	-0.07	5.94	5.96	+0.02			
T ₇	7.22	7.50	+0.28	5.74	5.74	0.00	1.35	1.68	+0.33
A ₈	8.28	8.39	+0.11	6.30	6.33	+0.03	7.64	7.63	-0.01
C ₉	7.28	7.43	+0.15	6.06	6.12	+0.06			
Strand 2									
G ₁₀	7.97	8.05	+0.08	5.99	6.06	+0.07			
T ₁₁	7.44	7.53	+0.09	5.78	5.75	-0.03	1.42	1.46	+0.04
A ₁₂	8.36	8.27	-0.09	6.31	6.22	-0.09	7.57	7.52	-0.05
C ₁₃ /G ₁₃	7.40	7.75	+0.35	5.78	5.61	-0.17			
C ₁₄	7.49	7.35	-0.14	6.18	6.09	-0.09			
C ₁₅ /G ₁₅	7.45	7.82	+0.37	5.58	5.68	+0.10			
A ₁₆	8.41	8.22	-0.19	6.32	6.31	-0.01	7.69	7.86	+0.17
T ₁₇	7.17	7.28	+0.11	5.96	6.06	+0.10	1.50	1.41	-0.09
C ₁₈	7.48	7.67	+0.19	6.22	6.34	+0.22			

^a Chemical shifts are given in ppm (± 0.01 ppm). The residual HOD resonance is referenced to 4.85 ppm (20 °C) for duplex I, and to 5.01 ppm (8 °C) for duplex II.

seen in ^{15}N -filtered 1D experiments, in which only resonances of ^{15}N -bound protons are observed.

In 4- $^{15}\text{NH}_2$ - C_{14} labeled duplex I, two proton resonances are observed in ^{15}N -filtered 1D experiments at 9.16 ppm and 6.26 ppm, and in labeled duplex II at 8.94 ppm and 5.98 ppm. The line widths of these resonances are dependent on both temperature and pH. In 2- $^{15}\text{NH}_2$ - AP_5 labeled duplex II, two AP proton resonances are observed in the ^{15}N -filtered 1D spectrum at 8.47 ppm and 6.83 ppm. These resonances are broadened into the baseline at 25 °C, but sharpen with decreasing temperature. The line widths of the AP 2-amino proton resonances do not appear to be pH-dependent. No evidence is seen for any extra imino-type protons which would suggest that one of the bases has a significant population of imino tautomeric form. Because labeled $^{15}\text{NH}_2$ - AP_5 duplex I was not available, the AP_5 NH_2 resonances were not assigned in duplex I.

The specific assignment of the two 4-amino protons of C_{14} is determined by the relative intensities of the NOE between the amino protons and the C_{14} H5 proton. In both duplexes, the stronger NOE is observed between the H5 and the upfield amino proton resonance (near 6 ppm in both duplexes). This information dictates that the upfield amino resonance be assigned to the proton nearest to the H5, and the downfield amino resonance (near 9 ppm in both duplexes) be assigned to the amino proton furthest from the H5, on the base pairing edge of the base. The C_{14} amino resonance near 9 ppm is downfield relative to single-stranded cytosine amino protons (7.0 to 8.0 ppm), and is even downfield relative to the hydrogen bonded cytosine amino protons in standard Watson-Crick base pairs (8.0 to 8.6 ppm). This unusual downfield shift suggests that the C_{14} amino proton is involved in a hydrogen bond, and may also be downfield shifted due to ring current shift by AP_5 .

The AP_5 H6 resonance was assigned in both duplexes I and II (see Table 5-1). This proton was assigned through the use of spectra acquired in H_2O , since there are no other nonexchangeable protons nearby. In both duplexes I and II, the H6 has an intense NOE cross peak to the downfield C_{14} amino proton, a weaker NOE cross peak to the

upfield C₁₄ amino proton, and an even weaker NOE to the C₁₄ H5. This pattern of NOE intensities to the two C₁₄ NH₂ protons is consistent with the assignment of AP₅ H6, since it is the only unassigned nonexchangeable proton nearby, and it has a more intense cross peak to the C₁₄ NH₂ proton on the base pairing edge of the cytosine base than to the upfield (non-hydrogen bonded) C₁₄ NH₂ proton.

In the 2D NOESY spectra of both duplexes taken in H₂O without ¹⁵N-decoupling during the acquisition, a distinctive pattern identifies the cross peak between the two C₁₄ amino proton doublets. This cross peak pattern occurs because the NOE arises between two protons which are bound onto the same nitrogen. The chemical shifts of two protons which are bound to the same ¹⁵N nucleus are affected in the same way, being shifted either upfield or downfield, depending on the spin state of the ¹⁵N nucleus. Therefore, the downfield components of the two doublets give rise to a cross peak, since the NOE must arise between pairs of protons bound to the same ¹⁵N nucleus. Similarly, the NOE observed between the two upfield components is due to pairs of protons which are bound to an ¹⁵N nucleus of opposite spin.

In both duplexes, it is evident that the intensities of the C₁₄ H6:C₁₄ H1' and AP₅ H8:AP₅ H1' cross peaks are weak and of similar intensity as the other aromatic:H1' cross peaks, suggesting that AP₅ and C₁₄ are *anti*, which excludes unusual geometries such as those found in Hoogsteen or reverse-Hoogsteen base pairs.

5.4.3 Comparison of NOE Patterns in the Mismatch Region. At neutral and high pH, the 2D NOESY spectrum of duplex I contains weak cross peaks from AP₅ H8 to G₄ H1' and AP₅ H1', and the AP₅ H8 resonance is sharp. In duplex II, at neutral pH the AP₅ H8 cross peaks to C₄ H1' and AP₅ H1' are both very weak in intensity, and the AP₅ H8 and H6 resonances have somewhat broader line widths. Also, all cross peaks to the C₄ H6 resonance (from T₃ H1', C₄ H1', and C₄ H5) are broad, suggesting some conformational dynamics are occurring in the region of the mispair. The G₁₅ H8-C₁₄ H1' and G₁₅ H8-

G₁₅ H1' cross peaks are both very weak. At pH 9.0, these resonances are sharp and give stronger cross peaks in the NOESY spectrum.

5.4.4 ¹⁵N NMR Resonance Assignments. The direct-detected ¹⁵N spectra of 4-¹⁵NH₂-C₁₄ labeled duplexes I (at pH 6, see Figure 5-5) and II (at pH 9) consist of one triplet, indicating that there are two protons bound on the cytosine N4. The 4-¹⁵NH₂-C₁₄ nitrogen resonates at 100.6 ppm in duplex I, and at 99.4 ppm in duplex II, referenced to ¹⁵NH₃. These values are consistent with the value of 99.6 ppm measured previously (Sowers et al., 1989), originally referenced to ¹⁵N-aniline which is downfield of ¹⁵NH₃ by 59.6 ppm (Witanowski et al., 1986). The chemical shift of the 2-¹⁵NH₂-AP₅ nitrogen resonance was not measured.

5.4.5 ¹H NMR pH Titrations. The results from the 1-dimensional ¹H NMR pH titrations of duplexes I and II are shown in Figure 5-6. Titrations were performed at identical salt conditions and temperature. In both duplexes, the appearance of a very broad resonance at 11.0 ppm is concurrent with the broadening and disappearance of the doublet assigned to the hydrogen bonded 4-amino proton of C₁₄. In the pH titration of duplex I, the broad resonance at 11.0 ppm is detectable at pH 6.0 and below, and in duplex II at pH 6.4 and below. In the 9.5-11.5 ppm region of the NMR spectrum, broad resonances typically appear from aromatic ring nitrogen-bound protons which are protected from exchange with bulk solvent, but are not hydrogen-bonded (i.e. for imino protons of thymine in hairpin loops). Either the AP N1, with a pK_a of 4.8 in a DNA polymer (Janion & Shugar, 1973), or cytosine N3, with a pK_a of 4.6 (Fasman, 1975) are likely sites for protonation at low pH.

Broadening of the C₁₄ amino proton doublet is evident at pH values lower than 6.4 for duplex I, and at pH values lower than 7.0 for duplex II. The broadening of the C₁₄ amino resonance may be due to increased exchange with solvent, or to exchange with another structural form on the intermediate NMR time scale, or both.

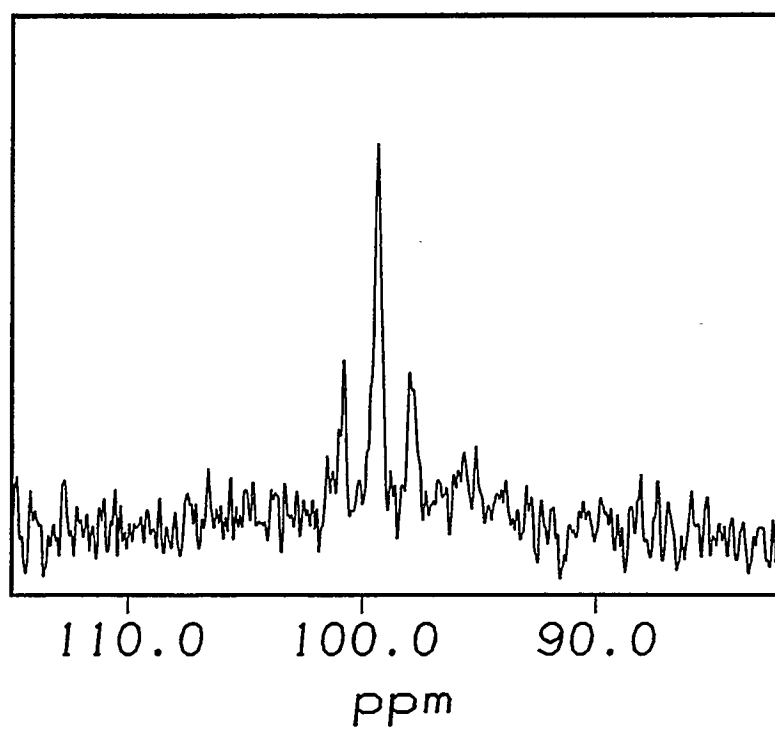


Figure 5-5. ^{15}N NMR spectrum of $^{15}\text{NH}_2\text{-C}_{14}$ labeled duplex I, at pH 6 and 10 °C.

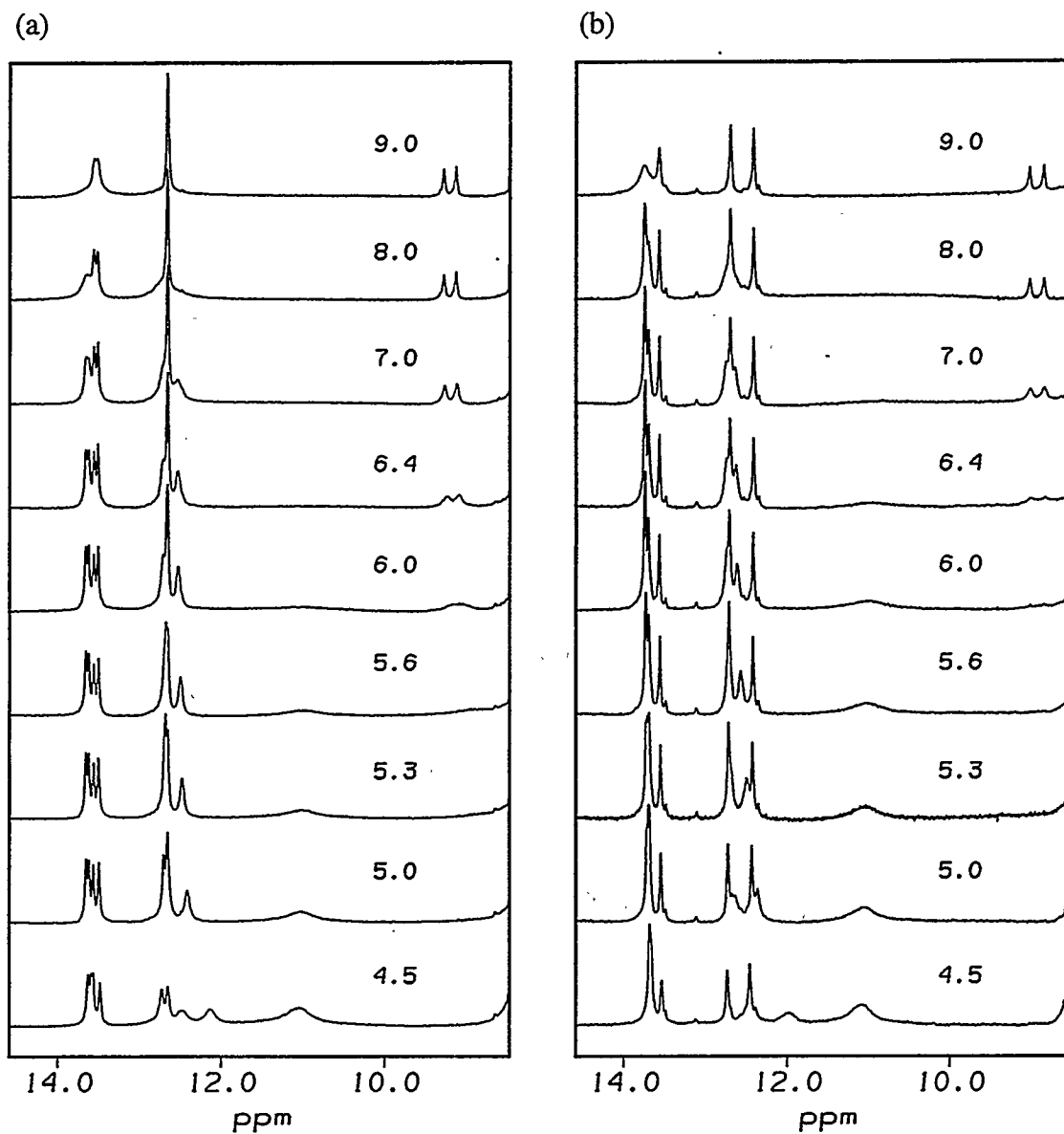


Figure 5-6. ^1H NMR spectra acquired at 15 °C during a pH titration of (a) DNA duplex I and (b) DNA duplex II. Samples were 1.0 mM in duplex, 150 mM NaCl.

