Mechanistic studies of bleomycin-mediated double-stranded DNA cleavage and structural studies of DNA containing normal and 4'-oxidized abasic sites

by

Jingyang Chen

B.S. Biochemistry
Nanjing University, Nanjing, 1999

Submitted to the Department of Chemistry in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy in Biological Chemistry

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

[February 2006]

January 2006

© 2006 Massachusetts Institute of Technology
All rights reserved

Signature of Author: ____________________________
Department of Chemistry
January 5, 2006

Certified by: ____________________________
Professor JoAnne Stubbe
Thesis Supervisor

Accepted by: ____________________________
Professor Robert W. Field
Chairman, Departmental Committee on Graduate Students
This doctoral thesis has been examined by a Committee of the Department of Chemistry as follows:

Professor John M. Essigmann

Chairman

Professor JoAnne Stubbe

Thesis Supervisor

Professor Leona Samson
To my wife Jing

and

my parents
Acknowledgements

First, I want to thank Prof. JoAnne Stubbe for giving me the opportunity to conduct my Ph.D. research in her laboratory. Her enthusiasm about science has been inspirational throughout my graduate career. Two things that I have learned from her—quantitative approaches of scientific problems and critical evaluation of results, will be the invaluable legacy that I will carry on for the rest of my career.

I would also like to thank people who have contributed to this thesis: Dr. Christopher J. Turner at the Francis Bitter Magnet Laboratory (MIT) who provided invaluable assistance in performing 1D and 2D NMR experiments. Prof. François-Yves Dupradeau and Prof. David A. Case at the Scripps Research Institute who collaborate with us on the molecular modeling. Previous and present members of the Stubbe group: Dr. Silvia Hoehn, Dr. Manas Ghorai, Naoko Abe, Shannon McElhinney, Dr. Dapeng Sun, and Grace Kenny who have worked on the bleomycin project.

I would also like to thank people who I have a wonderful intellectual interaction with during my graduate research: Prof. John Essigmann and Prof. Leona Samson who serve on my thesis committee and have provided a lot of insight into this research; Prof. Peter Dedon who shares the same interest in oxidative DNA damage; Prof. Stuart Licht and Prof. Catherine Drennan who are always available for discussion.

I would also like to thank other members of the Stubbe group who have made this laboratory a special place for me for the last five years, especially Aaron Hoskins who is like a brother in this “big family”. I will never forget their friendship and support. I also want to thank Susan Brighton, Lenore Rainey, and other people at the chemistry department for taking care of the administrative part of the research.

Finally, I would like to thank my family—my wife Jing, my parents, my bother Jingyan, and my sister Jingling who have been supporting me with love that I cherish the most.
Mechanistic studies of bleomycin-mediated double-stranded DNA cleavage and structural studies of DNA containing normal and 4’-oxidized abasic sites
by
Jingyang Chen
Submitted to the Department of Chemistry
in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Biological Chemistry

Abstract
In order to examine the role of partial intercalation in double-stranded (ds) DNA cleavage mediated by a single bleomycin (BLM), a bulky group (β-cyclodextrin) was chemically attached to the polyamine tail of BLM A5 to prevent the intercalation of the bithiazole tail. The ability of this analog (CD-BLM) to effect ds-DNA cleavage was quantitatively analyzed using the internally $[{}^{32}P]$-labeled hairpin DNA technology and the supercoiled DNA relaxation assay. CD-BLM can mediate both ss and ds-DNA cleavage, although 5-fold less efficient than BLM A5. Analysis of DNA cleavage by CD-BLM with competitive BLM-Co(III)-OOH indicates that the ds-DNA cleavage is mediated by two CD-BLM molecules, suggesting that the partial intercalation is essential for one BLM molecule to mediate ds-DNA cleavage.

A “hot spot” for blunt-ended ds-cleavage by BLM (5’-GTCA-3’/3’-CAGT-5’) has been identified in an effort to obtain structural insights into the mechanism of the blunt-ended ds-DNA cleavage. A 3’-phosphoglycolate/5’-phosphate (3’PG/5’P) gapped lesion in this “hot spot” has been synthesized to probe the structural basis for re-organization of BLM to the second cleavage site. This lesion was titrated with BLM-Co(III)-OOH. The resulting mixture of complexes was in fast exchange on the NMR time scale, which precluded further structural characterization.

In order to understand the mechanism(s) of recognition and repair of DNA lesions generated by BLM, a duplex DNA containing a 4’-oxidized abasic site (X) in d(CCAAGXACCGGG)-d(CCCGGTACTTTGG) (1) was synthesized and characterized by 2D NMR spectroscopy and molecular modeling. The results indicate that the 4’-oxidized abasic site adopts an intrahelical conformation, in contrast to a normal abasic site in the same sequence context, which is partially extrahelical.

A systematic structural characterization was performed using 2D NMR methods and molecular modeling on an oligonucleotide containing a normal abasic site (Y) with four different bases (A, G, T, or C) opposite the lesion in d(CCAAGGYACCGGG). The results suggest that the conformation in the abasic site region is more perturbed with a pyrimidine opposite the abasic site. This study provides the first structural insight into the dynamics of abasic sites that are intrinsically modulated by the opposite and neighboring bases.

Thesis Supervisor: JoAnne Stubbe
Title: Novartis Professor of Chemistry and Biology
Preface

Figure 1.7 in Chapter 1 is reproduced with permission from Figure 5 and Figure 6 in “Chemical and structural characterization of the interaction of bleomycin A2 with d(CGCGAATTTCGCG)₂. Efficient, double-strand DNA cleavage accessible without structural reorganization. Keck, M.V.; Manderville, R.A.; Hecht, S.M.; J. Am. Chem. Soc.; 2001, 123 (36), 8690-8700. Copyright (2001) American Chemical Society.
# Table of Contents

Acknowledgements 4  
Abstract 5  
Preface 6  
Table of Contents 7  
List of Figures 10  
List of Tables 17  

## Chapter 1. Introduction

1.1 Overview 20  
1.2 Structures of metallo-BLMs 21  
1.3 Chemistry of DNA cleavage by activated BLM 27  
1.4 BLM-mediated dsDNA cleavage 30  
1.5 Models of BLM-mediated dsDNA cleavage 34  
1.6 Synthesis and structures of DNA lesions: Implications for the mechanism of ds-DNA cleavage by a single BLM 42  
1.7 Repair of DNA lesions generated by BLMs 45  
1.8 Structures of DNA containing abasic sites and their interaction with Apel 49  
1.9 References 53  

## Chapter 2. Mechanistic studies of BLM-mediated dsDNA cleavage using CD-BLM

2.1 Introduction 63  
2.2 Experimental Section 67  
2.3 Results and Discussion 76
Chapter 3. Mechanistic studies of BLM-mediated blunt-ended ds-DNA cleavage

3.1 Introduction
3.2 Experimental section
3.3 Results and discussion
3.4 Conclusion
3.5 References

Chapter 4. Synthesis and Characterization of duplex DNA containing a 4’-oxidized abasic

4.1 Introduction
4.2 Experimental section
4.3 Results
4.4 Discussion
4.5 References

Chapter 5. Structures and dynamics of duplex DNAs containing a 4’-oxidized and a normal abasic site

5.1 Introduction
List of Figures

Figure 1.1 Structures of BLMs. 21
Figure 1.2 Ligand geometries of BLM-Co(III)-OOH and BLM-Cu(II). 24
Figure 1.3 DNA cleavage pathways by “activated” BLM. 29
Figure 1.4 Proposed sequence specificity for BLM mediated ds-DNA cleavage. 31
Figure 1.5 Solution structure of BLM-Co(III)-OOH bound to a ds-cleavage “hot spot”, 5’-GTAC-3’. 36
Figure 1.6 Model of BLM-mediated ds-DNA cleavage. 37
Figure 1.7 Structural model of BLM-Zn bound to Dickerson-Drew dodecamer. 40
Figure 1.8 Structure of BLM-Co(III)-OOH bound to a 3’PG/5’P-containing duplex DNA determined by 2D NMR methods. 44
Figure 1.9 Proposed base-excision repair pathway for repairing ss-DNA lesions generated by BLM. 48
Figure 1.10 The structure of human Apel bound to a duplex DNA containing THF site. 50

Figure 2.1 Structures of BLM A2, A5, Phleomycin D1 (PLM), and CD-BLM. 66
Figure 2.2 Pathways of BLM-mediated DNA damage. 67
Figure 2.3 Sequence of GT-2 and the distribution of the cleavage products. 68
Figure 2.4 Quantitation of ss and ds cleavage products analyzed by DNA sequencing PAGE. 74
Figure 2.5 Cleavage of internally [32P] labeled GT-2 by CD-BLM. 77
Figure 2.6 Concentration-dependent cleavage at each ss cleavage site in GT-2. 79
Figure 2.7 Concentration-dependent cleavage at the GT-2 ds-cleavage site (T13+T38). 81
Figure 2.8 Log/log plot of the percentage of T13+T38 ss-cleavage versus T13/T38 ds-cleavage. 82
Figure 2.9 Cleavage of internally [32P] labeled GT-2 by CD-BLM. 83
Figure 2.10 Competitive inhibition of CD-BLM-mediated DNA cleavage by the chemically inert BLM-Co(III)-OOH.

Figure 2.11 Quantitation of CD-BLM mediated cleavage and its attenuation by BLM-Co(III)-OOH.

Figure 2.12 BLM and CD-BLM induced relaxation of supercoiled plasmid pUC18.

Figure 2.13 Quantitation of the cleavage products of pUC18 by BLMs, PLM, and CD-BLM using the supercoiled DNA relaxation assay.

Figure 3.1 DNA cleavage pathways by BLM.

Figure 3.2 Molecular modeling of BLM-Co(III)-OOH bound to duplex DNA containing a 3’PG/5’P gap in a blunt-ended ds-cleavage event.

Figure 3.3 Sequences of hairpin DNA containing internal blunt-end cleavage sites.

Figure 3.4 Cleavage of internal $[^{32}\text{P}]-\text{GTCA}$ hairpin by BLM.

Figure 3.5 Cleavage of internal $[^{32}\text{P}]-\text{GTCC}$ hairpin by BLM.

Figure 3.6 Cleavage of internal $[^{32}\text{P}]-\text{CGCC}$ hairpin by BLM.

Figure 3.7 Cleavage of internal $[^{32}\text{P}]-\text{GTTA}$ hairpin by BLM.

Figure 3.8 Cleavage of internal $[^{32}\text{P}]-\text{GTTT}$ hairpin by BLM.

Figure 3.9 Sequence of duplex DNA(3) containing a hot spot for blunt-ended cleavage and its analog containing a 3’-PG/5’-P gap (4).

Figure 3.10 Analytical anion exchange HPLC analysis of 3 and 1 and 1D NMR spectrum showing the base proton (H8/H6) region of 3.

Figure 3.11 Cleavage of 3 by Fe-BLM. The duplex was 5’-[$^{32}$P] labeled at either 1 (first strand) or 2 (second strand).

Figure 3.12 NOESY (200 ms mixing time) spectrum of 3 showing the connectivity of methyl protons of T to the base proton (H8 or H6).

Figure 3.13 The NOESY (200 ms mixing time) spectrum of 3 showing the NOE connectivity between base (H8/H6) protons and H1’.

Figure 3.14 Titration of 3 with BLM-Co(III)-OOH monitored by 1D NMR.
Figure 3.15 Titration of 3 with BLMA2-Co(III)-OOH monitored by 1D NMR.

Figure 3.16 NOEs between His-H2 (9.13ppm) to Thre-Me (1.21ppm) and to αMe (0.65ppm) of BLMA-Co(III)-OOH in the complex.

Figure 3.17 NOESY (200 ms mixing time) spectrum of 3•BLM-Co(III)-OOH.

Figure 3.18 NOESY spectrum (200 ms mixing time) of 3• BLM-Co(III)-OOH showing the region of base protons (H6/H8) to H2’/H2”.

Figure 3.19 NOESY spectrum (200 ms mixing time) of 3•BLM-Co(III)-OOH showing the region of base protons (H6/H8) to H1’.

Figure 3.20 Oxidation of 5 to 4 monitored by analytical anion exchange HPLC.

Figure 3.21 Ion-pair RP-HPLC analysis of complete enzymatic digestion mixture of 4.

Figure 3.22 Analytical anion exchange HPLC analysis of the NMR sample of 4.

Figure 3.23 NOESY spectrum (200 ms mixing time) of 4 showing the NOE connectivity between base protons (H8/H6) to H1’ protons.

Figure 3.24 Titration of 4 with BLM-co(III)-OOH monitored by 1D NMR spectroscopy.

Figure 4.1 BLM-mediated DNA cleavage pathways.

Figure 4.2 Synthesis of a 4’-oxidized abasic site lesion from 4’-N3-dUTP using HIV-1 RT and UDG.

Figure 4.3 Characterization of oligonucleotides containing an abasic site (1) or a 4’-oxidized abasic site (7) by enzymatic digestion followed by GC/MS.

Figure 4.4 Monitoring incorporation of 4’-N3dU into duplex DNA using primer [5’-32P]-3 and template 4 with HIV-1 reverse transcriptase.

Figure 4.5 Denaturing PAGE allows separation of ss-DNA containing 4’-N3dU from its complement subsequent to treatment of the duplex DNA with NgoM IV.

Figure 4.6 Characterization of 6 via digestion with Nuclease P1 and AP followed by separation of the resulting nucleosides by RP-HPLC.
Figure 4.7 Conversion of 6 into 7 concomitant with formation of uracil release as monitored by reverse phase HPLC.

Figure 4.8 Quantitation of azide release during the conversion of 6 to 7 by the FDH inhibition assay.

Figure 4.9 Ion-pair RP-HPLC (System B) of the products generated when 7 was reduced by NaBH$_4$ followed by digestion with Nuclease P1 and AP.

Figure 4.10 GC/MS of trimethylsilylated 2-deoxy-pentitol generated from NaBH$_4$ and NaBD$_4$ reduced abasic sites or 4'-oxidized abasic sites.

Figure 4.11 Stability of a duplex DNA containing a 4'-oxidized abasic site (8) monitored by anion exchange HPLC.

Figure 4.12 4'-N$_3$dU incorporation and primer extension reaction by HIV-1RT monitored by anion-exchange HPLC.

Figure 4.13 4'-N$_3$dU incorporation and primer extension reaction by HIV-1RT monitored by 5'-[32P] labeled primer, 12.

Figure 4.14 Incorporation of 4'-N$_3$-dU confirmed by nucleoside composition analysis using complete enzymatic digestion of oligonucleotides followed by RP-HPLC analysis.

Figure 4.15 Conversion of 15 to 16 monitored by analytical RP-HPLC.

Figure 4.16 Ion-pairing RP-HPLC analysis of the nucleosides from the duplex DNA 17.

Figure 4.17 Proposed mechanism to generate 4'-oxidized abasic sites from 4'-N$_3$dU via the reduction of the 4'-azido group.

Figure 4.18 Reduction of 4'-N$_3$dU-containing DNA by THMP monitored by RP-HPLC.

Figure 4.19 Ion-pairing RP-HPLC analysis of the nucleosides from the duplex 14.

Figure 4.20 Proposed mechanisms for the reduction reaction of 4'-N$_3$dU-containing DNA by THMP.

Figure 5.1 Pathways of BLM-mediated DNA cleavage.

Figure 5.2 Duplex DNA formation analyzed by analytical anion-exchange HPLC.
Figure 5.3 NOESY (200 ms mixing time) spectra of 1 at 750 MHz showing the connectivities between the base (H6/H8) and the H1’ protons.

Figure 5.4 NOESY (200 ms mixing time) spectra of 4 at 750 MHz showing the connectivities between the base (H6/H8) and the H1’ protons.

Figure 5.5 Assignments of imino and amino protons of 1 and 4 using Watergate-NOESY spectra (200 ms mixing time) at 591 MHz in 90% H2O/10% D2O (v/v) at 4 °C.

Figure 5.6 Assignments of the protons associated with the abasic sites using TOCSY spectra (60 ms mixing time) at 750 MHz.

Figure 5.7 1H-13C GE-HSQC spectrum of 1.

Figure 5.8 Determining the anomeric state using the relative intensity of NOE cross peaks associated with the protons in the abasic sites.

Figure 5.9 Coupling constants between H1’ and H2’/H2’’ measured by E-COSY experiments of the abasic sites in 4.

Figure 5.10 NOE interactions associated with the protons of the abasic site and the adjacent 3’-base proton (A8H8) studied by NOESY (100 ms mixing time) at 750 MHz.

Figure 5.11 NMR structural refinement of 1, 4α, and 4β.

Figure 5.12 The stereo view of the abasic site and the neighboring base pairs in the final structures.

Figure 5.13 The conformation of the 4’-oxidized abasic site in 1 compared to that in a theoretical B-form DNA.

Figure 5.14 The conformation of the normal abasic site in 4α and 4β compared to that in a theoretical B-form DNA.

Figure 5.15 Time-dependent extrahelical motion of the abasic site in 1, 4α, and 4β.

Figure 6.1 Sequences of the duplex DNA constructs containing a single abasic site (X) with four different base (in bold).

Figure 6.2 The NOESY spectrum (200 ms mixing time) of Ab/G showing the NOE connectivities between base (H8/H6) protons and H1’.
Figure 6.3  The NOESY spectrum (200 ms mixing time) of Ab/C showing the NOE connectivity between base (H8/H6) protons and H1’.

Figure 6.4  Chemical shift assignments of base protons of C20, C21, T22, T23, and T24 for Ab/C.

Figure 6.5  The NOESY spectrum (200 ms mixing time) of Ab/T showing the NOE connectivity between base (H8/H6) protons and H1’ protons.

Figure 6.6  The NOESY (200 ms mixing time) spectrum showing the connectivities associated with T methyl protons (solid lines).

Figure 6.7  Watergate-NOESY spectra (200 ms mixing time) of the four duplex DNAs containing an abasic site showing the cross peaks involving T-imino to A-H2 and G-imino to C-amino protons.

Figure 6.8  The Watergate-NOESY spectrum (200 ms mixing time) of Ab/G as an example to illustrate the strategy to assign the chemical shifts of 2 amino protons of C.

Figure 6.9  Chemical shift assignments of protons associated with the abasic site deoxyribose by TOCSY experiments (60 ms mixing time).

Figure 6.10 The NOESY spectrum (200 ms mixing time) of Ab/G showing the NOE connectivities between abase protons (H1’ or H3’ to H2'/H2”).

Figure 6.11 The NOESY spectrum (200 ms mixing time) of Ab/C showing the NOE connectivities between abase protons (H1’ or H3’ to H2'/H2”).

Figure 6.12 The NOESY spectrum (200 ms mixing time) of Ab/T showing the NOE connectivities between abase protons (H1’ or H3’ to H2'/H2”).

Figure 6.13 Coupling constants between H1’ and H2'/H2” measured by E-COSY experiments for the abasic sites.

Figure 6.14 H3’-selective PH-COSY experiment to measure the backbone H3’-C3’-O3’-P coupling constants of Ab/C.

Figure 6.15 H3’-selective PH-COSY experiment to measure the backbone H3’-C3’-O3’-P coupling constants of Ab/T.

Figure 6.16 The overlay of ten final structures of Ab/A, Ab/G, Ab/C, and Ab/T.

Figure 6.17 A stereo view of the conformation of the abasic site region in the averaged structure of Ab/G.
Figure 6.18 A stereo view of the conformation of the abasic site region in the averaged structure of Ab/C. 270

Figure 6.19 The stereo view of the conformation of the abasic site region in the averaged structure of Ab/T. 271

Figure 6.20 Molecular dynamics simulation of Ab/G showing the conformational flexibility at the abasic site. 273

Figure 6.21 Molecular dynamics of Ab/C (α anomer) suggesting transient H-bonding interactions between C20 and A8 formation. 275

Figure 6.22 Quantitative analysis of the transient H-bond formation between A8 and C20 in Ab/C in molecular dynamics simulation. 276

Figure 6.23 Molecular dynamics of Ab/T (α anomer) suggesting transient H-bonding interactions between T20 and A8. 277

Figure 6.24 Quantitative analysis of the transient H-bond formation between A8 and T20 in Ab/T in molecular dynamics simulation. 278
List of Tables

**Table 3.1** Summary of ss to ds cleavage ratios in different sequence contexts that lead to a blunt-ended products.

**Table 3.2** Proton chemical shift assignments of non-exchangeable protons of 3.

**Table 3.3** Proton Chemical shift assignments of 3 in 3•BLM-Co(III)-OOH.

**Table 3.4** Intermolecular NOE interactions between 3 and two protons associated with signals at 6.89 and 6.86 ppm.

**Table 3.5** Proton chemical shifts assignment of non-exchangeable protons of 4.

**Table 4.1** Sequences of the oligonucleotides used for developing synthetic methodology.

**Table 4.2** Sequence information of oligonucleotides used for preparing the NMR samples containing a 4’-oxidized abasic site.

**Table 5.1** Sequence information of oligonucleotides in this study.

**Table 5.2** Summary the statistical analysis of the final 10 structures: root-mean-square deviation (RMSD) (Å), NMR restraints, and NMR violations.

**Table 5.3** $^1$H chemical shift assignments (in ppm) of 1.

**Table 5.4** $^1$H chemical shift assignments (in ppm) of 4.

**Table 5.5** $^{13}$C chemical shift assignments (in ppm) of 1.

**Table 5.6** Coupling constants (in Hz) for 1 (A) and 4 (B) derived from E-COSY experiments.

**Table 5.7** NOE restraints involving the protons associated with the abasic sites to their adjacent base protons.

**Table 6.1** Summary of root-mean-square deviation (RMSD), NMR restraints and NMR violations of the final 10 structures of Ab/G, Ab/C, and Ab/T.

**Table 6.2** Proton chemical shift assignments of Ab/G.
Table 6.3 Proton chemical shift assignments of Ab/C. 254

Table 6.4 Proton chemical shift assignments of Ab/T. 255

Table 6.5 Summary of the NOE interactions between protons around the abasic sites. 266
Chapter 1. Introduction
1.1 Overview

The bleomycins (BLMs, Figure 1.1) are a family of natural glycopeptides isolated from *Streptomyces verticillus* by Umezawa *et al.* in 1966(1). The BLMs have been an essential component in combination chemotherapy against head and neck cancer, and certain types of lymphomas(2). BLM is also part of a regiment involving etoposide and cisplatin (BEP Regiment) that is 90% curative for testicular cancer(3). The clinically administered form of the drug, Bleonoxane (Bristol-Myers Squibb), is composed of ~60% BLM A2, ~30% BLM B2 (Figure 1.1) and additional minor components(2).

The BLM’s cytotoxicity against cancer cells is proposed to be related to their ability to cause both single-stranded (ss) and double-stranded (ds) DNA damage, which requires cofactors including a reduced transition metal, O₂, and a one-electron reductant(4-6). Ss-cleavage occurs predominantly at pyrimidines 3' to a guanine (5'-GC-3' ~ 5'-GT-3' >> 5'-GA-3', _ denotes cleavage site)(7). Ds-DNA cleavage is a rare event compared to ss-cleavage. The average ratios of ss to ds cleavage by BLMs are reported to range from 6:1 to 20:1 measured by several groups using the supercoil DNA relaxation assay(8-12). Ds-cleavage is thought to play a major role in BLMs-induced cell death due to its severe intracellular consequences(13, 14).

BLMs are attractive therapeutics because of their unique low myelo-suppressive(15) and low immuno-suppressive(16) properties compared to other chemotherapy agents. However, the therapeutic efficacy of the BLMs is limited by drug resistance and dose-dependent development of pneumonitis that can lead to fatal pulmonary fibrosis(17).

Our group has recently been interested in understanding the mechanism of BLM-mediated dsDNA cleavage, determining the structures of DNA lesions generated by
BLMs, and understanding the mechanism(s) of BLM uptake. This thesis is focused on the first two problems. The long-range goal of our research is to make a more effective, less toxic BLM.

Figure 1.1 Structures of BLMs. The proposed equatorial ligands for metal binding are in red and the proposed axial ligands are in green. A consensus has not been reached on the identity of the axial ligands.

1.2 Structures of metallo-BLMs

The BLM family members share the same core structure but differ in their positively charged tails (Figure 1.1). The BLM molecule can be divided into four functional domains—the metal-binding domain, the disaccharide moiety, the bithiazole tail, and the linker region connecting the bithiazole tail with the metal-binding domain. Synthetic methods have been used extensively by the Boger(18-24) group and the Hecht group(25-32) to dissect the function of each domain. The role of the sugars is least well
understood. *In vitro* and *in vivo* studies have shown that deglycoBLM, which lacks the disaccharide moiety, has reduced affinity for DNA and is much less effective at mediating ds-DNA cleavage, suggesting that the disaccharide moiety is involved in DNA binding(33, 34). Furthermore, studies on monosaccharide BLM analogs by Boger and Hecht *et al.* suggest that the nature of the first carbohydrate (L-gulose) is important for efficient ss and ds-DNA cleavage(24, 31). The sugar moiety has also been speculated to be involved in the drug uptake by interacting with cell surface receptors(35). However, recent genetic studies in yeast suggest that the positively charged tail might be responsible for the uptake of BLM into the cell (See the discussion below) (36, 37).

The metal-binding domain is the warhead and contains the ligands involved in octahedral complex formation with the transition metals (Figure 1.1, the proposed equatorial ligands are in red and the proposed axial ligands are in green). The organization of the ligands around the metal center has important implications for the recognition of the DNA binding site. As discussed in Section 1.3, the cleavage chemistry requires activation of O₂ in the presence of a transition metal. The ligand geometry around the metal center has been studied with iron, copper, cobalt, or zinc-BLM complexes (reviewed by Claussen and Long (38)). A consensus has been reached regarding the equatorial ligands, which include the secondary amine of the β-aminoalanine, the pyrimidine nitrogen, the deprotonated histidine amide, and the imidazole of the histidine (colored in red, Figure 1.1). However, the identity of the axial ligand(s) (colored in green, Figure 1.1) and consequently, the screw sense of the complex remain controversial. Based on the DNA cleavage chemistry as discussed in Section 1.3, one axial ligand is provided by exogenous O₂. The other axial ligand has been proposed
to be either the primary amine of the β-aminoalanine or the carbamoyl group on the mannose. A model structure of BLM-Co(III)-OOH (an “activated BLM” analog, Section 1.3) determined by 2D NMR methods by our group suggests that the primary amine of β-aminoalanine is the axial ligand and is on the same face as the disaccharide moiety with OOH as the second axial ligand (Figure 1.2 A)(39). Recent X-ray structural studies on BLM-Cu(II) in a complex with BLM-binding protein from the BLM-producing strain (S. verticillus) also suggests the same ligand geometry and screw sense (Figure 1.2 B) with the exogenous axial ligand as Cl⁻ (colored in cyan, Figure 1.2 B)(40). However, studies on BLM-Zn(II) suggest a different ligand geometry. In the model structure of BLM-Zn(II), both the primary amine of the β-aminoalanine and the carbamoyl group on the mannose are axial ligands to the metal center(41-43). Other studies (Petering) suggest that the ligands in BLM-Zn(II) are in rapid exchange on the NMR time scale. In no case has the screw sense isomer of the BLM-Zn(II) complex been established. The binding of BLM to DNA and the chemistry of BLM-mediated cleavage (which is initiated by 4’-H atom abstraction as discussed in Section 1.3) suggest that the screw sense observed for BLM-Co(III)-OOH site is sufficient for the close proximity between the exogenous ligand and the 4’-H atom of DNA(44). Since we don’t have a model for BLM-Zn(II) binding, we don’t know if it is consistent with 4’-H atom abstraction.

In addition to providing ligands to coordinate metal ions, the metal binding domain has also been proposed to play a key role in the sequence specific binding of DNA by BLM. The 2D NMR structure of BLM A2-Co(III)-OOH bound to several duplexes of DNA containing a single binding site (5’-GGCC-3’, 5’-GTAC-3’, _ denotes the cleavage site) reveals that the N3 and 4-amino groups of the pyrimidine within the metal binding
domain form two H-bonds with the N3 and 2-amino group of the guanine 5' to the pyrimidine cleavage site(45) (see the discussion in Section 1.5).

Figure 1.2 Ligand geometries of BLM-Co(III)-OOH and BLM-Cu(II) A: The solution structure of BLM-Co(III)-OOH determined by 2D NMR methods (39). B: The 3D structure of BLM-Cu(II) bound to the active site of a “resistant protein” from the host organism (S. verticillus) determined by X-ray crystallography (PDB No. 1JIF)(40). In A and B, the atoms are colored as follows: N, blue; O, red; C, green; Co(III) or Cu(II) ion, magenta sphere; chloride ion, cyan sphere. The bithiazole tail is omitted for clarity. C: The superimposition of the metal binding domains of Cu(II)-BLM (in green) and the BLM-Co(III)-OOH (in gray).
The bithiazole tail is also in part responsible for DNA binding. The binding can occur either by partial intercalation\cite{46} or by interactions with the minor groove\cite{47,48}. In support of the intercalative mode of binding, Povirk et al. have studied relaxation of supercoiled DNA and reversal of supercoiling in the presence of BLM at pH 5.5 using sedimentation methods\cite{46}. No metal ions were used in those studies due to the lack of knowledge of the metal ion requirement at that time and the low pH was utilized to avoid nicking of supercoiled DNA. The DNA unwinding angle caused by BLM binding was determined to be 12° per BLM\cite{46}. In addition, Povirk et al. also measured the lengthening of linear DNA molecules by BLM binding by determining the characteristic time (μs) for rod-like DNA to orient in an electric field monitored by linear dichroism\cite{46}. The observed lengthening effect caused by BLM was measured to be 3.1 Å/BLM, which is within the range of other intercalators (1.8-4.5 Å). However, other studies have supported binding of the bithiazole tail to DNA by interaction with the minor groove. Chien et al. measured the lengthening of DNA by BLM using viscometric methods. In contrast to the expected increase in viscosity for an intercalator, no change in viscosity was observed for BLM\cite{48}. In addition, Huang et al. showed that binding to DNA caused two types of quenching of the fluorescence signal of the bithiazole moiety\cite{49}. One type was sensitive to the ionic strength of the solution and the other type was not. Huang et al. proposed that the ionic-sensitive type was associated with binding of the bithiazole in the minor groove by electrostatic interaction with DNA backbone and that the ionic-insensitive type was associated with binding by intercalation. Thus, previous studies suggest that there are multiple binding modes for the bithiazole tail of BLM to DNA.
The key question is which binding mode is responsible for efficient ss and ds DNA cleavage \textit{in vivo}. As will discussed in Section 1.5, based on biochemical and structural studies, our group proposes that the partial intercalation of the bithiazole tail is essential for BLM-mediated ds-cleavage\cite{50,51}.

The positively charged C-terminal tail of BLMs has also been proposed to contribute to DNA binding by electrostatic interactions with negatively charged DNA backbones. Members of the BLM family that differ in this tail demonstrate similar DNA cleavage efficiencies and sequence specificities. Recent studies from the Ramotar’s group also suggest that the positively charged group might be involved in drug uptake\cite{36,37}. Using genome-wide screening in yeast, Ramotar \textit{et al.} identified BLM-resistant strains using a fluorescently labeled BLM that appear to have altered drug-uptake \cite{37}. One mutant gene, Agp2, encodes a known L-carnitine transporter and two other mutant genes, Ptk2 and Sky2, are involved in polyamine transport. These results suggest that the positively charged group on the tail might be responsible for BLM’s uptake into the cell.

The linker region, which contains a methylvalerate-threonine dipeptide fragment, connects the metal binding domain and the bithiazole tail (Figure 1.1). Studies from the Boger group in collaboration with our group have demonstrated the essential role of the linker region in ds-DNA cleavage events\cite{21,22}. The Boger lab synthesized a number of BLM analogs with variations in the linker region. The efficiency of the ds-DNA cleavage by these analogs was examined quantitatively using the supercoiled DNA relaxation assay and the internal $^{[32P]}$-hairpin DNA assay. The results show that tinkering with any position within the linker significantly lowers ds cleavage efficiency. The linker in the
BLM appears to have been optimized to pre-organize the BLM for efficient ss and, especially, ds cleavage.

1.3 Chemistry of DNA cleavage by activated BLM.

The BLMs require a redox-active transition metal, $O_2$, and a reductant to mediate oxidative DNA damage\(^{(4)}\). In *vitro*, both Fe(II)-BLM and Cu(I)-BLM can mediate DNA cleavage. The “active species” of Cu(I)-BLM that initiates DNA cleavage has never been identified and its efficiency to mediate DNA cleavage is still controversial\(^{(52, 53)}\). Most of the studies have been focused on Fe(II)-BLM, which has long been believed to be the active form inside the cell. In the presence of Fe(II), $O_2$ and a one electron reductant, an “activated” BLM (BLM-Fe(III)-OOH) is generated (Figure 1.3)\(^{(54, 55)}\). Formation of activated BLM is fast ($k = 0.1 \text{ s}^{-1}$) and the rate of its decomposition is slow ($t_{1/2} \sim 2\text{min}$ at 4°C)\(^{(55)}\). The identity of this activated species as BLM-Fe(III)-OOH was established by rapid freeze-quench (RFQ) Mössbauer spectroscopy, RFQ-EPR spectroscopy, and ESI mass spectrometry\(^{(55)}\). This species is the last intermediate detected spectroscopically prior to the DNA cleavage event.

The “activated” BLM mediates both ss and ds cleavage, directly or indirectly, by first abstracting the 4’-hydrogen atom from C4’ of the deoxyribose moiety of a pyrimidine 3’ to a guanine (Figure 1.3A)\(^{(56)}\). The identity of the BLM species that actually generates the putative 4’-nucleotide radical remains unclear. It has been proposed based on model inorganic complexes, that BLM-Fe(III)-OOH can undergo either heterolytic O-O bond cleavage to generate BLM-Fe(V)=O (and OH⁻) or homolytic O-O bond cleavage to generate BLM-Fe(IV)=O and HO⁻\(^{(57, 58)}\) (Figure 1.3 A). The latter scenario seems
unlikely for BLM-Fe(III)-OOH because the hydroxyl radical generated concomitant with BLM-Fe(IV)=O would result in non-specific cleavage patterns on analysis by SDS-PAGE of DNA cleavage products that are not experimentally observed. Recently, the Solomon group proposed a third mechanism in which the peroxide is the activated BLM resulting in the cleavage of the O-O bond of the BLM-Fe(III)-OOH concerted with 4’-hydrogen atom abstraction (59)(Figure 1.3 A). Further experimental evidence is required to distinguish between those mechanistic options.

Depending on the availability of O₂, the 4’-nucleotide radical intermediate can partition between two pathways (Pathway A and B, Figure 1.3B). In the presence of oxygen (Pathway A), the 4’-radical intermediate reacts with O₂ to form a 4’-peroxy radical(60, 61). The resulting 4’-peroxy radical intermediate may then be reduced to the 4’-hydroperoxide, which undergoes a series of chemical transformations, ultimately generating a gapped DNA with a 3’-phosphoglycolate/5’ phosphate (3’-PG/5’-P) ends and a pyrimidine propenal(60). Under oxygen-limited conditions (Pathway B), the 4’-radical intermediate is proposed to be oxidized to a 4’-carbocation to which H₂O adds, generating a 4’-oxidized abasic site and a free nucleic acid base(62). In Pathway B, no direct strand cleavage occurs.