The imino resonances are essentially unperturbed from pH 9.0 down to pH 5.0, with only minor changes in chemical shift, and no apparent broadening. Below 5.0, the iminos shift and broaden. This indicates that the standard Watson-Crick base pairs in both duplexes are stable down to pH 5.0.

5.4.6 ^{15}N -filtered pH and Temperature Series. In preparation for the one-dimensional NOE experiments, sample pH and temperature were optimized in order to sharpen line widths of the cytosine and 2-aminopurine amino proton resonances in duplex II. The ^{15}N -filtered pH and temperature series are shown in Figures 5-7 and 5-8. Interestingly, increasing both pH and temperature affect the C and AP line widths differently. pH was varied between 7 and 9. The double stranded C amino proton resonances sharpen, the AP amino resonances remain unaffected, and the single-stranded C amino resonances (resulting from an excess of the C-containing strand) broaden with increased pH. The temperature was varied between 10 and 25 °C. The double stranded C amino proton resonances sharpen and the AP amino proton resonances broaden with increased temperature.

These contradictory effects are likely due to competing processes occurring in the two different bases. Exchange of protons with solvent water, rotation of amino groups about the nitrogen-carbon bond, and the overall tumbling rate of the molecule in solution are major determinants of amino resonance line widths. It is difficult to separate the contributions from these two processes, since pH and temperature may affect some or all of these rates.

5.4.7 One dimensional NOE Measurements. To identify the base pairing in the AP mismatches, duplex II was annealed using strands containing 4- $^{15}\text{NH}_2$ -C₁₄, and 2- $^{15}\text{NH}_2$ -AP₅. In a wobble conformation, the two hydrogen bonded amino protons on the labeled nitrogens are within 4.0 Å, and therefore should give rise to an NOE. When the downfield (hydrogen bonded) amino proton resonance on AP₅ is saturated, a weak NOE is observed on the downfield (hydrogen bonded) C₁₄ amino proton resonance (Figure 5-9). However,

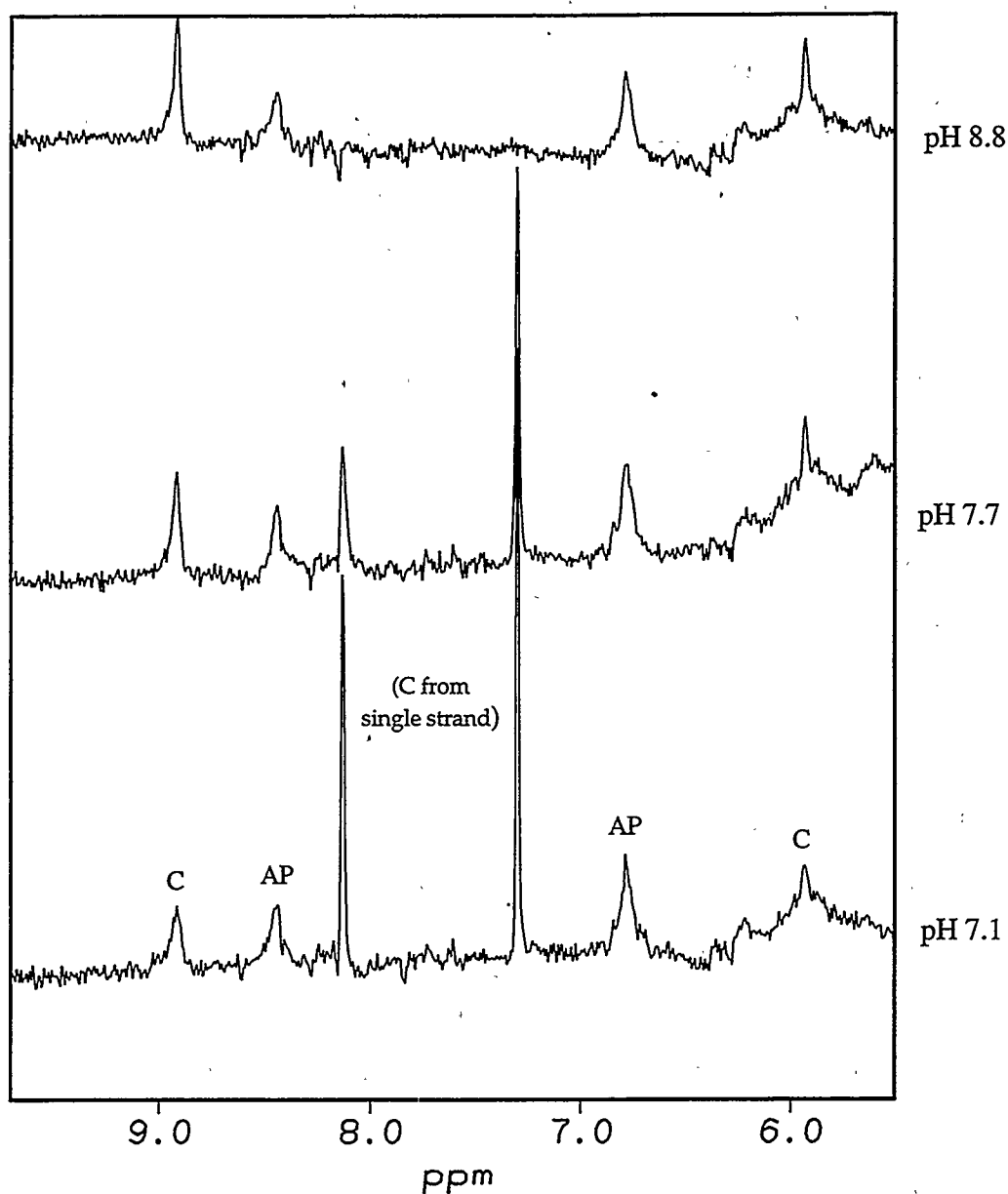


Figure 5-7. 1D ^{15}N -filtered ^1H NMR pH titration of duplex II containing 4- $^{15}\text{NH}_2\text{-C}_{14}$ and 2- $^{15}\text{NH}_2\text{-AP}_5$, acquired at 15 $^\circ\text{C}$. This sample contained excess of the second strand, as evidenced by the sharp single-stranded 4- $^{15}\text{NH}_2\text{-C}_{14}$ resonances.

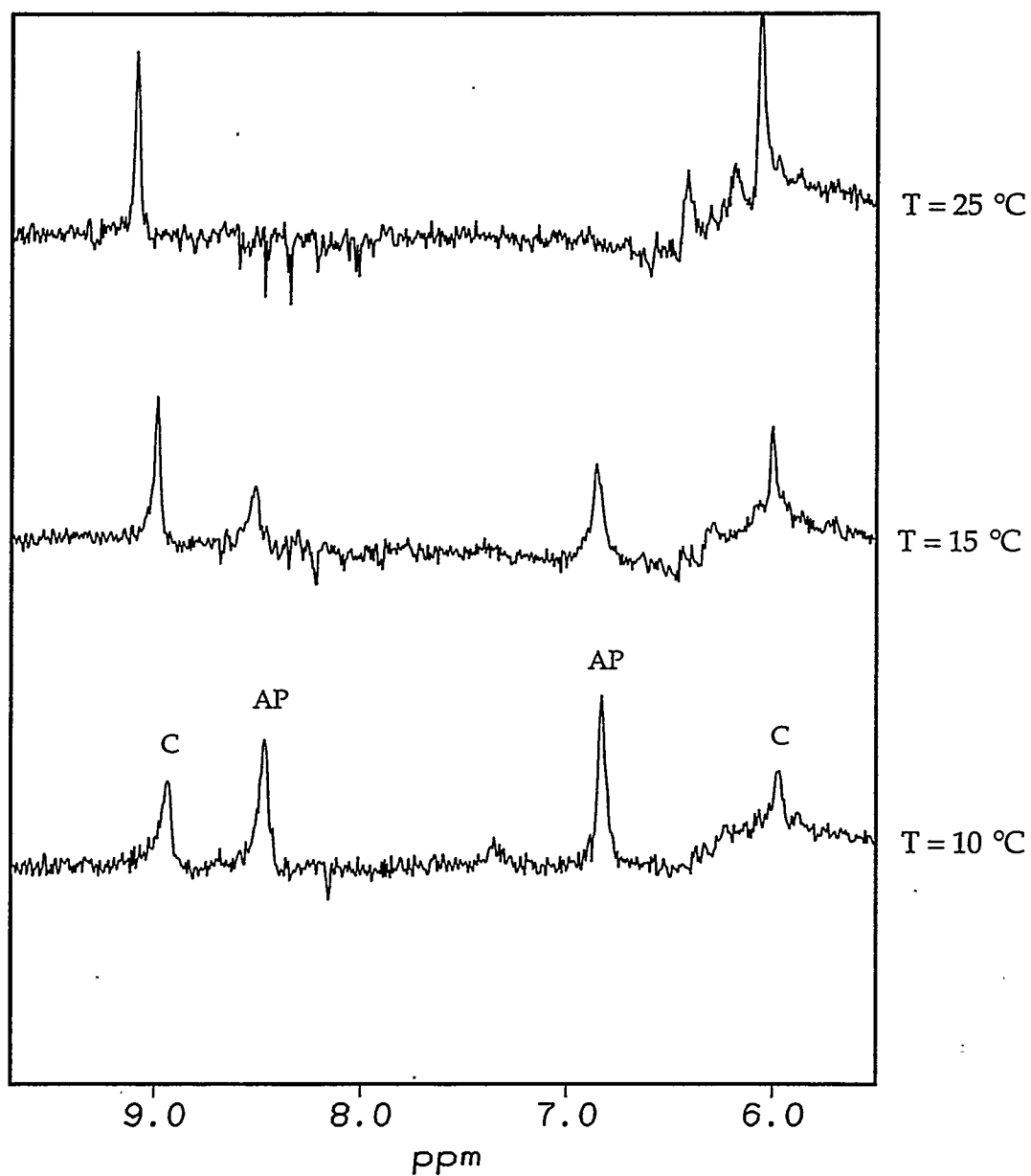


Figure 5-8. One-dimensional ^{15}N -filtered ^1H NMR temperature series of duplex II containing $4\text{-}^{15}\text{NH}_2\text{-C}_{14}$ and $2\text{-}^{15}\text{NH}_2\text{-AP}_5$, acquired at pH 8.8.