In addition to initiating chemical and sequence specific cleavage of DNA, the “activated” BLM (BLM-Fe(III)-OOH) can destroy itself(55), oxidize lipids(63), hydrolyze amide bonds of proteins(64), cleave RNA(65), as well as decompose to hydroxyl radicals that react rapidly and non-specifically with any molecule they encounter. Results from extensive in vitro and in vivo studies have demonstrated that
DNA cleavage is responsible for the observed cytotoxicity induced by BLM (recently reviewed by Chen and Stubbe(14)).

Figure 1.3 DNA cleavage pathways by “activated” BLM. A: Activation of BLM and three proposed mechanisms for 4'-H atom abstraction. B: DNA cleavage pathways by “activated” BLM. Two types of DNA lesions result from the common 4'-radical intermediate: Pathway A, a direct strand break generates a 3'-phosphoglycolate/5'-phosphate (3'PG/5'P) gap and a base propenal; Pathway B, an oxidized abasic site with an intact backbone. Ds-cleavage is proposed to occur through Pathway A.
1.4 BLM-mediated dsDNA cleavage

Povirk and other researchers have demonstrated that BLM causes both ss- and ds-DNA cleavages using the supercoiled plasmid relaxation assay(8-12). By quantitating the amounts of ss and ds cleavage products (nicked and linear plasmids, respectively), Povirk and other group measured the average ss to ds ratios for BLM to be 6.0-20 to 1, depending on the assay conditions(8-12). Thus, ds cleavage by BLM is a rare event compared to ss cleavage. In addition, in his seminal studies, Povirk also noted that the apparent ss to ds ratio (6.0 to 1) did not change over a wide range (20 folds) of BLM concentrations under the single-hit condition, which led Povirk to propose that the ds-cleavage is not due to two random ss-cleavage events(8).

Povirk et al. have also studied the chemistry and the sequence selectivity of dsDNA cleavage using native and denaturing DNA gel electrophoresis(66, 67). Their studies revealed that termini of both strands in the ds cleavage product possess 3'PG/5'P ends, indicating that cleavage on both strands involves 4'-H atom abstraction(66). In addition, Povirk et al. also observed that some ds cleavage sites contain a 3'-PG/5'-P gap with oxidized abasic sites on the opposite strand. No lesions with abasic sites on both strands were observed. These results led Povirk to a model in which the 3'-PG/5'-P lesion or its precursor is required for all ds-DNA cleavage events.

Based on examination of 30 ds-cleavage sites in a limited sequence space, Povirk et al. proposed a set of rules for prediction of ds-cleavage sites (Figure 1.4) (66). The rules for ds-cleavage specificity that result from these studies are that the initial cleavage site is a good ss cleavage site (5'-G-Py-3' site). If the base 3' to the Py is a purine, the second cleavage site on the opposite strand is 5'-staggered to the cleavage site on the first strand;
if the base 3’ to the Py is a pyrimidine, then the cleavage site on the second strand is
directly opposite the first cleavage site, resulting a blunt-ended cleavage (Figure 1.4).

\[
\begin{align*}
&\text{5’-G-Py-Pu-Pu-3’} & \text{5’-G-Py-Py-3’} \\
&\text{3’-C-Pu-Py-Py-5’} & \text{3’-C-Pu-Py-Pu-5’}
\end{align*}
\]

\text{Staggered-ended cleavage}

\[
\begin{align*}
&\text{5’-G-Py-Py-Pu-3’} & \text{5’-G-Py-Py-Py-3’} \\
&\text{3’-C-Pu-Pu-Py-5’} & \text{3’-C-Pu-Pu-Pu-5’}
\end{align*}
\]

\text{Blunt-ended cleavage}

**Figure 1.4** Proposed sequence specificity for BLM mediated ds-DNA cleavage. The primary cleavage site (1°, in bold) is a good ss cleavage site and is indicated by large arrows. The secondary (2°) cleavage site is governed by the residue (purines versus pyrimidines) 3’ to the pyrimidine in the primary cleavage site and is indicated by small arrows. (Py, pyrimidine; Pu, purine)

The chemistry and sequence-selectivity of BLM-mediated ds-cleavage led Povirk et al. to propose that a single BLM molecule is responsible for the cleavage at both strands of DNA(66). In this model, BLM is reactivated after the first cleavage event and “without
dissociating from DNA, undergoes a rearrangement in binding and effects a secondary
attack in the complementary strand, at a site whose position is determined primarily by
the position of the first break”(66).

An alternative model of ds-cleavage by BLM was proposed by Keller and
Oppenheimer, in which the DNA lesion formed at the first cleavage site potentiates
binding of a second molecule of BLM to the second cleavage site(68). Keller and
Oppenheimer showed that a one-base gap with 3’-P/5’P termini (which mimics the 3’-PG/5’P lesion) increased the extent of the cleavage directly opposite the gap(68).

Although the second cleavage site in this study is not consistent with the sequence-selectivity rules observed for BLM, these results raise the possibility that ds-cleavage could be effected by cooperative binding of a second BLM. This cooperative binding model also explains the concentration-independent ss to ds cleavage ratio.

Our group has taken a different approach to study BLM-mediated ds cleavage. Absalon developed internal[^32P]-hairpin DNA technology that allows quantitation of ss and ds cleavage site-specifically(69, 70). Our studies have been focused on a “hot spot” for ds-cleavage, 5’-GTAC-3’/3’-CATG-5’, which was identified by examining 17 ds cleavage sites in a 120-bp DNA fragment using native DNA gel electrophoresis(69). The ss to ds cleavage ratio for the 5’-GTAC-3’ site using this technology was measured to be 3.4 to 1(69). In addition, the ss to ds ratio for the 5’-GTAC-3’ sequence appeared to remain constant with 50 fold changes in BLM concentration(69). Finally, Absalon et al. also have demonstrated that ds DNA cleavage by BLM is never observed under anaerobic conditions(69). This observation is consistent with Porvirk’s previous studies showing that a 3’-PG end is present at the primary cleavage site in all ds cleavage products.

The internal[^32P]-hairpin DNA technology also enabled the study of sequence-specific isotope effects at the 5’-GTAC-3’ site(70). Isotope effects on ds-cleavage were observed when the C-4’ hydrogen at either cleavage site (5’-GTAC-3’/3’-CATG-5’) was substituted with deuterium, consistent with Porvirk’s studies indicating that C-4’ chemistry is involved in both cleavage events(70). The size of the isotope effects (6 to 8)
for ss and ds cleavage events were indistinguishable, suggesting that pathways leading to both cleavage events partition from a common intermediate subsequent to the 4’-H atom abstraction at the first cleavage site(70).

In order to distinguish between the single-BLM model proposed by Povirk et al. and the cooperative binding model proposed by Keller and Oppenheimer, Absalon et al. showed that the presence of a chemically unreactive BLM-Co(III)-OOH suppressed both ss- and ds-DNA cleavage to approximately the same extent, indicating that BLM-Co(III)-OOH did not compete for binding at the second cleavage site. These results are inconsistent with the prediction based on the cooperative binding model(70). Therefore, these results and the observation that the ss:ds cleavage ratio remains unchanged over a broad concentration range support the model that a single BLM molecule is responsible for ds-DNA cleavage.

In order for one BLM molecule to mediate two cleavage events, the BLM molecule must be reactivated after the initial 4’-H atom abstraction. A mechanism for the reactivation of BLM has been proposed(70). The reactivated BLM then has to re-position itself to the second cleavage site that is 15-18 Å away from the first cleavage site based on the observed sequence selectivity rule assuming B-form DNA structure. The relative rate of BLM dissociation versus reorganization after the first cleavage event is proposed to dictate the observed ss:ds cleavage ratio. A model to explain the large re-organization of the BLM molecule has been proposed by our group(50) and is discussed in the next section.
1.5 Models of BLM-mediated dsDNA cleavage

A structural model has been proposed by our group to explain how a single molecule of BLM can cause ds-cleavage without dissociation from the DNA(50). This model is based on the solution structure of the 1:1 complex of BLM-Co(III)-OOH (a diamagnetic analog for the activated BLM, BLM-Fe(III)-OOH) and duplex DNA containing a single BLM cleavage site (5’-GTAC-3’ or 5’-GGCC-3’) (Figure 1.5)(51). The structure of the BLM•DNA complex reveals that the metal-binding domain is located in the minor groove of DNA (Figure 1.5A). The pyrimidine moiety of BLM forms two H-bonds with guanine 5’ to the cleavage site (indicated by two yellow lines in Figure 1.5A and red dashed lines in Figure 1.5B). The dipeptide linker folds under the HOO ligand and the disaccharide moiety is on the opposite face of the linker. The bithiazole tail is inserted 3’ to the cleavage site in a partial intercalative mode of binding (Figure 1.5C). This binding mode positions the OOH ligand about 4 Å from the 4’-H of the pyrimidine to be cleaved (Figure 1.5A, grey line). The solution structure of BLM-Co(III)-OOH•DNA and a number of additional complexes(71-73) solved independently of our work offer a structural basis for the site specific cleavage and the chemistry of 4’-H abstraction.

We have proposed a model for BLM-mediated ds-DNA cleavage based on the structures of BLM•5’-GTAC-3’ and BLM•5’-GGCC-3’ (Figure 1.6A)(50). This model requires the partial intercalative mode of binding of the bithiazole tail of BLM, rotation of 180° around the C-C bond connecting the two thiazolium rings once the damage is effected on the first stand, and further rotation of 117° around the axis perpendicular to the plane of the bithiazole rings (Figure 1.6B). The BLM at the first cleavage site has two options. It can dissociate from the duplex DNA or it can re-organize to the second strand.
The relative rates of these two processions are the key to identifying and understanding hot spots for ds-cleavage in this model.

(Figure 1.5 To be continued)
Figure 1.5 Solution structure of BLM-Co(III)-OOH bound to a “hot spot” (5’-GTAC-3’) for ds-cleavage. A: The overall structure of the complex. The metal-binding domain is located in the minor groove with the bithiazole tail adopting a partial intercalative conformation. The yellow lines indicate two H-bonds between the 4-NH$_2$ and N3 groups of BLM and N3 and 2-NH$_2$ of guanine 5’ to the cleavage site. The sugar moieties are disordered with no defined interactions with the DNA. The gray line indicates the proximity of the OOH ligand to the 4’-H atom to be abstracted. B: H-bond interactions between the pyrimidine residue of BLM and G indicated by the red dashed lines. C: Partial intercalative conformation of the bithiazole moiety (colored in CPK) between neighboring base pairs (in gray).
Figure 1.6 Model of BLM-mediated ds-DNA cleavage. A: The re-organization of BLM from the first cleavage site (yellow) to the second cleavage site (red). The white arrows indicate the 4'-H to be abstracted on each strand, the yellow arrows indicate the HOO moiety and the red arrow indicates the bithiazole tail partially intercalated. B: A cartoon rendition of the re-organization, which is effected by rotation of 180° around the C-C bond connecting the thiazolium rings and rotation of 117° perpendicular to the plane of these rings. The metal-binding domain is depicted as a gray oval. Note the protons (in red) with the bithiazole tail change from a trans to a cis conformation during this process.
The structural element that triggers BLM to reorganize from the first cleavage site to second is not understood. We have postulated that it involves the 3'-PG/5'-P lesion or a precursor to this lesion(74). The lesion on the first strand could increase the flexibility of the DNA at the lesion site and thus facilitates the reorganization process. However, studies by Burger et al. suggest that formation of the 3'-PG/5'-P lesion (t\(_{1/2}\): 6.7 ± 0.3 min at 4 °C) is too slow to account for the observed rate of ds-DNA cleavage(75, 76), which led Povirk to propose that a precursor to the 3'PG/5'P is involved in the reorganization(77). However, the methods used by Burger et al. to measure rate of base propenal release/3'-PG/5'-P formation (\(^{31}\)P and \(^1\)H NMR) are insensitive and not site specific. Therefore, the question is raised as to whether there are certain sites that effect this transformation much more rapidly. In addition, the intermediates that precede the 3'-PG/3'-P lesion are chemically labile and could never be recapitulated in a stable DNA duplex that could be examined by NMR. Therefore, our group has focused on the stable 3'-PG/5'-P lesion and its effect to trigger the re-organization of BLM to the other strand. Using 2D NMR spectroscopy, our group has shown that the 3'-PG/5'-P gapped lesion in the 5'-G_AC-3'/3'-CATG-5' sequence is sufficient to direct BLM to the second strand(74) (see the discussion in Section 1.6).

Recently, the Hecht group reported a ds cleavage site (5'-CGCGA-3'/3'-GCGCT-5') in the Dickerson-Drew dodecamer. This site is distinct from any previously reported ds-cleavage sites and does not conform to Povirk’s rules(78). The observed secondary cleavage site is staggered in the 3’ direction (Figure 1.7A) and is two base pairs removed from the first cleavage site. NMR structural studies of a BLM-Zn(II) bound to the Dickerson-Drew dodecamer were interpreted to suggest that the metal binding domain of
BLM-Zn(II) is located halfway between the two cleavage sites with the bithiazole tail in the minor groove (Figure 1.7 B and C). The coordination of the Zn(II) atom is octahedral with all 6 ligands provided by BLM and with no binding site for the exogenous O₂. The organization of the ligands around the Zn(II) atom and thus, the structure of the BLM-Zn(II) complex has never been defined. Additional problems associated with the structural determination result from the BLM-Zn(II)/DNA complex being in a fast exchange on the NMR time scale and inability to assign the chemical shifts of the bithiazole protons (H5 and H5’, Figure 1.1). The scarcity of the NOE interactions (12 NOEs) between BLM-Zn(II) and the Dickerson dodecamer (Figure 1.7 A) has prevented accurate modeling of a single conformation of the bithiazole tail (compared to 60 NOEs between BLM-Co(III)-OOH and d(CCAGTACTGG)₂ (51)). Our attempt to reproduce the ds-cleavage pattern on the Dickerson-Drew dodecamer using the hairpin DNA technology has been unsuccessful. Therefore, the relevance of this unusual ds-cleavage site and the model structure proposed by Hecht et al. is called into question.

Several studies using BLM analogs have demonstrated that alternative binding modes to partial intercalation of the bithiazole tail can also lead to ss-DNA cleavage. Studies on a BLM analog with a chlorinated bithiazole (the H5 and H5’ of the bithiazole were substituted with chlorine atoms) showed that this analog could mediate a sequence-selective DNA cleavage, similar to nature BLMs. Furthermore, the chlorinated bithiazole tail was bound to DNA in the minor groove, demonstrated by the light-mediated cleavage pattern of DNA by the chlorinated bithiazole group (79).
Figure 1.7 Structural model of BLM-Zn bound to Dickerson-Drew dodecamer*. A: Sequence of the Dickerson-Drew dodecamer with the observed intermolecular NOE interactions with Zn-BLM by arrows. Two proposed cleavage sites (A5 ad C11) are in boxes. B: Model structure of BLM-Zn(II)-DNA viewed along the minor groove. The two 4'-H atoms to be abstracted are colored in yellow and purple. C: Side view of the structure. The two 4'-H atoms are indicated by two orange arrows. In B and C, the Zn-BLM molecule is in red and the Zn atom is shown as a white sphere. The DNA is colored in green. (* Figure 5 and 6 from Keck et al., J.Am.Chem.Soc. 2001(123) 8690-8700. Because a structural file has not been deposited in PDB, we are not able to present a clearer picture depicting the detailed interaction between BLM-Zn and DNA.)
The Hecht group has studied the DNA cleavage by BLM A5 tethered to either controlled pore glass beads or a dendrimer at the polyamine tail (80-82). These modifications have been proposed to be sufficiently large to prevent binding of the bithiazole tail by partial intercalation. DNA cleavage was examined with these analogs and both the efficiency and sequence selectivity was shown to be similar to untethered BLMs. These results suggest that a non-intercalative mode of binding can also lead to DNA cleavage. Although dsDNA cleavage was observed for the both dendrimer and CPG-tethered BLM using supercoiled plasmid relaxation assays, the ss to ds cleavage ratios were not reported. In addition, with the CPG-tethered BLM, the structure of BLM and the effective concentration in solid support is hard to quantitate. Therefore, it is not clear whether the observed ds cleavage resulted from a single tethered BLM or two tethered BLMs.

In order to re-examine the role of partial intercalation of the bithiazole tail in the ds DNA cleavage process and building on the approach used by Hecht and coworkers, we synthesized a BLM A5 analog with a β-cyclodextrin group chemically attached to the polyamine tail of BLM A5. This analog (CD-BLM) has been well characterized spectroscopically. The ability of this CD-BLM-Fe to effect ss and ds cleavage was examined quantitatively using the hairpin DNA technology and the supercoiled DNA relaxation assay. The mechanistic studies on the CD-BLM are the focus of Chapter 2.
1.6 Synthesis and structures of DNA lesions: Implications for the mechanism of ds-DNA cleavage by a single BLM

Our efforts to examine the structural basis for ds-cleavage require the synthesis and structural characterization of the 3’-PG/5’-P gapped lesion and the 4’-oxidized abasic lesion generated by BLM. This structural information and its interaction with BLM-Co(III)-OOH is requisite to understand the mechanism of dsDNA cleavage by a single BLM molecule.

As discussed in Section 1.5, based on our studies and Povirk’s studies, we focused on the 3’PG/5’P lesion as the trigger to initiate structural re-organization of BLM to the second cleavage site. Our group synthesized a 3’PG/5’P lesion in a duplex DNA suitable for 2D NMR studies(83). The synthetic procedure involved two oxidation steps in a one-pot reaction to convert a precursor, 3’-glycerol, to a 3’-PG end(84). Two hexaethyleneglycol linkers were incorporated at each end of duplex by solid-phase synthesis in order to increase the thermostability of the gaped DNA(83). The incorporation of the 3’PG/5’P lesion was confirmed by complete enzymatic digestion of the duplex DNA followed by quantitation of individual nucleotides(83).

The ability to generate the 3’PG/5’P lesion allowed us to test the hypothesis that the 3’PG/5’P lesion could trigger the re-organization of BLM to the second cleavage strand. A 3’PG/5’P gap was introduced into a 13mer duplex DNA containing a hot spot for ds cleavage, 5’-G_AC-3’ (_ denotes the 3’PG/5’P gap, Figure 1.8A). Interaction between the 5’-G_AC-3’ duplex and BLM-Co(III)-OOH was studied by titrating the 5’-G_AC-3’ duplex with BLM-Co(III)-OOH monitored by 1D NMR(74). The titration led to a 1:1
complex formation, indicating a high affinity between the 3'PG/5'P lesion in the 5'-G_AC-3' sequence and BLM. The structure of the 1:1 complex was solved by 2D NMR(74). The structural model revealed that BLM binds to minor groove of the gapped lesion with its bithiazole tail partially intercalated between base pairs 3' to the cleavage site and its metal binding domain (reactive center) H-bonded to G18 in the second cleavage site and its OOH group poised close to the H4' atom to be abstracted in T19 (Figure 1.8B). ROESY 2D NMR experiments also indicated that the trans and cis conformations of the bithiazole rings were rapidly interconverting, but that the trans conformation dominated in solution(74). These results provide strong support for the reorganization model in which BLM reorganization could be triggered by the lesion on the first strand, to move to the second cleavage site.

Our reorganization model further predicts that 4'-oxidized abasic lesion would not lead to sequence specific binding of BLM to the second cleavage site. Using a normal abasic site as a structural analog for the 4'-oxidized abasic site, Silvia et al. synthesized the lesion in the same sequence context and titrated with BLM-Co(III)-OOH, monitoring the complex formation by 1D NMR spectroscopy(85). The results indicated the presence of multiple conformations in rapid exchange on the NMR time scale, consistent with our model.
Figure 1.8 Structure of BLM-Co(III)-OOH bound to a 3'PG/5'P-containing duplex DNA determined by 2D NMR methods. A: Sequence of the 13mer duplex DNA containing a 3'PG/5'P gap and tethered ends. Our model predicts that BLM would bind to the second cleavage site (T19) with two H-bonding interactions with G18 and the bithiazole moiety would intercalate between T19 and A20. B: the structure of the complex. The H-bond interactions between BLM and G18 are indicated by the red dashed lines. The black dashed line indicates the close proximity of the OOH group to the 4'-H atom of T19 (white sphere) to be abstracted.
The success with the 3’PG/5’P DNA lesion to study the structural basis for the reorganization in the staggered-end ds cleavage site (5’-GTAC-3’/3’-CATG-5’) has prompted us to apply this methodology in an effort to identify a similar structural transformation in a blunt end cleavage site. Chapter 3 describes our studies to identify a hot spot (5’-GTCA-3’) for blunt-ended ds cleavage from 5 sequence candidates generated using Povirk’s specificity rules. We determined the ss to ds cleavage ratios using the internally labeled hairpin DNA method. Based on these results, duplex DNA containing an intact 5’-GTCA-3’ sequence or 5’-G_CA-3’ ( denotes a 3’PG/5’P lesion) was synthesized and its interaction with BLM-Co(III)-OOH was studied by 1D and 2D NMR spectroscopy (Chapter 3).

In contrast to our efforts to obtain the structure of the 3’G/5’P lesion, efforts to gain structural insight into the 4’-oxidized abasic site lesion (Figure 1.3) have been hindered by the synthetic challenges to make this chemical labile species. Recently, we have developed a convenient method to generate the 4’-oxidized abasic site(86). The detailed synthesis of this lesion in any sequence context and its characterization using 2D NMR methods is reported in Chapter 4. In Chapter 5, we report the first structural study on a duplex DNA containing a 4’-oxidized abasic site at the ds cleavage hot spot (5’-GXAC-3’) and compare this structure to a normal abasic site in the same sequence context.

1.7 Repair of DNA lesions generated by BLMs

Our ability to synthesize DNA lesions generated by BLM and characterize these lesions structurally offers the opportunity to study the mechanism of repair of these lesions in vitro and potentially, in vivo. At present, it is not clear why certain tumors are
more susceptible to BLM treatment than other tumors or normal tissues. Besides the potential differences in drug uptake and metabolism, another possible factor governing effectiveness could be the repair of BLM-induced DNA damage. Many tumor cells have mutated signal transducers and effectors that can alter the efficiency of DNA repair or the signals that lead to cell death or apoptosis. Thus, it is likely that understanding the DNA repair pathways in normal cells and identifying which of BLM-induced DNA lesions are most difficult to repair, could provide insight into how BLM’s cytotoxicity is enhanced in specific tumor cells. As noted in Figure 1.3, the BLMs cause both ss- and ds-DNA damage. Recent studies are beginning to provide insight into the proteins involved in ss break repair and their similarities and differences from the proteins involved in the ds-break repair pathways.

**ss-DNA Damage Repair** The ss-DNA lesions generated by BLM, as noted above, are either an oxidized abasic site or a 3’-PG/5’-P gapped DNA (Figure 1.3). The ss-lesions are thought to be repaired by the base-excision repair (BER) pathway using Ape1 and DNA polymerase β (polβ) (Figure 1.9)(87). Demple and coworkers by studying a single site have suggested that repair of the 4’-oxidized abasic site by Ape1 occurs with a similar rate to normal abasic sites by generating an oxidized abasic site using BLM-Fe(88). Repair of a 3’-PG site by Ape1 (which possesses a weak 3’-exonuclease activity), however, occurs at 1/400 the rate of the 4’-oxidized abasic site in the same sequence context(89, 90). The slow rate of repair suggested that other intracellular nucleases might be responsible for the removal of the 3’-PG lesion. Studies by Povirk et al. have shown that human tyrosine-DNA phosphatase (hTDP1) can remove the 3’-PG lesion in vitro(91). More recent studies suggest that in HeLa cell extracts, Ape1 is the major
nuclease involved in repairing the 3'-PG lesion(92). Thus, the role of Ape1 in vivo in the removal of the 3'-PG end and also the oxidized abasic site requires further analysis. The repair of both of these ss-lesions is template-driven and should lead to accurate restoration of genetic information. However, base substitution mutations can occur due to the lack of the proofreading function of polβ(93).

The proposed role of Ape1 in repairing BLM-induced ssDNA lesions has been demonstrated in several cell culture studies. In one study, elevated levels of Ape1 in germ cell lines have been shown to correlate with BLM resistance(94). In addition, HeLa cells and primary fibroblasts treated with BLM showed an adaptive response in which Ape1 levels were up-regulated and Ape1 was translocated into the nucleus(95). Transforming Apn1 (the S. cerevisiae Ape1 homolog) into lung epithelial cells made them more BLM resistant(96).

ds-DNA Damage Repair  As discussed in Section 1.4, BLMs also generate ds-lesions containing 5'-staggered and blunt ends (Figure 1.4). In mammalian cells, two pathways exist to repair ds DNA breaks, non-homologous end joining (NHEJ) and homologous recombination repair (HRR)(97). To repair BLM-generated ds-breaks, exonucleases like Ape1 or TDP1 are required to remove 3'-PG ends. In the staggered ends, a polymerase can be used for formation of a phosphodiester bond using the information on the opposite strand. Once the appropriate nucleotide is incorporated, ligation can occur. With blunt-ended lesions, however, the information on the opposite strand has been lost. Thus, the fidelity of the repair becomes challenging. The NHEJ pathway for repair is thus of special interests with respect to BLM-induced ds-breaks because of its error-prone nature(98).
Figure 1.9 Proposed base-excision repair pathway for repairing ss-DNA lesions generated by BLM. The repair pathway is initiated by Ape1, which catalyzes the hydrolysis of the phosphodiesterase bond 5’ to the oxidized abasic site or the 3’-PG site to generate a 3’-OH. In the case of the oxidized abasic site, the remaining 5’-deoxyribose is removed by the lyase activity of polymerase β (pol β) that then also fills in the gap with the appropriate nucleotide (Short patch pathway). The final ligation is catalyzed by XRCC1 and Ligase III. The one-nucleotide-gap intermediate can also be repaired by the long patch pathway without the involvement of pol β. In this pathway, PCNA, RFC, and pol δ/ε work together to catalyze strand-displacing DNA synthesis 3’ to the gap. The resulting flapping DNA is cleaved by FEN1 and the nicked DNA is ligated by Ligase I.
1.8 Structures of DNA containing abasic sites and their interaction with Ape1.

Because of the importance of DNA repair in understanding BLM-induced cytotoxicity in vivo, our group has been interested in using the structural information obtained about the DNA lesions generated by BLMs to gain insight into the recognition elements required by DNA repair enzymes. We have focused our attention on the interaction between DNA lesions and Ape1 because of Ape1’s essential role in repair BLM-induced DNA lesions as well as normal abasic DNAs.

The X-ray structure of Ape1 has been solved by the Tainer group in a complex with duplex DNA containing an abasic site analog (tetrahydrofuran, or THF)(99) and provides a paradigm to think about structural basis for lesion recognition by Ape1. In this structure, Ape1 interacts with the abasic site (THF) using residues from a flexible loop and a helix-turn-helix motif, which bind the DNA from both major and minor grooves (Figure 1.10A). The duplex DNA is bent ~35° at the site of the lesion. The most striking feature of the structure is that the THF moiety is flipped out of the phosphodiester backbone into a binding pocket formed by protein side chains that exclude intact bases (Figure 1.10B). It remains unclear whether Ape1 recognizes complete or partially flipped-out abasic sites or Ape1 actively flips the abasic site into the binding pocket.

The flexibility of the abasic site has been proposed to be exploited by the repair enzymes for the recognition and repair of the abasic lesions. The conformational dynamics of duplex DNA around the abasic site have been studied using time-resolved fluorescence spectroscopy on THF-containing DNAs with fluorescent base analogs incorporated opposite to or on either side of THF site(100-102). The fluorescence
relaxation times ranging from 50 ps to 10 ns have been detected in these studies. However, it is still largely speculative as to whether the fluorescent relaxation times correlate with specific conformational changes in the abasic site region. Recently, the Berg group has measured the time-dependent Stokes-shift of coumarin-labeled THF DNA in a complex with Apel and compared it to the free DNA(103). No significant differences were observed for the Stokes-shift between free THF DNA and its complex with Apel, suggesting that the fluorescent probe is insensitive to conformational changes due to the interaction with Apel.

Figure 1.10 The structure of human Apel bound to a duplex DNA containing a THF site. A: Apel binds the THF site by clamping it from the major groove (front) using a helix-turn-helix motif (residue 175-185, in magenta) and from the minor (back) groove using a coil region (residue 266-280, in orange). B: The front helix-turn-helix motif is removed to show the extrahelical conformation of the THF moiety in the complex (indicated by the arrow).
Given the technical challenges to study the structure and dynamics of abasic DNAs on a fast time scale, we(85) and Bolton’s group(104-106) have been using 2D NMR spectroscopy to study the averaged structures of abasic DNA lesions in solution. As with fluorescence studies, there are several limitations for 2D NMR structural studies. First, information obtained from 2D NMR spectroscopy reflects the averaged structures in solution and can only be deconvoluted if multiple conformations are in slow exchange. Second, NMR experiments are not sensitive to minor conformations in the ensemble of solution structures that are often mechanistically relevant. Third, the accuracy of the final structural modeling relies on the richness of the NOE interactions and coupling constants around the abasic site, which thus far have been very limited due to the flexible nature of the lesion. Despite the above limitations, 2D NMR spectroscopy remains a major tool to obtain structural insight into the flexability of the lesions and the neighboring base pairs.

Hoehn et al. have determined the solution structures of a piece of duplex DNA containing the 5’-GXAC-3’/3’-CATG-5’ sequence (X denotes an abasic site) (85). Four NOE interactions involving the abasic site and nearby residues were used to define the extrahelical conformation of the abasic site. The Bolton group has also studied oligonucleotides containing an abasic sites in a defined sequence context using 2D NMR(104-106). The consensus for the abasic DNA structures studied by our group and by Bolton’s group is that the abasic site does not perturb the overall B-form structure of duplex DNA and that the base opposite to the abasic site stacks well between neighboring base pairs. However, sequence-dependent variations have been observed for the conformation of the deoxyribose at the abasic site. In the structure of 5’-GAXAC-3’ with an adenine opposite to the abasic site, the abasic site is extrahelical for the α anomer and
intrahelical for the β anomer (105). Surprisingly, in the same sequence (5'-GAXAC-3') with a cytosine opposite to the abasic site, the structural modeling showed that the abasic site existed only as the β anomer in two distinct conformations attributed to two different hydration states (104). The observation that the base opposite the lesion could alter its anomerization state is contradictory to early studies that showed that two anomers existed in abasic sites with all four opposite bases (107, 108). However, in these early studies, the ambiguity of the chemical shift assignments and the lack of any NOE information between the abasic site and the adjacent residues require that the model structures are solely the results of molecular modeling.

In addition to studies on normal abasic sites discussed above, most researchers have focused on THF analogs of the abasic sites due to their chemical stability (109-112). There are several studies on solution structures of the THF-containing duplex DNA. A consensus has been reached from these studies that the THF does not change the overall B-form structure, similar to model structures determined for normal abasic sites. However, the conformations of the THF moiety (sugar pucker and intra- versus extrahelicity) and of the bases opposite to the lesion remain controversial. In most studies, the THF moiety is intrahelical and the opposite base stacks well between neighboring base pairs. In one study, however, both the THF and the pyrimidine opposite to this lesion are proposed to be extrahelical (110). However, no final structures were determined in that study due to the lack of molecular modeling techniques at that time. It has been proposed that the conformations of THF and/or the opposite base could be sequence-dependent. However, no systematic studies have been performed to analyze the role of neighboring bases in determining the local conformation and no generalized rules
have been postulated to predict the influence on abasic site conformations by sequence contexts. Furthermore, because there is no structural comparison of abasic-site and THF-containing duplex DNA in the same sequence context, it remains unclear whether the THF represents a good structural model for the abasic site.

Structural studies on normal abasic sites and THF analogs summarized above raise two interesting questions: how does the sequence context (including opposite and neighboring bases) affect the abasic site conformation and how does local conformation affect the interaction of abasic sites with DNA repair enzymes. As a first step to address the modulation of abasic site conformation in different sequence contexts, we studied solution structures of duplex DNA containing a normal abasic site in the 5'-GXAC-3'/3'-CNTG-5' sequence (N= A, T, G, or C) by 2D NMR methods. These studies, presented in Chapter 6, provide insight into structural variations at the abasic site caused by the identity of the opposite base, which could be exploited by DNA repair enzymes for sequence-specific recognition of abasic sites.

1.9 References


deglycobleomycin A2 analogues containing a tertiary N-methyl amide and simple
ester replacement for the L-histidine secondary amide: Direct functional
characterization of the requirement for secondary amide metal complexation.

Bleomycin A2 That Directly Address the Effect and Role of the Disaccharide:
Demannosylbleomycin A2 and a-D-Manopyranosyldeglycobleomycin A2. *J.


Role for the Metal-Binding Domain in Determining the DNA-Sequence

269, 10899-10904.

(28) Katano, K., An, H. Y., Aoyagi, Y., Overhand, M., Sucheck, S. J., Stevens, W. C.,
group antibiotics. Total syntheses of bleomycin demethyl A(2), bleomycin A(2),
and decarbamoyl bleomycin demethyl A(2). *J. Am. Chem. Soc.* 120, 11285-
11296.


M. (2002) Deglycobleomycin: solid-phase synthesis and DNA cleavage by the

(31) Thomas, C. J., Chizhov, A. O., Leitheiser, C. J., Rishel, M. J., Konishi, K., Tao,
Z. F., and Hecht, S. M. (2002) Solid-phase synthesis of bleomycin A(5) and three
monosaccharide analogues: exploring the role of the carbohydrate moiety in RNA

synthesis of deamido bleomycin A2, the major catabolite of the antitumor agent

ratio of single- to double-strand DNA breaks and their absolute values determine

Co•deglycoBleomycin A2 green and its complex with d(CCAGGCCTGG). *J. Am.

subcellular distribution of the copper complexes of bleomycin-A2 and
deglycobleomycin-A2. *Anticancer Drugs* 5, 199-201.


Chapter 2. Mechanistic studies of BLM-mediated dsDNA cleavage using CD-BLM
2.1 Introduction

The bleomycins (BLMs, Figure 2.1) are a group of natural glycopeptides produced by *Steptomyces verticillus* that have potent antitumor activity against lymphomas, head and neck cancers, and testicular cancer(1, 2). The BLMs’ therapeutic efficacy is proposed to be related to their ability to cause both single-stranded (ss) and double-stranded (ds) DNA damage in the presence of the required cofactors (Fe(II), O₂, and a one-electron reductant) (3-8). Figure 2.2 shows the pathways proposed for BLM-mediated DNA damage. The ds-DNA cleavage events have long been believed to be the major source of BLMs’ cytotoxicity(8, 9). This chapter will focus on the mechanism of BLM-mediated dsDNA cleavage.

A model in which a single BLM molecule, which has a single site that can be activated for cleavage, can effect two cleavage events without dissociation from DNA has been proposed by Povirk *et al.*(10). This model was initially proposed based on the analysis of the ss to ds cleavage ratios using the supercoiled DNA relaxation assay(11), which revealed a constant ss:ds cleavage ratio (6:1) over a 20-fold range of BLM concentrations. Povirk *et al.* also studied the chemistry and sequence selectivity of the ds DNA cleavage by analyzing 30 ds-cleavage sites using native DNA gel electrophoresis(10, 12). These studies led to a set of sequence selectivity rules for BLM-mediated ds cleavage (Figure 1.4, Chapter 1), which could not be explained by two random ss cleavage events(10). Using the internally [³²P]-labeled hairpin DNA technology(13, 14), Abaslon in our group has determined the ss to ds cleavage ratios at a hot spot for ds cleavage (5'-GTAC-3', _ indicates the cleavage site). The results indicated that the ss to ds cleavage ratio remained constant over a 70-fold range of the BLM
concentration(13). In addition, the ratio was not perturbed by the competitive binding of a chemically unreactive BLM-Co(III)-OOH, which supported the model of a single BLM molecule to effect dsDNA cleavage.

Structural models of BLM-Co(III)-OOH bound to several oligonucleotides containing a single BLM-binding site have been determined using 2D NMR studies by our group(15, 16), which provided a working model to explain the mechanism by which a single BLM molecule can effect ds-cleavage without dissociation from the DNA. This model requires binding of the bithiazole tail of BLM by partial intercalation at the first cleavage site (Figure 1.5, Chapter 1). Subsequent to damage initiation on the first stand, the BLM molecule is repositioned to the second strand by rotation of 180° around the C-C bond connecting the two thiazolium rings and by further rotation of 117° around the axis perpendicular to the bithiazole ring (Figure 1.6, Chapter 1). The mechanism of triggering the re-organization to the second cleavage site and the relative rate of this process versus dissociation from the DNA play a key role in identifying which sites are hot spots for ds-cleavage.

Several studies using BLM analogs have demonstrated that alternative binding modes to the partial intercalation can lead to ss-DNA cleavage. A BLM analog with a chlorinated bithiazole (the H5 and H5' of the bithiazole were substituted with chlorine atoms) was shown to mediate a sequence-selective DNA cleavage similar to BLM(17, 18). The chlorinated bithiazole tail was shown to be located in the minor groove of DNA by light-mediated DNA cleavage(18). The Hecht group has shown that BLM A5 tethered through the polyamine tail to either controlled pore glass (CPG) beads(19, 20) or a dendrimer(21) can also cleave DNA. These attachments were designed be sufficiently
large to prevent binding of the bithiazole tail by partial intercalation. The efficiency and sequence selectivity of cleavage by these tethered BLMs has been reported to be similar to that observed with the untethered BLM, demonstrating that non-intercalative modes of binding can also lead to DNA cleavage \((19-21)\). Unfortunately, the relationship of the alternative binding mode to the ds-DNA cleavage was not quantitatively investigated in these studies \((19, 21)\).

In this chapter, we extend Hecht’s approach to directly test the significance of the partial intercalation in the model of one BLM molecule to effect ds-cleavage. A bulky BLM analog with a β-cyclodextrin attached to the terminal amine of BLM A5 was synthesized (Figure 2.1). This modification is sufficiently large to prevent the binding of BLM by partial intercalation of the bithiazole tail. This β-cyclodextrin-tethered BLM analog (CD-BLM) has been thoroughly characterized. CD-BLM’s ability to effect dsDNA cleavage is reported using the internally \([32P] \) labeled hairpin DNA method \((13)\) and the supercoiled plasmid relaxation assay. The results indicate that CD-BLM can mediate both ss and ds-DNA cleavage in the same sequence-selective fashion as BLM A5, although 5 fold less efficient. Competition assays of the CD-BLM-mediated ds-DNA cleavage in the presence of chemically unreactive BLM-Co(III)-OOH indicate that the ds DNA cleavage is mediated by two CD-BLM molecules. The current study suggests that the partial intercalation is essential for one BLM molecule to mediate ds cleavage. Minor groove binding of the bithiazole tail can also lead to dsDNA cleavage, however, it requires cooperative binding of a second BLM molecule.
Figure 2.1 Structures of BLM A2, A5, Phleomycin D1 (PLM), and CD-BLM.
2.2 Experimental Section

**Material** The oligonucleotides (full-length GT-2 and GT-2 self-primer, Figure 2.3), were synthesized by Invitrogen Inc. on a 1-μmol scale. BLM A2 and A5 were purchased from CalBiochem in a metal-free form. BLM-Co(III)-OOH was prepared and purified as previously described. BLM CD-BLM was synthesized by Dr. Manas K. Ghorai of our group. The detailed synthesis and characterization of CD-BLM is reported in Appendix A. The concentration of CD-BLM was determined by the extinction coefficient at 292 nm (14,500 M⁻¹cm⁻¹), assumed to be identical to BLM A5. DNA-sequencing gel solutions
were purchased from National Diagnostics Co. *E. coli* DNA polymerase I-Klenow fragment (Klenow Fragment) was purchased from New England Biolabs (Catalogue No. M0210S, 1 U is defined as the amount of enzyme required to convert 10 nmol dNTPs into an acid-insoluble form in 30 min at 37 °C). T4 Polynucleotide Kinase (T4 PNK) was purchased from New England Biolabs (Catalogue No. M0201S, 1 U is defined as the amount of enzyme to incorporate 1 nmol acid-insoluble [32P] in 30 min at 37 °C). [γ-32P] Adenosine triphosphate (ATP) (Catalogue No.BLU502H, 3000 Ci/mmol) and [α-32P] deoxyguanosine triphosphate (dGTP) (Catalogue No.BLU514H, 3000 Ci/mmol) were purchased from Perkin Elmer Life Sciences. Other chemicals were purchased from Sigma-Aldrich Co.

---

![Diagram of GT-2 and the distribution of the cleavage products](image)

<table>
<thead>
<tr>
<th>Cleavage site</th>
<th>Predicted length</th>
<th>5'-End</th>
<th>3'-End</th>
<th>Observed length</th>
</tr>
</thead>
<tbody>
<tr>
<td>C10ss</td>
<td>41</td>
<td>Phosphate</td>
<td>OH</td>
<td>42-43</td>
</tr>
<tr>
<td>C43ss</td>
<td>42</td>
<td>OH</td>
<td>Phosphoglycolate (PG)</td>
<td>42-43</td>
</tr>
<tr>
<td>T13</td>
<td>38</td>
<td>Phosphate</td>
<td>OH</td>
<td>39</td>
</tr>
<tr>
<td>T36</td>
<td>15</td>
<td>Phosphate</td>
<td>OH</td>
<td>16</td>
</tr>
<tr>
<td>T38</td>
<td>37</td>
<td>OH</td>
<td>PG</td>
<td>37-38</td>
</tr>
<tr>
<td>A32</td>
<td>19</td>
<td>Phosphate</td>
<td>OH</td>
<td>20</td>
</tr>
<tr>
<td>T13+T38ds</td>
<td>24</td>
<td>phosphate</td>
<td>PG</td>
<td>24</td>
</tr>
</tbody>
</table>

**Figure 2.3** Sequence of GT-2 and the distribution of the cleavage products. The predicted ss-cleavage sites of GT-2 are numbered. The G in bold denotes the internal [32P] labeling site. The GT-2 self-primer contains the first 36 nucleotides of the full-length GT-2 (5'-CGA ATT CTG CTG TAC ACT TTC CCA AAA AGG GAA AGT-3').
Purification of the GT-2 primer by polyacrylamide gel electrophoresis (PAGE) The PAGE experiments to purify oligonucleotides were carried out on a Bio-Rad Protean Xl II Cell apparatus with the gel dimensions of 14 cm × 15.5 cm × 1.5mm. Each gel contained 40 mL 20% acrylamide/bis-acrylamide (19:1 cross-linking) in TBE buffer (89 mM Tris, 89 mM Boric acid and 2 mM EDTA pH 8.3) with 8 M urea. Crude GT-2 primer (200 nmol) from solid phase synthesis was dissolved in 500 µL loading buffer containing 8 M urea in TBE buffer with bromophol blue (0.01%) and Xylene FF (0.01%) as the indicator dyes (under these conditions, bromophol blue and Xylene FF co-migrated with 22-mer and 6-mer ss-DNA, respectively). The PAGE gel was run under a constant voltage of 290V for 2 h with circulating 50°C water. The apparatus was then dissembled and the gel was removed from the glass plates and wrapped with plastic film (SaranWrap). To visualize the region containing the GT-2 primer, the gel was laid on top of a Kodak Intensifying plate that had a fluorescent surface. A handheld UV light was used to detect the shadow on the Kodak Intensifying plate, which corresponded to the DNA bands in the gel. The regions of gel containing the DNA bands were excised soaked in the TBE buffer (5 mL) at 4 °C for 16 h. The purified GT-2 primer was then precipitated from the extraction buffer by mixing with 2 mL 3 M sodium acetate (pH 5.3) and 13 mL ethanol. The mixture was incubated at -80°C for 1 h and was centrifuged for 30 min at 13,000 rpm at 4 °C. The supernatant was discarded and the precipitate was washed once by mixing with 1 mL 75% ethanol and centrifuging for 30 min at 13,000 rpm at 4 °C. The supernatant was discarded and the precipitate was dried in vacuum. The yield of the GT-2 primer was 110 nmol (55%).

5'-[32P] labeling of full length GT-2. The 5'-[32P] labeling reaction was carried out in a 50
µL reaction mixture containing 5 µM full length GT-2, 5 µM [γ-32P]ATP (250 µCi) and 30 U T4 PNK in T4 PNK buffer (50mM Tris-HCl, 10mM MgCl₂, 5mM DTT, pH 7.6). The reaction mixture was incubated at 37 °C for 1 h. T4 PNK was deactivated by heating the reaction mixture at 75 °C for 5 min. The reaction mixture was then ethanol-precipitated by adding 20 µL 3 M sodium acetate (pH 5.3) and 120 µL ethanol. The ethanol precipitation mixture was kept at -80°C for 1 h and was centrifuged at 13,000 rpm for 30 min at 4°C. The supernatant was discarded and the precipitate was washed once by re-suspending the pellet with 1 mL 75% ethanol and centrifuging for 30 min at 13,000 rpm at 4°C. The supernatant was discarded and the precipitate was dried in vacuum. The labeling product was further purified by PAGE as described above. This labeling procedure yielded approximately 120 pmol 5'-[32P] GT-2 (2.1 x 10⁶ cpm/µmol).

Maxam-Gilbert sequencing of 5'-[32P] GT-2 The “G+A” reaction of the Maxam-Gilbert sequencing method was performed according to the published method(22) with slight modification. The reaction mixture of 30 µL contained 50,000 cpm 5'-[32P] GT-2, 3 µg calf thymus DNA, and 6 µL 98% formic acid. The reaction mixture was incubated at 37 °C for 9 min. The reaction was quenched by adding 10 µL 1 mg/mL calf thymus DNA and was dried in vacuum. The DNA pellet was then re-dissolved in 100 µL piperidine and the mixture was incubated at 95 °C for 30 min. The piperidine was removed by evaporating the reaction mixture in vacuum in a SpeedVac concentrator (Savant Co.). The final sample was dissolved in 50 µL DNA loading buffer containing 8 M urea in the TBE buffer.

Internal [32P]-labeling of GT-2 The internal labeling procedure was carried out in two steps in a one-pot reaction: incorporation of [32P]-dGMP followed by further extension of
the primer-template (13). For the incorporation of $^{32}$P-dGMP, the reaction mixture (50 μL) contained 2 μM GT-2 self-primer (Figure 2.3), 2 μM [$\alpha^{32}$P]-dGTP (84.5 μCi), and 10 U Klenow Fragment in Klenow reaction buffer (10mM Tris-HCl, 5mM MgCl$_2$, and 7.5mM DTT, pH 7.5). Before adding [$\alpha^{32}$P]-dGTP and Klenow Fragment, the mixture was heated to 95 °C for 2 min in a heating block and was cooled slowly to 25 °C for 1 h. The reaction mixture was incubated at 25 °C for 15 min before the dNTP mixture and an additional 10 U of Klenow Fragment were added. The final primer extension reaction contained 1.5 μM partially extended GT-2 primer, 400 μM dNTP and 20 units Klenow Fragments in 75 μL Klenow buffer. The reaction mixture was incubated at 25 °C for 20 min and was quenched by adding 30 μL of 3 M sodium acetate (pH 5.3) and 220 μL ethanol. The ethanol precipitation mixture was kept at -80°C for 1 h and was centrifuged at 13,000 rpm for 30 min at 4°C. The supernatant was discarded and the precipitate was re-suspended in 1 mL 75% ethanol. The mixture was centrifuged for 30 min at 13,000 rpm at 4 °C. The supernatant was discarded and the precipitate was dried in vacuum. The labeling product was further purified by PAGE as described above. The final yield of internal-$^{32}$P GT-2 was approximately 60 pmol (2 x 10$^7$ cpm/μmol).

Cleavage of internal $[^{32}P]$-GT-2 by BLMs The reaction mixture of 80 μL contained internal $[^{32}P]$-GT-2 (100,000 cpm), 0.6 μg/μL calf thymus DNA, 50 mM NaCl, and 50 mM HEPES pH 7.5. Before adding BLM, the mixture was heated to 95 °C for 2 min in a heating block and was cooled to 4 °C over 1 h. BLM A5 or CD-BLM was activated ex situ by mixing 500 μM BLM with equal concentration and volume of iron(II) ($\text{(NH}_4\text{)}_2\text{Fe(SO}_4\text{)}_2$) for 60 ± 2 s at 4 °C (13). For the reactions with low BLM concentrations (2.5 to 10 μM), the activation mixture was diluted 10-fold with deionized
water to make a 25 μM solution. For the reactions with high BLM concentrations (10 to 50 μM), the activation mixture was used without further dilution. An appropriate aliquot of the activation mixture (250 μM or the 25 μM after dilution) was added to the reaction mixture to achieve the desired final concentration. The reaction mixture was incubated at 4 °C for 10 min and was quenched by adding 50 μL 3 M NH₄OAc (pH 5.2) and 370 μL ethanol at 4 °C. The ethanol precipitation followed the same procedure described above. The DNA pellets were dried in vacuo and were dissolved in 10 μL DNA loading buffer (containing 90% formamide (v/v)).

In reactions that contained chemically inert BLMA2-Co(III)-OOH, the appropriate amount of BLM-Co(III)-OOH was mixed with the activated BLMs before they were added to the final reaction mixture.

**Analyzing DNA Cleavage Products by PAGE** The sequencing gel (35 cm × 43 cm × 0.6 mm) contained 65 mL 20% acrylamide/bis-acrylamide (19:1 cross-linking) in 8 M urea and TBE buffer (89 mM Tris-HCl, 89 mM boric acid, and 2 mM EDTA pH 8.3). Electrophoresis was carried out on a DNA sequencing apparatus (Life Technology Co., Model S2). The gel plate was pre-run for 30 min before the samples were loaded in order to heat the gel apparatus to 55 °C. After loading the samples, the gel was run at a constant power of 90 W for 2.5 h. The gel was then removed from the glass plates, dried in a gel dryer (Labconco, Model 100) at 70 °C for 45 min, and was exposed to a phosphorimage screen (35 cm × 43 cm) for 12 h. The DNA gels were visualized by the Phosphorimager Storm 840 (Amersham Biosciences) and quantitation was carried out as described below.

**Quantitation of the DNA cleavage bands by ImageQuant 5.0.** After the exposure of the sequencing gel to the phosphorimage screen, the screen was scanned on a Phosphoimager.
The resolution of the scanning pixel was set to be 200 µm. The gel image was saved either as a GIF file for immediate visualization or as a GEL file for the quantitative analysis. The manufacturer’s software, ImageQuant5.0, was used to quantitate the intensity of the DNA bands in the gel. The procedure of the quantitation has been described previously(13). First, a histogram is generated for each lane. The x axis of the histogram indicates the relative position of the gel and the y axis indicates the intensity of the band. The peaks and the base line are then selected based on the noise level of 0.5 and the sensitivity level of 1 (default settings). The noise level and the sensitivity level are kept constant for all experiments. If two peaks are overlapping, a vertical line is drawn from the valley to separate the two peaks(13). In Figure 2.4, two examples are presented to illustrate this peak picking process. The area and percentage of each peak are used to calculate ss to ds cleavage ratios at T13 and T38.

**Preparation and purification of pUC18 supercoiled plasmid.** The pUC18 plasmid (2686bp, Stratagen) with the ampicillin-resistant gene was transformed into a wild-type *E. coli* DH5α strain. The transformed bacteria were grown in 200 mL Luria-Bertani media overnight and were harvested by centrifugation at 8,000 rpm for 20 min to yield 2 g cells (wet weight). The plasmid was purified from the cell pellet using the Qiagen HiSpeed Plasmid MaxiKit following the manufacturer’s protocol. The final yield of the pUC18 plasmid was 350 µg.
Figure 2.4 Quantitation of ss and ds cleavage products analyzed by DNA sequencing PAGE. A, internal $[^{32}P]$-labeled GT-2 cleaved by 7.5 μM BLM A5; B, internal $[^{32}P]$-labeled GT-2 by 50 μM CD-BLM.
Relaxation of supercoiled plasmid DNA by BLMs: The supercoiled DNA relaxation assay has been widely used to measure the ss:ds cleavage ratio by BLM and other DNA-cleaving agents. The cleavage reaction was carried out using the activated BLM generated ex situ, similar to the procedure used in the internally labeled hairpin experiments described above. The final reaction mixture of 20 μL contained 1 μg pUC18 plasmid DNA, 50 mM HEPES, pH 7.5, and 0.2-0.8 μM activated BLM. The BLM was activated by mixing 50 μM BLM (10 μL) with equal volume and concentration of iron(II). The activation mixture was incubated at 4 °C for 1 min. An appropriate aliquot of the activated BLM was added into the reaction mixture to achieve the desired final concentration. The reaction mixture was incubated at 4 °C for 10 min and the reaction was quenched by adding 5 μL of loading buffer containing 50% glycerol.

The type I (supercoiled), II (nicked), and III (linear) forms of plasmid DNA in the reaction mixture were separated by a 1% agarose gel electrophoresis in the TBE buffer at a constant voltage of 80 V for 90 min. The bands were visualized using a Bio-Rad ChemiDoc XRS UV-trans-illuminator and was quantitated by the manufacturer’s software Bio-Rad Quality One. The numbers of the ss cleavage events \( n_1 \) and the ds cleavage events \( n_2 \) can be calculated from the Poisson distribution (11): \( f_1 = \exp(-n_1 n_2) \) and \( f_{III} = n_2 \exp(-n_2) \). (\( f_1 \) and \( f_{III} \): fractions of Type I (supercoiled) and Type III (linear) plasmid).
2.3 Results and Discussion

*Sequence-specific cleavage of GT-2 by CD-BLM.* The internally $^{32}$P-labeled hairpin DNA technology developed in our group has enabled us to quantitate the site-specific ss and ds cleavage simultaneously(13, 14). The hairpin DNA sequence used in this study, GT-2, contains a “hot spot” for ds-DNA cleavage by BLM (T13 and T38 in the 5'-GTAC-3' sequence, T denotes the cleavage site), as well as several good ss-cleavage site, including C10, C43, and T36 (Figure 2.3)(13). The ss and ds cleavages of internally labeled GT-2 by BLM result in DNA fragments with characteristic lengths for each cleavage site (Table in Figure 2.3), which can be separated and quantitated using the DNA sequencing gel electrophoresis(13). The results of a cleavage study of the internally $^{32}$P-labeled GT-2 using various concentrations of CD-BLM (0 to 50 μM) are shown Figure 2.5. The intensities of cleavage products mediated by 50 μM CD-BLM (Lane 6 in Figure 2.5) are analyzed in Figure 2.4B and are compared to the cleavage patterns by BLM A5 (7.7 μM BLM A5, Figure 2.4A). Under these assay conditions, 50 μM CD-BLM produced approximately the same extent of overall cleavage (12%) as 7.5 μM BLM A5 (10%), indicating that CD-BLM is at least 5 fold less efficient in the cleavage of GT-2 than BLM A5. The reduced overall DNA cleavage ability could be attributed to the cyclodextrin group, which would be expected to reduce the affinity of CD-BLM for DNA. It might also be related to the slower activation and/or faster degradation rate of the activated species, CD-BLM-Fe(III)-OOH. At present, a distinction cannot be made between these two possibilities.
Although CD-BLM is less efficient in overall DNA cleavage than BLM A5, the results reveal similar ss-DNA cleavage patterns (Figure 2.4). Thus, the bulky CD-BLM substituent and its presumed hindrance to intercalation do not seem to interfere with its ability to H-bond with G, 3' to the pyrimidine cleavage site.

The relative intensities of the cleavage products at each ss cleavage site (T13, T13, C10, C43, A32, and T36 of GT-2) for CD-BLM are also quite similar to BLM A5 (Figure 2.4) and the previously studied BLM A2(13). The predominant cleavage sites in GT-2 are...
T13 and T38, with C10/C43 >> T36 > A32. A notable difference between CD-BLM and BLM A5 is the relative intensity at T13 and T38. CD-BLM produced almost equal amounts of cleavage products at these two sites, while BLM A5 and A2(13) cleaved the T38 site with twice the efficiency relative to the T13 site (Figure 2.4). The difference of the cleavage intensities between T13 and T38 for natural BLMs was used by Absalon et al. to suggest that T38 might be the primary ss-cleavage site in the dsDNA cleavage(13). The change of the relative cleavage intensity at T13 and T38 for CD-BLM may reflect the affinity difference at these two cleavage sites between CD-BLM and normal BLMs.

Site-Specific dsDNA cleavage mediated by CD-BLM In addition to ss DNA cleavage, CD-BLM mediated dsDNA cleavage at T13 and T38 (Figure 2.4B and Figure 2.5). To ensure that the ds-cleavage is not associated with two independent ss cleavage events, the cleavage assay was carried out in the presence of 0.6 μg/μL unlabeled calf thymus DNA to ensure “single-hit” conditions. Quantitative analysis of the cleavage product intensity at each ss-cleavage site indicates that the amount of the ss-cleavage product increased linearly with the concentration of CD-BLM (Figure 2.6). In addition, the amount of the ds-cleavage also increases linearly with the CD-BLM concentration (denoted by filled circles (●) in Figure 2.7). The ss to ds cleavage ratio at T13 and T38 sites was determined in triplicate to be 6.7±1.2, which is higher than the ratio for BLM A2 (3.4, Absalon) or A5 (3.1±0.3) in the same sequence context. The ss to ds cleavage ratio for CD-BLM remains constant over the range of 12.5 to 50 μM CD-BLM (Figure 2.8), indicating that the observed dsDNA cleavage at T13 and T38 can not come from two coincidental ss cleavage events.
Figure 2.6 Concentration-dependent cleavage at each ss cleavage site in GT-2.
The ss to ds cleavage ratio by CD-BLM (6.7±1.2) is both statistically and mechanistically significant. The separation of the T13 and T38 ss cleavage bands (Figure 2.4) enabled the quantitative estimation of the amount of ds cleavage due to two coincidental ss cleavage events. At 50 μM CD-BLM concentration, the percentages of ss-cleavage at T13 (2.01%) and T38 (2.44%) give rise to the calculated random ds-cleavage of 0.054% (2.01% × 2.44%), which is 17 times smaller than the observed ds cleavage (0.92%). The calculated random ds-cleavage percentages at different CD-BLM concentrations were plotted in Figure 2.7 (■). This calculation, together with the concentration-independent ss to ds cleavage ratio, suggests that the observed ds cleavage by CD-BLM is mediated either by a single CD-BLM molecule, or by the cooperative binding of a second CD-BLM.

The total extent of cleavage in the above study is 3.3 to 14% (12.5 to 50 μM CD-BLM). To investigate whether the observed ss to ds cleavage ratio would change under conditions of <1% cleavage, additional studies were carried out with calf thymus DNA increased from 0.6 μg/μL to 5 μg/μL. The cleavage reaction mixture was analyzed by the DNA sequencing gel and is shown in Figure 2.9. The overall cleavage at 20 μM CD-BLM is 0.6%. The ss to ds cleavage ratio determined in this experiment, 7.8±2.4, (in a separate measurement, 7.3±1.4) is consistent with our previous studies (6.7±1.0). The error in the analysis largely comes from the background, which is likely due to a combination of non-specific cleavage by hydroxyl radicals generated by CD-BLMs, radiation damage of GT-2 from [32P] decay, and spontaneous depurination. The non-specific cleavage by CD-BLM is perhaps the major source of the background because the amount of the laddering (non-specific cleavage) increases with the CD-BLM
concentration. The presence of the background limits of the ss to ds cleavage ratio that can be determined to ~20:1. The previously observed limit of detection by Absalon et al. for the hairpin DNA method was 7:1(14), which was largely determined by ds cleavage generated by two ss cleavage events. When the overall cleavage is less than 1%, the random ds cleavage at T13 and T38 is estimated to be approximately 0.002%, far less than the background, which is usually 0.01%. As a result, the background becomes the predominant factor to set the limit of the quantitation in the present studies.

**Figure 2.7** Concentration-dependent cleavage at the GT-2 ds-cleavage site (T13+T38). The filled circles (●) are observed ds cleavage in percentage at T13 and T38. The filled squares (■) represent the calculated percentage of random ds-cleavage. The number is the product of the percent cleavage at the two ss-cleavage sites (T13 and T38).
Figure 2.8 Log/log plot of the percentage of T13+T38 ss-cleavage versus T13/T38 ds-cleavage using the data from Figure 2.5 with the concentrations of CD-BLM from 12.5 to 50 μM. The linear least-square fit of data (Y = 0.862 + 0.992X) indicates that the ratio of ss to ds cleavage remains constant over a 4-fold range in CD-BLM concentration. The value of the Y intercept, 0.862 from the equation, is consistent with the log value of the ss:ds cleavage ratio (6.7±1.2). ( %ss/%ds = c ⇒ log (%ss) = log c + log(%ds), c: the ss:ds cleavage ratio )
Figure 2.9 Cleavage of internally \(^{32}\text{P}\) labeled GT-2 by CD-BLM. Lane 1-4, 0, 5, 10, 20 μM CD-BLM. The total amount of DNA cleavage was 0.6% for Lane 4. The extent of cleavage has been quantitated by the phosphorimage method for 20 μM CD-BLM (Lane 4). The data are summarized in the table.

<table>
<thead>
<tr>
<th>[20 μM] CD-BLM</th>
<th>ss (T13+T38)</th>
<th>ds (T13 + T38)</th>
<th>Average background</th>
<th>ss to ds ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Band Intensity</td>
<td>213</td>
<td>36.9</td>
<td>9.8</td>
<td>7.8 ± 2.1</td>
</tr>
</tbody>
</table>
Ds-cleavage of GT-2 by CD-BLM is mediated through cooperative binding. To determine whether the observed dsDNA cleavage by CD-BLM is due to a single CD-BLM molecule or due to the cooperative binding of a second CD-BLM molecule, the cleavage of internal $[^{32}P]^{-}\text{GT-2}$ by CD-BLM was further investigated in the presence of the chemically inert BLM-Co(III)-OOH. Originally used by Absalon et al. (14), this competition assay was designed to provide additional support for the model that a single BLM molecule can mediate dsDNA cleavage. In this model, the second cleavage is effected by the bound BLM via re-organization and the ss to ds cleavage ratio is determined by the relative rate of dissociation versus re-organization after the first cleavage event. Thus, the single molecule model predicts that the presence of the unreactive BLM-Co(III)-OOH will not be able to compete for the second cleavage site and that the ss to ds ratio should remain the same. The cooperative binding model requires binding of a second BLM molecule to effect the ds cleavage and it predicts that BLM-Co(III)-OOH can compete for the second binding event and would increase the observed ss to ds ratio.

The cleavage of the internally labeled GT-2 by CD-BLM was analyzed by DNA sequencing gel and is shown in Figure 2.10 with 0 (Lane 2-4), 0.5 (Lane 5-7), or 0.25 (Lane 8-10) equivalent BLM-Co(III)-OOH/CD-BLM. The presence of BLM-Co(III)-OOH suppresses cleavage at both ss and ds sites. However, the ds cleavage is affected much more than the ss cleavage. This effect is best demonstrated by analysis of Lane 4 (20 μM CD-BLM), Lane 10 (20 μM CD-BLM + 5 μM BLM-Co(III)-OOH), and Lane 7 (20 μM CD-BLM + 10 μM BLM-Co(III)-OOH) (Histogram A, B, and C in Figure 2.11) quantitatively. The results (Table in Figure 2.11) indicated that BLM-Co(III)-OOH
suppressed the ds cleavage much more than the ss cleavage. The ss to ds cleavage ratio increases from 7.3±1.4 (no BLM-Co(III)-OOH) to 12.5±2.5 (0.25 equivalent BLM-Co(III)-OOH/CD-BLM) to 21±14.2 (0.5 equivalent BLM-Co(III)-OOH/CD-BLM). The ss to ds cleavage ratio measured with 0.5 equivalent of BLM-Co/CD-BLM, 21±14.2, is close to the lower limit of our detection and represents the background cleavage process (Histogram C in Figure 2.11). For comparison, the competition assays using BLM A5 were also performed with 0 or 5 equivalent of BLM-Co(III)-OOH (Lane 11 and 12 in Figure 2.10). The analysis is shown in Histogram D and E in Figure 2.11. The ss to ds cleavage ratio remains largely unchanged (3.6±0.9) in the presence of 5 equivalent BLM-Co(III)-OOH compared to BLM A5 alone (3.1±0.3). These results suggest that CD-BLM mediates the ds-DNA cleavage by cooperative binding of a second CD-BLM molecule after the first cleavage events, while BLM A5 mediates ds-cleavage by one BLM molecule, similar to the results previously observed by Absalon et al. with BLM A2(14).

**Ss:ds cleavage ratio determined by the supercoiled plasmid relaxation assay.** The supercoiled plasmid relaxation assay has been widely used to determine the ss to ds cleavage ratios by BLM(11, 23-25). The number of the ss and ds cleavage events can be calculated from the fraction of Type II (relaxed) and Type III (linear) plasmid using the Poisson distribution as described in the experimental section(11). The ss to ds cleavage ratio measured using this method is not sequence-specific and reflects the averaged ratio for the whole plasmid DNA.
Figure 2.10 Competitive inhibition of CD-BLM-mediated DNA cleavage by the chemically inert BLM-Co(III)-OOH. Lane 1: Control. Lane 2-4: 5, 10, 20 µM CD-BLM. Lane 5-7: 5, 10, 20 µM CD-BLM with 0.5 equivalent BLM-Co(III)-OOH. Lane 8-10: 5, 10, 20 µM CD-BLM with 0.25 equivalent BLM-Co(III)-OOH. Lane 11: 5 µM BLM A5. Lane 12: 5 µM BLM A5 with 5 equivalents BLM-Co(III)-OOH (25 µM).
Figure 2.11 Quantitation of CD-BLM mediated cleavage (from Figure 2.10) and its attenuation by BLM-Co(III)-OOH. A, 20 μM CD-BLM (Lane 4); B, 20 μM CD-BLM + 5 μM BLM-Co(III)-OOH (Lane 10); C, 20 μM CD-BLM + 10 μM BLM-Co(III)-OOH (Lane 7); D, 5 μM BLM A5 (Lane 11); E, 5 μM BLM A5 + 25 μM BLM-Co(III)-OOH (Lane 12). (* The intensities of the bands are calculated in arbitrary units) For comparison, Linegraph D (20 μM CD-BLM) and E (5 μM BLM A5) are normalized to the intensity of the T38 ss cleavage band.
CD-BLM is also shown to mediate dsDNA cleavage using the supercoiled plasmid relaxation assay. Figure 2.12 shows the image the agarose gel separating the two types of products (linear and nicked) from the starting pUC18 supercoiled plasmid. CD-BLM showed 50% of the cleavage efficiency compared to BLM A5 in this assay. The UV densitometry analysis of the gel is presented in Figure 2.13. Surprisingly, the ss to ds cleavage ratio determined for CD-BLM is 2.8 to 1, higher than that observed for BLM A5 (5.8 to 1) or A2 (7.3 to 1). For comparison, PLM, which is a poor ds cleavage agent, shows a ss : ds ratio of 47 to 1. However, the mechanistic interpretation of these ratios is complicated by the fact that the ss to ds ratio measured in the supercoiled DNA relaxation assay is not sequence specific. It is likely that there exist some hot spots or secondary structures that are especially sensitive for CD-BLM mediated ds cleavage. In addition, the supercoiled DNA relaxation assay cannot distinguish site-specific cleavage from non-specific hydroxyl radical cleavage. Therefore, it is also likely that the observed ds cleavage by CD-BLM is over-estimated by the presence of closely spaced random ss-cleavage sites.

Figure 2.12 BLM and CD-BLM induced relaxation of supercoiled plasmid pUC18. Lane 1, 1.0 μg pUC18; Lane 2, pUC18 + 1 μM Fe(II); Lane 3, pUC18 + 0.2 μM BLM A2; Lane 4, pUC18 + 0.2 μM BLM A5; Lane 5-6, pUC18 + 0.2 and 0.4 μM CD-BLM; Lane 7-8, pUC18 + 0.2 and 0.4 μM PLM.
Figure 2.13 Quantitation of the cleavage products of pUC18 by BLMs, PLM, and CD-BLM using the supercoiled DNA relaxation assay (Figure 2.12). A, 0.2 µM BLM (Lane 3); B, 0.2 µM BLM A5 (Lane 4); C, 0.4 µM CD-BLM (Lane 6); D, 0.4 µM PLM (Lane 8). The data are summarized in the table.
Interaction of CD-BLM with duplex DNA CD-BLM was designed to preclude the intercalation of the bithiazole tail. It has been technically challenging to demonstrate intercalation in ligand-DNA interaction. The classic method is to detect unwinding of the supercoiled DNA using agarose electrophoresis(26). This technique requires a gram-scale material that is unavailable for CD-BLM. In addition, this method is not capable of detecting weak intercalators that are in fast exchange or only intercalate at limited sequences. For example, results from studies with PLM were interpreted to indicate a non-intercalative binding mode(27-29). However, it is later demonstrated by NMR spectroscopy that PLM can intercalate in a duplex DNA containing the 5’-GTAC-3’ sequence(30). At this stage, the only way to address whether CD-BLM intercalates or not is to titrate it with DNA and monitor the chemical shift changes of the bithiazole protons, if a 1:1 complex of CD-BLM/DNA can be obtained in slow exchange on the NMR time scale.

The cooperative binding model of CD-BLM-mediated ds-cleavage requires the potentiation of binding of a second CD-BLM molecule after the first cleavage event. This model predicts that CD-BLM would have a higher affinity for ss-lesioned DNA than for the intact DNA. The identity of the ss-lesioned DNA that is responsible for the potentiation of binding is not clear at this stage. The 3’-phosphoglycolate/5’-phosphate lesion (Figure 2.2) was shown to be present at the first cleavage site in all BLM-mediated ds-cleavage and might play a role in recruiting the second CD-BLM(31) (Section 1.6, Chapter 1). Alternatively, a precursor to this lesion could also be involved in potentiating binding of BLM to the second strand(32, 33). The dissociation constant \( (K_D) \) of CD-BLM with 3’-PG/5’-P lesioned DNA can be measured by monitoring the fluorescence
quenching of the bithiazole upon binding to DNA and would provide additional support for the cooperative binding model.

*Alternative sites for BLM-mediated ds-DNA cleavage?* Recently, an unusual ds cleavage site was identified by the Hecht group in the Dickerson-Drew dodecamer (34) (Figure 1.7, Chapter 1). This site does not conform to the postulated selectivity rules. Structural studies using 2D NMR methods of BLM-Zn(II) bound to this dodecamer were interpreted to indicate that the metal binding domain of BLM-Zn(II) was located halfway between the two cleavage sites with the bithiazole tail in the minor groove (Figure 1.7 B and C, Chapter 1). In this study, however, BLM-Zn(II) is in fast exchange with the DNA on the NMR time scale and the chemical shifts of bithiazole protons could not be assigned. The scarcity of the NOE information between BLM-Zn(II) and DNA (total 12 NOE distance restraints, 3 of them are involved in the bithiazole tail as shown in Figure 1.7 A, Chapter 1) prevented modeling of a unique conformation of the bithiazole tail. Furthermore, our attempt to reproduce the ds-cleavage pattern observed by Hecht *et al.* on the Dickerson-Drew dodecamer has been unsuccessful (Appendix B). Therefore, the significance of this unusual ds-cleavage site and the model structure proposed by Hecht *et al.* remains unclear.