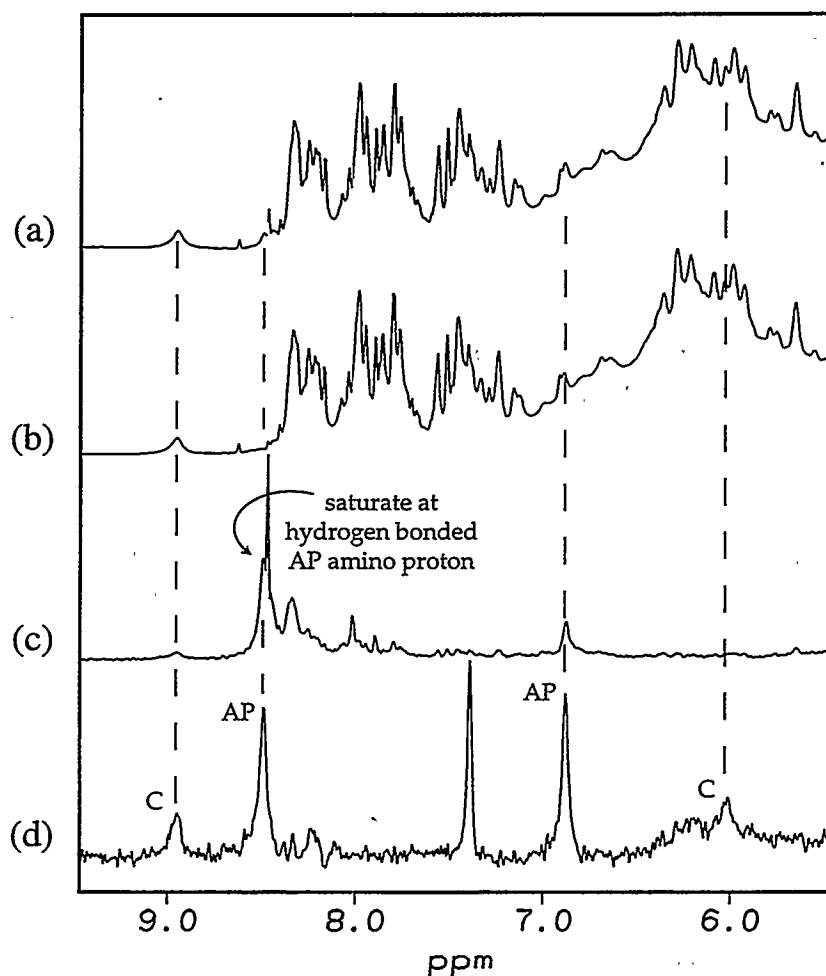


Figure 5-9. One-dimensional NOE data taken on duplex II with ^{15}N decoupling during acquisition. (a) ^1H reference spectrum, (b) ^1H spectrum with 1 sec saturation at the hydrogen-bonded (downfield) AP₅ amino proton resonance, (c) 1-dimensional NOE difference spectrum generated by subtracting (b) from (a). A strong NOE is visible to the AP₅ geminal amino proton resonance, and a weaker NOE is observed to the hydrogen-bonded C₁₄ resonance. The saturated AP resonance is almost degenerate with a small, sharp resonance which is due to excess single-strand present in the sample. (d) ^{15}N -filtered 1D ^1H spectrum. The amino proton resonances from the duplex are labeled. The unlabeled resonance is assigned to one of the amino protons on 4- $^{15}\text{NH}_2$ -cytosine due to the presence of excess of the second strand (see Figure 5-8).

when the reverse experiment is done, an NOE is not observed. A possible reason for this asymmetry in the NOE is that the two protons undergo different rates of exchange with solvent. Such a difference in rates might explain the differing pH- and temperature-dependences observed in the line widths of the C₁₄ and AP₅ amino proton resonances.

5.5 Discussion

5.5.1 Structure of the AP•C Mispair.

5.5.1.1 Imino vs. Amino Tautomers. The imino tautomeric form of DNA bases has long been used to explain spontaneous and base analog-induced mutation (Watson & Crick, 1953a; Watson & Crick, 1953b; Freese, 1959; Topal & Fresco, 1976; Saenger, 1983) (see Figure 5-3b). ¹⁵N-filtered ¹H NMR experiments show clear evidence that in 4-¹⁵NH₂-C₁₄ labeled duplexes I and II and 2-¹⁵NH₂-AP₅ labeled duplex I, two protons are bound to the labeled nitrogens in the AP•C base pair. No evidence is seen for any extra imino-type protons in the downfield region of the ¹H NMR spectrum which, if observed, would suggest that one of the bases has a significant population of imino tautomeric form.

Moreover in direct-detected ¹⁵N spectra of 4-¹⁵NH₂-C₁₄ labeled duplexes I and II, one triplet is observed, indicating that one tautomeric form is detectable in which the nitrogen resonance is split by two directly attached protons. The chemical shift of the cytosine ¹⁵NH₂ resonance also provides information about the tautomeric state of the base. The amino-imino tautomerization of 2-aminopyridine results in an estimated 115 ppm upfield shift of the exocyclic ¹⁵N resonance (Witanowski et al., 1986). A similar shift can be expected for the tautomerization of cytosine. Therefore, the measured chemical shift is sensitive to the presence of even a small (i.e. 1%) population of imino tautomer in the fast exchange limit. The ¹⁵NH₂-C₁₄ chemical shift in both AP-containing duplexes (100.6 ppm in duplex I, and 99.4 ppm in duplex II) is within 2 ppm of that found in duplex DNA (101.3 ppm, originally referenced to 2 M ¹⁵NH₄Cl in 2 M HCl, 29.4 ppm downfield of

neat $^{15}\text{NH}_3$ in James et al., 1981). Therefore, the AP•C mispair does not seem to shift the equilibrium at all toward the imino tautomer compared to a G•C pair.

Taken together, these data indicate that both C_{14} and AP_5 are amino tautomers, since two protons are observed on the exocyclic amino nitrogens, and no extra imino protons were observed in the NMR spectrum. We have determined that the amino tautomer is prevalent by at least 10- to 20-fold relative to any other tautomeric form, with the lower bound set by the limited sensitivity of the NMR experiment. To date, no structural study has produced evidence for any substantial population of the rare imino tautomeric form of bases in DNA, consistent with our findings here.

5.5.1.2 Watson-Crick Geometry. A standard Watson-Crick geometry pairing between AP and C would result in a mispair stabilized by one hydrogen bond (Figure 5-3a). Both ^1H and ^{15}N chemical shift data argue against this pairing geometry. Since amino protons of cytosines involved in Watson-Crick base pairing with guanine typically resonate between 8.0 and 8.6 ppm, the downfield chemical shift (9.16 ppm in duplex I and 8.94 ppm in duplex II) of the 4-amino proton of C_{14} is highly suggestive that it is hydrogen bonded to the ring nitrogen. The C_{14} -amino to AP-H6 NOE is weaker than would be expected for the very short distance in the W-C geometry, and is more consistent with a wobble. Additionally, the ^{15}N chemical shift of the C_{14} N4 (100.6 ppm in duplex I and 99.4 ppm in duplex II, referenced to $^{15}\text{NH}_3$) is consistent with a hydrogen bond at the C_{14} 4-amino group (Sowers et al., 1989). However, in the standard W-C geometry for the AP•C mispair the C_{14} 4-amino group is not hydrogen bonded. This geometry is also inconsistent with previous work by Goodman and Ratliff (Goodman & Ratliff, 1983), which showed that the UV absorption spectra of polynucleotides containing 2-aminopurine were characteristically red-shifted in the presence of cytosine, which suggests that AP•C mispairs contain a hydrogen bond at the N1 position of AP.

Protonation at N1 of AP_5 or at N3 of C_{14} would result in a charged base pair with Watson-Crick geometry stabilized by two hydrogen bonds (see Figure 5-3c). A proton at

one of these positions would likely resonate in the downfield region of the exchangeable proton NMR spectrum (>10 ppm) due both to its involvement in a hydrogen bond and its positioning in between and in the plane of two aromatic bases, which would further deshield the proton through ring current shift. At pH below 7, the hydrogen-bonded C₁₄ amino proton resonance is broad and a resonance is visible at about 11 ppm, the region expected for base protonation. However, at neutral and high pH there is no longer a resonance at 11 ppm, and all resonances in the downfield region of the NMR spectrum have been otherwise assigned, indicating that protonation is not occurring. The proton NMR spectrum remains essentially unchanged from pH 7 to 9, indicating that no significant structural rearrangements or changes in charge state are occurring in this range. Therefore, I observe no evidence for protonation of AP₅ at the N1 position, or of C₁₄ at the N3 position in either duplex at neutral pH.

Previous NMR studies of a DNA heptamer duplex containing an AP•C mispair suggested that at neutral pH, the cytosine paired with AP in the local sequence GAPG:CCC is protonated at the N3 position (Sowers et al., 1986). With the use of selective ¹⁵N labeling I am able to make unambiguous assignments of the exchangeable protons directly involved in the AP•C mispair, which contradict this earlier assignment. Direct evidence from 1D ¹⁵N-filtered and 2D ¹H NOESY spectra in this study shows that the resonance assigned to a cytosine N3 proton in previous work (at 9.37 ppm) (Sowers et al., 1986) is actually a downfield shifted cytosine NH₂ proton (at 9.16 ppm in duplex I in this study). Additionally, a resonance previously assigned to an AP amino proton (at 6.60 ppm) is properly assigned to the upfield C₁₄ amino proton (at 6.26 ppm in duplex I in this study).

These corrected assignments are consistent both with our current observations and with the observed NOE patterns in the previous NMR study. The assignments qualitatively are more reasonable since in the previous NMR study, no assignments were made for the amino protons of the cytosine in the AP•C mispair. Cytosine amino proton resonances, in both double- and single-stranded DNA, typically are sharper and hence more readily

observed by NMR than guanine or adenine amino protons. Guanine and adenine amino groups often rotate at or near intermediate exchange on the NMR time scale, causing severe broadening of the amino proton resonances and making assignment of these protons difficult. In the study by Sowers and coworkers, although assignments were not made for the two cytosine amino protons in the AP•C mispair, assignments were made, however, for the 2-amino protons on 2-aminopurine which are more likely to behave similarly to guanine and adenine amino protons in the NMR spectrum. Therefore, I conclude that these previous assignments were in error, and led to the incorrect conclusion that the AP•C base pair is protonated and in Watson-Crick geometry.

5.5.1.3 Wobble Geometry. The presence of stable wobble pairs in DNA and RNA has been established by NMR spectroscopy (Patel et al., 1982; Hare et al., 1986) and X-ray crystallography (Brown et al., 1985; Kneale et al., 1985). A wobble geometry for the AP•C mispair in DNA (see Figure 5-3d) was first suggested by Janion and Shugar (1973) and evidence for this geometry was observed at high pH (Sowers et al., 1989). In these studies the geometry of the AP•C base pair was determined to be wobble at high pH by utilizing ^{15}N chemical shift data to identify base nitrogens which are directly involved in hydrogen bonding interactions.

Several lines of evidence in the present study lead to the conclusion that the AP•C base pair is in a wobble geometry at neutral and high pH. ^1H NMR spectra of the two duplexes are essentially identical at pH 9.0 and pH 7.0, with some resonances in the mismatch broadening somewhat at neutral pH. Therefore, it appears that the structure of the AP•C base pair present at pH 9.0 remains dominant at neutral pH. The structure at lower pH is less clear.

The unusual chemical shift (9.16 ppm in duplex I and 8.94 ppm in duplex II) of the 4-amino proton of C_{14} also supports that the local structure may not be that of typical B-form DNA, and that the base pair may not be in a normal Watson-Crick geometry. Since amino protons of cytosines involved in Watson-Crick base pairing with guanine typically

resonate between 8.0 and 8.6 ppm, the downfield chemical shift of this amino proton is highly suggestive that it is hydrogen-bonded. A wobble geometry is the only proposed geometry for the AP•C base pair in which the C₁₄ 4-amino group is hydrogen-bonded (see Figure 5-3). In fact, the chemical shift of this resonance (9.16 ppm) is downfield relative to cytosine amino protons which are involved in a G•C base pair. This is consistent with a wobble geometry, in that the cytosine amino proton is directly hydrogen-bonded to AP₅ N1. In this geometry, the hydrogen bonded cytosine amino proton is likely to be 1.7-2.1 Å from the edge of the 2-aminopurine base, and may be shifted further downfield than usual due to an aromatic ring current shift. In a G•C base pair, the cytosine amino protons are approximately 2.6 Å from the edge of the guanine base, and consequently should have a smaller ring current shift. Moreover, a wobble geometry is consistent with previous studies which showed that the UV absorption spectra of polynucleotides containing 2-aminopurine were characteristically red-shifted in the presence of cytosine (Goodman & Ratliff, 1983), which suggests that AP•C mispairs contain a hydrogen bond at the N1 position of AP. This type of wobble geometry has also been proposed by Patel and coworkers for the O⁶-ethylguanine•cytosine base pair which could pair in a Watson-Crick geometry stabilized by one hydrogen bond, but instead forms a wobble pair stabilized by two hydrogen bonds (Kalnick et al., 1989). Formation of this pairing geometry by the O⁶-ethylguanine•cytosine base pair is supported by NMR and crystallographic data (Sriram et al., 1992; Sriram et al., 1994). It appears that the additional hydrogen bond gained in going from a Watson-Crick geometry to a wobble geometry makes up for any loss of interbase stacking interactions.

1-dimensional NOE experiments have been used previously to establish the geometry of the G•T wobble pair (Patel et al., 1982; Quignard et al., 1987). The G•T wobble pair is well-suited for NMR studies because the interbase hydrogen bonds occur through imino protons (Figure 5-1). Imino proton resonances are relatively well-dispersed in the NMR spectrum and have reasonably narrow line widths, simplifying 1D NOE

experiments which provide direct evidence for the wobble geometry. A key difference between the AP•C wobble and the G•T wobble, however, is that the interbase hydrogen bonds in the AP•C wobble pair occur through amino protons, which tend to have broader line widths than imino protons and often overlap with aromatic proton resonances, making 1D NOE experiments difficult. In this study, the 1D NOE experiments taken on the AP•C mispair at high pH show direct evidence that the cytosine amino group is within 5.0 Å of the AP amino group (Figure 5-9). This requires that the AP•C mispair be in a wobble geometry. The NOE is weak, however, probably because of the large line width of the AP and C amino protons.

5.5.2 Sequence Dependence of Structure. Stacking interactions with nearest-neighbor bases affect the local structure and helical stability around a DNA base pair (Werntges et al., 1986; Modrich, 1987; Ke & Wartell, 1993; Doktycz et al., 1995). The sequence dependence on the stability in the region of the AP•C mispair is evidenced in two ways: first, through the pH dependence of the NMR spectra for the two duplexes, and second, through the differences in dynamics of the bases near the mispairs in duplexes I and II.

Below neutral pH (Figure 5-6), the mispair resonances in both duplexes appear to undergo a structural change which is localized to the mispair, since there are no major chemical shift or line width changes in resonances outside the mispair, and the eight imino protons observed from the eight flanking base pairs are relatively unaffected. The two duplexes exhibit somewhat different pH-dependence for this structural change, with the transition occurring approximately 0.5 pH units higher in duplex II than in duplex I.

Duplex II shows more evidence of local dynamics near the mispair at neutral pH than duplex I. Exchangeable and nonexchangeable protons in the AP₅•C₁₄ mispair and the adjacent C₄•G₁₅ base pair show evidence of dynamics since NOESY cross peaks involving these protons are weak, due at least in part to broader line widths. No such effect is seen for duplex I. The difference in dynamics observed suggests that the stacking stabilization

of the C₄-AP₅ step in duplex II is smaller than that of the G₄-AP₅ step in duplex I. These observations are consistent with calculated thermal stabilities of stacked DNA doublets containing AP paired with T, which predict that the doublet C-AP is far less stable than the doublet G-AP (Petruska & Goodman, 1985). It should be noted however, that the stacking stabilization of these doublets will differ somewhat when AP is paired with C rather than with T.

Therefore, it appears that the stability of the AP•C base pair may depend in part on interactions with and the stability of nearby base pairs. The difference in stacking energies between the purine-AP-purine sequence (duplex I) and the pyrimidine-AP-pyrimidine sequence (duplex II) may impart differing stability to the wobble base pair and local structure.

5.5.3 Implications for Misinsertion and Repair of AP. A model has been developed to describe the method of discrimination by DNA polymerase in selecting the correct base for insertion and in editing misinserted bases. In this model, the polymerase screens according to the base-pairing free energy of the pair and the precise geometry of the Watson-Crick base pair (Kornberg, 1980; Echols & Goodman, 1991), perhaps by measuring bond angles and the cross-strand C1'-C1' distance which are remarkably similar in A•T and G•C base pairs (Saenger, 1983). This selection process allows the polymerase to distinguish the proper base from all other substrates by a factor of 10³-10⁵, so that at low frequencies other bases are misinserted.

According to this model, the wobble structure of the AP•C mispair may account for the strong preference of DNA polymerase to insert AP opposite T rather than C (Watanabe & Goodman, 1981; Watanabe & Goodman, 1982). Although both AP•C and AP•T mispairs each consist of two hydrogen bonds, the polymerase can distinguish the two base pairs based on their differences in geometry, and inserts the Watson-Crick-like AP•T mispair (Sowers et al., 1986) at a higher frequency than the wobble AP•C mispair (Watanabe & Goodman, 1981).

Likewise, the increased rate of insertion of AP opposite C relative to A (Mhaskar & Goodman, 1984) is consistent with the wobble structure of the AP•C mispair. In both AP•A (Fazakerley et al., 1987) and AP•C mispairs, two hydrogen bonds are formed with amino protons to generate wobble base pairs, with both bases in the anti conformation. However, the AP•A mispair consists of two bulky purines which deviate further from Watson-Crick geometry compared to the AP•C mispair, which consists of one purine and one pyrimidine. Therefore, the AP•C wobble pair may be more easily recognized as a substrate by the polymerase.

The results from this study indicate that the structure and stability of the AP•C mispair is somewhat dependent on local sequence. This has potential implications for the observed dependence of single-site mutation rates with nearest-neighboring bases (Pless & Bessman, 1983; Mendelman et al., 1989; Joyce et al., 1992; Mitra et al., 1993). The mutagenicity of AP has been shown to be strongly dependent on the local DNA sequence at the mutation site (Ronen, 1979). The local DNA sequence, specifically the extent to which the primer-3'-nearest neighbor can stabilize an incoming dNTP substrate at the polymerase active site, has been shown to influence efficiency of AP misinsertion opposite T by DNA polymerase (Pless & Bessman, 1983; Petruska & Goodman, 1985; Bloom et al., 1993). These effects are attributed to base stacking interactions between the incoming AP nucleotide and the 3'-primer terminus.

In a similar manner, local DNA sequence has a profound effect on the editing efficiencies of polymerase-associated proofreading exonuclease, which is sensitive to the degree of single-strandedness at a primer-3'-terminus. A striking indication of the effect of local sequence context on the editing of base pairs involving AP was demonstrated in a presteady state kinetic study of exonucleolytic proofreading by bacteriophage T4 DNA polymerase (Bloom et al., 1994). As expected, excision of AP located at a primer-3'-terminus opposite each of the four template bases, within the same surrounding sequence, occurs at rates that are consistent with calculated thermal stabilities of each base pair

(Petruska & Goodman, 1985). AP opposite T is excised at a relatively slow rate, while excision of AP opposite A and C occurs more rapidly. AP is removed most rapidly when it is opposite G, since no hydrogen bonds are formed. However, a Watson-Crick-like AP•T mispair in an A,T-rich environment is removed relatively rapidly compared to an AP•C mispair in a G,C-rich environment, demonstrating the importance of localized sequence context on proofreading specificities. Studies which utilize 2-aminopurine to measure rates of DNA base insertion, proofreading, or repair must now consider the AP•C wobble base pairing. The dependence of the stability of the AP•C mispair on the local sequence context highlights the need for consideration of this effect in such studies.

Chapter 6

Summary and Conclusions

Much can be learned about DNA and ligand-DNA complexes at the molecular level through the use of NMR spectroscopy and molecular modeling. In the preceding chapters, the binding behavior of distamycin and conjugated distamycin analogs containing reactive agents was examined. One- and two-dimensional NMR experiments were used to characterize the binding site, orientation, stoichiometry, and qualitative affinity of ligand binding to several short DNA oligomers.

The observed mode of distamycin binding can be correlated with sequence and structural features of the DNA binding sites. It was found that distamycin binds to A-tract sequences with high affinity in the 1:1 mode. This result suggests that A-tracts have narrowed minor grooves as has been observed in crystallographic studies, and that there is an energetic cost to widening the groove of such sequences. Other A,T-rich sequences, such as the alternating 5'-ATATAT-3' site, bind a single distamycin with low affinity, but bind a second distamycin with high affinity, resulting in a cooperative 2:1 ligand:DNA complex. The minor groove of a 2:1 complex is 3 to 4 Å wider than that of a 1:1 ligand:DNA complex. This implies that the minor groove is inherently wider in these DNA sequences, and there is an energetic cost to reducing the minor groove width to form a high affinity 1:1 complex.

These observations are supported by the studies of inosine-containing sites. The results from distamycin binding studies clearly indicate that I,C- and A,T-rich regions have different structural features despite the fact that the functional groups in the minor groove are identical in these two sequences. The I,C regions are likely to have a wider minor groove, leading to higher cooperativity of distamycin binding. Although the groove may be wide enough to accommodate two ligands side-by-side, the absence of 2:1 complex

formation in IICC strongly suggests that a four base pair site is too short to accommodate a 2:1 complex. Additionally, comparison between distamycin and netropsin binding to the asymmetric site AAAA indicates a large difference in the orientational preference of the two ligands. Netropsin, perhaps because of the presence of a second positive charge, binds in either orientation equally well. This result highlights the need for further study of the effects of end groups and placement of charged groups on ligand sequence specificity. Understanding of these issues is critical for the rational design of ligands which bind target DNA sequences.

Molecules which direct chemistry to DNA can be and are used as anticancer treatments, and molecules which direct chemistry to DNA sequence-specifically have many possible applications including use as tools in molecular biology, as antiviral treatments or as diagnostic probes. Two conjugated ligands were studied which contain reactive agents which alone have little or no sequence specificity. The reactive groups were tethered to a netropsin or distamycin moiety, with the intention of increasing sequence specificity of the reactive agent. The first ligand studied, netropsin-diazene, contains a TMM diyl precursor which has been shown to cleave DNA. It was found that this ligand binds at the predicted site AAAA, but the precise orientation of the diazene moiety could not be determined from the data. The presence of the diazene reduced both the binding affinity and the orientational binding preference of the ligand compared to the parent compound distamycin. These effects are important considerations in rational ligand design, since it appears that ligand binding behavior cannot necessarily be predicted by the known behavior of the parent molecule. The second conjugated ligand studied, XL-Dst, contains a 2,3-bis(hydroxymethyl)pyrrole moiety which covalently cross-links the two DNA strands. The design of this ligand was based on the pharmacophore of mitomycin C, and the covalent complex exhibits many similarities to mitomycin. It was found that the tether bridges one base pair, which may have implications for future design.

In Chapter 5, the geometry of the 2-aminopurine•cytosine mispair in DNA was determined. At least four models have been proposed for the geometry of this mispair. Through the use of NMR spectroscopy with selective ^{15}N -labeling of exocyclic amino nitrogens on bases of interest, this study resolves ambiguities in previous work. I show here that, in two different DNA sequences, the AP•C mispair at neutral and high pH is in a wobble geometry. Results from NMR pH titrations and analysis of NOESY data indicate that the structure and stability of this base mispair is dependent upon the local DNA sequence. These results will contribute to the overall understanding of polymerase fidelity and repair processes in mutagenesis. Researchers who use 2-aminopurine as a fluorescent probe to measure rates of DNA base insertion, proofreading, or repair must now consider the AP•C wobble base pairing.

In summary, the studies presented here examine a DNA mispair and several ligand-DNA complexes by NMR spectroscopy. The forces contributing to sequence specific binding are complicated, and are often difficult to separate and examine individually. By studying one ligand (distamycin) binding to various DNA sequences, or by comparing the binding of related ligands (netropsin, netropsin-diazene, XL-distamycin) to a particular DNA sequence, insights are gained into the subtle structural features of DNA and the requirements for ligand binding. Knowledge of the unusual structure of a base mispair will contribute to understanding interactions of polymerases with DNA, and DNA recognition by repair enzymes. Together, these studies contribute to a better understanding of intermolecular interactions in biology.

Bibliography

- Abu-Daya, A., Brown, P.M. & Fox, K.R. (1995) "DNA sequence preferences of several AT-selective minor groove binding ligands." *Nucleic Acids Res.* 23, 3385-3392.
- Acedo, M., Fàbrega, C., Aviño, A., Goodman, M., Fagan, P., Wemmer, D. & Eritja, R. (1994) "A simple method for ¹⁵N-labelling of exocyclic amino groups in synthetic oligodeoxynucleotides." *Nucleic Acids Res.* 22, 2982-2989.
- Baker, B.F. & Dervan, P.B. (1985) "Sequence-specific cleavage of double-helical DNA. N-Bromoacetyldistamycin." *J. Am. Chem. Soc.* 107, 8266-8268.
- Baker, B.F. & Dervan, P.B. (1989) "Sequence-specific cleavage of DNA by N-bromoacetyldistamycin. Product and kinetic analyses." *J. Am. Chem. Soc.* 111, 2700-2712.
- Berson, J.A. (1982) *Diradicals*, Wiley, New York.
- Bessman, M.J., Muzyczka, N., Goodman, M.F. & Schnaar, R.L. (1974) "Studies on the biochemical basis of spontaneous mutation. II. The incorporation of a base and its analogue into DNA by wild type, mutator and antimutator DNA polymerases." *J. Mol. Biol.* 88, 409-421.
- Billera, C.F. & Little, R.D. (1994) "Hydrogen atom transfer reactions to trimethylenemethane diyls - a new reactivity pattern leading to bicyclic ring systems." *J. Am. Chem. Soc.* 116, 5487-5488.
- Bloom, L.B., Otto, M.R., Beechem, J.M. & Goodman, M.F. (1993) "Influence of 5'-nearest neighbors on the insertion kinetics of the fluorescent nucleotide analog 2-aminopurine by Klenow fragment." *Biochemistry* 32, 11247-11258.
- Bloom, L.B., Otto, M.R., Eritja, R., Reha-Krantz, L.J., Goodman, M.F. & Beechem, J.M. (1994) "Pre-steady-state kinetic analysis of sequence dependent excision of the fluorescent nucleotide analog 2-aminopurine by T4 DNA polymerase." *Biochemistry* 33, 7576-7586.
- Borgias, B.A. & James, T.L. (1989) "Two-dimensional nuclear Overhauser effect: Complete relaxation matrix analysis." *Meth. Enzymol.* 176, 169-183.
- Borgias, B.A. & James, T.L. (1990) "MARDIGRAS - A procedure for matrix analysis of relaxation for discerning geometry of an aqueous structure." *J. Magn. Reson.* 87, 475-487.
- Borowy-Borowski, H., Lipman, R. & Tomasz, M. (1990) "Recognition between mitomycin C and specific DNA sequences for cross-link formation." *Biochemistry* 29, 2999-3006.