**2.4 Conclusion**

A BLM analog (CD-BLM) with cyclodextrin attached to the bithiazole tail, designed to prevent binding to DNA by partial intercalation, was synthesized to investigate the importance of this binding mode in dsDNA cleavage. Fe(II)-CD-BLM mediates the same sequence-specific DNA cleavage pattern as BLM A2 and A5, although 5 fold less efficient. Internal [³²P]-labeled hairpin DNA technology was used to demonstrate that
Fe(II)-CD-BLM mediates dsDNA cleavage by cooperative binding of a second CD-BLM molecule, in contrast with BLM A5 and A2. The results support the essential role of partial intercalation of the bithiazole tail in a single BLM-mediated dsDNA cleavage.

2.5 References


Appendix A. Synthesis of β-cyclodextrin-tethered BLM A5 (CD-BLM) (Figure 1)

**Material** Bleomycin A₅ was purchased from CalBioChem. β-cyclodextrin and other chemicals were purchased from Sigma-Aldrich. Dioxane was purified by refluxing over KOH, followed by distillation over Na. NMR spectra were measured on Bruker 400 MHz or Varian XL 500 MHz spectrometers. ¹H and ¹³C NMR chemical shifts are expressed in parts per million using tetramethyilsilane as a reference.

![Synthesis of CD-BLM (1)](attachment:image)

**Figure 1** Synthesis of CD-BLM (1)

**Synthesis of mono-6-deoxy-6-(p-tolylsulfonyl)-β-cyclodextrin (3) from β-cyclodextrin (2)**

(a) **Preparation of p-toluenesulfonic anhydride(Ts₂O)**. A mixture of p-toluenesulfonyl chloride (8 g, 0.043 mol) and p-toluenesulfonic acid (2 g, 0.011 mol) in 50 mL of CH₂Cl₂ was stirred at 20 °C overnight. The reaction mixture was then filtered through 100 g silica gel (particle size 40-63 μm, EM Science). The product in the filtrate was precipitated
from hexane to give 6 g of Ts$_2$O (75% yield, reported 88%).

(b) **Synthesis of 3 from 2 using Ts$_2$O.** Ts$_2$O (2.45 g, 7.5 mmol) was added to a suspension of β-cyclodextrin (2) (5.75 g, 5 mmol) in 125 mL of H$_2$O. The reaction mixture was stirred at 20 °C for 2 h. A solution of NaOH (2.5 g, in 25 mL of H$_2$O) was then added to the reaction mixture. The unreacted Ts$_2$O was removed by filtration. The filtrate was brought to pH ~8 by the addition of 6.7 g of solid NH$_4$Cl. The solution was kept at 4 °C overnight to precipitate out the mono-6-deoxy-6-(p-tolylsulfonyl)-β-cyclodextrin as white solid 2.58 g (35%). m.p. 184 °C (reported: 179 °C). $^1$H NMR (Bruker 400 MHz, DMSO-$d_6$): $\delta$ 7.75 (d, $J = 8.1$ Hz, 2H), 7.43 (d, $J = 8.08$ Hz, 2H), 5.89-5.65 [m, 14H, (OH$_2$, OH$_3$)], 4.83 [br s, 5H (H$_1$)], 4.76 [br s, 2H (H$_1$)], 4.55-4.38 [m, 6H (OH$_6$)], 4.32 [m, 1H (H$_{6b}$)], 4.17 [m, 1H (H$_{6a}$)], 3.75-3.43 [m, 28 H (H$_3$, H$_5$, H$_{6a}$, H$_{6b}$)], 3.42-3.15 [m, overlaps with HOD (H$_2$, H$_4$)], 2.42 [s, 3H (CH$_3$)]. ESI-MS for C$_{49}$H$_{76}$O$_{37}$SNa [M + Na]$^+$: calcd: 1311.3678; obsd: 1311.3715.

**Synthesis of mono-6-deoxy-6-azido-β-cyclodextrin (4).**

Mono-6-deoxy-6-(p-tolylsulfonyl)-β-cyclodextrin (3) (300 mg, 0.232 mmol) was suspended in dry DMF (1.0 mL) and warmed to 63 °C over a period of 15 min to obtain a homogeneous solution. Potassium iodide (19 mg, 0.115 mmol) and sodium azide (151 mg, 2.32 mmol) were added and the reaction mixture was stirred at 60-65 °C for 24 h. The reaction mixture was then cooled to 20 °C and was treated with Amberlite MB-3 resin (Sigma) to remove salts. Acetone (50 mL) was added to the effluent providing a white precipitate. The precipitate was collected by suction filtration and dried under vacuum to yield 242 mg of the title compound (90%) as a single product. m.p. 204 °C (reported 206 °C, slow decomposition). TLC (silica): R$_f$ 0.71, (i-PrOH:H$_2$O:EtOA:conoc.
NH₄OH = 5:3:1:1. ¹H NMR (Bruker 400 MHz, DMSO-d₆): δ 5.85-5.60 [m, 14H, (OH, 2 & 3 positions)], 4.87 [br s, 1H (H₁)], 4.82 [br s, 6H (H₁)], 4.60-4.45 [m, 6H (OH, 6-position)], 3.80-3.45 [m, 28 H (H₃, H₅, H₆a, H₆b)], 3.42-3.22 [m, overlaps with HOD (H₂, H₄)]. ESI-MS for C₄₂H₆₉N₃O₃₄Na, [M + Na]+: calcd: 1282.3655, obsd: 1182.3653.

**Synthesis of mono-6-deoxy-6-amino-β-cyclodextrin (5).**

Mono-6-deoxy-6-azido-β-cyclodextrin (4) (0.23 g, 0.198 mmol) and triphenylphosphine (115 mg, 0.438 mmol) were dissolved in DMF (4 mL) and 30% NH₃ (1.0 mL) was added. The mixture was stirred at 20 °C for 5 h and 30 mL acetone was added to give the crude product as white precipitate. The crude product (80 mg) was dissolved in 6 mL H₂O and was loaded on a Sephadex CM-25 column (2.5 x 8 cm). The column was eluted with 100 ml of water then 0.1 N NH₄OH (100 mL). The fractions were collected and the solvent was removed under vacuum to give 50 mg 5 (62%). ¹H NMR (Varian 500 MHz, D₂O): δ 4.92 [m, 7H, (H₁)], 3.86-3.78 [m, 7H (H₃)], 3.77-3.66 [m, 19 H (H₅, H₆a, H₆b)], 3.54-3.46 [m, 7H (H₂)], 3.46-3.39 [m, 6H (H₄)], 3.33 [m, 1H (H₄)], 2.98 [d, 1 H (H₆b)], 2.74 [dd, 1 H (H₆a)]. ¹³C NMR (400 MHz, D₂O): 102.1 (C₁), 101.8 (C₁), 83.1 (C₄), 81.3 (C₄), 81.1, 73.3 (C₃), 72.3 (C₅), 72 (C₂), 71.5, 67.6 (C₆), 60.5 (C₆). ESI-MS: calculated for C₄₂H₇₁N₃O₃₄ [M + H]+ 1134.3930, found 1134.3956.

**Synthesis of 6 by coupling of 5 with adipic acid monomethyl ester.**

Adipic acid monomethyl ester (5.23 mg, 0.033 mmol) was added to a solution of 5 (35 mg, 0.031 mmol) in 0.6 mL of DMF. After the solution was cooled to -10 °C, DCC (7.63 mg, 0.037 mmol) and HOBt (5 mg, 0.037 mmol) were added. The resulting reaction mixture was kept for stirring at 20 °C for 24 h. Insoluble materials were removed by filtration and 10 mL acetone was added to the filtrate. The precipitate was collected and
dried *in vacuo* to give 30 mg of crude product (6). Crude product was purified by silica gel column chromatography (1.5 x 10 cm column). 25 mg of the crude product was dissolved in *n*-BuOH/EtOH/H₂O (5:4:3) and silica gel (1g) was added, solvents were removed to make the mixture anhydrous. The dried silica gel adsorbed with the compound was loaded on the top of a column prepacked with silica gel (dry packing) with a funnel. The compound was eluted with the same solvent, *n*-BuOH/EtOH/H₂O (5:4:3) under pressure. 5 mL fractions were collected. Fractions were collected based on TLC [silica, *n*-BuOH/EtOH/H₂O (5:4:3)]. Solvents were removed under vacuo to give 23 mg of 29 (70%). ¹H NMR (Bruker 400 MHz, DMSO-*d₆*): 6.07-5.61 [m, 14H, (OH₂, OH₃)], 4.94-4.73 [m, 7H (H₁)], 4.63-4.5 [m, 6H (OH₆)], 3.81-3.52 [m, 29H (H₃, H₅, H₆a, H₆b, OMe)], 3.47-3.11 [m, overlaps with HOD(H₂, H₄)], 2.30-2.28 [m, 2H (H₄)], 2.20-2.08 [m, 2H (H₅)], 1.58-1.30 [4H, (H₆a, H₆b)]. ESI-MS for C₄₉H₈₁NO₃₇ [M + H]⁺: 1276.4560, found 1276.4565.

**Synthesis of 7 by hydrolysis of the methyl ester derivative of β-cyclodextrin (6)**

To a solution of 6 (39 mg, 30.58 µmol) in 0.8 mL of dioxane/H₂O (3:1), 25.6 µL 2 M NaOH (51.2 µmol) was added and the reaction mixture was stirred at 20 °C for 6 h. The reaction mixture was neutralized with glacial acetic acid and the solvent was removed to give 40 mg crude solid product. The crude product was dissolved in 2 mL water and was precipitated by adding 50 mL acetone to remove sodium acetate. The crude product (40 mg) was then purified by reverse phase C18 column chromatography (10 mL column). The column was run with 100 mL of water and followed by 5% MeOH/H₂O. Fractions were collected and solvent was removed to give 20 mg 7 as white solid in 52% yield. ¹H NMR (Bruker 400 MHz, D₂O): 5.05-4.95 [m, 7H (H₁)], 3.95-3.68 [m, 27H (H₃, H₅, H₆a, H₆b, OMe)], 2.38-2.29 [m, 2H (H₄)], 1.60-1.42 [4H, (H₆a, H₆b)].
H_{6a}), 3.60-3.45 [m, 13H (H_{2}, H_{4})], 3.40-3.3 [m, 1H, H_{4}], 3.26 [dd, J_{5,6a} = 6.4 Hz, J_{6a,6b} = 14.3, 1 H (H_{6a}) 1H, H_{6a}], 2.25-2.16 [m, 2H (H_{a})], 2.15-2.08 [m, 2H (H_{b})], 1.59-1.45 [4H, (H_{b}, H_{c})]. ESI-MS: calculated for C_{48}H_{78}NO_{37} [M - H]^{-} 1260.4247, found 1260.4278.

**Coupling 30 with Cu(II)BLMA_{5}**:  

To a solution of 30 (2.5 mg, 1.98 μmol) in 0.3 mL of DMF at rt was added diisopropylethylamine (0.51 mg, 3.96 μmol, 10 μL from a solution of 34.5 μL in 500 mL of DMF) followed by BOP reagent (0.88 mg, 1.98 μmol, 10 μL from a solution of 44 mg in 500 μL of DMF). The reaction mixture was stirred for 2 min at rt and the resulting solution was added to a anhydrous mixture of Cu(II)BLMA_{5} (1.98 μmol) in 0.2 mL of DMF at 0 °C. Amount of Cu(II)BLMA_{5} was measured from its aqueous solution by UV (ε of Cu(II)BLMA_{5} = 17400) and it was made anhydrous by removing water and repeated evaporation from dry DMF (4 x 1 mL) followed by drying under vacuum. The reaction mixture was stirred at 4 °C for 36 h.

Solvent was removed in vacuo to give a blue product which was dissolved in 520 μL of water and purified by reverse phase HPLC. Column: Econosil C18 10 μ (Altech, l = 250 mm, id = 4.6 mm), solvent: linear gradient of 0-50% acetonitrile in 0.1 M NH_{4}OAc (pH 6.8) over 60 min; flow rate 1.0 mL/min.

Cu(II)BLMA_{5} was eluted at 34.9 min and the major product was eluted at 27.8 min (Fig 40). Fractions were collected by monitoring the absorption at 292 nm. Fraction of the major product was lyophilized to give 31 as a bluish solid. Yield of the product 48% (the fraction having RT = 27.8, considering same ε = 17400 as of Cu(II)BLMA_{5} at λ292).

UV and ^1H NMR spectra were recorded. Cu(II) being a paramagnetic species, the
product could not be characterized by \textsuperscript{1}H NMR; it gave only some broad signals. UV spectrum looks exactly similar to that of Cu(II)BLMA\textsubscript{3} and having absorption at 292 and 592 nm. High resolution mass (ESI): calculated for C\textsubscript{105}H\textsubscript{166}N\textsubscript{20}O\textsubscript{57}S\textsubscript{2}Cu [M\textsuperscript{2+}] 1372.9716, found 1372.9767 (Corresponding M\textsuperscript{+} = 2745.9534).

**Synthesis of metal-free β-cyclodextrin-tethered BLM A5 (CD-BLM, 1)**

The copper complex form of CD-BLM (8, 0.81 \textmu mol) was dissolved in 0.6 mL 15\% Na\textsubscript{2}EDTA (pH 4.65) and the reaction mixture was stirred at 4 °C for 20 h. The reaction mixture was then purified by RP-HPLC using a Econosil C18 column (4.6 x 250 mm) with a linear gradient of 20-50\% methanol in 0.1 M NH\textsubscript{4}OAc (pH 6.8) over 50 min. The flow rate is 1.0 mL/min. Compound, retention time, yield: 1, 28.2 min, 61\%. The chemical shifts of 1 was listed in Table 1. ESI-MS: calculated for C\textsubscript{105}H\textsubscript{168}N\textsubscript{20}O\textsubscript{57}S\textsubscript{2} [M + 2H]\textsuperscript{2+} 1342.5146, found 1342.5163 (Corresponding M\textsuperscript{+} = 2683.0326).
<table>
<thead>
<tr>
<th>Protons</th>
<th>BLMA2</th>
<th>Tethered BLMA₅</th>
<th>Protons</th>
<th>BLMA2</th>
<th>Tethered BLMA₅</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td></td>
<td></td>
<td>B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CαH</td>
<td>2.62</td>
<td>2.61</td>
<td>CαH</td>
<td>3.24</td>
<td>3.25</td>
</tr>
<tr>
<td>CαH'</td>
<td>2.69</td>
<td>2.66</td>
<td>CβH</td>
<td>3.60</td>
<td>3.61</td>
</tr>
<tr>
<td>CβH</td>
<td>3.96</td>
<td>3.99</td>
<td>C5H</td>
<td>8.20</td>
<td>8.27</td>
</tr>
<tr>
<td>CH₃</td>
<td>2.01</td>
<td>1.92</td>
<td>C5'H</td>
<td>8.00</td>
<td>8.06</td>
</tr>
<tr>
<td>H</td>
<td></td>
<td></td>
<td>G</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CαH</td>
<td>5.05</td>
<td>5.06</td>
<td>C1H</td>
<td>5.25</td>
<td>5.26</td>
</tr>
<tr>
<td>CβH</td>
<td>5.26</td>
<td>5.28</td>
<td>C2H</td>
<td>4.01</td>
<td>4.00</td>
</tr>
<tr>
<td>C2H</td>
<td>7.79</td>
<td>7.79</td>
<td>C3H</td>
<td>4.09</td>
<td>4.00</td>
</tr>
<tr>
<td>C4H</td>
<td>7.26</td>
<td>7.26</td>
<td>C4H</td>
<td>3.84</td>
<td>3.82</td>
</tr>
<tr>
<td>A</td>
<td></td>
<td></td>
<td>M</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CαH</td>
<td>3.84</td>
<td>3.92</td>
<td>C1H</td>
<td>4.99</td>
<td>4.98</td>
</tr>
<tr>
<td>CβH</td>
<td>2.83</td>
<td>2.68</td>
<td>C2H</td>
<td>4.04</td>
<td>4.04</td>
</tr>
<tr>
<td>CβH'</td>
<td>2.88</td>
<td>2.76</td>
<td>C3H</td>
<td>4.65</td>
<td>4.65</td>
</tr>
<tr>
<td>V</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>αCH3</td>
<td>1.10</td>
<td>1.07</td>
<td>a, d</td>
<td>2.2</td>
<td></td>
</tr>
<tr>
<td>γCH3</td>
<td>1.12</td>
<td>1.09</td>
<td>b, c</td>
<td>1.52</td>
<td></td>
</tr>
<tr>
<td>CαH</td>
<td>2.45</td>
<td>2.41</td>
<td>e</td>
<td>3.17</td>
<td></td>
</tr>
<tr>
<td>CβH</td>
<td>3.70</td>
<td>3.70</td>
<td>f</td>
<td>1.56</td>
<td></td>
</tr>
<tr>
<td>CyH</td>
<td>3.86</td>
<td>3.86</td>
<td>g</td>
<td>1.68</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CH₃</td>
<td>1.08</td>
<td>1.01</td>
<td>h</td>
<td>3.04</td>
<td></td>
</tr>
<tr>
<td>CαH</td>
<td>4.21</td>
<td>4.19</td>
<td>i</td>
<td>3.1</td>
<td></td>
</tr>
<tr>
<td>CβH</td>
<td>4.08</td>
<td>4.06</td>
<td>j</td>
<td>2.02</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>k</td>
<td>3.52</td>
<td></td>
</tr>
</tbody>
</table>
Appendix B. Re-examination of the proposed ds-cleavage site in the Dickerson-Drew dodecamer (Figure 1).

Figure 1. Sequence of Dickerson-Drew dodecamer (1) and a hairpin DNA (2) containing the sequence of 1. The proposed ds cleavage site in 1 is indicated by the solid arrows. The dashed arrow indicates the symmetry-related ds cleavage site. The proposed ss-cleavage sites (A5, C31, and 111) in 2 is indicated by the arrows. The A5/C31 cleavage sites in 2 are equivalent to the ds-cleavage site in 1.

The cleavage pattern of the [5′-32P]-hairpin DNA (2, Figure 1) containing the Dickerson-Drew dodecamer sequence (1) by BLM was re-examined under two assay conditions. The first assay condition followed exactly the same procedure as described by Keck et al. (J. Am. Chem. Soc., 2001, 123, 8690-8700). The reaction mixture of 100 μL contained 5×10^6 cpm (2 pmol) [5′-32P]-2 in 10 mM NaCl, 50 mM sodium cacodylate, pH 6.5. An equal amount of Fe(II) (as (NH₄)₂Fe(SO₄)₂) and BLM A2 was added separately into the reaction mixture from a 100 μM stock solution to achieve final Fe(II)-BLM concentration of 0, 1, 2, 5 μM. The reaction mixture was incubated at 0°C for 30 min. The reaction mixture was then mixed with 100 μL of 2 M piperidine. The mixture (200 μL) was heated at 90 °C for 30 min and the DNA cleavage products were precipitated from the solution by adding 3 mL of 2% LiClO₄ in acetone. The solution was centrifuged at 22,000 g for 30 min at 4 °C. The supernatant was discarded and the pellet was washed with 0.5 mL 95% ethanol at 4 °C. The solution was centrifuged again at 22,000 g for 30 min at 4 °C. The pellet was dried in vacuum and was re-dissolved in 20 μL loading buffer.
containing 8 M urea. The samples were analyzed by 20% DNA sequencing gel and was visualized by a phosphorimager. The gel analysis are shown in Figure 2A and is compared to the results obtained by Keck et al. (Figure 2B). Our results reveals very little cleavage at the expected A5 cleavage site, contrast to that observed by Keck et al. We also observed extensive non-specific cleavage products under this assay condition. This observation is possibly due to the random hydroxyl radical damage to DNA, which could result in extensive strand breaks after the alkaline treatment.

Figure 2 DNA sequencing gel (20%) analysis of the cleavage reaction of [5'-32P]-2 with BLM A2. A, the result from our group. B, the result from the Hecht group (Adapted from Figure 8A in Keck et al., J. Am. Chem. Soc. (2001) 123, 8690-8700). The assay conditions were exactly the same for A and B. The sequence of 2 is shown with the proposed cleavage sites (A5, I11, and C31) indicated by the arrows.
The cleavage pattern of [5'-32P]-2 by BLM A2 was also analyzed using the procedure described in Section 2.2, Chapter 2. There are several differences between the two assay conditions. Under our assay conditions, BLM A2 was activated \textit{ex situ} and the reaction mixture contained 0.6 μg/μL calf thymus DNA. After the reaction, no alkaline treatment was carried out. The DNA sequencing gel analysis is shown in Figure 3. Consistent with the observation in Figure 2, our results indicate no ss cleavage at A5 within the limit of the detection.

\textbf{Figure 3.} DNA sequencing gel (20%) analysis of the cleavage reaction of [5'-32P]-2 with BLM A2. The sequence of 2 is shown with the proposed cleavage sites (A5, I11, and C31) indicated by the arrows. The assay conditions are described in Section 2.2, Chapter 2.
Chapter 3. Mechanistic studies on BLM-mediated blunt-ended cleavage
3.1 Introduction

Bleomycins (BLMs) are a family of natural glycopeptides that have potent anti-tumor activity against a variety of cancers(1, 2). The cytotoxicity of BLMs is proposed to be related to their ability to cause both single-stranded (ss) and double-stranded (ds) DNA damage in the presence of required cofactors (Fe(II), O₂ and reductant)(3-7). As discussed in detail in Chapter 1 (Section 1.3), BLM initiates DNA damage by generating a 4′-radical intermediate(8), which leads to two types of ss-DNA lesions depending on the availability of O₂(Figure 3.1). Although in vitro studies have showed that ssDNA cleavage by BLMs occur much more frequently than dsDNA cleavage(9-13), the ds DNA cleavage has long been believed to be the major source of BLMs’ cytotoxicity because of the severe cellular consequences associated with ds DNA breaks(7, 14). These consequences include mutagenesis, necrosis, and apoptosis. The mechanism of BLM-mediated ds-DNA cleavage has been the focus of our research.

BLM mediates ss- and ds-DNA cleavage in a sequence selective fashion. Ss-cleavage occurs predominantly at pyrimidines 3′-to a guanine (5′-GC-3′ ~ 5′-GT-3′ >> 5′-GA-3′, _ denotes the cleavage site)(15). Povirk et al. also proposed a set of sequence-selectivity rules (Figure 1.4, Chapter 1) for ds-cleavage by analyzing 30 ds-cleavage sites in a limited sequence space using native and denaturing DNA gel electrophoresis(16, 17). The rules are that the first cleavage site is a good ss cleavage site. Depending on the local sequence context, the secondary cleavage site can be either staggered to the 5′ end or directly opposite to the first cleavage site (Figure 1.4, Chapter 1).
Figure 3.1 DNA cleavage pathways by BLM.

The dsDNA cleavage is proposed to be effected by a single BLM molecule, which has only one active site (Section 1.4, Chapter 1)(13, 16, 18). A structural model for the mechanism by which a single BLM molecule can mediate two cleavage events has been proposed by our group based on the structure of an activated BLM analog (BLM-Co(III)-OOH) bound to the “hot spot” for ds-cleavage (5'-GTAC-3'/3'-CATG-5', _ indicates the cleavage site)(19, 20). In this model, the bithiazole moiety of BLM intercalates itself 3' to the first cleavage site with the metal binding domain positioned adjacent to (and within H-bonding distance of) the guanine 5' to the cleavage site in the minor groove (Figure 1.5A, Chapter 1). To initiate the second cleavage event, BLM rotates around the bond connecting the two thiazole rings to re-position itself to the second cleavage site, which is 18 Å away (Figure 1.6, Chapter 1). The rotation is proposed to be triggered by the lesion
(3'PG/5'P)(20) or its precursor(21) on the first strand.

The reorganization model is straightforward in explaining the staggered-end cleavage product generated by activated BLM, in which the symmetry of the primary and secondary cleavage sites requires little re-adjustment of the linker region(19, 20). However, BLM also mediates blunt-ended ds-cleavage, in which the second cleavage site is directly opposite the first cleavage site(16) (Figure 1.4, Chapter 1). We have proposed that the reorganization model can explain both types of ds-cleavage. Computer modeling by docking BLM-Co(III)-OOH onto duplex DNA containing a 5'-GGP^P C-3' site (Figure 3.2A), a potential blunt-ended cleavage site whose structure has been solved by 2D NMR(22), suggests that with a proper adjustment of the linker region, BLM can be re-positioned to the second cleavage site that is directly opposite to the first cleavage site as required by the blunt-ended cleavage (Figure 3.2B).

Our model would be more appealing if experimental evidence for the reorganization required for the blunt-ended cleavage process could be observed. We have previously provided evidence for our model for ds-cleavage by examining the structure of the 3’PG/5’P lesion generated within a defined DNA duplex using 2D NMR methods and titrating it with BLM-Co(III)-OOH (Chapter 1, Section 1.6)(23). We have now applied this methodology in an effort to obtain structural information on the blunt-ended cleavage. In this chapter, we first identified a “hot spot”, 5’-GTCA-3’/3’-CAGT-5’ (indicates the cleavage site), for blunt-ended cleavage using the internally [32P]-labeled hairpin DNA technology(24). Duplex DNAs containing this intact “hot spot” and the 5’-G^P C-3’ lesioned “hot spot” were synthesized and their interactions with BLM-Co(III)-OOH were studied using 1D and 2D NMR methods. Titration of the duplex DNA
containing the 5'-GTCA-3' sequence with BLM resulted in formation of multiple complexes as revealed by 1D NMR. Analysis of the major species (~60%) by 2D NMR suggests that BLM binds to the predicted ss-cleavage site by partial intercalation, consistent with our model.

**Figure 3.2** Molecular modeling of BLM-Co(III)-OOH bound to duplex DNA containing a 3'PG/5'P gap in a blunt-ended ds-cleavage event. **A**, A cartoon rendition showing the reorganization in blunt-ended cleavage and the sequence of the 13mer duplex DNA containing a 3'PG/5'P lesion used in the computer modeling. **B**, the computer-modeled structure of the complex in Figure 3.2A (right) after reorganization. The yellow arrow indicates the OOH group and the white arrow indicates the 4'-H atom (white sphere) of T19 to be abstracted.
Titration the $5'-G^pC-A-3'$ lesion with BLM-Co(III)-OOH, however, resulted in a mixture of complexes in fast exchange on the NMR time scale, making chemical shift assignments and structural calculation not possible. These results reflect the complexity in studying the blunt-ended cleavage event by NMR, which requires tight binding and homogeneous species in solution. Nevertheless, this chapter presents the first mechanistic study on the blunt-ended cleavage mediated by BLM. Searching for more favorable blunt-ended cleavage sites in a larger sequence space and identifying alternative structural triggers for reorganization are required for future structural studies.

3.2 Experimental section

Materials The oligonucleotides were purchased from Invitrogen Inc. or from Midland Certified Reagents Co. BLM A2 was purchased from CalBiochem as a metal-free form. DNA-sequencing gel solutions were purchased from National Diagnostics Co. BLM A2-Co(III)OOH was prepared and purified following the published procedure(25). The concentration of BLM-Co(III)-OOH was determined by $A_{290} (\varepsilon_{290} = 2.1 \times 10^4 \text{ M}^{-1}\text{cm}^{-1})$ (25). D$_2$O (99.96%) was purchased from Cambridge Isotopes Lab. The semi-preparative reverse-phase C18 column (EconoSil, 250×10 mm, 10 μm) was purchased from Alltech Co. Other chemicals were purchased from Sigma-Aldrich Co.

Determination of ss to ds cleavage ratio using hairpin DNA technology. Five hairpin DNA oligomers containing blunt-ended cleavage sites (Figure 3.3) were internally $[^32\text{P}]$ labeled (dG in bold, Figure 3.3) following previously published procedures(24). The ss to ds cleavage ratios by BLM at these sites were determined as described in Chapter 2.
Figure 3.3 Sequences of hairpin DNA containing internal blunt-end cleavage sites. The predicted ds-cleavage sites are indicated by large arrows and other ss-cleavage sites are indicated by small arrows. The [32P] labeled dG is in bold. * denotes the putative secondary cleavage site in the 5'-GTTT-3' sequence.
Purification of 5'-d(CCAAAGTCATGGG)-3’ (1) and 5'-d(CCATGACTTTGG)-3’ (2).

Compound 1 and 2 were synthesized on a 10 μmol scale by Invitrogen Inc. with the 5’-dimethoxytrityl (DMT) group retained to facilitate the purification by reverse phase HPLC (RP-HPLC). The crude oligonucleotides were purified by RP-HPLC using a semipreparative C18 column (Econosil C18,10×250 mm, 10 μm from Alltech). Solution A contained 0.1 M ammonium bicarbonate, pH 8.0. Solution B contained 100% acetonitrile. The elution was performed with a linear gradient from 10 to 40% B over 30 min and a flow rate 3 mL/min. The fraction containing 5’-DMT oligonucleotides (retention time: 25 min) were collected and lyophilized to dryness. The 5’-DMT group was cleaved from the oligonucleotides by dissolving the pellet in 5 mL 80% acetic acid at 25 °C for 1 h. DMT was removed from the reaction mixture by extracting the aqueous reaction solution twice with 5 mL ether. The organic layer containing DMT was discarded and the aqueous layer was lyophilized to dryness. The purified oligonucleotides were then re-dissolved in 1 mL deionized water. The yield of 1 (ε_{260}: 129.7 mM⁻¹cm⁻¹) was 4.2 μmol and the yield of 2 (ε_{260}: 118.2 mM⁻¹cm⁻¹) was 5.7 μmol.

NMR sample preparation of duplex (3) composed of 1 and 2. Equal amounts (1.0 μmol) of 1 and 2 were mixed in 1 mL solution containing 20 mM HEPES and 1 mM EDTA, pH 7.5 (HE buffer) and were annealed by heating to 95 °C, followed by cooling to 20 °C over 1 h. The resulting duplex (3) was further purified by the anion-exchange HPLC using a DNAPac 100 semi-prep column (Dionex, 10 × 250 mm). The elution involved solution A (25 mM HEPES, pH 8.0) and solution B (200 mM NaClO₄, 25 mM HEPES, pH 8.0) and a linear gradient from 10 to 90% B over 20 min with a flow rate of 3
mL/min. The fractions containing 3 (retention time: 16.5 min, yield: 91%) were collected and dialyzed using the Spectra/Por 7 dialysis membrane (molecular weight cutoff: 1000 Da) against 4 \times 2 \text{ L} deionized water changed periodically over 2 days before being lyophilized to dryness. The sample was then exchanged into D$_2$O by lyophilization from 2 \times 1 \text{ mL} 99.96\% D$_2$O and was re-dissolved in 630 \mu L 99.96\% D$_2$O. The pH of the solution was adjusted by adding 70 \mu L 0.5 M sodium phosphate, pH 6.8 in 99.96\% D$_2$O. The final NMR sample (700 \mu L) contained 1.27 mM 3 in 50 mM sodium phosphate, pH 6.8.

*Titration of 3 with BLM-Co(III)-OOH monitored by 1D NMR spectroscopy.* For the titration of 3, BLM-Co(III)-OOH (10 mM in D$_2$O) was added to 3 (1.27 mM, 700 \mu L) in small aliquots. After each addition, a 1D NMR spectrum was acquired as described below.

*NMR Experiments.* All NMR experiments were performed in D$_2$O at 20 °C on a custom-built 750 MHz or 591 MHz spectrometer at the Francis Bitter Magnet Laboratory. The raw data were transferred to a Silicon Graphic workstation and were processed using Felix 2001 (Acceryls Inc.). Proton chemical shifts were referenced to an internal standard, sodium 3-(trimethylsilyl)-1-propanesulfonate (TSP) at 0.00 ppm. One-dimensional NMR experiments were performed on a 591 MHz NMR spectrometer with a spectrum width of 7022.5 Hz. The 1D spectra were acquired in 32,768 complex points with 128 scans. The raw 1D data were exported to Felix 2001 and were processed with an exponential window function (line broadening of 2 Hz). For 2D NMR experiments, NOESY (100 and 200 ms mixing times) and TOCSY (60 and 110 ms mixing times) spectra were recorded on either the 750 MHz or 591 MHz spectrometer. For NOESY and
TOCSY experiments, datasets of 4096 (t2) × 512 (t1) complex points were acquired with
the spectral width of 8012.4 Hz (750 MHz) or 7022.5 Hz (591 MHz) in both dimensions
and 32 scans per t1 increment. A presaturation pulse (2 s) was applied during the recycle
delay period to suppress solvent signals in all NMR experiments. In all datasets, t1
dimensions were linear-predicted to 4096 points and the data were processed with an
exponential window function.

*Synthesis of 5 (Figure 3.9).* The precursor (5, Figure 3.9) to the 3′-PG/5′-P lesioned
duplex DNA (4) with hexaethyleneglycol linkers on each end was synthesized by
Midland Certified Reagents Co. on a 1 µmol scale. Compound 5 was purified by anion-
exchange HPLC as described below (retention time: 20.1 min). The yield of 5 was 0.535
µmol (53.5 %). Compound 5 was analyzed by the MALDI-TOF MS ([M-H] : 8585.7;
calculated: 8589.8).

*Oxidation of 5 to 6* Compound 5 was converted to the 3′-PG/5′-P lesioned duplex DNA
following previously published procedures(22, 26). The reaction mixture in a final
volume of 1 mL contained 535 µM 5 and 20 mM NaIO4 and was incubated at 0 °C. The
reaction of 5 was monitored by anion-exchange HPLC using a Dionex DNA Pac-100
analytical ion exchange column (4 × 250 mm). Buffer A contained 25 mM sodium
acetate in 90% water/10% acetonitrile, pH 6.0. Buffer B contained 1 M sodium chloride
in Buffer A. The starting material (5) and the product (6, Figure 3.9) were separated by a
linear gradient from 40 to 55% B over 30 min with a flow rate of 1mL/min). Compound 5
and 6 eluted at 20.1 and 20.6 min, respectively. The reaction was completed in 2.5 h and
was quenched at 0 °C by addition of 0.25 mL of 0.2 M methionine in H2O. The pH of the
reaction mixture was adjusted to 5.0 by adding 150 µL 0.2M NaH2PO4 in H2O. To this
reaction mixture was added 25 μL of 0.2 M freshly prepared sodium hypochlorite and the reaction mixture was incubated at 25 °C. The reaction was monitored by anion-exchange HPLC as described above. Compound 6 and the product, 4, eluted at 20.6 and 24.2 min respectively. The reaction was completed in 2 h and was quenched by addition of 25 μL of 0.2 M Na₂SO₃ in H₂O. The pH of the reaction mixture was adjusted to 6.0 by adding 100 μL of 0.1 M NaOH. The final product, 4, was purified from the reaction mixture by the anion-exchange HPLC as described above with a retention time of 24.2 min (yield: 0.324 μmol, 65% starting from 5). The product was collected and dialyzed using the Spectra/Por 7 dialysis membrane (molecular weight cutoff: 1000 Da) against 4 × 2 L deionized H₂O changed periodically over 2 d before being lyophilized to dryness. The sample was then exchanged into D₂O by lyophilization in 2 × 1 mL 99.96% D₂O. The final NMR sample (300 μL) contained 0.5 mM 2 in 50 mM sodium phosphate, pH 6.8. The MALDI-TOF MS analysis gave the m/s ([M-H]⁻) of 8565.8 (Calculated: 8567.3).

Characterization of 4 by enzymatic digestion. Compound 4 (5.6 nmol) was incubated in a final reaction mixture of 100 μL containing 0.02 U of snake venom phosphodiesterase I (PDE-I, Sigma, Catalogue No. P4631, 1 U is defined as the amount of enzyme to hydrolyze 1.0 μmol bis(ρ-nitrophenyl)phosphate per min at 37 °C), 100 U DNase I (Roche, Catalogue No. 10776785001, 1 U is defined as the amount of enzyme to effect an absorbance increase of 0.001 per min at 260 nm at 25 °C in 1 mL solution containing 40 μg/mL calf thymus DNA) in 25 mM Tris-HCl, 10 mM MgCl₂, pH 7.4. The reaction mixture was incubated at 37 °C for 2 h and the reaction was quenched by heating the solution at 95 °C for 5 min. The protein was removed by centrifugation and 20 U calf intestine alkaline phosphatase (Roche, Catalogue No. 1097075, 1 U is defined as the
amount of enzyme to hydrolyze 1 μmol 4-nitrophenyl phosphate in 1 min at 37 °C) was added to the solution. The reaction mixture was incubated at 37 °C for 30 min and the digestion mixture was analyzed by ion-pair RP-HPLC using an Alltech Adsorbosphere Nucleotide-Nucleoside column (4.6 x 250 mm). Buffer A contained 60 mM KH₂PO₄, 5 mM tetrabutylammonium phosphate, pH 4.3 and Buffer B contained 5 mM tetrabutylammonium phosphate in methanol. The elution was performed with a linear gradient from 0 to 50% B over 20 min and 50% to 90% B from 20 to 30 min (flow rate 1 mL/min). Compound, retention time: dC, 11.64 min; dG, 14.40 min; T, 14.82 min; dA 16.70 min; dG-phosphoglycolate, 21.71 min; dG-hexaethyleneglycol, 22.84 min. The peaks were quantified by comparison to four external nucleoside standards (dC, dG, T, and dA).

**Titration of 4 with BLM-Co(III)-OOH monitored by 1D NMR spectroscopy.** For the titration of 4, BLM-Co(III)-OOH(4 mM in D₂O) was added to 4 (0.5 mM, 300 μL) in small aliquots. After each addition, a 1D NMR spectrum was acquired as described above.

### 3.3 Results and Discussion

**Choice of sequences for identification of a “hot spot” for blunt-ended cleavage.** To test our model for BLM-mediated blunt-ended cleavage, we sought to identify a DNA sequence that is optimal (high ds:ss ratio) for blunt-ended cleavage by BLM. Previous studies by Povirk’s group have generated a set of rules for blunt-ended cleavage mediated by BLM(16, 17). The first cleavage event occurs at a hot spot for ss cleavage and requires a pyrimidine 3’ to the pyrimidine cleavage site, e.g., 5’-G-Py-Py-3’ (Py denotes the first
cleavage site). The second cleavage site is a purine directly opposite to the first cleavage site and is usually not a good ss cleavage site for BLM. The 5'-CGCC-3' sequence was suggested by Povirk(16, 21) to be a potential hot spot for blunt-ended cleavage and this sequence was initially investigated using our hairpin DNA technology(24), which allows quantitative assessment of the ds:ss cleavage ratio (Figure 3.3). However, in general, the 5'-G-C-Py-3' sequences are not good candidates for studying the second cleavage event because on the second strand, 3'-C-G-Pu-5', there are two potential cleavage sites. One is the G opposite the first cleavage site and the second is the C 3' to the G as a ss cleavage site. Therefore, in addition to the 5'-CGCC-3' sequence suggested by Povirk, we focused most of our attention to the 5'-G-T-Py-3' sequences. Five such sequences (5'-CGCC-3', 5'-G-T-C-A-3', 5'-G-T-C-C-3', 5'-G-T-T-A-3', 5'-G-T-T-T-3') were synthesized and characterized, and the ss : ds cleavage ratios were determined.

Table 3.1 Summary of ss to ds cleavage ratios in different sequence contexts that lead to a blunt-ended products. The underlined base denotes the cleavage site.

<table>
<thead>
<tr>
<th>Sequence</th>
<th>5'-CGCC-3'</th>
<th>5'-GTCA-3'</th>
<th>5'-GTCC-3'</th>
<th>5'-GTTA-3'</th>
<th>5'-GTTT-3'</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3'-GCCG-5'</td>
<td>3'-CAGT-5'</td>
<td>3'-CAAGG-5'</td>
<td>3'-CAAT-5'</td>
<td>3'-CAAA-5'</td>
</tr>
<tr>
<td>Ss:ds ratio</td>
<td>n.o.a</td>
<td>4.0</td>
<td>5.2</td>
<td>8.9</td>
<td>7.8</td>
</tr>
</tbody>
</table>

(a n.o. not observed)

_Determination of ss to ds cleavage ratios._ The internally [32P]-labeled hairpin DNAs were generated as previously described(24). The sequences of hairpin DNAs investigated containing putative blunt-ended cleavage sites are listed in Figure 3.3. The results of DNA sequencing gel analysis of BLM-mediated cleavage on these internally [32P] labeled
hairpins are shown in Figure 3.4 to Figure 3.8 and the ss to ds cleavage ratios are summarized in Table 3.1. One sequence, 5'-GTCA-3', has a ss to ds cleavage ratio of 1 to 4.0, which is close to the ss:ds ratio for 5'-GTAC-3' (3.4 to 1), a "hot spot" for the staggered-ended cleavage that we previously studied. Therefore, the 5'-GTCA-3' sequence was used in further studies on the blunt-ended cleavage process. The second cleavage site for this "hot spot" (5'-GTCA-3') is 5'-TGAC-3'. It has been observed that 5'-G-A-3' can also be a weak ss cleavage site for BLM(15, 16, 24) with the guanine 5' to this site potentially mediating H-bonding interactions with the pyrimidine group of BLM (Figure 1.5B, Chapter 1). This result is consistent with our interpretation of previous studies on the staggered-ended cleavage "hot spot" (5'-GTAC-3'/5'-GTAC-3') suggesting that having a good ss-cleavage site on each strand promotes the ds-DNA cleavage by BLM(24).

Figure 3.4 Cleavage of internal [32P]-GTCA hairpin by BLM.
Figure 3.5 Cleavage of internal $^{32}$P-GTCC hairpin by BLM.
Figure 3.6 Cleavage of internal $[^{32}P]$-CGCC hairpin by BLM.
Figure 3.7 Cleavage of internal $[^{32}P]$-GT TA hairpin by BLM.
Figure 3.8 Cleavage of internal [\(^{32}\)P]-GT TT hairpin by BLM.

Choice of duplex DNA (3) containing the blunt-ended cleavage “hot spot” (5'-GTCA-3').

The 5'-GTCA-3' site was incorporated into a 13mer duplex DNA (3, Figure 3.9). The sequence flanking the “hot spot” (5'-CCA AAG TCA TGG G-3') has been used in the previous NMR studies and the chemical shifts information obtained in those studies would facilitate the assignments in this study (22, 27). Compound 3 was prepared by annealing equal amounts of 1 and 2 (Figure 3.9). The purity of 3 was confirmed by anion
exchange HPLC analysis, which showed a single species with a retention time of 13.2 min (Figure 3.10A). Under the same conditions, the ss 13mer 1 has a retention time of 10.4 min (Figure 3.10B). Furthermore, the 1D NMR spectrum of 3 showed sharp peaks in the base proton region, indicating formation of a stable duplex (Figure 3.10C).

**Figure 3.9** Sequence of duplex DNA(3) containing a hot spot for blunt-ended cleavage and its analog containing a 3’-PG/5’-P gap (4). Compound 3 was prepared by annealing 1 (5'-CCAAAGTCATGGG-3') and 2 (5'-CCCATGACTTTGG-3') in a 1:1 ratio. Arrows indicate the proposed primary (large arrow) and secondary (small arrow) cleavage site. Compound 4 was synthesized from its precursor, 5, by two oxidation steps via an intermediate, 6.
Figure 3.10  Analytical anion exchange HPLC analysis of 3 (A) and 1 (B) and 1D NMR spectrum showing the base proton (H8/H6) region of 3 (C).
Sequence-specific cleavage of 3 by activated BLM was studied on [5'-32P]-3 with the [32P] label on either strand. Sequencing gel analysis (Figure 3.11) shows that T7 on the first strand and A7 on the second strand were the predominant cleavage sites, consistent with the blunt-ended cleavage pattern. Several additional cleavage bands are also observed under the assay conditions (indicated by * in Figure 3.11). These minor cleavage sites are either due to hydroxyl radicals or to non-specific cleavage by BLM.

![Figure 3.11 Cleavage of 3 by Fe-BLM. The duplex was 5'-[32P] labeled at either 1 (first strand) or 2 (second strand).* indicates minor cleavage sites.](image)

**Figure 3.11** Cleavage of 3 by Fe-BLM. The duplex was 5'-[32P] labeled at either 1 (first strand) or 2 (second strand).* indicates minor cleavage sites.

**Assignment of proton chemical shifts of 3.** Chemical shift assignments of non-exchangeable protons in 3 were carried out using standard TOCSY and NOESY experiments in order to facilitate the interpretation of future 1D NMR spectra in which 3 is titrated with BLM-Co(III)-OOH. The deoxyribose protons (H1', H2'/2'', and H3') of each residue were assigned using TOCSY experiments. The H4' and H5'/5'' were not
assigned due to signal crowding. The sequential assignments were made using the NOE connectivities between the base protons (H8/H6, 6.8-8.5 ppm) and sugar protons (H1’ or H2’/2’ or H3’). The dispersion of NOE cross peaks and the chemical shift information from previous studies on duplex DNA with similar sequences(22, 27) facilitated the assignments. Two typical regions of the NOESY (200 ms mixing time) spectrum are shown in Figure 3.12 (base protons to methyl protons of T) and Figure 3.13 (base protons to H1’ protons) to demonstrate the quality of the spectrum. The proton chemical shifts of 3 are summarized in Table 3.2.

![Figure 3.12](image)

**Figure 3.12** NOESY (200 ms mixing time) spectrum of 3 showing the connectivity of methyl protons of T to the base proton (H8 or H6).

![Diagram](image)
Figure 3.13 The NOESY (200 ms mixing time) spectrum of 3 showing the NOE connectivity between base (H8/H6) protons and H1'. The sequence of 3 is shown above the spectra. The proposed BLM cleavage site, 5'-GTCA-3', is underlined (the large arrow indicates the primary cleavage site and the small arrow indicates the secondary cleavage site). The NOE walk was carried out on both top and bottom strands.
<table>
<thead>
<tr>
<th></th>
<th>H1'</th>
<th>H2'</th>
<th>H2&quot;</th>
<th>H3'</th>
<th>H8/H6</th>
<th>H5/Me</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>5.90</td>
<td>2.00</td>
<td>2.43</td>
<td>4.62</td>
<td>7.69</td>
<td>5.90</td>
</tr>
<tr>
<td>C2</td>
<td>5.21</td>
<td>1.95</td>
<td>2.21</td>
<td>4.76</td>
<td>7.48</td>
<td>5.65</td>
</tr>
<tr>
<td>A3</td>
<td>5.71</td>
<td>2.71</td>
<td>2.80</td>
<td>5.00</td>
<td>8.19</td>
<td></td>
</tr>
<tr>
<td>A4</td>
<td>5.82</td>
<td>2.58</td>
<td>2.80</td>
<td>5.00</td>
<td>8.05</td>
<td></td>
</tr>
<tr>
<td>A5</td>
<td>5.91</td>
<td>2.58</td>
<td>2.76</td>
<td>4.94</td>
<td>7.93</td>
<td></td>
</tr>
<tr>
<td>G6</td>
<td>5.71</td>
<td>2.30</td>
<td>2.59</td>
<td>4.73</td>
<td>7.30</td>
<td></td>
</tr>
<tr>
<td>T7</td>
<td>5.94</td>
<td>2.02</td>
<td>2.42</td>
<td>4.80</td>
<td>7.11</td>
<td>1.02</td>
</tr>
<tr>
<td>C8</td>
<td>5.53</td>
<td>2.06</td>
<td>2.39</td>
<td>4.81</td>
<td>7.47</td>
<td>5.55</td>
</tr>
<tr>
<td>A9</td>
<td>6.16</td>
<td>2.63</td>
<td>2.86</td>
<td>4.97</td>
<td>8.28</td>
<td></td>
</tr>
<tr>
<td>T10</td>
<td>5.59</td>
<td>1.75</td>
<td>2.13</td>
<td>4.75</td>
<td>7.01</td>
<td>1.37</td>
</tr>
<tr>
<td>G11</td>
<td>5.36</td>
<td>2.29</td>
<td>2.59</td>
<td>4.89</td>
<td>7.75</td>
<td></td>
</tr>
<tr>
<td>G12</td>
<td>5.72</td>
<td>2.52</td>
<td>2.67</td>
<td>4.91</td>
<td>7.66</td>
<td></td>
</tr>
<tr>
<td>T13</td>
<td>5.60</td>
<td>2.28</td>
<td>2.41</td>
<td>4.57</td>
<td>7.68</td>
<td></td>
</tr>
<tr>
<td>C14</td>
<td>5.92</td>
<td>2.14</td>
<td>2.50</td>
<td>4.62</td>
<td>7.76</td>
<td>5.91</td>
</tr>
<tr>
<td>C15</td>
<td>5.66</td>
<td>2.12</td>
<td>2.43</td>
<td>4.81</td>
<td>7.58</td>
<td>5.61</td>
</tr>
<tr>
<td>C16</td>
<td>5.34</td>
<td>2.11</td>
<td>2.36</td>
<td>4.80</td>
<td>7.49</td>
<td>5.65</td>
</tr>
<tr>
<td>A17</td>
<td>6.21</td>
<td>2.66</td>
<td>2.88</td>
<td>4.99</td>
<td>8.30</td>
<td></td>
</tr>
<tr>
<td>T18</td>
<td>5.57</td>
<td>1.77</td>
<td>2.19</td>
<td>n.d.</td>
<td>6.98</td>
<td>1.39</td>
</tr>
<tr>
<td>G19</td>
<td>5.39</td>
<td>2.58</td>
<td>2.64</td>
<td>4.92</td>
<td>7.75</td>
<td></td>
</tr>
<tr>
<td>A20</td>
<td>6.14</td>
<td>2.62</td>
<td>2.83</td>
<td>4.96</td>
<td>8.09</td>
<td></td>
</tr>
<tr>
<td>C21</td>
<td>5.70</td>
<td>1.91</td>
<td>2.44</td>
<td>4.60</td>
<td>7.17</td>
<td>5.11</td>
</tr>
<tr>
<td>T22</td>
<td>5.98</td>
<td>2.11</td>
<td>2.54</td>
<td>4.82</td>
<td>7.34</td>
<td>1.45</td>
</tr>
<tr>
<td>T23</td>
<td>6.08</td>
<td>2.11</td>
<td>2.53</td>
<td>4.85</td>
<td>7.42</td>
<td>1.58</td>
</tr>
<tr>
<td>T24</td>
<td>5.68</td>
<td>1.91</td>
<td>2.26</td>
<td>4.82</td>
<td>7.26</td>
<td>1.66</td>
</tr>
<tr>
<td>G25</td>
<td>5.60</td>
<td>2.62</td>
<td>2.66</td>
<td>4.93</td>
<td>7.82</td>
<td></td>
</tr>
<tr>
<td>G26</td>
<td>6.09</td>
<td>2.30</td>
<td>2.45</td>
<td>n.d.</td>
<td>7.75</td>
<td></td>
</tr>
</tbody>
</table>
Titration of 3 with BLM-Co(III)-OOH. Compound 3 was titrated with BLM-Co(III)-OOH in D₂O and the progress was monitored by 1D NMR spectroscopy. The methyl proton region of the 1D NMR spectra (1.0-2.0 ppm) was particularly useful for monitoring the progress of the titration (Figure 3.14). The disappearance of T7-Me (Line A, Figure 3.14) and T24-Me (Line C, Figure 3.14) of 3 was observed concomitantly with the appearance of a new peak at 1.38 ppm (Line B, Figure 3.14), which was later assigned as T7-methyl protons in the 3•BLM-Co(III)-OOH complex. The large chemical shift change associated with T7 (in 5’-GTCA-3’) methyl protons is consistent with its expected close proximity to BLM-Co(III)-OOH. The titration of 3 with BLM-Co(III)-OOH was continued until the complete disappearance of the peak associated with the T7 methyl protons (Line A, Figure 3.14). At the end of the titration, most of the methyl peaks appeared broader and the spectrum of the 3•BLM-Co(III)-OOH complex was much more complicated than that of 3 before titration (Figure 3.14).

The titration process was also monitored in the region of DNA base protons and aromatic protons of BLM-Co(III)-OOH (6.6-9.5 ppm, Figure 3.15). During the course of titration, several new peaks became apparent outside of the usual chemical shift envelope for DNA base protons (7.0-8.5 ppm) and are labeled by upward arrows in Figure 3.15 (middle). The feature (9.13 ppm) in the downfield region was assigned to the His H2 proton of BLM in the 3•BLM-Co(III)-OOH complex based on previous titration studies. The His H2 signal consists of a major sharp peak and two minor peaks (Figure 3.15), indicating that there were more than one species in slow exchange on the NMR time scale. The formation of the minor complexes was not due to over-titration because they were present during the entire titration process.
Figure 3.14 Titration of 3 with BLM-Co(III)-OOH monitored by 1D NMR. The up-field regions of the spectra are shown. The assignments of methyl proton chemical shifts in 3 are labeled on the top of the spectra. TEA: Triethylammonium. * indicates the existence of the multiple complexes. Positions A, B, and C (indicated by the red dashed line) are used to monitor the progress of the titration (see the discussion in the text).
Figure 3.15 Titration of 3 with BLMA2-Co(III)-OOH monitored by 1D NMR. The downfield regions of the spectra are shown. The assignments of base proton chemical shifts in 3 are labeled on the top of the spectra. Arrows indicate four new peaks that appear during the course of titration. The assignments of bithiazole H5 and H5’ protons are tentative. (His, Hydroxyhistidine, Bit, Bithiazole)
Characterization of the major species of 3•BLM-Co(III)-OOH. TOCSY and NOESY experiments were performed on the final titration mixture containing a major 3•BLM-Co(III)-OOH species and 2 minor components. Cross peaks associated with minor conformations were observed in many regions of the NOESY spectra. The presence of multiple species, severe line-broadening, and peak-overlap complicated the analysis of chemical shifts and NOE cross peaks. Only 60% of the non-exchangeable protons in 3 could be unambiguously assigned for the major 3•BLM-Co(III)-OOH species (Table 3.3). In addition, only 10 BLM-Co(III)-OOH protons could be assigned due to spectral overlap. Figure 3.16 shows two NOE cross peaks associated with His H2 of BLM-Co(III)-OOH in 3•BLM-Co(III)-OOH. These two cross peaks (His H5-Thr Me and His H5-Val αMe) have also been present in previous structures of BLM-Co(III)-OOH•DNA(23, 28), but not in free BLM-Co(III)-OOH(25). In addition, cross peaks that are associated with two other conformations are also present (indicated by arrows in Figure 3.16). Quantitation of the cross peak volumes suggests that the relative abundance of the major conformation to the two other conformations is approximately 60 : 30 : 10.

The presence of multiple conformations and our inability to assign the majority of protons in 3•BLM-Co(III)-OOH preclude further structural studies.

Despite the complexity, we hoped that NOESY experiments might provide some insight into the conformation of the bithiazole moiety in the complex. Our model predicts that the bithiazole group will partially intercalate between T7 and C8 of 3. Previous studies suggest that this partial intercalative binding mode will produce characteristic NOE cross peaks between bithiazole H5’ and protons associated with T7 and between bithiazole H5 and protons associated with G19 and/or A20 on the other strand.
Table 3.3 Proton Chemical shift assignments of 3 in 3•BLM-Co(III)-OOH. Only 60% of the non-exchangeable protons could be unambiguously assigned due to line broadening and signal overlapping.

<table>
<thead>
<tr>
<th>Base</th>
<th>H1'</th>
<th>H2'</th>
<th>H2''</th>
<th>H3'</th>
<th>H8/H6</th>
<th>H5/Me</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>5.97</td>
<td></td>
<td></td>
<td></td>
<td>7.75</td>
<td>5.94</td>
</tr>
<tr>
<td>C2</td>
<td>5.36</td>
<td>2.10</td>
<td>2.38</td>
<td>4.84</td>
<td>7.57</td>
<td>5.71</td>
</tr>
<tr>
<td>A3</td>
<td>5.96</td>
<td>2.77</td>
<td>2.96</td>
<td>5.07</td>
<td>8.25</td>
<td></td>
</tr>
<tr>
<td>A4</td>
<td>5.78</td>
<td>2.65</td>
<td>2.74</td>
<td>5.05</td>
<td>8.16</td>
<td></td>
</tr>
<tr>
<td>A5</td>
<td>6.08</td>
<td>2.45</td>
<td>2.86</td>
<td>5.04</td>
<td>7.90</td>
<td></td>
</tr>
<tr>
<td>G6</td>
<td>5.44</td>
<td>2.23</td>
<td>2.61</td>
<td></td>
<td>7.75</td>
<td></td>
</tr>
<tr>
<td>T7</td>
<td>5.88</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7.21</td>
</tr>
<tr>
<td>C8</td>
<td>5.33</td>
<td>2.18</td>
<td>2.39</td>
<td></td>
<td>7.54</td>
<td>5.70</td>
</tr>
<tr>
<td>A9</td>
<td>6.23</td>
<td>2.67</td>
<td>2.89</td>
<td>5.02</td>
<td>8.34</td>
<td></td>
</tr>
<tr>
<td>T10</td>
<td>5.65</td>
<td></td>
<td></td>
<td></td>
<td>6.75</td>
<td>1.38</td>
</tr>
<tr>
<td>G11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C14</td>
<td>5.98</td>
<td></td>
<td></td>
<td></td>
<td>7.80</td>
<td>5.95</td>
</tr>
<tr>
<td>C15</td>
<td>5.71</td>
<td></td>
<td></td>
<td></td>
<td>7.61</td>
<td>5.65</td>
</tr>
<tr>
<td>C16</td>
<td>5.12</td>
<td></td>
<td></td>
<td></td>
<td>7.58</td>
<td>5.92</td>
</tr>
<tr>
<td>A17</td>
<td>6.31</td>
<td>2.79</td>
<td>2.95</td>
<td></td>
<td>8.35</td>
<td></td>
</tr>
<tr>
<td>T18</td>
<td>5.33</td>
<td>2.18</td>
<td>2.34</td>
<td></td>
<td>7.04</td>
<td>1.43</td>
</tr>
<tr>
<td>G19</td>
<td>5.29</td>
<td>2.53</td>
<td>2.59</td>
<td></td>
<td>7.52</td>
<td></td>
</tr>
<tr>
<td>A20</td>
<td>6.16</td>
<td>2.65</td>
<td>2.92</td>
<td></td>
<td>8.10</td>
<td></td>
</tr>
<tr>
<td>C21</td>
<td>5.41</td>
<td>2.05</td>
<td>2.16</td>
<td></td>
<td>7.21</td>
<td>5.24</td>
</tr>
<tr>
<td>T22</td>
<td>5.99</td>
<td></td>
<td></td>
<td></td>
<td>7.47</td>
<td>1.53</td>
</tr>
<tr>
<td>T23</td>
<td>5.71</td>
<td>1.85</td>
<td>2.44</td>
<td></td>
<td>7.29</td>
<td>1.71</td>
</tr>
<tr>
<td>T24</td>
<td>5.72</td>
<td>1.93</td>
<td>2.29</td>
<td></td>
<td>7.27</td>
<td>1.67</td>
</tr>
<tr>
<td>G25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 3.16 NOEs between His-H2 (9.13 ppm) to Thre-Me (1.21 ppm) and to Val-αMe (0.65 ppm) of BLM-Co(III)-OOH in the complex. (The arrows denote the peaks associated with two other conformations).
Figure 3.17 NOESY (200 ms mixing time) spectrum of 3•BLM-Co(III)-OOH. The sequence of 3 and the predicted binding mode of BLM-Co(III)-OOH are shown above the spectrum. The region of base protons (H6/H8) to T-methyl protons is shown. A, the cross peak between T7-methyl protons and the signal at 6.86 ppm, which is assigned as the bithiazole H5' proton.
Figure 3.18 NOESY spectrum (200 ms mixing time) of 3• BLM-Co(III)-OOH showing the region of base protons (H6/H8) to H2'/H2''. The sequence of 3 and the predicted binding mode of BLM-Co(III)-OOH are shown above the spectrum. Vertical lines are drawn to connect H2' and H2''. A and B, cross peaks associated with the signal at 6.89 ppm.
Figure 3.19 NOESY spectrum (200 ms mixing time) of 3•BLM-Co(III)-OOH showing the region of base protons (H6/H8) to H1'. The sequence of 3 and the predicted binding mode of BLM-Co(III)-OOH are shown above the spectrum. Cross peak A and B, between the signal at 6.89 ppm and G19 H1' or G19 H3'. Crosspeak C, between G19 H8 and G19 H3'. (Bit, bithiazole)
It is challenging to assign H5 and H5’ protons of the bithiazole moiety because they are isolated from other spin systems, i.e., they have no through-bond connectivities with other protons. Previous studies have shown that intercalation of bithiazole between base pairs causes a large up-field shift of bithiazole protons from 7.82 ppm (H5’) and 8.17 ppm (H5) to 6.8-6.9 ppm (20, 28), which is upfield from the chemical shift window of most base protons (7-8.5 ppm). The base proton region in the 1D NMR spectrum of the 3•BLM-Co(III)-OOH reveals three peaks that have chemical shifts less than 7.0 ppm, all of which are associated with formation of the 3•BLM-Co(III)-OOH complex (Figure 3.15, in a red box). The signal at 6.75 ppm was assigned as the base proton (H6) of T10 based on the NOE connectivities in the base-methyl proton region (Figure 3.17). The other two peaks at 6.86 ppm and 6.89 ppm are tentatively assigned as H5’ and H5 of the bithiazole protons. Because not all base protons have been assigned, it is possible but unlikely that these two peaks are associated with the unassigned bases at the termini of 3 (G11, G12, G13, G25, G26). Since these base protons are not adjacent to the proposed binding site, it is unlikely that their chemical shifts are perturbed from 7.7-7.8 ppm in free DNA to 6.8-6.9 ppm in the complex. The intensities of the two peaks at 6.86 ppm and 6.89 ppm (Figure 3.15) also suggest that they are unlikely to be associated with the two minor conformations.

The assignments of the two peaks at 6.89 ppm and 6.86 ppm (Figure 3.15, red box) as the bithiazole H5 and H5’ protons are supported by a number of intermolecular NOE cross peaks associated with the two signals (Table 3.4). The 6.86 ppm signal shows a cross peak with T7 methyl protons (cross peak A in Figure 3.17). In addition, the 6.89 ppm signal shows cross peaks with G19 H2'/H2'' (cross peak A and B in Figure 3.18) and
with G19 H1' and H3' (cross peak A and B in Figure 3.19). It is very unlikely that any of the unassigned base protons (G11, G12, G13, G25, and G26) would produce this NOE pattern. Based on previous studies of oligonucleotide•BLM-Co(III)-OOH(20, 28), this NOE pattern is best explained by an intercalative binding mode of the bithiazole tail between T7-A20 and C8-G19 base pairs.

**Table 3.4** Intermolecular NOE interactions between 3 and two protons associated with signals at 6.89 and 6.86 ppm, which are tentatively assigned as the bithiazole H5 and H5' protons, respectively.

<table>
<thead>
<tr>
<th></th>
<th>Intensity</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.89 ppm (H5)</td>
<td></td>
</tr>
<tr>
<td>G19 H1'</td>
<td>strong</td>
</tr>
<tr>
<td>G19 H2'/H2''</td>
<td>medium</td>
</tr>
<tr>
<td>G19 H3'</td>
<td>weak</td>
</tr>
<tr>
<td>A20 H8</td>
<td>medium</td>
</tr>
<tr>
<td>6.86 ppm (H5')</td>
<td></td>
</tr>
<tr>
<td>T7-Me</td>
<td>medium</td>
</tr>
</tbody>
</table>

*Synthesis and characterization of 4* The titration of BLM-Co(III)-OOH with 3 indicated formation of ~60% of a 1:1 complex. NOE information on this 3•BLM-Co(III)-OOH complex suggests that BLM-Co(III)-OOH binds 3 with the bithiazole tail intercalating between T7 and C8 as predicted by our model. To test whether the 3'PG lesion in 5'-GTCA-3' sequence is sufficient to re-position BLM at the cleavage site on the opposite strand as our model predicts, we prepared compound 4, a duplex DNA containing a 3'PG/5'P gap (Figure 3.9) from the precursor, 5, by a two-step oxidation procedure. The difference in retention times of 4, 5, and 6 (Figure 3.9) on anion exchange HPLC chromatography was used to monitor the course of the reaction (Figure 3.20). The final product of 4 was purified further by anion exchange HPLC. The nucleoside composition of 4 was quantitated by complete enzymatic digestion of 4 with DNase I, PDE-1, and alkaline phosphatase followed by ion-pairing RP-HPLC (Figure 3.21). The relative amount of
each nucleoside was quantitated and was consistent with the expected ratio (dC : dG : T : dA : dG-PG : dG-hexaethyleneglycol, 7.3 : 4.0 : 5.3 : 6.3 : 1.0 : 1.8; calculated: 7 : 4 : 5 : 6 : 1 : 2). The purity and the duplex formation of 4 in the NMR sample were confirmed by the analytical anion-exchange HPLC, in which a single peak was observed, and by the 1D NMR, in which the base proton region showed sharp peaks with good dispersion (Figure 3.22). In addition, in the less crowded region of the 1D NMR spectrum in Figure 3.22 (6.9 ppm – 7.5 ppm, 8.1 ppm – 8.5 ppm), single peaks are observed for each residue, which further confirms the purity of the sample.

Proton chemical shifts assignment of 4 TOCSY and NOESY experiments were used to assign the non-exchangeable protons in 4 to help to interpret the subsequent 1D NMR titration data. The sequential assignment is exemplified in Figure 3.23, which shows the NOESY spectrum in the region of base protons (7.0-8.5 ppm) to H1’ protons (5.0-6.5 ppm). Analysis of sequential NOE connectivity and cross peak intensities indicated that 4 adopts typical B-type conformation and that a single conformation is observed. Sequential NOE connectivity was detected from C1 to G6 and from C8 to G13 and the 3’PG/5’P gap introduced the expected break into the NOE walk (Figure 3.23). Sequential NOE connectivity was also observed on the opposite strand (C14 to G26), especially from G19 to A20 to C21, in which A20 is the base opposite to the 3’PG/5’P gap. This result suggests that the unpaired base, A20, stacks between adjacent bases as in B-form DNA, similar to what we have previously observed for 3’PG/5’P lesions in other sequence contexts(22). The proton chemical shift assignments of 4 are summarized in Table 3.5.
Figure 3.20 Oxidation of 5 to 4 monitored by analytical anion exchange HPLC. A, aliquot from the first oxidation reaction mixture containing 1 nmol 5. (* denotes a small amount of 6 formed before HPLC analysis.) B, aliquot from the second oxidation reaction mixture containing 1 nmol 6. (the arrow denotes a small amount of 4 formed before HPLC analysis.) C, aliquot from the second oxidation reaction mixture containing 1 nmol 4.
Figure 3.21 Ion-pair RP-HPLC analysis of complete enzymatic digestion mixture of 4 by PDE-1, DNase I followed by alkaline phosphatase treatment. (EG)$_6$, hexaethylene glycol; PG, phosphoglycolate.
Figure 3.22 Analytical anion exchange HPLC analysis of the NMR sample of 4 (2 nmol) (top) and the 1D NMR spectrum of 4 showing the base proton region (bottom).
**Figure 3.23** NOESY spectrum (200 ms mixing time) of 4 showing the NOE connectivity between base protons (H8/H6) to H1' protons. The sequence of 4 is shown above the spectra.
Table 3.5 Proton chemical shifts assignment of non-exchangeable protons of 4.

<table>
<thead>
<tr>
<th></th>
<th>H1'</th>
<th>H2'</th>
<th>H2''</th>
<th>H3'</th>
<th>H8/H6</th>
<th>H5/Me</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>6.010</td>
<td>2.323</td>
<td>2.482</td>
<td>4.843</td>
<td>7.752</td>
<td>6.017</td>
</tr>
<tr>
<td>C2</td>
<td>5.288</td>
<td>2.002</td>
<td>2.269</td>
<td>4.811</td>
<td>7.515</td>
<td>5.769</td>
</tr>
<tr>
<td>A3</td>
<td>5.791</td>
<td>2.817</td>
<td>2.881</td>
<td>5.073</td>
<td>8.286</td>
<td></td>
</tr>
<tr>
<td>A4</td>
<td>5.819</td>
<td>2.635</td>
<td>2.812</td>
<td>5.066</td>
<td>8.154</td>
<td></td>
</tr>
<tr>
<td>A5</td>
<td>5.957</td>
<td>2.488</td>
<td>2.776</td>
<td>5.006</td>
<td>7.996</td>
<td></td>
</tr>
<tr>
<td>G6</td>
<td>5.826</td>
<td>2.460</td>
<td>2.502</td>
<td>4.963</td>
<td>7.666</td>
<td></td>
</tr>
<tr>
<td>PG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C8</td>
<td>5.822</td>
<td>2.103</td>
<td>2.484</td>
<td>4.880</td>
<td>7.824</td>
<td>6.019</td>
</tr>
<tr>
<td>A9</td>
<td>6.279</td>
<td>2.786</td>
<td>2.955</td>
<td>5.050</td>
<td>8.432</td>
<td></td>
</tr>
<tr>
<td>T10</td>
<td>5.674</td>
<td>1.869</td>
<td>2.237</td>
<td>4.834</td>
<td>7.148</td>
<td>1.468</td>
</tr>
<tr>
<td>G11</td>
<td>5.495</td>
<td>2.632</td>
<td>2.673</td>
<td>4.970</td>
<td>7.822</td>
<td></td>
</tr>
<tr>
<td>G12</td>
<td>5.645</td>
<td>2.539</td>
<td>2.644</td>
<td>5.000</td>
<td>7.722</td>
<td></td>
</tr>
<tr>
<td>G13</td>
<td>6.129</td>
<td>2.699</td>
<td>2.591</td>
<td>4.993</td>
<td>7.928</td>
<td></td>
</tr>
<tr>
<td>C14</td>
<td>6.054</td>
<td>2.382</td>
<td>2.523</td>
<td>4.866</td>
<td>7.786</td>
<td>6.017</td>
</tr>
<tr>
<td>C15</td>
<td>6.004</td>
<td>2.198</td>
<td>2.479</td>
<td>4.887</td>
<td>7.572</td>
<td>5.732</td>
</tr>
<tr>
<td>C16</td>
<td>5.429</td>
<td>2.191</td>
<td>2.430</td>
<td>4.878</td>
<td>7.579</td>
<td>5.735</td>
</tr>
<tr>
<td>A17</td>
<td>6.298</td>
<td>2.757</td>
<td>2.935</td>
<td>5.059</td>
<td>8.387</td>
<td></td>
</tr>
<tr>
<td>T18</td>
<td>5.707</td>
<td>1.796</td>
<td>2.212</td>
<td>4.820</td>
<td>7.074</td>
<td>1.547</td>
</tr>
<tr>
<td>G19</td>
<td>5.707</td>
<td>2.456</td>
<td>2.503</td>
<td>4.921</td>
<td>7.705</td>
<td></td>
</tr>
<tr>
<td>A20</td>
<td>6.144</td>
<td>2.628</td>
<td>2.667</td>
<td>4.919</td>
<td>8.098</td>
<td></td>
</tr>
<tr>
<td>C21</td>
<td>5.834</td>
<td>2.106</td>
<td>2.492</td>
<td>4.869</td>
<td>7.574</td>
<td>5.757</td>
</tr>
<tr>
<td>T22</td>
<td>6.139</td>
<td>2.273</td>
<td>2.629</td>
<td>4.905</td>
<td>7.540</td>
<td>1.680</td>
</tr>
<tr>
<td>T23</td>
<td>6.160</td>
<td>2.182</td>
<td>2.641</td>
<td>4.915</td>
<td>7.538</td>
<td>1.739</td>
</tr>
<tr>
<td>T24</td>
<td>5.735</td>
<td>1.950</td>
<td>2.260</td>
<td>4.878</td>
<td>7.350</td>
<td>1.776</td>
</tr>
<tr>
<td>G25</td>
<td>5.503</td>
<td>2.654</td>
<td>2.681</td>
<td>4.880</td>
<td>7.879</td>
<td></td>
</tr>
<tr>
<td>G26</td>
<td>6.171</td>
<td>2.723</td>
<td>2.632</td>
<td>5.004</td>
<td>7.980</td>
<td></td>
</tr>
</tbody>
</table>

Titration of 4 with BLM-Co(III)-OOH monitored by 1D NMR. Titration of 4 with BLM-Co(III)-OOH did not lead to an apparent formation of 1:1 complex. Figure 3.24 shows the downfield region of 1D NMR spectra containing base protons of DNA and aromatic protons of BLM-Co(III)-OOH (6.5-9.5 ppm). During the course of titration, the 1D NMR spectra showed evidence of several conformations without a predominant one. This is apparent in the region around 9.2 ppm, in which the chemical shift of His H2 proton of BLM-Co(III)-OOH in the DNA•BLM-Co(III)-OOH complex is located. A sharp peak and two broad peaks are apparent (Figure 3.24). Another region of interest is between 6.5
and 7 ppm, where the two bithiazole protons (Bit H5 and Bit H5’) are expected to appear upon complex formation with DNA. Several broad features are observed in contrast to the titration of BLM-Co(III)-OOH with 3 (Figure 3.15). None of these features could be assigned to specific protons due to lack of NOE interactions with assignable protons from the 2D NMR experiments.