- Boulard, Y., Cognet, J.A.H., Gabarro-Arpa, J., Le Bret, M., Sowers, L.C. & Fazakerley, G.V. (1992) "The pH dependent configurations of the C.A mispair in DNA." *Nucleic Acids Res.* 20, 1933-1941.
- Bouziat, M., Ketterle, C., Helissey, P., Herfeld, P., Le Bret, M., Giorgi-Renault, S. & Auclair, C. (1995) "Sequence-directed single strand cleavage of DNA by a netropsin-flavin hybrid molecule." *Biochemistry* 34, 14051-14058.
- Bregant, T.M., Groppe, J. & Little, R.D. (1994) "New class of DNA-cleaving agents based on trimethylenemethane." *J. Am. Chem. Soc.* 116, 3635-3636.
- Breslauer, K.J., Remeta, D.P., Chou, W.-Y., Ferrante, R., Curry, J., Zaunczkowski, D., Snyder, J.G. & Marky, L.A. (1987) "Enthalpy-entropy compensations in drug-DNA binding studies." *Proc. Natl. Acad. Sci. USA* 84, 8922-8926.
- Broggini, M., Erba, E., Ponti, M., Ballinari, D., Geroni, C., Spreafico, F. & D'Incalci, M. (1991) "Selective DNA interaction of the novel distamycin derivative FCE 24517." *Cancer Res.* 51, 199-204.
- Brown, D.G., Sanderson, M., Garman, E. & Neidle, S. (1992) "Crystal structure of a berenil-d(CGCAAATTTGCG) complex. An example of drug-DNA recognition based on sequence-dependent structural features." *J. Mol. Biol.* 226, 481-490.
- Brown, T., Kennard, O., Kneale, G. & Rabinovich, D. (1985) "High-resolution structure of a DNA helix containing mismatched base pairs." *Nature* 315, 604-606.
- Bruice, T.C., Mei, H.-Y., He, G.-X. & Lopez, V. (1992) "Rational design of substituted tripyrrole peptides that complex with DNA by both selective minor-groove binding and electrostatic interaction with the phosphate backbone." *Proc. Natl. Acad. Sci. USA* 89, 1700-1704.
- Burkhoff, A.M. & Tullius, T.D. (1987) "The unusual conformation adopted by the adenine tracts in kinetoplast DNA." *Cell* 48, 935-943.
- Capobianco, M.L., Colonna, F.P., Forni, A., Garbesi, A., Iotti, S., Moretti, I., Samori, B. & Tondelli, L. (1991) "Interactions of nucleic acids with distamycins. Binding of Dst-3 to d(CGTTTAAACG)₂ and d(CGTACGTACG)₂." *Nucleic Acids Res.* 19, 1695-1698.
- Chandrakumar, N. & Subramanian, S. (1987) *Modern Techniques in High-Resolution FT-NMR*, Springer-Verlag, New York.
- Chang, W.-H. (1993) "The interaction of distamycin A with a TATA-containing oligomer by two dimensional NMR spectroscopy." Master's thesis, University of California, Berkeley.
- Chen, F.-X., Zhang, Y., Church, K., Bodell, W.J. & Gold, B. (1993) "DNA crosslinking, sister chromatid exchange and cytotoxicity of N-2-

- chloroethylnitrosoureas tethered to minor groove binding peptides." *Carcinogenesis* 14, 935-940.
- Chen, S.M., Leupin, W., Rance, M. & Chazin, W.J. (1992) "Two-dimensional NMR studies of d(GGTTAATGCGGT)-d(ACCGCATTAACC) complexed with the minor groove binding drug SN-6999." *Biochemistry* 31, 4406-4413.
- Chen, X., Ramakrishnan, B., Rao, S.T. & Sundaralingam, M. (1994) "Binding of two distamycin A molecules in the minor groove of an alternating B-DNA duplex." *Nature Struct. Biol.* 1, 169-175.
- Chen, Y.-H. & Lown, J.W. (1994) "A new DNA minor groove binding motif: Cross-linked lexitropsins." *J. Am. Chem. Soc.* 116, 6995-7005.
- Cheng, J.W., Chou, S.H. & Reid, B.R. (1992) "Base pairing geometry in GA mismatches depends entirely on the neighboring sequence." *J. Mol. Biol.* 228, 1037-1041.
- Chou, S.H., Cheng, J.W. & Reid, B.R. (1992) "Solution structure of [d(ATGAGCGAATA)]₂. Adjacent G-A mismatches stabilized by cross-strand base-stacking and BII phosphate groups." *J. Mol. Biol.* 228, 138-155.
- Chuprina, V.P. (1985) "Regularities in formation of the spine of hydration in the DNA minor groove and its influence on the DNA structure." *FEBS Lett.* 186, 98-102.
- Chuprina, V.P. (1987) "Anomalous structure and properties of poly(dA)-poly(dT). Computer simulation of the polynucleotide structure with the spine of hydration in the minor groove." *Nucleic Acids Res.* 15, 293-311.
- Church, K.M., Wurdeman, R.L., Zhang, Y., Chen, F.-X. & Gold, B. (1990) "N-(2-chloroethyl)-N-nitrosoureas covalently bound to nonionic and monocationic lexitropsin dipeptides. Synthesis, DNA affinity binding characteristics, and reactions with ³²P-end-labeled DNA." *Biochemistry* 29, 6827-6838.
- Coll, M., Aymami, J., van der Marel, G.A., van Boom, J.H., Rich, A. & Wang, A.H.-J. (1989) "Molecular structure of the netropsin-d(CGCGATATCGCG) complex: DNA conformation in an alternating AT segment." *Biochemistry* 28, 310-320.
- Coll, M., Fredrick, C.A., Wang, A.H.-J. & Rich, A. (1987) "A bifurcated hydrogen-bonded conformation in the d(A·T) base pairs of the DNA dodecamer d(CGCAAATTTGCG) and its complex with distamycin." *Proc. Natl. Acad. Sci. U.S.A.* 84, 8385-8389.
- Czarny, A., Boykin, D.W., Wood, A.A., Nunn, C.M., Neidle, S., Zhao, M. & Wilson, W.D. (1995) "Analysis of van der Waals and electrostatic contributions in the interactions of minor groove binding benzimidazoles with DNA." *J. Am. Chem. Soc.* 117, 4716-4717.

- Dabrowiak, J.C., Goodisman, J. & Kissinger, K. (1990) "Thermodynamic data from drug-DNA footprinting experiments." *Biochemistry* 29, 6139-6145.
- Dervan, P.B. (1986) "Design of sequence-specific DNA-binding molecules." *Science* 232, 464-471.
- Dickerson, R.E. (1989) "Definitions and nomenclature of nucleic acid structure parameters." *Nucleic Acids Res.* 17, 1797-1803.
- Dickerson, R.E. (1990) in *Structure & Methods* (R.H. Sarma & M.H. Sarma, ed.) 1-38, Adenine Press, Schenectady, New York.
- Dickerson, R.E. (1992) "DNA structure from A to Z." *Meth. Enzymol.* 211, 67-111.
- Dickerson, R.E. & Drew, H.R. (1981) "Structure of a B-DNA dodecamer. II. Influence of base sequence on helix structure." *J. Mol. Biol.* 149, 761-786.
- Diekmann, S., von Kitzing, E., McLaughlin, L., Ott, J. & Eckstein, F. (1987) "The influence of exocyclic substituents of purine bases on DNA curvature." *Proc. Natl. Acad. Sci. USA* 84, 8257-8261.
- Discover User Guide, version 2.8 (1992) Biosym Technologies, San Diego.
- Doktycz, M.J., Morris, M.D., Dormady, S.J., Beattie, K.L. & Jacobson, K.B. (1995) "Optical melting of 128 octamer DNA duplexes: Effects of base pair location and nearest neighbors on thermal stability." *J. Biol. Chem.* 270, 8439-8445.
- Drake, J.W., Allen, E.F., Forsberg, S.A., Preparata, R.-M. & Greening, E.O. (1969) "Spontaneous mutation." *Nature* 221, 1128-1132.
- Drew, H.R. & Travers, A.A. (1984) "DNA structural variations in the *E. coli tyrT* promoter." *Cell* 37, 491-502.
- Drobny, G., Pines, A., Sinton, S., Weitekamp, D.P. & Wemmer, D.E. (1978) "Fourier transform multiple quantum nuclear magnetic resonance." *Faraday Div. Chem. Soc. Symp.* 13, 49-55.
- Dwyer, T.J., Geierstanger, B.H., Bathini, Y., Lown, J.W. & Wemmer, D.E. (1992) "Design and binding of a distamycin A analog to d(CGCAAGTTGGC)-d(GCCAACTTGCG): Synthesis, NMR studies, and implications for the design of sequence-specific minor groove binding oligopeptides." *J. Am. Chem. Soc.* 114, 5911-5919.
- Dwyer, T.J., Geierstanger, B.H., Mrksich, M., Dervan, P.B. & Wemmer, D.E. (1993) "Structural analysis of covalent peptide dimers, bis(pyridine-2-carboxamidonetropsin)(CH₂)₃₋₆, in complex with 5'-TGACT-3' sites by two-dimensional NMR." *J. Am. Chem. Soc.* 115, 9900-9906.
- Echols, H. & Goodman, M.F. (1991) "Fidelity mechanisms in DNA replication." *Ann. Rev. Biochem.* 60, 477-511.

- Edwards, K.J., Brown, D.G., Spink, N., Skelly, J.V. & Neidle, S. (1992) "Molecular structure of the B-DNA dodecamer d(CGCAAATTTGCG)₂: An examination of propeller twist and minor groove water structure at 2.2 Å resolution." *J. Mol. Biol.* 226, 1161-1173.
- Fasman, G.D. (Ed.) (1975) *Handbook of Biochemistry and Molecular Biology*, Vol. I. CRC Press, Cleveland, Ohio.
- Fazakerley, G.V., Sowers, L.C., Eritja, R., Kaplan, B.E. & Goodman, M.F. (1987) "NMR studies on an oligodeoxynucleotide containing 2-aminopurine opposite adenine." *Biochemistry* 26, 5641-5646.
- Fede, A., Billeter, M., Leupin, W. & Wüthrich, K. (1993) "Determination of the NMR solution structure of the Hoechst 33258-d(GTGGAATTCCAC)₂ complex and comparison with the X-ray structure." *Structure* 1, 177-186.
- Fish, E.L., Lane, M.J. & Vournakis, J.N. (1988) "Determination of equilibrium binding affinity of distamycin and netropsin to the synthetic deoxyoligonucleotide sequence d(GGTATACC)₂ by quantitative DNase I footprinting." *Biochemistry* 27, 6026-6032.
- Fox, K.R. (1992) "Probing the conformations of eight cloned DNA dodecamers: CGCGAATTCGCG, CGCGTTAACGCG, CGCGTATACGCG, CGCGATATCGCG, CGCAAATTTGCG, CGCTTTAAAGCG, CGCGGATCCGCG and CGCGGTACCGCG." *Nucleic Acids Res.* 20, 6487-6493.
- Freese, E. (1959) "The specific mutagenic effect of base analogues on phage T4." *J. Mol. Biol.* 1, 87-105.
- Frey, M.W., Sowers, L.C., Millar, D.P. & Benkovic, S.J. (1995) "Nucleotide analog 2-aminopurine as a spectroscopic probe of nucleotide incorporation by the Klenow fragment of *Escherichia coli* Polymerase I and bacteriophage T4 DNA polymerase." *Biochemistry* 34, 9185-9192.
- Gao, X. & Patel, D.J. (1988) "G(syn)-A(anti) mismatch formation in DNA dodecamers at acidic pH: pH-dependent conformational transition of G-A mispairs detected by proton NMR." *J. Am. Chem. Soc.* 110, 5178-5182.
- Gao, Y.G., Sriram, M., Denny, W.A. & Wang, A.H.-J. (1993) "Minor groove binding of SN6999 to an alkylated DNA: Molecular structure of d(CGC[e⁶]AATTCGCG)-SN6999 complex." *Biochemistry* 32, 9639-9648.
- Geierstanger, B.H., Dwyer, T.J., Bathini, Y., Lown, J.W. & Wemmer, D.E. (1993) "NMR characterization of a heterocomplex formed by distamycin and its analog 2-ImD with d(CGCAAGTTGGC):d(GCCAACTTGCG): Preference for the 1:1:1 2-