Previous studies by Oppenheimer et al. (29) and recent studies by our group (23) suggest that lesioned DNA could increase binding affinity of BLM-Co(III)-OOH relative to a similar intact sequence. If the 3’PG/5’P lesion made a tighter binding to BLM-Co(III)-OOH, then the titration of 4 by BLM-Co(III)-OOH would have worked well. Compared to the 5’-GPG pAC-3’/5’-GTAC-3’ sequence that we used previously in titration studies to obtain a 1:1 complex (23), the second cleavage site in the blunt-ended cleavage hot spot, 5’-GPG pCA-3’/5’-GAATC, is a weaker ss cleavage site. The ss to ds ratio measured for 5’-GTCA-3’, 4.0 to 1, indicates that the cleavage at the second strand is less frequent than for the 5’-GTAC-3’ sequence (ss to ds ratio 3.4 to 1). In addition, our ds-DNA cleavage model predicts that the observed ss to ds cleavage ratio reflects the relative rates of re-organization versus dissociation from DNA. Thus, the blunt-ended cleavage with a pyrimidine 3’ to the 5’-G-Py-3’ cleavage site is inherently a rarer event in comparison to the staggered ds cleavage sites. Our data reveals that the affinity of between 4 and BLM-Co(III)-OOH is not strong enough to form a 1:1 complex on the NMR time scale. Thus, no structural model is accessible. The “hot spot” for blunt-ended cleavage in this study was chosen from a very limited sequence space, it is possible that there are other hot spots that would be superior for structural studies. New methods to screen for ds-cleavage “hot spots”, such as those recently developed by Boger et al. (30),
would allow the analysis of the robustness of Povirk’s rules for ds-cleavage and may
provide an oligonucleotide sequence amenable to the method described in this chapter.

Figure 3.24 Titration of 4 with BLM-co(III)-OOH monitored by 1D NMR spectroscopy. The regions containing the base protons are shown below. * indicates the existence of multiple conformations
3.4 Conclusion

A “hot spot” for blunt-ended ds-cleavage by BLM (5'-GTCA-3'/3'-CAGT-5') was identified by the internal [32P]-hairpin DNA technology and was used in an effort to obtain structural insight into the mechanism of the blunt-ended ds-DNA cleavage that is proposed to be mediated by a single BLM molecule. Titration of 3 with BLM-Co(III)-OOH resulted in formation of multiple complexes as revealed by 1D NMR methods. Analysis of the NOE pattern of the major species suggests that BLM-Co(III)-OOH binds to the predicted cleavage site by partial intercalation. Titration of 4 with BLM-Co(III)-OOH also resulted in a mixture of complexes. However, in this case, the complexes were in fast exchange on the NMR time scale, which precluded structural determination. We propose that the problems encountered in obtaining a stable 1:1 complex of DNA•BLM-Co(III)-OOH are related to the low frequency of the ds-cleavage events relative to the ss-cleavage. The minor amounts of “complex of interest” remain challenging to characterize. Identifying more favorable blunt-ended cleavage sites by exploring a larger sequence space could benefit future mechanistic studies on the blunt-ended ds-cleavage by BLM.

3.5 References


oligonucleotides containing an internal 3'-phosphoglycolate, 5'-phosphate gapped lesion. *Nucleic Acids Res* 30, 5497-508.


Chapter 4. Synthesis and Characterization of duplex DNA containing a 4′-oxidized abasic site
4.1 Introduction

Bleomycins (BLMs) are natural products used clinically in combination with other drugs in the treatment of a variety of cancers(1). The cytotoxicity of the BLMs is thought to be related to their ability to bind to duplex DNA and generate both single-stranded (ss) and double-stranded (ds) DNA lesions via oxidative damage to the deoxyribose backbone(2, 3). Cleavage of DNA requires the presence of both ferrous iron and O$_2$, the cofactors thought to be required for BLM’s activity in vivo. The difficulty in repairing the ds-lesions has suggested that they are predominantly responsible for the therapeutic efficacy of the BLMs(4, 5). The sugar damage generated by BLMs is initiated by 4'-hydrogen atom abstraction from a pyrimidine 3' to a guanine(6). The resulting 4'-radical can lead to two types of damage depending on the availability of O$_2$ (Figure 4.1). Under anaerobic conditions, a 4'-oxidized abasic site is generated which remains part of an intact DNA strand(7, 8). In the presence of O$_2$, a 4'-peroxy radical is generated which eventually leads to a gapped 3'-phosphoglycolate/5'-phosphate lesion (3'-PG/5'-P) and elimination of a pyrimidine propenal (base propenal, Figure 4.1)(9). Both types of lesions are also generated by ionizing radiation and by other natural products such as the enediynes(10).

DNA lesions result from altered nucleic acid bases as well as deoxyribose sugars. DNA lesions generated by BLM are specifically localized to the sugars (Figure 4.1). If the lesions are left unrepaired, they may give rise to mutations and ultimately disease(11). Most of the work in this area has focused on base lesions(12-14). However, estimations based on radiolysis of DNA suggest that 10 to 20% of the damage is associated with the deoxyribose backbone and subsequently, frank or indirect DNA strand cleavage(15, 16). The structures of a number of oxidatively damaged sugars resulting from ionizing radiation(17, 18), chemical oxidants(19), and a variety of antitumor agents(4, 20) have been characterized. The recent development of more
sensitive analytical and mass spectrometric methods suggest that additional sugar damage will likely be identified, a requirement to understand the physiological consequences of the lesions and the mechanisms of their repair(21).

![BLM-mediated DNA cleavage pathways](image)

**Figure 4.1** BLM-mediated DNA cleavage pathways.

Our interest in the mechanism of cytotoxicity of the BLMs has recently led us to develop synthetic methods to generate gapped 3'-PG/5'-P lesions and to study their structures by 2D NMR methods(22). As noted in Figure 4.1, BLM also generates 4'-oxidized abasic sites. However, synthesis of this type of lesion has been challenging due to their chemical instability. In this chapter, a convenient synthesis of the 4'-oxidized abasic site lesion has been developed using a combination of synthetic and enzymatic methods (Figure 4.2), and the characterization of its stability under physiological conditions is reported. Furthermore, this synthetic method has

---

153
been used to incorporate a 4'-oxidized abasic site into a 13-mer duplex DNA on a μmol scale that is suitable for structural studies (Chapter 5) by 2D NMR methods. Our ability to generate these lesions in any sequence context and knowledge of their structure should facilitate our understanding of the mechanism of recognition by DNA repair enzymes.

![Synthesis of a 4'-oxidized abasic site lesion from 4'-dNTP using HIV-1 RT and UDG. The sequences of 3, 4, 5, 6 and 7 are listed in Table 4.1. N: 4'-dU.](image)

**Figure 4.2** Synthesis of a 4'-oxidized abasic site lesion from 4'-dUTP using HIV-1 RT and UDG. The sequences of 3, 4, 5, 6 and 7 are listed in Table 4.1. N: 4'-dU.

### 4.2 Experimental Section

**Materials** The oligonucleotides were synthesized by the MIT Biopolymers Laboratory or by Invitrogen Inc. The following extinction coefficients were calculated as reported (23) (Compound, ε₂₆₀): 1, 124.3 mM⁻¹cm⁻¹; 2, 132.5 mM⁻¹cm⁻¹; 3, 144.8 mM⁻¹cm⁻¹; 4, 274.0 mM⁻¹cm⁻¹; 5, 333 mM⁻¹cm⁻¹; 6, 210 mM⁻¹cm⁻¹; 7, 201 mM⁻¹cm⁻¹; 8 or 9, 242 mM⁻¹cm⁻¹; 12, 62 mM⁻¹cm⁻¹; 13, 115.9 mM⁻¹cm⁻¹; 14, 184.5 mM⁻¹cm⁻¹; 15, 129.3 mM⁻¹cm⁻¹; 16, 121.2 mM⁻¹cm⁻¹; 17, 174.7 mM⁻¹cm⁻¹ (Table 4.1 and 4.2). N,O-bis(trimethylsilyl)trifluoroacetamide (BSTFA),
trimethylchlorosilane (TMCS), and tris-(2-carboxyethyl)-phosphine (TCEP) were purchased from Pierce. All other chemicals were purchased from Sigma-Aldrich Co.

**Table 4.1** Sequences of the oligonucleotides used for developing synthetic methodology.

<table>
<thead>
<tr>
<th>Sequence information *</th>
</tr>
</thead>
<tbody>
<tr>
<td>1  5’-AAA AXA AAA-3’</td>
</tr>
<tr>
<td>2  5’-AAA AUA AAA-3’</td>
</tr>
<tr>
<td>3  5’-CTG AGC TCC AAA G-3’</td>
</tr>
<tr>
<td>4  5’-CTG CCG GCC CGG TAC TTT GGA GCT CAG-3’</td>
</tr>
<tr>
<td>5  5’-CTG AGC TCC AAA GNA CCG GGC CGG CAG-3’</td>
</tr>
<tr>
<td>6  5’-CTG AGC TCC AAA GNA CCG GG-3’</td>
</tr>
<tr>
<td>7  5’-CTG AGC TCC AAA GYA CCG GG-3’</td>
</tr>
<tr>
<td>8  5’-CTG AGC TCC AAA GYA CCG GG-3’</td>
</tr>
<tr>
<td>9  5’-CTG AGC TCC AAA GXA CCG GG-3’</td>
</tr>
</tbody>
</table>

* X: a normal abasic site. N: 4’-N₃-dU. Y: a 4’-oxidized abasic site. The underlined sequence denotes the NgoM IV recognition sequence and ↓ indicates the cleavage site.
Table 4.2 Sequence information of oligonucleotides used for preparing the NMR samples containing a 4'-oxidized abasic site. *: N: 4'-N,3dU. X: a 4'-oxidized abasic site. **: the calculated values are in parentheses).

<table>
<thead>
<tr>
<th></th>
<th>Sequence*</th>
<th>MALDI-TOF MS**</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>5'-CCA AAG-3'</td>
<td>1786.31 (1786.23)</td>
</tr>
<tr>
<td>13</td>
<td>5'-CCC GGT ACT TTG G-3'</td>
<td>3942.47 (3942.62)</td>
</tr>
<tr>
<td>5'-biotin-13</td>
<td>5'-Biotin-CCC GGT ACT TTG G-3'</td>
<td>4348.10 (4347.62)</td>
</tr>
<tr>
<td>14</td>
<td>5'-CCA AAG NAC CGG G-3' 3'-GGT TTC ATG GCC C-biotin-5'</td>
<td>3995.74 (3996.60)</td>
</tr>
<tr>
<td>15</td>
<td>5'-CCA AAG NAC CGG G-3'</td>
<td>3995.74 (3996.60)</td>
</tr>
<tr>
<td>16</td>
<td>5'-CCA AAG XAC CGG G-3'</td>
<td>3877.47 (3877.52)</td>
</tr>
<tr>
<td>17</td>
<td>5'-CCA AAG XAC CGG G-3' 3'-GGT TTC ATG GCC C-5'</td>
<td></td>
</tr>
</tbody>
</table>

*E. coli* uracil-DNA glycosylase (UDG, 1 U releases 60 pmol uracil per min from ds uracil-containing DNA), DNA polymerase I-Klenow fragment (1 U converts 10 nmol dNTPs to an acid-insoluble form in 30 min at 37 °C), and NgoM IV (1 U digests 1μg adenovirus-2 DNA in 1 h at 37 °C) were purchased from New England Biolabs. Alternatively, UDG was purified subsequent to expression from pET21a-UDG kindly supplied by Dr. James Stivers, Johns Hopkins University, as previously described(24). Nuclease P1 (1U catalyzes the hydrolysis of 1 μmol phosphodiester linkage in yeast RNA in 1 min at 37 °C) and alkaline phosphatase from calf intestine (AP, 1 U catalyzes the hydrolysis of 1 μmol of 4-nitrophenyl phosphate in 1 min at 37 °C) were purchased from Roche. HIV-1 reverse transcriptase (HIV-1 RT, 1 U incorporates 1 nmol dNTP in 20 min at 37 °C, pH 8.3) and snake venom phosphodiesterase-1 (PDE-1, 1 U
hydrolyzes 1 μmol p-nitrophenyl thymidine-5-phosphate in 1 min at 25 °C, pH 8.9) were purchased from Worthington Biochemical Co. Alternatively, HIV-1 RT was purified following the published protocol(25) from E. coli DH5α containing the p66(his)/51 plasmid (provided by Dr. Paul Boyer and Dr. Stephen Hughes, National Cancer Institute). AMV reverse transcriptase (AMV RT, 1 U incorporates 1 nmol TMP into acid-insoluble product in 10 min at 37 °C) was purchased from USB. RAV-2 reverse transcriptase (RAV-2 RT, 1 U incorporates 1 nmol TMP into acid-insoluble product in 10 min at 37 °C with poly (rA)/polyT as the primer/template) was purchased from Amersham Pharmacia Biotech. Human polymerase β (1 U incorporates 1 nmol dNTP into acid-insoluble product in 1 h at 37 °C) was purchased from Trevigen. Yeast formate dehydrogenase (FDH, 1 U oxidizes 1.0 μmol formate to CO₂ per min using at pH 7.5, 37 °C) was purchased from Sigma.

Ethanol precipitation of oligonucleotides was carried out in a volume of 1 mL by adding 0.3 M sodium acetate, pH 5.2 and 71.4% (v/v) ethanol. The mixture was incubated at −80 °C for 1 h and was centrifuged at 10,000 g for 30 min. The supernatant was decanted and the pellet was washed with 500 μL ice-cold 75 % ethanol. The mixture was centrifuged again at 10,000 g for 30 min. The supernatant was removed and the pellet was dried in vacuo.

Reverse-phase HPLC (RP-HPLC) was carried out using a Nucleotide-Nucleoside Column (Alltech, 250 mm × 4.6 mm). Two different chromatographic elution programs were used. The first (System A) involved solution A (50 mM ammonium acetate, pH 7.0) and solution B (methanol) using a linear gradient of 0-50%B over 30 min with a flow rate of 1 mL/min. The second (System B) involved solution A (60 mM KH₂PO₄, 5 mM tetrabutylammonium phosphate, pH 4.7) and solution B (5 mM tetrabutylammonium phosphate in methanol) using a linear
gradient of 0-50% B over 20 min and 50-90% over an additional 10 min with a flow rate of 1 mL/min.

Anion-exchange HPLC was carried out using a DNAPac 100 column (Dionex, 250 mm x 4 mm). The elution (System C) involved solution A (25 mM HEPES, pH 8.0) and solution B (200 mM NaClO₄, 25 mM HEPES, pH 8.0) using a linear gradient of 25-100% B over 15 min with a flow rate of 1 mL/min.

The GC/MS experiments were carried out on a Hewlett-Packard Agilent 5973N with an HP Restek Rtx-1 column (30 m x 250 µm x 1 µm). The oven temperature was maintained at 100 °C for 5 min followed by a linear gradient of 100 °C to 250 °C over the next 15 min.

Matrix-assisted laser desorption/ionization (MALDI)-mass spectrometry (MS) experiments were carried out on a Bruker Daltonics OminiFlex MALDI-TOF mass spectrometer. The MALDI matrix solution was prepared by mixing 2 volumes of 52.5 mg/ml 3-hydroxyl-picolinic acid in 50% acetonitrile/water (v/v) with 1 volume of 0.1 M diammonium citrate. Samples containing ~200 pmol oligonucleotide were loaded onto 10 µL C18 ZipTip (MilliPore) and were desalted by washing twice with 20 µL H₂O. The desalted oligonucleotides were eluted with 1 µL MALDI matrix solution and were spotted on the target. The instrumental parameters were: ion source 1, 19 kV; ion source 2, 17 kV; lens, 8.6 kV; extraction time, 200 ns; sample rate, 1 per ns. The detector was set in a linear, positive mode.

*Synthesis of 5'-AAA AXA AAA-3' (I, X denotes an abasic site) as a control.* 1 (Table 4.1) was synthesized by treating crude 5'-AAA AUA AAA-3' (2) with UDG. The reaction mixture contained in a final volume of 200 µL: 178 nmol 2, 20 mM Tris-HCl (pH 8.0), 1 mM EDTA, 1 mM DTT, and 100 U UDG. The reaction was allowed to proceed at 37°C for 4.5 h. 1 was
purified by RP-HPLC System A. (Compound, retention time, yield: uracil, 7.8 min, 114 nmol; 1, 20.2 min, 110 nmol).

Reduction of 1 to 1a by NaBH₄ or NaBD₄ and enzymatic digestion of 1a with Nuclease P1 and AP (Figure 4.3). 1 (20 nmol) was incubated with 50 mM NaBH₄ in 100 mM Tris-HCl (pH 8.3), in a total volume of 200 μL at 4 °C for 30 min. Acetic acid (20 μL) was added to neutralize the reaction mixture and 1a was separated by RP-HPLC System A. (Compound, retention time, yield: 1a, 20.3 min, 19.2 nmol).

![Chemical structures and reactions]

Figure 4.3 Characterization of oligonucleotides containing an abasic site (1) or a 4′-oxidized abasic site (7) by enzymatic digestion followed by GC/MS. Similar experiments have been performed with NaBD₄. 1d and 7d were subjected to GC/MS analysis.

Nuclease P1 digestion of 1a was carried out in a final volume of 100 μL containing: 15 nmol 1a, 0.1 M NaOAc, pH 5.3, 0.1 mM ZnCl₂, and 15 U Nuclease P1. The reaction mixture was incubated at 37 °C for 1 h. The mixture was then diluted to 200 μL containing 50 mM Tris-HCl (pH 8.5), 0.1 mM EDTA, and 20 U AP. The incubation was carried out at 37 °C for 30 min. The reaction was stopped by heating the reaction mixture at 95 °C for 10 min. The protein was removed by centrifugation and the products were separated by RP-HPLC System A to give the compound, retention time, and yield: 1b, 17.4 min, 12 nmol; dA, 21.0 min, 85 nmol.
Alternatively, the products of the enzymatic digestion of 1a (5 nmol) were analyzed by RP-HPLC System B to give the compound, retention time, and yield: dA, 14.9 min, 27.4 nmol; 1b, 19.1 min, 4.3 nmol.

The reduction of 1 by NaBD$_4$ and the subsequent enzymatic digestion were carried out under the same conditions as described above. The ESI-MS ([M–H]$^-$) for 1b (NaBH$_4$): calcd, 448.1239, found, 448.1231; 1b (NaBD$_4$): calcd, 449.1302; found 449.1301.

**GC/MS analysis of 2-deoxy-pentitols (1c) derived from 1b.** 1b (8 nmol) was incubated in 50 mM Tris-HCl (pH 8.3), 10 mM MgCl$_2$ and 2 U PDE-1 in a total volume of 20 µL. The reaction was allowed to proceed at 37 °C for 2 h. The reaction was stopped by heating the reaction mixture at 95 °C for 10 min. The protein was removed by centrifugation. Half of the reaction mixture was subjected to RP-HPLC (System B) and the product was found to co-elute with 5'-dAMP. The other half was lyophilized and transferred to a 0.3 mL reaction vial (Pierce) by dissolving in methanol followed by evaporating to dryness in vacuo. 1c (4 nmol) was silylated with a 20 µL 1:1 acetonitrile:BSTFA (with 1% TMCS) at 100 °C for 30 min to produce 1d (Figure 4.3). An aliquot (1 µL) of the silylation mixture was then analyzed by GC/MS. 1d was eluted at 17.4 min (224 °C).

**Synthesis of 4'-azido-2'-deoxyuridine-5'-monophosphate (4'-N$_3$-dUMP).** 4'-Azido-2'-deoxyuridine (4'-N$_3$-dU) was synthesized as previously reported(26). It was converted to 4'-N$_3$-dUMP using a modification of the method of Yoshikawa et al.(27). 4'-N$_3$-dU (50 mg, 0.186 mmol) was dissolved in 0.5 mL of freshly distilled triethylphosphate. Freshly distilled POCl$_3$ (74.6 µL, 0.80 mmol) was then added dropwise with stirring at 4 °C. The reaction mixture was slowly warmed to room temperature and stirred for an additional 12 h. The mixture was then diluted into 50 mL of ice cold water and titrated with NaOH until the pH was between 7 and 7.5.
The 4'-N₃-dUMP was purified on a DEAE Sephadex A-25 column (2.5 cm x 22 cm) using a linear gradient of 0-0.4 M triethylammonium bicarbonate (TEAB, 1L x 1L). The flow rate was 2.7 mL/min and the fraction size was 28 mL. 4'-N₃-dUMP eluted at 0.25 M TEAB and the fractions were pooled and concentrated in vacuo until the triethylamine was removed. The product was characterized by ¹H NMR: (D₂O, HDO = 4.81 ppm, δ): 2.48 (2H, m, H₂'/H₂''), 4.09 (2H, m, H₅'/H₅''), 4.59 (1H, m, H₃'), 5.81 (1H, d, J = 7.9 Hz, H₅), 6.33 (1H, m, H₁), 7.71 (1H, d, J = 7.9 Hz, H₆). The overall yield was 50%.

Synthesis of 4'-azido-2'-deoxyuridine-5'-triphosphate (4'-N₃-dUTP). 4'-N₃-dUTP was prepared from 4'-N₃-dUMP via the method by Hoard and Ott(28). 4'-N₃-dUMP (34 rmol) was dissolved in 0.5 mL DMF and tributylamine (8.1 μL, 34 μmol) was added. The solvent was removed in vacuo and the salt was rendered anhydrous by evaporation with dry DMF (4 × 0.25 mL), dry pyridine (2 × 0.25 mL), and dry DMF (2 × 0.25 mL). The resulting product was then dissolved in 0.3 mL DMF and 27.5 mg 1,1'-carbonyldiimidazole (171 μmol) in 0.3 mL of DMF was added. The reaction mixture was stirred for 12 h at room temperature under argon. Then, 272 μmol of H₂O in 9.8 μL H₂O/DMF (1:1) was added and the reaction mixture was stirred for another 30 min at room temperature. Tributylammonium pyrophosphate (169.9 μmol) in 2 mL DMF, which was prepared following the published procedure(28), was added and the reaction mixture was stirred for an additional 24 h. The supernatant was collected and the precipitate was washed 5 times with 0.5 mL DMF. The combined supernatant and washes were mixed with 5 ml methanol and the solvent was removed in vacuo. The pellet was dissolved in 15 mL cold water. The product was purified on a DEAE Sephadex A-25 column (2.5 cm x 15 cm) with a linear gradient (0.65 L x 0.65 L, 0 to 0.8 M) of TEAB, pH 8.5. The flow rate was 3 mL/min and the fraction size was 17.5 mL. The major peak eluted at 0.6-0.68 M TEAB. The fractions were
pooled and concentrated *in vacuo*. The product was characterized by \(^1\)H NMR (D\(_2\)O, HDO = 4.81 ppm, \(\delta\)): 2.52 (2H, m, H2'/H2”), 4.25 (2 H, m, H5'/H5”), 4.83 (1 H, m, H3’), 5.93 (1 H, d, J = 8.1 Hz, H5), 6.42 (1 H, m, H1’), 7.85 (1 H, d, J= 8.1 Hz, H6). The yield of 4'-N\(_3\)-dUTP was 50%.

**Incorporation of 4'-N\(_3\)-dU into duplex DNA using a primer/template extension method.** The reaction mixture contained in a final volume of 1 mL: 10 \(\mu\)M primer (3, Table 4.1), 10 \(\mu\)M template (4, Table 4.1), 1 mM dNTPs (dATP, dGTP and dCTP), 0.1 mM 4'-N\(_3\)-dUTP, 50 mM Tris-HCl (pH 8.3), 50 mM KCl, 10 mM MgCl\(_2\), 0.5 mM DTT and 300 U HIV-1 RT. 3 and 4 were annealed by heating at 100 °C for 5 min followed by cooling to room temperature over 1 h before dNTPs and HIV-1 RT were added. The reaction mixture was incubated at 37 °C and the reaction was monitored using [5'-32P]-3. Aliquots were taken (2 \(\mu\)L) at various incubation times and analyzed by a 12% sequencing gel (35 cm × 42 cm). The gel was run at constant power of 90 W for 3 h and was then dried, exposed to a storage phosphor screen (Molecular Dynamics, now Amersham Biosciences), and visualized on a Storm 840 PhosphorImager (Molecular Dynamics, now Amersham Biosciences). The reaction was complete in 2 h and the product, 5, was separated from the reaction mixture by ethanol precipitation.

**Isolation of ss oligonucleotide containing 4'-N\(_3\)-dU (6, Table 4.1) from ds 5.** ss-DNA is required for UDG to remove uracil from 4'-N\(_3\)-dU. 5 was designed with an NgoM IV restriction site (Table 4.1). The reaction mixture of 200 \(\mu\)L contained 5 (5 nmol), 120 U NgoM IV in a buffer (50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate, and 1 mM DTT, pH 7.9) and was incubated at 37 °C for 12 h. 6 (Table 4.1, 20-mer) was separated from its complimentary strand (24-mer) on a 12% polyacrylamide gel (20 cm × 20 cm) at constant
voltage of 350 V for 2 h. The gel was visualized by UV shadow. 6 was recovered from the gel in 50% yield.

Removal of uracil from 6 using UDG. The reaction mixture of 160 μL contained 10 μM 6, 160 U UDG, 20 mM HEPES, pH 8.0, and 1 mM EDTA. The reaction was incubated at 37 °C and uracil release was monitored by RP-HPLC System A (Compound, retention time: uracil, 7.8 min, 6, 23.0 min, 7, 21.5 min). The reaction was completed in 2 h.

Quantitation of azide release using formate dehydrogenase (FDH). The release of azide was monitored using the procedure by Blanchard and Cleland(29). The assay mixture in 500 μL contained: 1.7 mM sodium formate, 10 mM NAD⁺, 100 mM HEPES (pH 7.5) and 0.025 U FDH. The rate of the reaction was measured by change in A₃₄₀nm (ε = 6.2 mM⁻¹ cm⁻¹). A standard curve was generated by addition of various amounts of azide (50 to 200 nM) to the reaction mixture and the decrease in rate was plotted against the azide concentration. To quantitate the azide release during the conversion of 6 to 7, 6 (3.2 nmol) was treated with UDG under the conditions described above. The reaction mixture was analyzed by RP-HPLC System A. Uracil was quantitated and the flow-through of the column (retention time 2.5 min to 3.5 min, 1 mL) was collected. Aliquots were added to the FDH assay mixture and the azide concentration was quantitated by the inhibition based on the standard curve.

Reduction of 7 to 7a by NaBH₄ or NaBD₄ and digestion of 7a by Nuclease P1 and AP.

Compound 7 (2.2 nmol) was reduced by NaBH₄ under the conditions described above for the abasic site control. The reaction mixture was neutralized by addition of 20 μL acetic acid and the reduced product, 7a, was purified by RP-HPLC System A to give the compound, retention time, yield: 7a, 21.7 min, 2.0 nmol.
7a was then digested by Nuclease P1 and AP as described above (Figure 4.3). 7b has a retention time identical to thymidine using RP-HPLC System A. Therefore, 7b was isolated using ion-pairing RP-HPLC (System B). 7b was lyophilized and desalted using RP-HPLC System A. Its retention times were identical to 1b in both RP-HPLC System A and System B. The reduction of 7 by NaBD₄ and the subsequent enzymatic digestion were carried out under the same conditions as described above.

Conversion of 7b to 2-deoxypentitol (7c) and 5'-dAMP followed by GC/MS analysis. 7b (2.0 nmol) was digested with PDE-1 and silylated to give 7d as described above for the abasic site oligonucleotide. 7d was eluted at 17.4 min (224 °C).

Stability of the duplex DNA (8) containing a 4'-oxidized abasic site. 8 (10 μM) was generated by annealing 7 with its complementary strand in 100 mM NaCl and 10 mM sodium phosphate (pH 7.0) at 70°C for 1 min followed by cooling to 20°C in a heat block over 1 h. Compound 8 was incubated at 37 °C and aliquots were analyzed by anion exchange HPLC at 10 h intervals from 0 to 70 h. The stability of the abasic site lesion (9, Table 4.1) in the same sequence context was monitored under identical conditions.

Incorporation of 4'-N₃dU into a 13mer duplex DNA (14 in Table 4.2) by HIV-1 RT The reaction mixture of 25 mL contained 40 μM primer (12, Table 4.2), 40 μM template (5'-biotin-13, Table 4.2), 1 mM dNTPs (dATP, dGTP and dCTP), 40 μM 4'-N₃-dUTP, 50 mM Tris-HCl (pH 8.3), 50 mM KCl, 10 mM MgCl₂, 0.5 mM DTT, and 1 μM HIV-1 RT. The primer (12) and the template (5'-biotin-13) were annealed by heating the reaction mixture at 55 °C for 5 min followed by cooling to 4 °C over 1 h before dNTPs and HIV-1 RT were added. The reaction mixture was incubated at 10 °C and the reaction progress was monitored by the analytical anion-exchange HPLC using 5 μL aliquots as described above (Compound, retention time: 12, 6.5 min; 5'-
biotin-13, 13.8 min; partially extended primer/template, 15.0 min; 14, 16.2 min). The reaction was complete in 48 h.

Isolation of 13-mer single-stranded (ss) DNA containing 4'-N$_3$dU (15, Table 4.2) The primer-extension reaction mixture from the previous step (25 mL) was loaded onto a column (1.5 cm x 6 cm) containing 10 mL streptavidin-agarose resin (Sigma-Aldrich Co.) equilibrated with HE buffer (20 mM HEPES, 1 mM EDTA, pH 7.0) at 4 °C. The column was washed with 20 mL HE buffer twice at 4 °C. The ss-product, 15, was then eluted with 20 mL HE buffer at 75 °C in an incubator. The average yield of 15 from the starting primer/template is ~ 95%. MALDI-TOF MS analysis of 15 indicated the incorporation of 4'-N$_3$dU ([M+H]$^+$ observed, 3995.74; calculated, 3996.60).

Conversion of 15 to 13-mer ss-DNA containing a 4'-oxidized abasic site (16) by UDG. The reaction mixture of 2 mL contained 400 nM 15 and 6 μM UDG in HE buffer. The reaction was incubated at 37 °C and uracil release was monitored by the preparative RP-HPLC (Compound, retention time: uracil, 6.2 min; 16, 16.8 min). The reaction was complete in 6 h and the average yield of 16 was 85-90%.

Preparation of duplex DNA (17, Table 4.2) for analysis by 2D NMR methods A 1:1 ratio of 13 (0.65 μmol) in 1 mL deionized water was added directly into the RP-HPLC fractions from the previous step (22.5 mL containing 0.65 μmol 16, 50 mM ammonium acetate, pH 7.0, and 20% acetonitrile). The mixture was heated to 55 °C for 1 min and was cooled to 4 °C over 1 h to give 17. The solution was then dialyzed extensively against 4 x 2 L deionized water periodically changed over 2 days. The solution containing 17 was evaporated in vacuum to reduce the total volume to 2 mL. Compound 17 was further purified by the preparative anion-exchange HPLC as described above. The fractions containing 17 were combined and dialyzed against 4 x 2 L
deionized water over 2 days. The sample was then exchanged into D$_2$O by lyophilization in 3 x 1 mL 99.996% D$_2$O (Cambridge Isotopes). Compound 17 (2.0 mM) in 325 µL of 10 mM sodium phosphate, 0.2 mM EDTA, pH 6.5 was used for NMR studies.

**Reduction of 14 and 15 by tris-(hydroxymethyl)-phosphine (THMP).** The reduction of 14 or 15 by THMP was carried out in a reaction mixture of 200 µL contained 20 µM 14 or 15 (4 nmol), 1 mM THMP, and 50 mM ammonium acetate, pH 6.8. The reaction was incubated at 37 °C for 2 h and uracil release was monitored by RP-HPLC System A (Compound, retention time: uracil, 7.8 min, 15, 16.3 min, product from 15, 14.9 min, product from 14, 18.3 min). The product of the reduction reaction from 15 was analyzed by MALDI-TOF MS.

### 4.3 Results

**Incorporation of 4'-N$_3$-dU into duplex DNA.** Two general methods have been developed to incorporate nucleotide analogs into oligonucleotides. The most widely used strategy is generation of the appropriately blocked phosphoramidite of the unnatural nucleoside and use of solid phase DNA synthesis methods(30) to generate ss-DNA. The second requires chemical synthesis of the nucleoside 5'-triphosphate analogs and use of a primer/template and a DNA polymerase to make duplex DNA. The latter method is often required when the nucleotide analog phosphoramidites are unstable to the solid phase methods. Both approaches have been reported for incorporation of 4'-azidothymine into oligonucleotides. Gibson et al. reported the synthesis of 4'-azidothymidine phosphoramidite and its incorporation into ss-DNA by the standard solid phase DNA synthesis method(31). Unfortunately, neither the starting material, 4'-azidothymidine phosphoramidite, nor the final oligonucleotide product was characterized. In our hands, 4'-N$_3$-dU is not stable to the concentrated ammonia conditions required to remove
oligonucleotide from the solid support. Therefore, this approach for incorporation of 4'-N₃-dU into DNA is problematic.