- ImD:Dst:DNA complex over the 2:1 2-ImD:DNA and the 2:1 Dst:DNA complexes." *J. Am. Chem. Soc.* 115, 4474-4482.
- Geierstanger, B.H., Jacobsen, J.P., Mrksich, M., Dervan, P.B. & Wemmer, D.E. (1994a) "Structural and dynamic characterization of the heterodimeric and homodimeric complexes of distamycin and 1-methylimidazole-2-carboxamide-netropsin bound to the minor groove of DNA." *Biochemistry* 33, 3055-3062.
- Geierstanger, B.H., Mrksich, M., Dervan, P.B. & Wemmer, D.E. (unpublished results)
- Geierstanger, B.H., Parks, M., Dervan, P. & Wemmer, D.E. (1994b) "Design of a GC-specific DNA minor groove-binding peptide." *Science* 266, 646-650.
- Geierstanger, B.H., Volkman, B.F., Kremer, W. & Wemmer, D.E. (1994c) "Short peptide fragments derived from HMG-I/Y proteins bind specifically to the minor groove of DNA." *Biochemistry* 33, 5347-5355.
- Gervais, V., Cognet, J.A., Le Bret, M., Sowers, L.C. & Fazakerley, G.V. (1995) "Solution structure of two mismatches A.A and T.T in the K-ras gene context by nuclear magnetic resonance and molecular dynamics." *Eur. J. Biochem.* 228, 279-290.
- Goldberg, I.H. (1993) in *Molecular Aspects of Anticancer Drug-DNA Interactions* (S. Neidle & M. Waring, ed.) 364, CRC Press, Boca Raton, Florida.
- Goodman, M.F. & Ratliff, R.L. (1983) "Evidence of 2-aminopurine-cytosine base mispairs involving two hydrogen bonds." *J. Biol. Chem.* 258, 12842-12846.
- Goodsell, D.S., Kopka, M.L., Cascio, D. & Dickerson, R.E. (1993) "Crystal structure of CATGGCCATG and its implications for A-tract bending models." *Proc. Natl. Acad. Sci. USA* 90, 2930-2934.
- Greene, K.L., Jones, R.L., Li, Y., Robinson, H., Wang, A.H., Zon, G. & Wilson, W.D. (1994) "Solution structure of a GA mismatch DNA sequence, d(CCATGAATGG)₂ determined by 2D NMR and structural refinement methods." *Biochemistry* 33, 1053-1062.
- Guéron, M., Charretier, E., Hagerhorst, J., Kochoyan, M., Leroy, J.L. & Moraillon, A. (1990a) in *Structure & Methods* (R.H. Sarma & M.H. Sarma, ed.) 113-137, Adenine Press, Schenectady, New York.
- Guéron, M., Charretier, E., Kochoyan, M. & Leroy, J.L. (1990b) in *Frontiers of NMR in Molecular Biology* (D. Live, I.M. Armitage & D. Patel, ed.) 225-238, Alan R. Liss, New York.
- Guest, C.R., Hochstrasser, R.A., Sowers, L.C. & Millar, D.P. (1991) "Dynamics of mismatched base pairs in DNA." *Biochemistry* 30, 3271-3279.
- Hagerman, P.J. (1986) "Sequence-directed curvature of DNA." *Nature* 321, 449-450.

- Hagerman, P.J. (1990a) "Pyrimidine 5-methyl groups influence the magnitude of DNA curvature." *Biochemistry* 29, 1980-1983.
- Hagerman, P.J. (1990b) "Sequence-directed curvature of DNA." *Ann. Rev. Biochem.* 59, 755-781.
- Hare, D., Shapiro, L. & Patel, D.J. (1986) "Wobble dG-dT pairing in right-handed DNA: Solution conformation of the d(CGTGAATTCGCG) duplex deduced from distance geometry analysis of nuclear Overhauser effect spectra." *Biochemistry* 25, 7445-7456.
- Hare, D.R., Wemmer, D.E., Chou, S.-H., Drobny, G. & Reid, B.R. (1983) "Assignment of the non-exchangeable proton resonances of d(CGCGAATTCGCG) using two-dimensional nuclear magnetic resonance methods." *J. Mol. Biol.* 171, 319-336.
- Harris, R.K. (1986) *Nuclear Magnetic Resonance Spectroscopy*, John Wiley & Sons, New York.
- Hatayama, T. & Goldberg, I.H. (1978) *Proc. Natl. Acad. Sci. USA* 75, 3603-3607.
- Hatayama, T. & Goldberg, I.H. (1979) *Biochim. Biophys. Acta* 563, 59-71.
- Heinemann, U. & Alings, C. (1989) "Crystallographic study of one turn of G/C-rich B-DNA." *J. Mol. Biol.* 210, 369-381.
- Hore, P.J. (1983) "Solvent suppression in Fourier transform nuclear magnetic resonance." *J. Magn. Reson.* 55, 283-300.
- Hosur, R.V., Govil, G. & Miles, H.T. (1988) "Application of two-dimensional NMR spectroscopy in the determination of solution conformation of nucleic acids." *Mag. Res. Chem.* 26, 927-944.
- Hoult, D.I. & Richards, R.E. (1975) "Critical factors in the design of sensitive high resolution nuclear magnetic resonance spectrometers." *Proc. Royal Soc. Lon. A* 344, 311-340.
- Hu, S.H., Weisz, K., James, T.L. & Shafer, R.H. (1992) "¹H-NMR studies on d(GCTTAAGC)₂ and its complex with berenil." *Eur. J. Biochem.* 204, 31-38.
- Huang, L., Quada, J.C. & Lown, J.W. (1995) "Design, synthesis, and sequence selective DNA cleavage of functional models of bleomycin. 1. Hybrids incorporating a simple metal-complexing moiety of bleomycin and lexitropsin carriers." *Bioconjugate Chem.* 6, 21-33.
- Hunter, W.N. (1992) "Crystallographic studies of DNA containing mismatches, modified and unpaired bases." *Meth. Enzymol.* 211, 221-231.
- Ikemoto, N., Kumar, R.A., Ling, T.T., Ellestad, G.A., Danishefsky, S.J. & Patel, D.J. (1995) "Calicheamicin-DNA complexes: warhead alignment and saccharide recognition of the minor groove." *Proc. Natl. Acad. Sci. USA* 92, 10506-10510.

- Iyer, V.N. & Szybalski, W. (1963) *Proc. Natl. Acad. Sci. USA* 50, 355-362.
- James, T.L., James, J.L. & Lapidot, A. (1981) "Structural and dynamic information about double-stranded DNA from nitrogen-15 NMR spectroscopy." *J. Am. Chem. Soc.* 103, 6748-6750.
- Janion, C. & Shugar, D. (1973) "Preparation and properties of poly-2-aminopurine ribotidylic acid." *Acta Biochim. Polon.* 20, 271-284.
- Jeener, J., Meier, B.H., Bachmann, P. & Ernst, R.R. (1979) *J. Chem. Phys.* 71, 4546-4553.
- Job, M. (1928) *Ann. Chim.* 9, 113.
- Joyce, C.M., Chen Sun, X. & Grindley, N.G. (1992) "Reactions at the polymerase active site that contribute to the fidelity of Escherichia coli DNA polymerase 1 (Klenow fragment)." *J. Biol. Chem.* 267, 24485-24500.
- Kalnick, M.W., Li, B.F.L., Swann, P.F. & Patel, D.J. (1989) "O6-ethylguanine carcinogenic lesions in DNA: An NMR study of O6-et-G:C pairing in dodecanucleotide duplexes." *Biochemistry* 28, 6182-6192.
- Katahira, M., Sugeta, H. & Kyogoku, Y. (1990) "A new model for the bending of DNAs containing the oligo(dA) tracts based on NMR observations." *Nucleic Acids Res.* 18, 613-618.
- Katahira, M., Sugeta, H., Kyogoku, Y., Fujii, S., Fujisawa, R. & Tomita, K. (1988) "One- and two-dimensional NMR studies on the conformation of DNA containing the oligo(dA)oligo(dT) tract." *Nucleic Acids Res.* 16, 8619-8632.
- Ke, S.-H. & Wartell, R.M. (1993) "Influence of nearest neighbor sequence on the stability of base pair mismatches in long DNA: Determination by temperature-gradient gel electrophoresis." *Nucleic Acids Res.* 21, 5137-5143.
- Keeler, J. (1988) in *Multinuclear Magnetic Resonance in Liquids and Solids - Chemical Applications* (P. Granger & R.K. Harris, ed.) 103-129, Kluwer Academic Publishers, Boston.
- Kessler, H., Gehrke, M. & Griesinger, C. (1988) "Two-dimensional NMR spectroscopy: Background and overview of the experiments." *Angew. Chem. Int. Ed. Engl.* 27, 490-536.
- Kim, J.L., Nikolov, D.B. & Burley, S.K. (1993a) "Co-crystal structure of TBP recognizing the minor groove of a TATA element." *Nature* 365, 520-527.
- Kim, Y., Geiger, J.H., Hahn, S. & Sigler, P.B. (1993b) "Crystal structure of a yeast TBP/TATA-box complex." *Nature* 365, 512-520.

- Kime, M.J. & Moore, P.B. (1983) "Nuclear Overhauser experiments at 500 MHz on the downfield proton spectrum of a ribonuclease-resistant fragment of 5S ribonucleic acid." *Biochemistry* 22, 2615-2622.
- Kintanar, A., Klevit, R.E. & Reid, B.R. (1987) "Two-dimensional NMR investigation of a bent DNA fragment: Assignment of the proton resonances and preliminary structure analysis." *Nucleic Acids Res.* 15, 5845-5862.
- Klevit, R.E., Wemmer, D.E. & Reid, B.R. (1986) "¹H NMR studies on the interaction between distamycin A and a symmetrical DNA dodecamer." *Biochemistry* 25, 3296-3303.
- Kneale, G., Brown, T., Kennard, O. & Rabinovich, D. (1985) "G-T base-pairs in a DNA helix: The crystal structure of d(GGGGTCCCC)." *J. Mol. Biol.* 186, 805-814.
- Kohwi, Y. & Panchenko, Y. (1993) "Transcription-dependent recombination induced by triple-helix formation." *Genes Develop.* 7, 1766-1778.
- Koo, H.-S. & Crothers, D.M. (1987) "Chemical determinants of DNA bending at adenine-thymine tracts." *Biochemistry* 26, 3745-3748.
- Kopka, M.L., Yoon, C., Goodsell, D., Pjura, P. & Dickerson, R.E. (1985a) "Binding of an antitumor drug to DNA: Netropsin and CGCGAATT^{Br}CGCG." *J. Mol. Biol.* 183, 553-563.
- Kopka, M.L., Yoon, C., Goodsell, D., Pjura, P. & Dickerson, R.E. (1985b) "The molecular origin of DNA-drug specificity in netropsin and distamycin." *Proc. Natl. Acad. Sci. USA* 82, 1376-1380.
- Kornberg, A. (1980) *DNA Replication*, Freeman, San Francisco.
- Kubinec, M.G. & Wemmer, D.E. (1992) "NMR evidence for DNA bound water in solution." *J. Am. Chem. Soc.* 114, 8739-8740.
- Larsen, T.A., Kopka, M.L. & Dickerson, R.E. (1991) "Crystal structure analysis of the B-DNA dodecamer CGTGAATTCACG." *Biochemistry* 30, 4443-4449.
- Lee, M., Rhodes, A.L., Wyatt, M.D., Forrow, S. & Hartley, J. (1993a) "GC base sequence recognition by oligo(imidazolecarboxamide) and terminus-modified analogues of distamycin deduced from circular dichroism, proton nuclear magnetic resonance, and methidiumpropylethylenediaminetetraacetate-iron(II) footprinting studies." *Biochemistry* 32, 4237-4245.
- Lee, M., Rhodes, A.L., Wyatt, M.D., D'Incalci, M., Forrow, S. & Hartley, J.A. (1993b) "In Vitro cytotoxicity of GC sequence directed alkylating agents related to distamycin." *J. Med. Chem.* 36, 863-870.
- Lee, M., Roldan, M.C., Haskell, M.K., McAdam, S.R. & Hartley, J.A. (1994) "In Vitro photoinduced cytotoxicity and DNA binding properties of psoralen and coumarin

- conjugates of netropsin analogues: DNA sequence-directed alkylation and cross-link formation." *J. Med. Chem.* 37, 1208-1213.
- Leijon, M. & Gräslund, A. (1992) "Effects of sequence and length on imino proton exchange and base pair opening kinetics in DNA oligonucleotide duplexes." *Nucleic Acids Res.* 20, 5339-5343.
- Leroy, J.L., Kochoyan, M., Huynh-Dinh, T. & Guéron, M. (1988) "Characterization of base-pair opening in deoxynucleotide duplexes using catalyzed exchange of the imino proton." *J. Mol. Biol.* 200, 223-238.
- Liepinsh, E., Leupin, W. & Otting, G. (1994) "Hydration of DNA in aqueous solution: NMR evidence for a kinetic destabilization of the minor groove hydration of d(TTAA)₂ versus d(AATT)₂ segments." *Nucleic Acids Res.* 22, 2249-2254.
- Liepinsh, E., Otting, G. & Wüthrich, K. (1992) "NMR observation of individual molecules of hydration water bound to DNA duplexes: direct evidence for a spine of hydration water present in aqueous solution." *Nucleic Acids Res.* 20, 6549-6553.
- Liquier, J., Mchami, A. & Taillandier, E. (1989) "FTIR study of netropsin binding to poly d(A-T) and poly dA.poly dT." *J. Biomol. Struct. Dyn.* 7, 119-126.
- Little, R.D., Masjedizadeh, M.R., Moeller, K.D. & Dannecker-Doering, I. (1992) "Factors affecting regioselectivity in the intramolecular diyl trapping reaction." *Synlett* 2, 107-113.
- Live, D.H., Dane, D.G., Agosta, W.C. & Cowburn, D. (1984) "Long range hydrogen bond mediated effects in peptides: ¹⁵N NMR study of gramicidin S in water and organic solvents." *J. Am. Chem. Soc.* 106, 1939-1941.
- Love, J.J., Li, X., Case, D.A., Giese, K., Grosschedl, R. & Wright, P.E. (1995) "Structural basis for DNA bending by the architectural transcription factor LEF-1." *Nature* 376, 791-795.
- Lown, J.W. (1994) "DNA recognition by lexitropsins, minor groove binding agents." *J. Mol. Recogn.* 7, 79-88.
- Lown, J.W., Krowicki, K., Bhat, U.G., Skorobogaty, A., Ward, B. & Dabrowiak, J.C. (1986) "Molecular recognition between oligopeptides and nucleic acids: Novel imidazole-containing oligopeptides related to netropsin that exhibit altered DNA sequence specificity." *Biochemistry* 25, 7408-7416.
- Marini, J.C., Levene, S.D., Crothers, D.M. & Englund, P.T. (1982) "Bent helical structure in kinetoplast DNA." *Proc. Natl. Acad. Sci. USA* 79, 7664-7668.

- Marion, D. & Wüthrich, K. (1983) "Application of phase sensitive two-dimensional correlated spectroscopy (COSY) for measurements of ^1H - ^1H spin-spin coupling constants in proteins." *Biochem. Biophys. Res. Comm.* 113, 967-974.
- Marky, L.A. (1986) *Polym. Prepr. (Am. Chem. Soc., Div. Polym. Chem.)* 27, 417-418.
- Marky, L.A. & Breslauer, K.J. (1987) "Origins of netropsin binding affinity and specificity: Correlations of thermodynamic and structural data." *Proc. Natl. Acad. Sci. USA* 84, 4359-4363.
- Marky, L.A., Curry, J. & Breslauer, K.J. (1985) in *Molecular Basis of Cancer, Part B: Macromolecular Recognition, Chemotherapy and Immunology* ed.) 155-173, Alan R. Liss, New York.
- Marky, L.A. & Kupke, D.W. (1989) "Probing the hydration of the minor groove of A-T synthetic DNA polymers by volume and heat changes." *J. Am. Chem. Soc.* 111, 9982-9988.
- McAuley-Hecht, K.E., Leonard, G.A., Gibson, N.J., Thomson, J.B., Watson, W.P., Hunter, W.N. & Brown, T. (1994) "Crystal structure of a DNA duplex containing 8-hydroxydeoxyguanine-adenine base pairs." *Biochemistry* 33, 10266-10270.
- McCarthy, J.G., Williams, L.D. & Rich, A. (1990) "Chemical reactivity of potassium permanganate and diethyl pyrocarbonate with B DNA: Specific reactivity with short A-tracts." *Biochemistry* 29, 6071-6081.
- Mendelman, L.V., Boosalis, M.S., Petruska, J. & Goodman, M.F. (1989) "Nearest neighbor influences on DNA polymerase insertion fidelity." *J. Biol. Chem.* 264, 14415-14423.
- Mhaskar, D.N. & Goodman, M.F. (1984) "On the molecular basis of transition mutations. Frequency of forming 2-aminopurine:cytosine base mispairs in the G-C $\rightarrow$ A-T mutational pathway by T4 DNA polymerase *in vitro*." *J. Biol. Chem.* 259, 11713-11717.
- Millard, J.T., Weidner, M.F., Raucher, S. & Hopkins, P.B. (1990) "Determination of the DNA cross-linking sequence specificity of reductively activated mitomycin C at single nucleotide resolution: deoxyguanosine residues at CpG are cross-linked preferentially." *J. Am. Chem. Soc.* 112, 3637-3641.
- Mitra, R., Pettitt, B.M., Rame, G.L. & Blake, R.D. (1993) "The relationship between mutation rates for the (C-G) $\rightarrow$ (T-A) transition and features of T-G mispair structures in different neighbor environments, determined by free energy molecular mechanics." *Nucleic Acids Res.* 21, 6028-6037.
- Modrich, P. (1987) "DNA mismatch correction." *Ann. Rev. Biochem.* 56, 435-66.
- Morris, G.A. & Freeman, R. (1978) *J. Magn. Reson.* 29, 433-462.

- Mrksich, M., Parks, M.E. & Dervan, P.B. (1994) "Hairpin peptide motif. A new class of oligopeptides for sequence-specific recognition in the minor groove of double-helical DNA." *J. Am. Chem. Soc.* 116, 7983-7988.
- Mrksich, M., Wade, W.S., Dwyer, T.J., Geierstanger, B.H., Wemmer, D.E. & Dervan, P.B. (1992) "Antiparallel side-by-side dimeric motif for sequence-specific recognition in the minor groove of DNA by the designed peptide 1-methylimidazole-2-carboxamide netropsin." *Proc. Natl. Acad. Sci. USA* 89, 7586-7590.
- Nadeau, J.G. & Crothers, D.M. (1989) "Structural basis for DNA bending." *Proc. Natl. Acad. Sci. USA* 86, 2622-2626.
- Nedderman, A.N.R., Stone, M.J., Williams, D.H., Lin, P.K.T. & Brown, D.M. (1993) "Molecular basis for methoxyamine-initiated mutagenesis: ¹H nuclear magnetic resonance studies of oligonucleotide duplexes containing base-modified cytosine residues." *J. Mol. Biol.* 230, 1068-1076.
- Nelsen, S.F., Blackstock, S.C. & Frigo, T.B. (1986) *Tetrahedron* 42, 1769.
- Nelson, H.C.M., Finch, J.T., Luisi, B.F. & Klug, A. (1987) "The structure of an oligo(dA)-oligo(dT) tract and its biological implications." *Nature* 330, 221-226.
- Neuhaus, D. & Williamson, M.P. (1989) *The Nuclear Overhauser Effect in Structural and Conformational Analysis*, VCH Publishers, New York.
- Nicolaou, K.C. & Dai, W.-M. (1991) "Chemistry and biology of the enediyne anticancer antibiotics." *Angew. Chem. Int. Ed. Engl.* 30, 1387-1416.
- Nicolaou, K.C., Smith, A.L. & Yue, E.W. (1993) "Chemistry and biology of natural and designed enediynes." *Proc. Natl. Acad. Sci. USA* 90, 5881-5888.
- Nightingale, K.P. & Fox, K.R. (1993) "DNA structure influences sequence specific cleavage by bleomycin." *Nucleic Acids Res.* 21, 2549-2555.
- Nikonowicz, E.P., Meadows, R.P., Fagan, P. & Gorenstein, D.G. (1991) "NMR structural refinement of a tandem GA mismatched decamer d(CCAAGATTGG)₂ via the hybrid matrix procedure." *Biochemistry* 30, 1323-1334.
- Norman, D., Live, D., Sastry, M., Lipman, R., Hingerty, B.E., Tomasz, M., Broyde, S. & Patel, D.J. (1990) "NMR and computational characterization of mitomycin cross-linked to adjacent deoxyguanosines in the minor groove of the d(TACGTA)-d(TACGTA) duplex." *Biochemistry* 29, 2861-2875.
- Otting, G., Liepinsh, E. & Wüthrich, K. (1991) "Protein hydration in aqueous solution." *Science* 254, 974-980.
- Palmer, A.G. (1993) "Dynamics properties of proteins from NMR spectroscopy." *Curr. Opin. Biotech.* 4, 385-391.

- Paloma, L.G., Smith, J.A., Chazin, W.J. & Nicolaou, K.C. (1994) "Interaction of calicheamicin with duplex DNA: role of the oligosaccharide domain and identification of multiple binding modes." *J. Am. Chem. Soc.* **116**, 3697-3708.
- Parkinson, J.A., Barber, J., Douglas, K.T., Rosamond, J. & Sharples, D. (1990) "Minor-groove recognition of the self-complementary duplex d(CGCGAATTCGCG)₂ by Hoechst 33258: a high-field NMR study." *Biochemistry* **29**, 10181-10190.
- Patel, D.J. (1979) "Netropsin-d(GGAATTCC) complex: Antibiotic binding at adenine-thymine base pairs in the minor groove of the self-complementary octanucleotide duplex." *Eur. J. Biochem.* **99**, 369-378.
- Patel, D.J. (1982) "Antibiotic-DNA interactions: Intermolecular nuclear Overhauser effects in the netropsin-d(CGCGAATTCGCG) complex in solution." *Proc. Natl. Acad. Sci. USA* **79**, 6424-6428.
- Patel, D.J., Kozlowski, S.A., Marky, L.A., Rice, J.A., Broka, C., Dallas, J., Itakura, K. & Breslauer, K.J. (1982) "Structure, dynamics, and energetics of deoxyguanosine-thymidine wobble base pair formation in the self-complementary d(CGTGAATTCGCG) duplex in solution." *Biochemistry* **21**, 437-444.
- Patel, D.J. & Shapiro, L. (1986) "Sequence-dependent recognition of DNA duplexes: Netropsin complexation to the AATT site of the d(GGAATTCC) duplex in aqueous solution." *J. Biol. Chem.* **261**, 1230-1240.
- Pelton, J.G. (1990) "Binding modes and dynamics of distamycin A with several A-T-rich oligomers by two-dimensional NMR." Ph.D. thesis, University of California, Berkeley.
- Pelton, J.G. & Wemmer, D.E. (1988) "Structural modeling of the distamycin A-d(CGCGAATTCGCG)₂ complex using 2D NMR and molecular mechanics." *Biochemistry* **27**, 8088-8096.
- Pelton, J.G. & Wemmer, D.E. (1989) "Structural characterization of a 2:1 distamycin A-d(CGCAAATTGGC) complex by two-dimensional NMR." *Proc. Natl. Acad. Sci. USA* **86**, 5723-5727.
- Pelton, J.G. & Wemmer, D.E. (1990a) "Binding modes of distamycin A with d(CGCAAATTTGCG)₂ determined by two-dimensional NMR." *J. Am. Chem. Soc.* **112**, 1393-1399.
- Pelton, J.G. & Wemmer, D.E. (1990b) "Structure and dynamics of distamycin A with d(CGCAAATTGGC):d(GCCAATTTGCG) at low drug:DNA ratio." *J. Biomol. Struct. Dyn.* **8**, 81-97.

- Perrin, C.L. & Dwyer, T.J. (1990) "Application of 2-dimensional NMR to kinetics of chemical exchange." *Chem. Rev.* 90, 935-967.
- Petruska, J. & Goodman, M.F. (1985) "Influence of neighboring bases on DNA polymerase insertion and proofreading fidelity." *J. Biol. Chem.* 260, 7533-7539.
- Pinto, A.L. & Lippard, S.J. (1985) "Binding of the antitumor drug cis-diamminedichloroplatinum(II) (cisplatin) to DNA." *Biochim. Biophys. Acta* 780, 167-180.
- Pjura, P.E., Grzeskowiak, K. & Dickerson, R.E. (1987) "Binding of Hoechst 33258 to the minor groove of B-DNA." *J. Mol. Biol.* 197, 257-271.
- Pless, R.C. & Bessman, M.J. (1983) "Influence of local nucleotide sequence on substitution of 2-aminopurine for adenine during deoxyribonucleic acid synthesis in vitro." *Biochemistry* 22, 4905-4915.
- Price, M.A. & Tullius, T.D. (1992) "Using hydroxyl radical to probe DNA structure." *Meth. Enzymol.* 212, 194-219.
- Price, M.A. & Tullius, T.D. (1993) "How the structure of an adenine tract depends on sequence context: a new model for the structure of T_nA_n DNA sequences." *Biochemistry* 32, 127-136.
- Privé, G.G., Heinemann, U., Chandrasegaran, S., Kan, L.-S., Kopka, M. & Dickerson, R.E. (1987) "Helix geometry, hydration, and G-A mismatch in a B-DNA decamer." *Science* 238, 498-504.
- Pullman, A. & Pullman, B. (1981) "Molecular electrostatic potential of the nucleic acids." *Quart. Rev. Biophys.* 14, 289-380.
- Pullman, B. (1989) in *Advances in Drug Research* (B. Testa, ed.) Vol. 1, Academic Publishers, London.
- Quignard, E., Fazakerley, G.V., van der Marel, G., van Boom, J.H. & Guschlbauer, W. (1987) "Comparison of the conformation of an oligonucleotide containing a central G-T base pair with the non-mismatch sequence by proton NMR." *Nucleic Acids Res.* 15, 3397-3409.
- Rajagopal, P. (1988) Ph.D. thesis, University of Washington, Seattle.
- Rentzeperis, D., Marky, L.A., Dwyer, T.J., Geierstanger, B.H., Pelton, J.G. & Wemmer, D.E. (1995) "Interaction of minor groove ligands to an AAATT/AATTT site: Correlation of thermodynamic characterization and solution structure." *Biochemistry* 34, 2937-2945.
- Ronen, A. (1979) "2-aminopurine." *Mut. Res.* 75, 1-47.
- Saenger, W. (1983) *Principles of Nucleic Acid Structure*, Springer-Verlag, New York.