The studies of Chen et al. showing that 4'-azidothymidine-5'-triphosphate can be incorporated into duplex DNA using HIV-1 reverse transcriptase(32), suggested that the primer/template/polymerase approach for incorporation of 4'-N₃-dU would be successful. Incorporation of unnatural nucleotides using polymerases often encounters problems with either slow incorporation or slow extension subsequent to incorporation. Recently, the simultaneous use of two DNA polymerases has been applied to overcome these shortcomings(33).

We investigated a number of DNA polymerases (including Klenow fragment, AMV RT, human DNA polymerase β, RAV-2 RT, and HIV-1 RT) for their efficiency of incorporation of 4'-N₃-dUTP and extension into duplex DNA. The Klenow fragment failed to incorporate 4'-N₃-dUTP under all conditions. AMV RT could incorporate 4'-N₃-dU, however, extension resulted in significant amounts of truncated products. Human DNA polymerase β, RAV-2 RT, and HIV-1 RT were all found to incorporate 4'-N₃-dUTP and to generate complete duplex DNA. The ability to isolate large amounts of HIV-1 RT from an over-expression system and its efficiency at incorporation and extension to generate intact duplex DNA was the basis for our choice of HIV-1 RT for further studies.

Using [5'-32P]-(3) as the primer and 4 as the template (Table 4.1), the time dependent incorporation of 4'-N₃-dUTP into an oligonucleotide by HIV-1 RT was monitored by PAGE (Figure 4.4). The gel reveals some stalling in extension of the primer subsequent to the incorporation of 4'-N₃-dU. However, in 2 h, all of the primer can readily be converted to product (Figure 4.5, lane 2). The overall yield of 5 was 95%. The actual primer extension product is one
nucleotides longer than the template due to the terminal nucleotide addition activity of HIV-1 RT (Figure 4.4, lane 6)(34).

Figure 4.4 Monitoring incorporation of 4'-N₃dU into duplex DNA using primer [5'-³²P]-3 and template 4 with HIV-1 reverse transcriptase. Lane 1, [5'-³²P]-3 (13mer); Lane 2-6, 0, 10, 20, 30 and 60 min incubation time; Lane 7, [5'-³²P]-27mer as a MW marker. At 2 h, all stalled oligonucleotides were converted to product (see Figure 4.5).

Isolation of 6 from 5 (Table 4.1). Our strategy to generate 4'-oxidized abasic site is shown in Figure 4.2. The key step is to use UDG to remove uracil, generating a 4'-azido abasic site intermediate that we predicted would rapidly lose azide to generate the desired lesion. Previous studies of Lindahl and Stivers have demonstrated that both ss- and ds-DNA containing deoxyuridine are substrates of UDG with ss-DNA being hydrolyzed 20 times more efficiently than ds-DNA(35, 36). Our efforts to use UDG to convert 4'-N₃-dU to a 4'-oxidized abasic site in duplex DNA, however, were unsuccessful. No uracil release was observed even with 150,000 times the amount of enzyme used in the cleavage of uracil from deoxyuridine and extended incubation times. Thus, to use UDG to release uracil, we needed to separate the two strands of the duplex.
To achieve this separation, one end of the duplex DNA (5) was designed to contain an NgoM IV cleavage site (Table 4.1). Digestion of the duplex with this restriction endonuclease resulted in production of two oligonucleotides: a 24-mer and the desired 20mer (6). The size difference was sufficient for strand separation using denaturing PAGE. The bands were visualized by “UV shadow” as shown in Figure 4.5. Compound 6 was purified from the gel by the “crush and soak” method in 50% overall recovery. NgoM IV digestion also eliminated the variable end of the extended primer, present due to HIV-1 RT’s non-templated nucleotide addition activity. The MALDI-TOF MS analysis of 6 showed [M+H]: 6155.65 (calculated: 6154.97). To further characterize 6, it was digested with nuclease P1 and AP. The resulting nucleosides were analyzed by RP-HPLC System A, giving ratios of dC/dG/T/4’-N₃-dU/dA as 5.5:6:2:1:1:1:4.9, consistent with the expected ratios (6:6:2:1:5, Figure 4.6).

![Figure 4.5](image)

**Figure 4.5** Denaturing PAGE allows separation of ss-DNA containing 4’-N₃-dU from its complement subsequent to treatment of the duplex DNA with NgoM IV. Lane 1: The primer (3) only. Lane 2: The primer extension reaction mixture containing the intact duplex (5) which is a 27mer. Lane 3: The two ss-DNA fragments generated by restriction enzyme digestion: a 24mer and one 20mer (6). In addition, unreacted duplex is also present in Lane 3.
Figure 4.6 Characterization of 6 via digestion with Nuclease P1 and AP followed by separation of the resulting nucleosides by RP-HPLC (Solvent system A).

Conversion of 6 into a 4'-oxidized abasic site containing oligonucleotide (7) by UDG. The isolated ss-6 was incubated with UDG to determine whether the desired 4'-oxidized abasic site containing oligonucleotide (7) could be generated. Uracil release, monitored by HPLC (Figure 4.7), was quantitative. However, the amount of UDG required for efficient uracil release was 3,000 fold greater than that used to quantitatively generate uracil from deoxyuridine in the same sequence context. The product, 7, was analyzed by MALDI-MS. The [M+H]+ observed was 6033.21, suggesting that the product is a ring-closed or a ring-opened 1’-hydrated aldehyde (calcd: 6033.03). No ring-opened, unhydrated species (calcd: 6015.02) was detected. Previous studies examining an abasic site labeled at Cl' with 13C and 17O revealed that the aldehydic (ring-opened) form and its hydrate constituted less than 1% of total species, with the majority of the abasic site being a ring-closed mixture of α/β anomers(37, 38). Based on the MALDI-MS analysis and these previous studies on the normal abasic site, we propose that the dominant form of 7 is the ring-closed hemiacetal.
Figure 4.7 Conversion of 6 into 7 concomitant with formation of uracil release as monitored by reverse phase HPLC. The void volume of the column was analyzed for azide release. (* denotes the partially degraded products)

Quantitation of azide release during the conversion of 6 to 7. Our synthetic scheme (Figure 4.2) predicts that a stoichiometric amount of azide should accompany uracil release. Previous studies of Blanchard and Cleland have shown that azide is a potent inhibitor of formate dehydrogenase with $K_i = 7 \text{nM}(29)$. A standard curve for azide inhibition of FDH was generated (Figure 4.8). The flow-through of the RP-HPLC column was used in the FDH assay to quantitate the azide release. Comparison of inhibition caused by the azide in the flow-through relative to the standard curve revealed that azide and uracil were released in a 1:1 ratio.

Enzymatic digestion of 7 and GC/MS analysis of the deoxypentitols (7c). The structure of the 4'-oxidized abasic site (7) was further characterized as outlined in Figure 4.3 following the strategy of Rabow et al.(8). An abasic site standard (1) was characterized as a control for the sequence of reactions described in Figure 4.3. 7 was converted to 7a by reduction with NaBH₄ (or NaBD₄). 7a was enzymatically digested into its nucleoside components using Nuclease P1 and AP. The
products were then analyzed by ion-pairing RP-HPLC (System B), giving ratios of

dC/dG/T/dA/7b as 5.5:6.0:2.0:4.0:0.9, consistent with the expected ratios of 6.0:6.0:2.0:4.0:1.0
(Figure 4.9). 7b co-migrated with an authentic sample of this species generated from the
digestion of the abasic site control (1b).

![Graph showing the quantitation of azide release during the conversion of 6 to 7 by the FDH inhibition assay.](image)

**Figure 4.8** Quantitation of azide release during the conversion of 6 to 7 by the FDH inhibition assay. *, Standard azide concentrations; ×, azide released via UDG reaction.

To further support the structural assignment of 7b, it was hydrolyzed by PDE-1 to produce 2-deoxypentitol (7c) and 5'-dAMP. 7c was silylated and GC/MS analysis revealed two products (7d) associated with the diastereomers resulting from reduction of the ketone group at C4'. The mass spectrum of both products revealed a fragmentation pattern identical to that observed with 1d (Figure 4.10). The mass spectrum of NaBD₄ reduced 7d unambiguously established the incorporation of two deuterium atoms: one at C1' and the second at C4' (Figure 4.10). The MS data in conjunction with the degradation studies establish that the 4'-oxidized abasic site lesion
can be generated in 50% yield from the starting primer/template. More recently, the overall yield has been improved to ~90% by using a biotinylated template to facilitate the separation of the template from the ss-product (see the discussion below).

Figure 4.9 Ion-pair RP-HPLC (System B) of the products generated when 7 was reduced by NaBH₄ followed by digestion with Nuclease P1 and AP.

Stability of a duplex DNA (8) containing a 4'-oxidized abasic site relative to an abasic site in the same sequence context (9). As with abasic sites, 4'-oxidized abasic sites are also chemically labile. The ring opened, aldehydic form of this compound has been proposed to be susceptible to elimination of the 3'-phosphate and, consequently, DNA strand scission(39). A preliminary experiment was carried out to examine the stability of this lesion under conditions that mimic physiological pH and ionic strength. The elimination process was monitored using anion exchange chromatography under conditions in which the duplex DNA remained intact. The stability of the duplex DNA containing a normal abasic site (9, Table 4.1) in the same sequence context as 8 was also examined under the same conditions. The result establishes that the 4'-oxidized abasic site, which has a half-life of 26 h, is chemically more labile than the abasic site, which has a half-life of 130 h (Figure 4.11).
Figure 4.10 GC/MS of trimethylsilylated 2-deoxy-pentitol generated from NaBH₄ and NaBD₄ reduced abasic sites (1d, top) or 4'-oxidized abasic sites (7d, bottom).
Figure 4.11 Stability of a duplex DNA containing a 4'-oxidized abasic site (8) monitored by anion exchange HPLC. A: HPLC profiles of 8 incubated at 37 °C in 50 mM NaPO4, 100 mM NaCl at pH 7.0 for 0, 10 and 20 h. * indicates the hexameric ss-oligonucleotide (10) and the arrow indicates the 13-mer/20-mer duplex (11, C). B: A plot of ln % remaining 8 versus time (h). The degradation of 8 is pseudo-first order with a $k_{apparent} = 2.7 \times 10^2$ h$^{-1}$ ($t_{1/2} = 26$ h). C: Proposed structures of degradation products of 8 as observed by anion exchange HPLC.
Synthesis of 13mer duplex DNA containing a 4'-oxidized abasic site (17, Table 4.2) for structural characterization by 2D NMR methods. To generate the 4'-oxidized abasic site in a 13mer duplex DNA, a hexameric oligonucleotide primer (12) was annealed to a 13-mer template (5'-biotin-13, Table 4.2) at 4°C. The melting temperature of the resulting primer/template pair was estimated to be ~15°C. Therefore, the 4'-N3dU incorporation and the subsequent primer extension were performed at 10°C. The reaction was monitored using [5'-32P]-12 and DNA sequencing gel analysis (Figure 4.12). Alternatively, the progress of the reaction was monitored by analytical anion-exchange HPLC. Under the assay conditions, the starting material, the partially extended product, and the final full-length product had distinct retention times, which allowed monitoring of the progress of the reaction (Figure 4.13). The reaction was completed in 48 h.

Figure 4.12 4'-N3dU incorporation and primer extension reaction catalyzed by HIV-1 RT and monitored by 5'-[32P] labeled primer, 12. Lanes 1-5, extension reaction quenched at 0, 3, 6, 17, 48 h. (* denotes the over-extended product, usually an extra adenosine at 3', due to HIV-1 RT’s terminal nucleotidyltransferase activity.)
Figure 4.13 4'-N₃dU incorporation and primer extension reaction catalyzed by HIV-1 RT monitored by anion-exchange HPLC. (* denotes the over-extended product, usually an extra adenosine at 3', due to HIV-1 RT’s terminal nucleotidyltransferase activity.)
Our previous method to separate two DNA strands in the ds product containing 4'-N₃dU involved time-consuming steps of restriction enzyme digestion, denaturing polyacrylamide gel electrophoresis, and extraction of ss DNA from the gel. This procedure suffered from a low yield and was only suitable for obtaining ~nmol amounts of ss-products. For the preparation of material on the μmol scale for structural studies, a biotin tag was incorporated at the 5' of the template strand. Subsequent to completion of chain extension, the mixture from the primer extension reaction was passed through a column containing streptavidin-agarose resin at 4 °C. The column was washed with HE buffer to remove HIV-1 RT and unincorporated dNTPs. The ss product (15) was isolated from 14 by eluting the column with HE buffer at 75 °C (the Tm for 14 was calculated to be ~45 °C). The incorporation of 4'-N₃dU into 14 and 15 was confirmed by complete enzymatic digestion with Nuclease P1 and alkaline phosphatase followed by the quantitation of individual nucleosides using RP-HPLC (Figure 4.14 B and C, respectively). The results showed that the incorporation of 4'-N₃dU is stoichiometric: the ratio of dC : dG : T : 4'-N₃dU : dA for 14 was 7.9 : 8.1 : 4.0 : 1.0 : 5.1 (calculated, 8 : 8 : 4 : 1 : 5) and for 15 was 4.0 : 4.1 : 0 : 0.95 : 4.1 (calculated, 4 : 4 : 0 : 1 : 4).

Conversion of 15 to 16 by UDG was monitored by RP-HPLC. The reaction was complete in 6 h and the conversion was quantitative (Figure 4.15). MALDI-TOF MS analysis of 16 indicates a species with a m/s ([M+H]+) value of 3877.47, which is consistent with a ring-closed hemiacetal (calculated: 3877.52). The final NMR sample of duplex DNA (17) was obtained by mixing a 1:1 ratio of 16 and 13, heating the mixture to 55 °C for 1 min, and cooling to 4 °C over 1 h. Compound 17 was further characterized by reduction with NaBH₄ followed by digestion with Nuclease P1 and alkaline phosphatase. The resulting nucleosides were analyzed by ion-
pairing RP-HPLC, giving ratios of dC/dG/T/dA/7b as 7.9:8.0:4.1:4.0:0.9, consistent with the expected ratios of 8.0:8.0:4.0:4.0:1.0 (Figure 4.16).

Figure 4.14 Incorporation of 4'-N₉dU confirmed by nucleoside composition analysis using complete enzymatic digestion of oligonucleotides followed by RP-HPLC analysis. A: Nucleoside standards. B: Nucleoside composition of 14. C: Nucleoside composition of 15. The sequences of 14 and 15 are also shown in each panel (N: 4'-N₉dU).
Figure 4.15 Conversion of 15 to 16 monitored by analytical RP-HPLC.

Figure 4.16 Ion-pairing RP-HPLC analysis of the nucleosides from the duplex DNA 17 after it was reduced by NaBH₄ followed by digestion with Nuclease P1 and AP.
4.4 Discussion

4'-Oxidized abasic sites are common DNA lesions generated by ionizing radiation damage, the BLMs, and the enediyynes. The preparation of these lesions has posed a synthetic challenge due to their chemical instability. During the course of the study, Kim et al. reported a chemical synthesis of a phosphoramidite stable to standard solid phase DNA synthesis, which upon photolysis, can generate a 4'-oxidized abasic site(39). This light-mediated method involves a seven-step synthesis of the phosphoramidite precursor from deoxyribose with an overall yield of 9%. The mixture of 4'-diastereomers makes spectroscopic analysis of each synthetic step difficult. While no details of oligonucleotide synthesis were presented, chain extension of the "unnatural C4' diastereomer" during the solid phase synthesis is expected to be less efficient. In addition, conversion of the precursor to the 4'-oxidized abasic site by photolysis at 300 nm causes non-specific photo-mediated damage to DNA, which makes this method problematic to generate the 4'-oxidized abasic site on a large scale or in vivo.

We also have focused on the generation of a precursor to the 4'-oxidized abasic site. Subsequent to its incorporation into DNA, this precursor could be transformed into the desired lesion by under mild conditions to avoid the problem of DNA strand scission or non-specific DNA damage. Ideally, we also want to generate the lesion inside the cell and study its physiological consequences.

The success of 3'-azido thymidine (AZT) as an AIDS therapeutic resulted in synthesis of a number of other azido modified nucleosides including 4'-azido nucleosides(26). These compounds appeared to be potential precursors to the 4'-oxidized abasic site if a polymerase could be identified to catalyze their incorporation into DNA. Previous studies of Chen et al. (32) and recent studies with other unusual nucleotides(40, 41) suggested that HIV-1 RT is sufficiently
promiscuous not only to incorporate the modified nucleotides, but also to catalyze chain extension to generate full-length duplex DNA. Once incorporated, several routes were envisioned that could lead to the desired product without phosphodiester bond cleavage. As noted above (Figure 4.6 and 4.14), incorporation proceeds efficiently and quantitatively.

To examine the generation of the desired lesion, we first investigated reductants such as thiolates and water-soluble phosphines(42). Reduction of 4'-azido group should result in the generation of a 4'-amine, which could potentially collapse to a 4'-imine abasic site (Figure 4.17). The 4'-imine could then be hydrolyzed to a ketone and ammonia. Reduction proceeded rapidly with a variety of thiolates. Unfortunately, due to the basic conditions required to generate thiolates, the rate of phosphodiester bond cleavage was comparable to the rate of the azide reduction. This approach was consequently abandoned.

![Figure 4.17 Proposed mechanism to generate 4'-oxidized abasic sites from 4'-N₃-dU via the reduction of the 4'-azido group.](image)

A water soluble phosphine, tris-(2-carboxyethyl)-phosphine (TCEP), was also examined as a potential reductant (the Staudinger reaction)(42). The reduction proceeded rapidly and efficiently on the nucleoside (4'-N₃-dU). When this nucleoside was within the context of DNA, however, the rate of reduction was drastically reduced, presumably due to the electrostatic clash between the DNA and the carboxylate groups of TCEP.
To circumvent this electrostatic problem, a non-charged water-soluble phosphine, tris-(hydroxymethyl)-phosphine (THMP), has also been tested for the direct reduction of the azide within the duplex. THMP was shown to release uracil from both 15 and 14 (Figure 4.18 B and C, respectively). However, MALDI-TOF analysis indicated that the THMP-reduced product ([M+H]+: 3957.02) from 15 is inconsistent with the expected product, 16 ([M+H]+: 3877.52, Table 4.2).

The identity of the THMP-reduced product from 14 was also investigated by treating the product with NaBH$_4$ followed by enzymatic digestion with Nuclease P1 and alkaline phosphatase. The nucleosides were analyzed with ion-pair RP-HPLC (Figure 4.19). The resulting HPLC profile, when compared with those from 4'-oxidized DNA prepared by the UDG treatment (Figure 4.9 and Figure 4.16), indicated that none of the expected Compound 7b (Figure 4.3) was present.

Figure 4.18 Reduction of 4'-N$_3$dU-containing DNA by THMP monitored by RP-HPLC. A: 15 alone. B: Reaction of 15 with THMP. C: Reaction of 14 with THMP.
These results prompt us to re-examine the mechanisms of the reduction by THMP (Figure 4.20). According to the mechanism of the Staudinger reaction(43, 44), THMP would react with the 4'-azido group first to generate a 4'-iminophosphorane-dU intermediate (I, Figure 4.20). Species I could collapse after releasing uracil to generate a 4'-iminophosphorane abasic site (II). Alternatively, species I could be hydrolyzed to generate 4'-amino-dU, which would collapse to a 4'-amino abasic site (IV and V) after uracil release. Both mechanisms could explain the uracil release observed experimentally (Figure 4.18). Species II might also be hydrolyzed to a 4'-amino abasic site. Ideally, the 4'-amino abasic site would be converted to the desired 4'-oxidized abasic site after hydrolysis of the 4'-imino residue in the ring-open form (IV).
The [M+H]^+ value measured for the reduction product from 15 ([M+H]^+: 3957.02) showed a +79.5 shift from the value expected for 16 ([M+H]^+: 3877.52), which currently cannot be assigned to any intermediate in this proposed mechanism. It is likely that species II, IV, or V could undergo further rearrangement. Further characterization of the reduction product generated with THMP is essential to modify the reaction conditions so that desired lesion can be generated.

Figure 4.20 Proposed mechanisms for the reduction reaction of 4'-N3dU-containing DNA by THMP (R: -CH2OH).

185
Because of the problems associated with chemical reduction, we focused our effort on the enzymatic conversion by UDG. This reaction leads directly to uracil release and subsequently azide release, generating the desired product (Figure 4.2). While previous studies have shown that both ss- and ds-DNA containing deoxyuridine are substrates of UDG, the reduced reactivity of 4'-N$_3$-dU with UDG has limited the substrate to ss-DNA. Therefore, we use the biothynylated template to facilitate the isolation of ss-DNA from the ds-DNA product using the streptavidin-agarose column and elution under denaturing conditions. With this technical improvement, we are able to generate this lesion \textit{in vitro} in quantities sufficient for structure determination and analysis with DNA repair enzymes.

The half-life of a 4'-oxidized abasic site was measured by Kim \textit{et al.} and found to be 7.8 h (37 \textdegree{}C, 100 mM NaCl at pH 7.5) in ss-DNA(39). The rate constant is faster by a factor of three relative to the one we have measured in ds-DNA (26 h at 37 \textdegree{}C, 100 mM NaCl at pH 7). Although the pH conditions and the sequences are different in the two studies, we believe duplex DNA stabilizes the lesion. A similar effect of DNA structure has recently been reported for the 2-deoxyribonolactone abasic site lesion(45). Its half-life was found to be 20 h in ss-DNA and 32-54 h in ds-DNA (37\textdegree{}C, 150 mM NaCl, 2 mM MgCl$_2$ at pH 7.5)(45). Stability of sugar lesions, not surprisingly, appears to be sensitive to both sequence and sequence context (ss versus ds).

It is desirable to generate the 4'-oxidized abasic site in duplex DNA \textit{in vivo}. As noted above, our attempts, thus far, have failed to directly generate the lesion in duplex DNA. It has not been reported whether the photolysis method can generate this lesion in duplex DNA. In addition, the photolysis would cause unwanted DNA damages and trigger UV-induced responded inside the cells. In our case, we are hopeful that, given the availability of a variety of 4'-azido-substituted deoxynucleotides (U, G, T, A, and C)(26), mismatch repair DNA glycosylases will allow
generation of 4'-oxidized abasic sites in duplex DNA(46). If one of these enzymatic approaches is successful, the lesion could be generated in vivo, which is not possible by the light-mediated approach. Otherwise, our method will require transformation of ss-DNA into the cells(47, 48).

Oxidative damage occurs to both the bases and the sugars of DNA(11). Damage to the bases and the mutagenic consequences of this damage have been much more extensively investigated(12-14) than damage to the deoxyribose sugars. Previous studies have been reported on the generation of 4'-oxidized abasic sites using BLM (Figure 4.1)(49, 50). Unfortunately, the cleavage of DNA by BLMs is not very sequence-specific and consequently, generation of the 4'-oxidized abasic site at the primary site also results in lesions at secondary sites. The difficulty of preparing this lesion homogeneously using BLMs has thus limited its examination with DNA repair enzymes. Studies of Demple et al. on a 36-mer containing a single 4'-oxidized abasic site generated with BLM have demonstrated that Apel can catalyze phosphodiester bond cleavage 5' to the lesion, resulting in a 3'-OH end and a 5'-phosphorylated-lesioned deoxyribose(50). Although a detailed kinetic analysis was not carried out, the efficiency of cleavage was suggested to be similar to Apel catalyzed reaction on an abasic site.

Recently, using the phosphoramidite approach, the 4'-oxidized abasic site has been studied with a variety of DNA polymerases(51). The results suggest that DNA polymerases in E. coli preferentially incorporate an A opposite to the 4'-oxidized abasic site. In addition, the 4'-oxidized abasic site also induced an unusual three-nucleotide deletion(52). Our ability to generate the 4'-oxidized abasic site lesion in any sequence context now allows us to understand the structure of the lesions (both ss- and ds-lesions) and the efficiency with which they are repaired by Apel and β-polymerase and to further characterize other repair proteins.
4.5 References


Chapter 5. Structures and dynamics of duplex DNA containing a 4’-oxidized and a normal abasic site
5.1 Introduction

Bleomycins (BLMs) are clinically important chemotherapeutic agents for the treatment of a variety of cancers(1). BLMs are proposed to cause cancer cell death by generating both single-stranded (ss) and double-stranded (ds) DNA lesions(2-4). Cleavage of DNA requires the presence of both ferrous iron and O$_2$, which generates the "activated BLM" that is responsible for the oxidative damage to the deoxyribose sugar of DNA (Figure 5.1). The sugar damage is initiated by 4'-H atom abstraction from a pyrimidine 3' to a guanine(5). The resulting 4'-radical intermediate can lead to two types of DNA damage products: a gapped 3'-phosphoglycolate/5'-phosphate lesion (3'-PG/5'-P)(6) under O$_2$-rich conditions and a 4'-oxidized abasic site under O$_2$-depleted conditions(7, 8)(Figure 5.1). The knowledge of how these lesions are recognized and repaired in vivo is essential to understanding the cytotoxicity and drug resistance of BLM in cancer cells.

Our group has been interested in the 4'-oxidized abasic site lesions, which is a common DNA damage product generated by a variety of agents, including BLM(7, 8), enediynes(9, 10), and reactive oxygen species(11). Recently, we have developed a synthetic method to generate the 4'-oxidized abasic site in any sequence context (Chapter 4)(12). In this method, a precursor, 4'-azido-deoxyuridine, is incorporated site specifically into the primer using a defined template by HIV-1 reverse transcriptase and, subsequent to strand separation of the duplex product, is converted to the 4'-oxidized abasic site using uracil-DNA glycosylase. During the course of our study, the Greenberg group reported a chemical synthesis of a phosphoramidite precursor to the 4'-oxidized abasic site that was stable to the standard solid phase DNA synthesis. Upon photolysis, this precursor can generate the 4'-oxidized abasic site (13). This light-mediated reaction is slow and can cause photolytic DNA damage, which limits its application to in vitro studies on a small scale. The ability to generate an 4'-oxidized abasic lesions in duplex DNA by our method thus offers the opportunity
to study their physiological consequences in DNA repair and mutagenesis in vitro and in vivo (14, 15).

**Figure 5.1** Pathways of BLM-mediated DNA cleavage. (3'PG/5'P, 3'-phosphoglycolate/5'-phosphate)

In humans, the 4'-oxidized abasic sites are thought to be recognized and repaired by Ape1, which is part of the base excision repair (BER) pathway (Figure 1.8, Chapter 1)(16). The predominant substrate for Ape1 in vivo is duplex DNA containing an abasic sites, generated by spontaneous base loss or repair of damaged bases by DNA glycosylases(17). Questions remains as to how DNA repair enzymes like Ape1 specifically recognize DNA lesions such as normal or 4'-oxidized abasic sites.

In this chapter, we present the first solution structural study using 2D NMR methods on duplex
DNA containing the 4′-oxidized abasic site (denoted by X), \( \text{d(CCAAAGXACCGGG)} \cdot \text{d(CCCGGTACTTTGG)} \) (1). As a comparison, the structure of the duplex DNA containing a normal abasic site (denoted by Y) in the same sequence context, \( \text{d(CCAAAGYACCGGG)} \cdot \text{d(CCCGGTACTTTGG)} \) (4), has also been determined. Using NMR derived restraints, a structural model of 1 has been determined, in which the deoxyribose is a ring-closed hemiacetal in the \( \alpha \) anomeric state. Both the abasic deoxyribose and the opposite base in 1 are intrahelical. Two structural models have also been determined for 4: one of the \( \alpha \) anomer at the abasic site (4\( \alpha \)) and the other of the \( \beta \) anomer at the abasic site (4\( \beta \)). With both 4\( \alpha \) and 4\( \beta \), the abasic site is partially extrahelical without significant distortion of the DNA backbone and the base opposite the abasic site is well stacked within the duplex DNA. Molecular dynamics calculations using the NMR structures suggest that the normal abasic site (4\( \alpha \) and 4\( \beta \)) is conformationally more flexible than the 4′-oxidized abasic site (1). These structural studies are the first step to understanding the mechanism(s) of recognition of these lesions by DNA repair enzymes.

### 5.2 Experimental Section

**Material**  The oligonucleotides were synthesized on a 10-\( \mu \)mol scale by Invitrogen Inc. and were further purified by the anion-exchange HPLC as described below. *E. coli* uracil-DNA glycosylase (UDG) was purified from expression plasmid pET21a-UDG kindly supplied by Dr. James Stivers, Johns Hopkins University, as previously described\(^\text{(18)}\). All other chemicals were purchased from Sigma-Aldrich Co.

*High performance liquid chromatography (HPLC) to analyze the purity of or to purify oligonucleotides* The analytical reverse-phase HPLC (RP-HPLC) was carried out using a Nucleotide-Nucleoside column (Alltech, 250 mm \( \times \) 4.6 mm). The standard chromatographic elution
program involved solution A (50 mM ammonium acetate, pH 7.0) and solution B (methanol) and a linear gradient from 0% to 50% B over 30 min with a flow rate of 1 mL/min. The preparative RP-HPLC was carried out using an Econosil C18 column (Alltech, 10 μm, 250 mm × 10 mm). The elution program involved solution A (50 mM ammonium acetate, pH 6.8) and solution B (acetonitrile) and a linear gradient from 0% to 40% B over 20 min with a flow rate of 3 mL/min.

Analytical anion-exchange HPLC was performed on a DNAPac 100 column (Dionex, 250 mm × 4 mm). The elution involved solution A (25 mM HEPES, pH 8.0) and solution B (1 M NaCl, 25 mM HEPES, pH 8.0) and a linear gradient from 30% to 60% B over 20 min with a flow rate of 1 mL/min. The preparative anion-exchange HPLC was carried out using a DNAPac 100 semi-prep column (Dionex, 250 mm × 10 mm). The elution was identical to that described above except that the flow rate was 3 mL/min.

**Preparation of 1 (Table 5.1) for analysis by 2D NMR methods** The synthesis and characterization of 1 was described in Chapter 4. Compound 1 (2.0 mM) in 325 μL of 10 mM sodium phosphate, 0.2 mM EDTA, pH 6.5 (D_2O or 10% D_2O/90% H_2O (v/v)) was used for NMR studies.

**Preparation of 4 (Table 5.1) for analysis by 2D NMR methods** Duplex DNA (4, Table 5.1) containing a normal abasic site in the same sequence context as 1 was prepared following the published procedure(19). Conversion of 2 to 3 was carried out in a reaction mixture of 2 mL contained 450 μM 2 and 2 μM UDG in 20 mM HEPES, 1 mM EDTA, pH 7.0. The reaction was incubated at 37 °C and uracil release was monitored by the analytical RP-HPLC (Compound, retention time: uracil, 6.2 min; 2, 20.2 min, 3, 16.8 min). The reaction was complete in 6 h. MALDI-TOF MS analysis of 3 indicates a species with a m/z ([M+H]^+) value of 3861.43 (Calculated:3861.52). Compound 3 was isolated from the reaction mixture by the preparative RP_HPLC with a yield of 94% (0.86 μmol). The NMR sample containing 4 was prepared following
the method described in Chapter 4. A 1:1 ratio of the template (d(CCCGGTACTTTGG), 0.86 μmol) in 1 mL deionized water was added directly into the RP-HPLC fractions from the previous step (22.5 mL containing 0.86 μmol 3, 50 mM ammonium acetate, pH 7.0, and 20% acetonitrile). The mixture was heated to 55 °C for 1 min and was cooled to 4 °C over 1 h to give 4. The solution was then dialyzed extensively against 4 × 2 L deionized water periodically changed over 2 days. The solution containing 4 was evaporated in vacuum to reduce the total volume to 2 mL. Compound 4 was further purified by the preparative anion-exchange HPLC. The fractions containing 4 were combined and dialyzed against 4 × 2 L deionized water over 2 days. The sample was then exchanged into D₂O by lyophilization in 3 × 1 mL 99.996% D₂O (Cambridge Isotopes). The final NMR sample contained 4 (2.7 mM) in 325 μL of 10 mM sodium phosphate, 0.2 mM EDTA, pH 6.5 in D₂O or 10% D₂O/90% H₂O (v/v).

Table 5.1 Sequence information of oligonucleotides in this study.

<table>
<thead>
<tr>
<th></th>
<th>Sequence*</th>
<th>HPLC Retention time (min)</th>
<th>MALDI-TOF MS***</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5'-CCA AAG XAC CGG G-3' 3'-GGT TTC ATG GCC C-5'</td>
<td>13.3a</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>5'-CCA AAG UAC CGG G-3'</td>
<td>20.2b</td>
<td>3955.46 (3955.63)</td>
</tr>
<tr>
<td>3</td>
<td>5'-CCA AAG YAC CGG G-3'</td>
<td>16.7b</td>
<td>3861.43 (3861.52)</td>
</tr>
<tr>
<td>4</td>
<td>5'-CCA AAG YAC CGG G-3' 3'-GGT TTC ATG GCC C-5'</td>
<td>13.4a</td>
<td></td>
</tr>
</tbody>
</table>

(* N, 4'-N3dU; X, the 4'-oxidized abasic site; Y, the normal abasic sites; ** The HPLC conditions are indicated as the superscripts: a, analytical anion-exchange HPLC; b, analytical RP-HPLC; *** the calculated values are in parentheses).
NMR experiments The NMR experiments were performed on custom-built 750 MHz and 591 MHz spectrometers at the Francis Bitter Magnet Laboratory. The raw data were transferred to a Silicon Graphic workstation and were processed using Felix 2001 (Acceryls Inc.). Proton chemical shifts were referenced to an internal standard, sodium 3-(trimethylsilyl)-1-propanesulfonate (TSP) at 0.00 ppm. NOESY (100 ms, 200 ms, and 400 mixing times), TOCSY (60 ms and 110 ms mixing times), and E-COSY were recorded at 25 °C in D_2O. For NOESY and TOCSY experiments, datasets of 4096 (t_2) × 512 (t_1) complex points were acquired with the spectral width of 8012.4 Hz (750 MHz) or 7022.5 Hz (591 MHz) in both dimensions and 32 scans per t_1 increment. For the E-COSY experiment, datasets of 4096 (t_2) × 1028 (t_1) complex points were acquired with 48 scans per t_1 increment. A pre-saturation pulse (2 s) was applied during the recycle delay period to suppress solvent signals in all NMR experiments carried out in D_2O. Watergate-NOESY experiments (200 ms and 400 ms mixing times) in 90% H_2O/10% D_2O were recorded at 4 °C. Datasets of 4096 (t_2) × 512 (t_1) complex points were acquired with spectral widths of 13020.8 Hz (591 MHz) in both dimensions and 32 scans per t_1 increment. For all the datasets, t_1 dimensions were linear-predicted to 4096 points and the data were processed with an exponential window function.

^1H-^13C GE-HSQC experiments(20) were acquired on the 591 MHz spectrometer at 25 °C with 1024 × 512 complex points and a spectrum width of 6009.6 Hz in the ^1H dimension and 16025.64 Hz in the ^13C dimension. For each t_1 increment, 64 transients were recorded. The ^13C dimension in the final datasets was linear-predicted to 1024 points. The spectrum was referenced indirectly through the transmitter frequency by external calibration on TSP.