- Sanders, J.K.M. & Hunter, B.K. (1987) *Modern NMR Spectroscopy*, Oxford University Press, Oxford.
- Schultz, P.G. & Dervan, P.B. (1983) "Sequence-specific double-strand cleavage of DNA by bis(EDTA-distamycin-Fe^{II}) and EDTA-bis(distamycin)·Fe^{II}." *J. Am. Chem. Soc.* 105, 7748-7750.
- Schultz, P.G. & Dervan, P.B. (1984) "Distamycin and penta-N-methylpyrrolicarboxamide binding sites on native DNA. A comparison of methidiumpropyl-EDTA-Fe(II) footprinting and DNA affinity cleavage." *J. Biomol. Struct. Dyn.* 1, 1133-1147.
- Schultz, P.G., Taylor, J.S. & Dervan, P.B. (1982) "Design and synthesis of a sequence-specific DNA cleaving molecule. (Distamycin-EDTA)iron(II)." *J. Am. Chem. Soc.* 104, 6861-6663.
- Schumacher, M.A., Choi, K.Y., Zalkin, H. & Brennan, R.G. (1994) "Crystal structure of LacI member, PurR, bound to DNA: Minor groove binding by alpha helices." *Science* 266, 763-770.
- Semmelhack, M.F., Gallagher, J.J., Ding, W.-D., Krishnamurthy, G., Babine, R. & Ellestad, G.A. (1994) "The effect on DNA cleavage potency of tethering a simple cyclic enediyne to a netropsin analog." *J. Org. Chem.* 59, 4357-4359.
- Sessa, C., Pagani, O., Zurlo, M.G., de Jong, J., Hofmann, C., Lassus, M., Marrari, P., Strolin, B.M. & Cavalli, F. (1994) "Phase I study of the novel distamycin derivative tallimustine (FCE 24517)." *Annals of Oncology* 5, 901-907.
- Shum, B. & Crothers, D.M. (1983) *Biopolymers* 22, 919-933.
- Sigurdsson, S.Th. & Hopkins, P.B. (1994) "Synthesis and reactions with DNA of a family of DNA-DNA affinity cross-linking agents." *Tetrahedron* 50, 12065-12084.
- Sigurdsson, S.Th., Rink, S.M. & Hopkins, P.B. (1993) "Affinity cross-linking of duplex DNA by a pyrrole-oligopeptide conjugate." *J. Am. Chem. Soc.* 115, 12633-12634.
- Sklenar, V., Brooks, B.R., Zon, G. & Bax, A. (1987) "Absorption mode two-dimensional NOE spectroscopy of exchangeable protons in oligonucleotides." *FEBS Lett.* 216, 249-252.
- Smith, F.W., Schultze, P. & Feigon, J. (1995) "Solution structures of unimolecular quadruplexes formed by oligonucleotides containing Oxytricha telomere repeats." *Structure* 3, 997-1008.
- Sowers, L.C., Eritja, R., Chen, F.M., Khwaja, T., Kaplan, B., Goodman, M.F. & Fazakerley, G.V. (1989) "Characterization of the high pH wobble structure of the 2-aminopurine cytosine mismatch by N15 NMR." *Biochem. Biophys. Res. Comm.* 165, 89-92.

- Sowers, L.C., Fazakerley, G.V., Eritja, R., Kaplan, B.E. & Goodman, M.F. (1986)
"Base pairing and mutagenesis: Observation of a protonated base pair between 2-aminopurine and cytosine in an oligonucleotide by proton NMR." *Proc. Natl. Acad. Sci. USA* 83, 5434-5438.
- Spielmann, H.P., Fagan, P.A., Bregant, T.M., Little, R.D. & Wemmer, D.E. (1995)
"The binding modes of a rationally designed photoactivated DNA nuclease determined by NMR." *Nucleic Acids Res.* 23, 1576-1583.
- Sriram, M., van der Marel, G.A., Roelen, H.L.P.F., van Boom, J.H. & Wang, A.H.-J. (1992) "Structural consequences of a carcinogenic alkylation lesion on DNA: effect of O⁶-ethylguanine on the molecular structure of the d(CGC[e⁶G]AATTCGCG)-netropsin complex." *Biochemistry* 31, 11823-11834.
- Sriram, M., Yang, D., Gao, Y.G. & Wang, A.H.-J. (1994) "Crystal and solution structures of d(CGC[e⁶G]AATTCGCG)-drug complexes reveal conformational polymorphism of O⁶-ethyl-guanine:cytosine base pair." *Ann. New York Acad. Sci.* 726, 18-43.
- States, D.J., Habekorn, R.A. & Ruben, D.J. (1982) "A two-dimensional nuclear Overhauser experiment with pure absorption phase in four quadrants." *J. Magn. Reson.* 48, 286-292.
- Stone, M.J., Nedderman, A.N.R., Williams, D.H., Lin, P.K.T. & Brown, D.M. (1991)
"Molecular basis for methoxyamine initiated mutagenesis: ¹H nuclear magnetic resonance studies of base-modified oligodeoxynucleotides." *J. Mol. Biol.* 222, 711-723.
- Tabernero, L., Verdaguer, N., Coll, M., Fita, I., van der Marel, G.A., van Boom, J.H., Rich, A. & Aymamí, J. (1993) "Molecular structure of the A-tract DNA dodecamer d(CGCAAATTTGCG) complexed with the minor groove binding drug netropsin." *Biochemistry* 32, 8403-8410.
- Teng, M.K., Usman, N., Frederick, C.A. & Wang, A.H.-J. (1988) "The molecular structure of the complex of Hoechst 33258 and the DNA dodecamer d(CGCGAATTCGCG)." *Nucleic Acids Res.* 16, 2671-2690.
- Teng, S.P., Woodson, S.A. & Crothers, D.M. (1989) "DNA sequence specificity of mitomycin cross-linking." *Biochemistry* 28, 3901-3907.
- Teplukhin, A.V., Poltev, V.I. & Chuprina, V.P. (1992) "Dependence of the hydration shell structure in the minor groove of the DNA double helix on the groove width as revealed by Monte Carlo simulation." *Biopolymers* 32, 1445-1453.
- Thurston, D.E. & Thompson, A.S. (1990) "The molecular recognition of DNA." *Chem. Britain* 26, 767-772.

- Topal, M.D. & Fresco, J.R. (1976) "Complementary base pairing and the origin of substitution mutations." *Nature* 263, 285-289.
- Travers, A.A. (1995) "Reading the minor groove." *Nature Struct. Biol.* 2, 615-618.
- Voet, D. & Voet, J.G. (1990) *Biochemistry*, John Wiley & Sons, New York.
- Wade, W.S. & Dervan, P.B. (1987) "Alteration of the sequence specificity of distamycin on DNA by replacement of an N-methylpyrrolicarboxamide with pyridine-2-carboxamide." *J. Am. Chem. Soc.* 109, 1574-1575.
- Wade, W.S., Mrksich, M. & Dervan, P.B. (1993) "Binding affinities of synthetic peptides, pyridine-2-carboxamidonetropsin and 1-methylimidazole-2-carboxamidonetropsin, that form 2:1 complexes in the minor groove of double-helical DNA." *Biochemistry* 32, 11385-11389.
- Ward, D.C., Reich, E. & Stryer, L. (1969) "Fluorescence studies of nucleotides and polynucleotides. I. Formycin, 2-aminopurine riboside, 2,6-diaminopurine riboside, and their derivatives." *J. Biol. Chem.* 244, 1228-1237.
- Warshaw, M. & Cantor, C. (1970) *Biopolymers* 9, 1079-1103.
- Watanabe, S.M. & Goodman, M.F. (1981) "On the molecular basis of transition mutations: Frequencies of forming 2-aminopurine:cytosine and adenine:cytosine base mispairs *in vitro*." *Proc. Natl. Acad. Sci. USA* 78, 2864-2868.
- Watanabe, S.M. & Goodman, M.F. (1982) "Kinetic measurement of 2-aminopurine-cytosine and 2-aminopurine-thymine base pairs as a test of DNA polymerase fidelity mechanisms." *Proc. Natl. Acad. Sci. USA* 79, 6429-6433.
- Watson, J.D. & Crick, F.H.C. (1953a) "Genetical implications of the structure of deoxyribonucleic acid." *Nature* 171, 964-967.
- Watson, J.D. & Crick, F.H.C. (1953b) "The structure of DNA." *Cold Spring Harbor Symp. Quant. Biol.* 18, 123-131.
- Weidner, M.F., Sigurdsson, S.Th. & Hopkins, P.B. (1990) "Sequence preferences of DNA interstrand cross-linking agents: dG-to-dG cross-linking at 5'-CG by structurally simplified analogues of mitomycin C." *Biochemistry* 29, 9225-9233.
- Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G.S., Profeta, J. & Weiner, P. (1984) "A new force field for molecular mechanical simulation of nucleic acids and proteins." *J. Am. Chem. Soc.* 106, 765-784.
- Weiner, S.J., Kollman, P.A., Nguyen, D.T. & Case, D.A. (1986) "An all atom force field for simulations of proteins and nucleic acids." *J. Comput. Chem.* 7, 230-252.
- Wemmer, D.E., Fagan, P.A. & Pelton, J.G. (1990) in *Molecular Basis of Specificity in Nucleic Acid-Drug Interactions* (B. Pullman & J. Jortner, ed.) 95-101, Kluwer Academic Publishers, Dordrecht, The Netherlands.

- Wemmer, D.E., Geierstanger, B.H., Fagan, P.A., Dwyer, T.J., Jacobsen, J.P., Pelton, J.G., Ball, G.E., Leheny, A.R., Chang, W.H., Bathini, Y., Lown, J.W., Rentzeperis, D., Marky, L., Singh, S.B., Ajay & Kollman, P.A. (1994) in *Structural Biology: The State of the Art, Proceedings of the Eighth Conversation on Biomolecular Stereodynamics* (R.H. Sarma & M.H. Sarma, ed.) 301-323, Adenine Press, Guilderland, NY.
- Werner, M.H., Huth, J.R., Gronenborn, A.M. & Clore, G.M. (1995) "Molecular basis of human 46X,Y sex reversal revealed from the three-dimensional solution structure of the human SRY-DNA complex." *Cell* 81, 705-714.
- Werntges, H., Steger, G., Riesner, D. & Fritz, H.-J. (1986) "Mismatches in DNA double strands: Thermodynamic parameters and their correlation to repair efficiencies." *Nucleic Acids Res.* 14, 3773-3790.
- Wing, R., Drew, H., Takano, T., Broka, C., Tanaka, S., Itakura, K. & Dickerson, R.E. (1980) "Crystal structure analysis of a complete turn of B-DNA." *Nature* 287, 755-758.
- Witanowski, M., Stefaniak, L. & Webb, G.A. (Ed.) (1986) *Annual Reports on NMR Spectroscopy*, Vol. 18. Academic Press, London.
- Woo, J., Sigurdsson, S.Th. & Hopkins, P.B. (1993) "DNA interstrand cross-linking reactions of pyrrole-derived, bifunctional electrophiles: Evidence for a common target site in DNA." *J. Am. Chem. Soc.* 115, 3407-3415.
- Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley & Sons, New York.
- Yanagi, K., Privé, G.G. & Dickerson, R.E. (1991) "Analysis of local helix geometry in three B-DNA decamers and eight dodecamers." *J. Mol. Biol.* 217, 201-214.
- Yoon, C., Privé, G.G., Goodsell, D.S. & Dickerson, R.E. (1988) "Structure of an alternating-B DNA helix and its relationship to A-tract DNA." *Proc. Natl. Acad. Sci. U.S.A.* 6332-6336.
- Youngquist, R.S. & Dervan, P.B. (1985) "Sequence-specific recognition of B-DNA by oligo(N-methylpyrrolocarboxamide)s." *Proc. Natl. Acad. Sci. USA* 82, 2565-2569.
- Yuan, H., Quintana, J. & Dickerson, R.E. (1992) "Alternative structures for alternating poly(dA-dT) tracts: The structure of the B-DNA decamer CGATATATCG." *Biochemistry* 31, 8009-8021.
- Zakrzewska, K., Lavery, R. & Pullman, B. (1983) "Theoretical studies of the selective binding to DNA of two non-intercalating ligands: netropsin and SN 18071." *Nucleic Acids Res.* 11, 8825-8839.

- Zhang, Y., Chen, F.-X., Mehta, P. & Gold, B. (1993) "Groove- and sequence-selective alkylation of DNA by sulfonate esters tethered to lexitropsins." *Biochemistry* 32, 7954-7965.
- Zimmer, C. & Wähnert, U. (1986) "Nonintercalating DNA-binding ligands: specificity of the interaction and their use as tools in biophysical, biochemical and biological investigations of the genetic material." *Prog. Biophys. Mol. Biol.* 47, 31-112.