Calculation of NMR restraints Distance restraints were derived from the volumes of the cross peaks in NOESY spectra. The peak volumes were calculated with the peak picking protocol in the Felix 2000 software package. Volumes of cross peaks between H2' and H2'' (1.772 Å) or between H5 and
H6 (2.460 Å) of cytosine were used to calibrate the proton-proton distances. The cross peak intensities were classified as weak, medium, or strong with the distance restraints of 1.5-3.0, 2.0-5.0, or 3.0-6.0 Å respectively. Alternatively, a more rigorous approach of iterative relaxation matrix analysis was applied to estimate the distance restraints using the program MARDIGRAS(21, 22). MARDIGRAS calculations were carried out on NOESY spectra acquired at various mixing times. For each NOESY cross peak, an upper and a lower distance restraint was calculated. The final distance restraints derived from MARDIGRAS calculation agreed well with manual classification. For 1, a total of 410 distance restraints were used for the subsequent structural calculation. Watson-Crick H-bond interactions (8) detected in the Watergate-NOESY experiments were included as restraints. A similar analysis on 4α and 4β gave 482 distance restraints. Dihedral angle restraints of the deoxyribose moieties were derived from the coupling constants measured in E-COSY experiments. A total of 43 sugar dihedral angle restraints were included for the structure refinement for 1 (a total of 57 for 4α and 4β, Table 5.2).

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>4α</th>
<th>4β</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pairwise RMSD (Å)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>starting structures</td>
<td>4.80</td>
<td>4.54</td>
<td>4.55</td>
</tr>
<tr>
<td>final 10 structures</td>
<td>1.72</td>
<td>1.70</td>
<td>1.56</td>
</tr>
<tr>
<td>Number of NMR Restraints</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NOE</td>
<td>410</td>
<td>482</td>
<td>482</td>
</tr>
<tr>
<td>Dihedral angles</td>
<td>43</td>
<td>57</td>
<td>57</td>
</tr>
<tr>
<td>Violations of NMR restraints</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NOE violations (Å)</td>
<td>0.056</td>
<td>0.14</td>
<td>0.11</td>
</tr>
<tr>
<td>Dihedral angle violations (°)</td>
<td>2.33</td>
<td>2.21</td>
<td>2.17</td>
</tr>
</tbody>
</table>

Table 5.2 Summary the statistical analysis of the final 10 structures: root-mean-square deviation (RMSD) (Å), NMR restraints, and NMR violations.
NMR structure calculation and refinement of 1, 4α, and 4β  All structural calculations were performed on an SGI Altix 3700 LINUX server with 128 Intel Itanium 1.3 GHz processors at the Scripps Research Institute (TSRI) using the SANDER module of AMBER 8.0(23). The starting model structures were built using the Biopolymer module of Insight 2000 (Accelrys Inc.). Three starting structures were constructed including an A-form, a B-form and a third form with arbitrary helical parameters(24). The root-mean square deviation (RMSD) for the three structures was ~ 4.5 Å. A generalized Born implicit solvent model was applied to simulate the solvent dielectric constant(25, 26). The initial model structures were energy minimized by 500 steps of conjugated gradient minimization followed by 500 steps of steepest descent minimization to remove high energy interactions introduced during the manual model building. The simulated annealing was carried out by heating the structure from 300 K to 900 K in 5 ps followed by cooling to 300K over the next 15 ps. During the cooling process, weak temperature coupling (slow cooling) was applied for the first 13 ps followed by strong temperature coupling (rapid cooling) for the last 2 ps(24). NMR restraints were enforced during the annealing process with an increasing force constant from 5 to 50 kcal/mol over the first 3 ps and was maintained at 50 kcal/mol for the rest of the calculation(24). The annealed structure was then subjected to a 100-ps constant temperature (300 K) molecular dynamics with NMR restraints. Ten structures for each starting structure (A-, B-, and the third form) were calculated by sampling conformations during the last 10 ps of the molecular dynamics at a 1-ps interval. The resulting 30 structures have an averaged RMSD of <2.0 Å, which indicates a structural convergence after refinement. Ten final structures were chosen from the 30 structures based on the fewest NMR violations to represent the ensemble of the averaged structure. The statistical analysis of the final ten structures is summarized in Table 5.2.

Molecular Dynamics Simulations with an Explicit Solvent Model The molecular dynamics
simulations were performed at a constant temperature (300 K) using the software AMBER 8.0 following the protocol described by Dupradeau et al. (27). The averaged structures were first solvated using the TIP3P solvent model with approximately 6,000 water molecules in an octahedral solvent box with a thickness of 10 Å. Twenty four sodium ions were also included in the calculations to neutralize the phosphates of the phosphodiester backbone. The molecular dynamics calculations were carried out on a 3-ns time scale. The conformation was sampled every 1 ps and was analyzed by the PTRAJ module in Amber 8.0.

5.3 Results

*Preparation of the NMR samples of 1 and 4* The synthesis and characterization of 1 have been described in detail in Chapter 4. The purity and the duplex formation of 1 were confirmed by anion-exchange HPLC analysis, which showed a single species with a retention time of 13.3 min (Figure 5.2 B), different from a 13mer ss-DNA control (10.2 min) (Figure 5.2 A).

To prepare the NMR sample of 4, ss DNA containing an abasic site (3) was synthesized by treating 2 with UDG following published procedures (19). Compound 3 was purified by semi-preparative RP-HPLC and was analyzed by MALDI-TOF MS (Table 5.1). A 1:1 ratio of 3 and the template, d(CCCGTTACTTTGG), was mixed, heated to 55 °C for 1 min, and annealed by cooling to 4 °C over 1 h to give 4. Analytical anion-exchange HPLC analysis of 4 showed a single species with a retention time of 13.4 min, similar to 1 (Figure 5.2 C).

*Assignments of proton chemical shifts* Standard TOCSY (60 and 110 mixing time) and NOESY (100, 200, and 400 ms mixing time) experiments were performed to assign non-exchangeable protons of DNA (28). The chemical shifts of protons associated with the deoxyribooses were assigned using TOCSY experiments and the results are summarized in Table 5.3 (1) and Table 5.4 (4).
Figure 5.2 Duplex DNA formation analyzed by analytical anion-exchange HPLC. A, a 13mer ss-DNA, d(CCCGGTACTTTTG), as a control; B, 1; C, 4.
### Table 5.3 

|       | H8/H6 | Me/H5 | H1' | H2' | H2'' | H3'  | H4'  | H5/5'' | -NH2, -NH, or H2 |
|-------|-------|-------|-----|-----|------|------|------|--------|-----------------
| C1    | 7.768 | 5.995 | 5.990 | 2.069 | 2.510 | 4.684 | 4.123 | 3.769 | 7.123/7.999    |
| A3    | 8.258 | 5.795 | 2.771 | 2.850 | 5.055 | 4.373 | 4.119/4.500 | 7.549          |
| A4    | 8.109 | 5.804 | 2.594 | 2.778 | 5.043 | 4.404 | 4.207 | 7.216          |
| X     |       | 5.155 | 2.104 | 1.827 | 4.559 | n.d.  | n.d.  |               |
| A8    | 8.385 | 6.335 | 2.815 | 2.948 | 5.020 | 4.411 | 3.993 |               |
| C10   | 7.355 | 5.515 | 5.515 | 1.848 | 2.224 | 4.871 | 4.207 | 4.106          |
| C14   | 7.820 | 5.996 | 6.010 | 2.205 | 2.554 | 4.687 | 4.364 | 3.804          |
| C16   | 7.443 | 5.672 | 5.587 | 2.036 | 2.378 | 4.856 | 4.207 | 4.102          |
| G18   | 7.637 | 5.925 | 2.204 | 2.474 | 4.874 | 4.212 | 4.080/4.056 | 12.787         |
| T19   | 7.155 | 5.843 | 2.074 | 2.397 | 4.815 | 4.850 | n.d.  |               |
| A20   | 8.144 | 6.112 | 2.562 | 2.711 | 4.931 | 4.369 | 4.141 |               |
| C21   | 7.552 | 5.540 | 5.936 | 2.515 | 2.717 | 4.916 | 4.162 | 4.086          |

(X, the 4'-oxidized abasic site; n.d., not determined due to either overlapping or weak signals).
Table 5.4 $^1$H chemical shift assignments (in ppm) of 4.

<table>
<thead>
<tr>
<th></th>
<th>H8/H6</th>
<th>Me/H5</th>
<th>H1'</th>
<th>H2'</th>
<th>H2''</th>
<th>H3'</th>
<th>H4'</th>
<th>H5''/-NH2, -NH, or H2</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>7.764</td>
<td>5.968</td>
<td>5.974</td>
<td>2.060</td>
<td>2.504</td>
<td>4.668</td>
<td>4.115</td>
<td>3.744/3.770</td>
</tr>
<tr>
<td>A3</td>
<td>8.264</td>
<td>5.773</td>
<td>2.770</td>
<td>2.843</td>
<td>5.047</td>
<td>4.365</td>
<td>4.399</td>
<td>7.342</td>
</tr>
<tr>
<td>A4</td>
<td>8.117</td>
<td>5.783</td>
<td>2.595</td>
<td>2.767</td>
<td>5.031</td>
<td>4.399</td>
<td>4.193</td>
<td>7.210</td>
</tr>
<tr>
<td>A5</td>
<td>7.926</td>
<td>5.906</td>
<td>2.428</td>
<td>2.692</td>
<td>4.971</td>
<td>4.360</td>
<td>4.188</td>
<td>4.148</td>
</tr>
<tr>
<td>A8-α</td>
<td>8.360</td>
<td>6.310</td>
<td>2.786</td>
<td>2.940</td>
<td>4.982</td>
<td>4.470</td>
<td>4.130</td>
<td>4.039</td>
</tr>
<tr>
<td>G11</td>
<td>7.830</td>
<td>5.441</td>
<td>2.632</td>
<td>2.671</td>
<td>4.935</td>
<td>4.271</td>
<td>n.d.</td>
<td>13.224</td>
</tr>
<tr>
<td>G12</td>
<td>7.736</td>
<td>5.786</td>
<td>2.590</td>
<td>2.739</td>
<td>4.976</td>
<td>4.378</td>
<td>4.143</td>
<td>4.115</td>
</tr>
<tr>
<td>C14</td>
<td>7.822</td>
<td>5.982</td>
<td>5.995</td>
<td>2.178</td>
<td>2.558</td>
<td>4.676</td>
<td>4.138</td>
<td>3.775/3.812</td>
</tr>
<tr>
<td>C15</td>
<td>7.657</td>
<td>5.694</td>
<td>5.995</td>
<td>2.211</td>
<td>2.470</td>
<td>4.861</td>
<td>4.206</td>
<td>4.062</td>
</tr>
<tr>
<td>G17</td>
<td>7.646</td>
<td>5.716</td>
<td>2.768</td>
<td>2.712</td>
<td>4.997</td>
<td>4.383</td>
<td>n.d.</td>
<td>13.192</td>
</tr>
<tr>
<td>G18-α</td>
<td>7.647</td>
<td>5.916</td>
<td>2.480</td>
<td>2.671</td>
<td>4.885</td>
<td>4.384</td>
<td>4.206</td>
<td>12.802</td>
</tr>
<tr>
<td>G18-β</td>
<td>7.628</td>
<td>5.916</td>
<td>2.480</td>
<td>2.671</td>
<td>4.885</td>
<td>4.384</td>
<td>4.206</td>
<td>12.802</td>
</tr>
<tr>
<td>T19-β</td>
<td>7.115</td>
<td>1.287</td>
<td>5.779</td>
<td>2.005</td>
<td>2.305</td>
<td>4.767</td>
<td>4.389</td>
<td>3.976/3.989</td>
</tr>
<tr>
<td>A20-α</td>
<td>8.133</td>
<td>6.091</td>
<td>2.553</td>
<td>2.679</td>
<td>4.916</td>
<td>4.363</td>
<td>4.123</td>
<td>4.125</td>
</tr>
<tr>
<td>C21-β</td>
<td>7.673</td>
<td>5.681</td>
<td>5.942</td>
<td>2.217</td>
<td>2.530</td>
<td>4.786</td>
<td>4.279</td>
<td>4.128</td>
</tr>
<tr>
<td>T23</td>
<td>7.518</td>
<td>1.720</td>
<td>6.102</td>
<td>2.149</td>
<td>2.566</td>
<td>4.890</td>
<td>4.245</td>
<td>4.130/4.190</td>
</tr>
</tbody>
</table>

(Y, an abasic site; n.d., not determined due to either overlapping or weak signals)
Thesequential connectivities between each base proton (H6 of pyrimidines, H8 of purines) and the deoxyribose protons (H1’, H2’/H2”’, H3’, H4’, and most of H5’) of the 5’ residue were established mainly by analyzing the 200 ms mixing time NOESY spectra. NOE connectivities between H5 and H6 protons of cytosines and between methyl and H6 protons of thymidines were also used to assist the assignments of NOE signals in overlapping regions.

The NOE connectivities between H1’ and the base protons of 1 are presented in Figure 5.3. The NOE connectivities on the 5’ and 3’ sides of the 4’-oxidized abasic site are shown in red lines and as expected, is disrupted between G6 and A8 (indicated by × and shown in dashed red lines). NOE connectivities are also observed on the other strand, with the connectivity from T19 to C21 opposite to the 4’-oxidized abasic site indicated by red lines (Figure 5.3). This result suggests that the base opposite to the 4’-oxidized abasic site (A20) adopts an intrahelical conformation, similar to that in B-form DNA.

The NOE connectivities between H1’ and the base protons of 4 are shown in Figure 5.4. The NOE connectivities adjacent to the abasic site are shown in red lines. Similar to 1, the NOE connectivity is disrupted by the abasic site between G6 and A8 (indicated by × and red dashed lines in Figure 5.4, top). Complete NOE connectivities are observed for residues on the opposite strand (Figure 5.4, bottom). However, in contrast to 1, two distinct sets of NOE cross peaks are observed for T19, A20, and C21 (red dashed lines in Figure 5.4, bottom), consistent with the α and β anomeric forms observed in previous studies with an abasic site in similar sequence context(19).
Figure 5.3 NOESY (200 ms mixing time) spectra of 1 at 750 MHz showing the connectivities between the base (H6/H8) and the H1' protons. Top spectrum, NOE connectivities on the first strand; bottom spectrum, NOE connectivities on the second strand. The connectivities adjacent to the lesion are highlighted in red. The disruption of the NOE connectivity on the first strand at the 4'-oxidized abasic site is denoted by X. Cross peak A, A8H8-C9H5; B, A20H8-C21H5.
Figure 5.4 NOESY (200 ms mixing time) spectra of 4 at 750 MHz showing the connectivities between the base (H6/H8) and the H1' protons. Top spectrum, the connectivities on the first strand; bottom spectrum, the connectivities on the second strand. The connectivities adjacent to the lesion are highlighted in red. The disruption of the NOE connectivity on the first strand at the abasic site is denoted by X (top). Two sets of signals on the second strand that are associated with 4α and 4β are indicated in red dashed lines. Cross peak A, A8H8-C9H5; B, A20H8-C21H5 (in 4α); C, A20H8-C21H5 (in 4β).
Exchangeable base protons (amino and imino protons) involved in the Watson-Crick base pair interaction were assigned by Watergate-NOESY experiments in 90% H$_2$O/10% D$_2$O at 4 °C. For both 1 and 4, 8 imino protons are detected and assigned based on their NOE interactions with their base pair partners (red dashed line, Figure 5.5). The chemical shifts of imino protons and the NOE patterns between the two abasic DNAs in the imino region (T-imino to A-H2 and G-imino to C-amino) are similar (Figure 5.5), which suggests that the two abasic DNAs adopt a similar global structure. The imino protons on each end of DNA (G13 and G26) and adjacent to the abasic site (G6 and T19) were not observed due to their fast exchange with solvent on the NMR time scale. These results suggest that the base pair interactions adjacent to the abasic site are weaker than those in regular B-form DNA.

**Figure 5.5** Assignments of imino and amino protons of 1 and 4 using Watergate-NOESY spectra (200 ms mixing time) at 591 MHz in 90% H$_2$O/10% D$_2$O (v/v) at 4 °C. The red dashed lines indicate the connectivities between the G-imino proton and C-amino protons. The cross peaks are assigned as: A, T24-imino/A3-H2; B, T22-imino/A5-H2; C, T23-imino/A4-H2; D, G12-imino/C15-amino; E, G11-imino/C16-amino; F, G17-imino/C10-amino; G, G25-imino/C2-amino; H, G18-imino/C9-amino.
Assignments of the 4'-oxidized abasic site of 4  Chemical shift assignments of protons associated with the abasic site are essential to identify NOE interactions between this site and adjacent residues. The proton chemical shift assignments have been obtained from standard TOCSY experiments(28). For 1, a proton with an unusual chemical shift was observed at 5.155 ppm was observed (X-H1', Figure 5.6 A). This chemical shift is considerably upfield from the envelop of H1' in B-DNA in which it ranges from 5.3 to 6.5 ppm (Table 5.3). The assignment of this proton to H1' of the 4'-oxidized abasic site was based on its through-bond connectivities with two other protons: one at 1.827 ppm and the other at 2.104 ppm (red line, Figure 5.6 A), which are assigned to H2'/H2'' of the same residue. Both the H2' and H2'' proton also show through-bond connectivity with a proton at 4.559 ppm, which is upfield shifted from all the H3' protons (Table 5.3). The inability to associate any of these protons with the completely assigned bases and their unusual chemical shifts suggest that they were associated with the 4'-oxidized abasic site. No evidence of a second anomer was apparent by any of the TOCSY experiments.

The chemical shift assignment of H1' at 5.155 ppm in the 4'-oxidized abasic site was also confirmed by an ¹H-¹³C GE-HSQC experiment (Figure 5.7 A). The proton signal at 5.155 ppm shows a cross peak with the carbon signal at 98.41 ppm. This carbon signal is consistent with the chemical shift previously assigned to C1’ in an abasic site(29) (Figure 5.7 A) and is downfield shifted relative to C1’ carbons in nucleotides (Table 5.5). The region containing cross peaks between base protons (H8 or H6) and their attached carbons (C8 or C6) is shown in Figure 5.7 B to show the quality of the ¹H-¹³C GE-HSQC spectrum of 1. The ¹³C chemical shift assignments of 1 are summarized in Table 5.5.
Figure 5.6 Assignments of the protons associated with the abasic sites using TOCSY spectra (60 ms mixing time) at 750 MHz. A, The through-bond connectivities between H1' and H2'/H2'' protons (red line) of the 4'-oxidized abasic site (X) in 1. B, The through-bond connectivities between H1' and H2'/H2'' protons (red line) of the 4'-abasic site (Y) in 4. The two sets of signals are associated with the α and β anomers revealed by NOESY experiments.
Figure 5.7 $^1$H-$^{13}$C GE-HSQC spectrum of 1. A, the region showing cross peaks between H5 and C5 of cytosines. The arrow indicates the cross peak between H1' and C1' of the 4'-oxidized abasic site (4'-oxAb). B, the regions showing cross peaks between base protons (H6 of pyrimidines or H8 of purines) and their attached carbon atoms (C6 of pyrimidine or C8 of purine).
Table 5.5 $^{13}$C chemical shift assignments (in ppm) of 1.

<table>
<thead>
<tr>
<th></th>
<th>Cl'</th>
<th>H8/H6</th>
<th>H5/Me</th>
<th></th>
<th>Cl'</th>
<th>H8/H6</th>
<th>H5/Me</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>87.03</td>
<td>143.27</td>
<td>99.56</td>
<td>G26</td>
<td>84.82</td>
<td>138.85</td>
<td></td>
</tr>
<tr>
<td>C2</td>
<td>86.60</td>
<td>143.59</td>
<td>98.93</td>
<td>G25</td>
<td>84.19</td>
<td>138.74</td>
<td></td>
</tr>
<tr>
<td>A3</td>
<td>n.d.</td>
<td>141.69</td>
<td></td>
<td>T24</td>
<td>85.24</td>
<td>139.80</td>
<td>14.61</td>
</tr>
<tr>
<td>A4</td>
<td>n.d.</td>
<td>141.06</td>
<td></td>
<td>T23</td>
<td>n.d.</td>
<td>139.90</td>
<td>14.61</td>
</tr>
<tr>
<td>A5</td>
<td>n.d.</td>
<td>140.96</td>
<td></td>
<td>T22</td>
<td>n.d.</td>
<td>137.90</td>
<td>14.61</td>
</tr>
<tr>
<td>G6</td>
<td>85.34</td>
<td>138.21</td>
<td></td>
<td>C21</td>
<td>88.08</td>
<td>142.96</td>
<td>98.72</td>
</tr>
<tr>
<td>X</td>
<td>98.41</td>
<td></td>
<td></td>
<td>A20</td>
<td>85.45</td>
<td>141.96</td>
<td></td>
</tr>
<tr>
<td>A8</td>
<td>n.d.</td>
<td>141.69</td>
<td></td>
<td>T19</td>
<td>85.35</td>
<td>138.64</td>
<td>14.29</td>
</tr>
<tr>
<td>C9</td>
<td>n.d.</td>
<td>142.22</td>
<td>98.41</td>
<td>G18</td>
<td>84.72</td>
<td>137.48</td>
<td></td>
</tr>
<tr>
<td>C10</td>
<td>86.61</td>
<td>143.27</td>
<td>98.83</td>
<td>G17</td>
<td>84.19</td>
<td>138.43</td>
<td></td>
</tr>
<tr>
<td>G11</td>
<td>84.29</td>
<td>138.53</td>
<td></td>
<td>C16</td>
<td>86.61</td>
<td>143.38</td>
<td>99.04</td>
</tr>
<tr>
<td>G12</td>
<td>84.30</td>
<td>138.11</td>
<td></td>
<td>C15</td>
<td>88.20</td>
<td>143.48</td>
<td>99.14</td>
</tr>
<tr>
<td>G13</td>
<td>84.82</td>
<td>138.74</td>
<td></td>
<td>C14</td>
<td>88.19</td>
<td>143.27</td>
<td>99.67</td>
</tr>
</tbody>
</table>

(Y, an abasic site; n.d., not determined due to either overlapping or weak signals)
Assignments of the abasic site protons of 4α and 4β. The analysis of the TOCSY spectra of 4 indicates two signals at 5.311 ppm and 5.385 ppm associated with the H1’ protons of the normal abasic site (Y-H1’, Figure 5.6 B). As in the case of the 4’-oxidized abasic site of 1, these protons are also upfield-shifted relative to all of the H1’ protons in B-form DNA (Table 5.4). These two signals show through-bond connectivities with two sets of H2'/H2” protons (red lines, Figure 5.6 B). These signals are also not associated with any base (all of which have been assigned, Table 5.4), suggesting that they are unique to the abasic site. The existence of two sets of distinct signals are consistent with two anomeric states of the abasic site in solution(29, 30).

Assignments of the α and β anomers of 1 and 4 using NOESY experiments. Our group has previously reported the assignments of each anomer of an abasic site based on the relative NOE intensities of H2'/H2” to H1’(19). The assignments of H2’ and H2” protons were based on the intensity of each cross peak with H3’, i.e., H2’ has a stronger cross peak than H2”. The assignment of the anomeric configuration is based on the relative cross peak intensities between H1’-H2’ and H1’-H2”. For example, with the α anomer, H1’ is closer to H2’ and produces a stronger NOE cross peak than to H2”. The reverse is true for the β anomer.

Analysis of the TOCSY spectra of 1 indicates that only one major species. The cross peak intensities in its NOESY spectrum suggest that the H1’ proton is closer to the H2’ proton (B, Figure 5.8A) than to the H2” proton (A, Figure 5.8A). This result is consistent with the α anomer. The assignment of a single conformation is also consistent with the NOE cross peak patterns observed in Figure 5.3, which show a single set of signals adjacent to the 4’-oxidized abasic site (solid red lines, Figure 5.3).
Figure 5.8 Determining the anomeric state using the relative intensity of NOE cross peaks associated with the protons in the abasic sites. A, for 1, one set of signals are present and the connectivities are highlighted in red. The intensity of the NOE cross peak between H1' and H2'' (cross peak A) is less than that between H1' and H2' (cross peak B), which suggests that this species is an α anomer. B, for 4, two sets of signals are present (labeled in red and blue, respectively). For the species labeled in red, the intensity of the NOE cross peak between H1' and H2'' (cross peak C) is less than that between H1' and H2' (cross peak D), suggesting that this species is an α anomer; for the species labeled in blue, the intensity of the NOE cross peak between H1' and H2'' (cross peak F) is stronger than that between H1' and H2' (cross peak E), suggesting that this species is a β anomer.
In contrast to 1, the TOCSY experiments suggest that there are $\alpha$ and $\beta$ anomers of the abasic site ($4\alpha$ and $4\beta$). These two species are assigned using the same strategy described for 1. The appropriate 200 ms NOESY spectrum is shown in Figure 5.8B. The $\alpha$ anomer is indicated in red and the $\beta$ anomer is indicated in blue. The chemical shift assignments of $4\alpha$ and $4\beta$ are very similar to the previously reported values for the protons associated with a normal abasic site in a similar sequence context(19).

*Establishing sugar conformations in 1 and 4.* E-COSY experiments at 750 MHz were performed to measure the coupling constants used to assign the sugar puckers for the deoxyribose moieties in 1, $4\alpha$ and $4\beta$ (Table 5.6 A and B, respectively). For 1, the coupling constants of H1'-H2' and H1'-H2'' were measured for all residues except for the 4'-oxidized abasic site. The observed coupling constants are consistent with a C2'-endo or S-type conformation for these residues, which are expected in B-form DNA(31). For the 4'-oxidized abasic site, neither $J_{\text{H1'H2'}}$ nor $J_{\text{H1'H2''}}$ could be determined accurately due to the low intensities of the cross peaks. Therefore, the sugar conformation of the 4'-oxidized abasic site is not constrained during the NMR structure calculation.

For $4\alpha$ and $4\beta$, the coupling constants were measured for all residues (Table 5.6 B). Except for the abasic site, the coupling constants are consistent with a C2'-endo or S-conformation (31) expected for a regular B-form DNA. The coupling constants associated with abasic site protons (Figure 5.9 and Table 5.6) are similar to those reported for a normal abasic site in a similar sequence context(19). Comparison of the experimentally determined coupling constants with the calculated values for different sugar conformations(31) suggests that the $\alpha$ anomer adopts a O1'-endo conformation and that the $\beta$ anomer adopts a near S-type C1'-exo conformation.
Figure 5.9 Coupling constants between H1' and H2'/H2'' measured by E-COSY experiments of the abasic sites in 4. The chemical shift assignments of these abasic protons are also shown.

Table 5.6 Coupling constants (in Hz) for 1 (A) and 4 (B) derived from E-COSY experiments*.

<table>
<thead>
<tr>
<th>Table 5.6A</th>
<th>C1</th>
<th>C2</th>
<th>A3</th>
<th>A4</th>
<th>A5</th>
<th>G6</th>
<th>X</th>
<th>A8</th>
<th>C9</th>
<th>C10</th>
<th>G11</th>
<th>G12</th>
<th>G13</th>
</tr>
</thead>
<tbody>
<tr>
<td>J_{H1',H2'}</td>
<td>8.2</td>
<td>10.4</td>
<td>9.6</td>
<td>8.8</td>
<td>8.7</td>
<td>9.3</td>
<td>n.d.</td>
<td>8.2</td>
<td>8.8</td>
<td>8.2</td>
<td>10.9</td>
<td>11.5</td>
<td>10.1</td>
</tr>
<tr>
<td>J_{H1',H2''}</td>
<td>6.6</td>
<td>7.4</td>
<td>7.2</td>
<td>6.8</td>
<td>7.0</td>
<td>6.7</td>
<td>n.d.</td>
<td>6.9</td>
<td>7.3</td>
<td>6.8</td>
<td>7.1</td>
<td>7.6</td>
<td>6.8</td>
</tr>
<tr>
<td>J_{H1',H2'}</td>
<td>G26</td>
<td>G25</td>
<td>T24</td>
<td>T23</td>
<td>T22</td>
<td>C21</td>
<td>A20</td>
<td>T19</td>
<td>G18</td>
<td>G17</td>
<td>C16</td>
<td>C15</td>
<td>C14</td>
</tr>
<tr>
<td>J_{H1',H2''}</td>
<td>9.5</td>
<td>11.2</td>
<td>10.7</td>
<td>10.2</td>
<td>8.5</td>
<td>8.6</td>
<td>8.8</td>
<td>9.4</td>
<td>10.5</td>
<td>10.9</td>
<td>9.5</td>
<td>8.9</td>
<td>8.0</td>
</tr>
<tr>
<td>J_{H1',H2''}</td>
<td>7.4</td>
<td>6.6</td>
<td>7.8</td>
<td>7.2</td>
<td>7.0</td>
<td>6.8</td>
<td>6.7</td>
<td>7.0</td>
<td>6.9</td>
<td>n.d.</td>
<td>7.2</td>
<td>6.8</td>
<td>6.5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 5.6B</th>
<th>C1</th>
<th>C2</th>
<th>A3</th>
<th>A4</th>
<th>A5</th>
<th>G6</th>
<th>Y</th>
<th>A8</th>
<th>C9</th>
<th>C10</th>
<th>G11</th>
<th>G12</th>
<th>G13</th>
</tr>
</thead>
<tbody>
<tr>
<td>J_{H1',H2'}</td>
<td>8.1</td>
<td>9.8</td>
<td>9.6</td>
<td>9.6</td>
<td>9.4</td>
<td>8.8</td>
<td>4.3(α)</td>
<td>5.8(β)</td>
<td>8.1</td>
<td>8.1</td>
<td>9.0</td>
<td>8.9</td>
<td>9.9</td>
</tr>
<tr>
<td>J_{H1',H2''}</td>
<td>6.5</td>
<td>6.0</td>
<td>6.6</td>
<td>6.3</td>
<td>6.2</td>
<td>6.6</td>
<td>1.4(α)</td>
<td>4.5(β)</td>
<td>6.4</td>
<td>7.4</td>
<td>7.6</td>
<td>7.3</td>
<td>7.6</td>
</tr>
<tr>
<td>J_{H1',H2'}</td>
<td>G26</td>
<td>G25</td>
<td>T24</td>
<td>T23</td>
<td>T22</td>
<td>C21</td>
<td>A20</td>
<td>T19</td>
<td>G18</td>
<td>G17</td>
<td>C16</td>
<td>C15</td>
<td>C14</td>
</tr>
<tr>
<td>J_{H1',H2''}</td>
<td>8.8</td>
<td>9.6</td>
<td>9.8</td>
<td>9.2</td>
<td>8.2</td>
<td>8.1(α)</td>
<td>8.0(β)</td>
<td>8.8(α)</td>
<td>8.9(β)</td>
<td>7.6</td>
<td>8.8</td>
<td>8.3</td>
<td>9.9</td>
</tr>
<tr>
<td>J_{H1',H2''}</td>
<td>6.1</td>
<td>6.0</td>
<td>7.7</td>
<td>6.0</td>
<td>7.0</td>
<td>7.7(α)</td>
<td>n.d.(β)</td>
<td>6.7(α)</td>
<td>6.8(β)</td>
<td>7.2</td>
<td>6.4</td>
<td>n.d.</td>
<td>7.2</td>
</tr>
</tbody>
</table>

(* X, the 4'-oxidized abasic site; Y, the normal abasic site; n.d., not determined due to overlapping or weak signals).
NOE interactions associated with abasic site protons of 1 and 4. A number of through-space interactions were detected in the NOE experiments (100 ms mixing time) between A8H8 and H1', H2'/H2'' and H3 of either the abasic site (4α and 4β) or the 4'-oxidized abasic site (1) (Figure 5.10 A-C and Table 5.7). These NOE interactions provide crucial insight into the conformation of the abasic sites. For 1, Figure 5.10A shows a medium and a weak NOE signal of A8H8-XH2'' and A8H8-XH2'', respectively. In B-form DNA, the distances between H2'/H2'' to the neighboring 3' base proton are 3.3Å/2.3 Å, respectively, which produce medium NOE signals. This is illustrated by the NOE signals between A3H8 and C2H2'/H2'' of 1 in Figure 10A (red line). The unusually weak interaction between A8H8 and the H2'' proton of the 4'-oxidized abasic site suggests that this lesion adopts a distinct conformation from that of B-form DNA.

For both 4α and 4β, weak NOE interactions are observed between A8H8 and YH2'/H2'' (Figure 5.10B, arrow). This result contrasts with that expected for the same interaction in generic B-form DNA compared to the similar cross peaks in B-form DNA where the signal is of medium strength. (Figure 5.10B, red line). In addition, a medium A8H8–YH3' NOE signal is observed for both 4α and 4β (Figure 5.10C). In B-form DNA, the distance between H3' to the adjacent 3' base proton is ~4.4 Å, which usually produces a weak NOE signal as demonstrated by the NOE cross peak between A3H8 and C2H3' in Figure 5.10C (the red arrow in Figure 5.10C). These unusual NOE interaction patterns suggest that the conformation of the abasic site sugar in 4 is distinct from B-form DNA.

Table 5.7 NOE restraints involving the protons associated with the abasic sites to their adjacent base protons.

<table>
<thead>
<tr>
<th>NOE cross peak</th>
<th>Intensity</th>
<th>Distance in final structure (Å)</th>
<th>NOE cross peak</th>
<th>Intensity</th>
<th>Distance in final structure (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>XH2'-A8 H8</td>
<td>medium</td>
<td>2.59</td>
<td>YH2'-A8 H8</td>
<td>weak</td>
<td>3.38(α) 4.24(β)</td>
</tr>
<tr>
<td>XH2''-A8 H8</td>
<td>weak</td>
<td>4.22</td>
<td>YH2''-A8 H8</td>
<td>weak</td>
<td>3.67(α) 4.15(β)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>YH3'-A8 H8</td>
<td>medium</td>
<td>2.69(α) 2.94(β)</td>
</tr>
</tbody>
</table>

(* X, the 4'-oxidized abasic site; Y, the normal abasic site)
Figure 5.10 NOE interactions associated with the protons of the abasic site and the adjacent 3'-base proton (A8H8) studied by NOESY (100 ms mixing time) at 750 MHz. A, A8H8 of 1 shows a medium NOE with the H2' proton of the 4'-oxidized abasic site (X-H2') and a weak NOE with the X-H2" proton (indicated by the arrows). The NOE between A3H8 and C2H2'/H2" are indicated by the red line. B, the A8H8 protons of 4α and 4β show weak NOE interactions with the H2'/H2" protons of the abasic site (Y-H2'/2") as indicated by the arrows. The NOE between A3H8 and C2H2'/H2" are indicated by the red line. C, the A8H8 protons of 4α and 4β show medium NOE interactions with the H3' proton of the abasic site (Y-H3') as indicated by the arrows. The NOE between A3H8 and C2H3' in 4 is indicated by the red arrow.
The averaged structures The experimentally determined distance and dihedral angle constraints were used to calculate an ensemble of ten structures of 1, 4α, and 4β. None of these structures contain any NOE distance violations larger than 0.5 Å or dihedral angle violations larger than 5°. Overlays of these ten structures for 1, 4α, and 4β are shown in Figure 5.11A. The averaged structures are shown in Figure 5.11B. The structures share two features in common: they are B-form DNA and the base (A20) opposite to the lesion (colored in blue, Figure 5.11A) is intrahelical and stacks between adjacent base pairs, as in B-form DNA (see also in Figure 5.12A-C).

Sugar conformations at the abasic sites Despite the overall similarities in global conformations of 1, 4α, and 4β, the sugar conformations at the abasic site are distinct as revealed in the 10 structures (Figure 5.11A, red) and the averaged structures (Figure 5.11B and 5.12). For 1, the 4'-oxidized deoxyribose adopts an intrahelical conformation (Figure 5.12 A). This conformation requires a relatively short (2.59 Å) H2’-A8 H8 and long (4.22 Å) H2’’-A8H8 distances (Figure 5.13 B, dashed red line) compared to those in a B-form DNA (Figure 5.13 A, dashed red line). This conformation is consistent with the intensities of NOE cross peaks involving these protons (Figure 5.10A).

In contrast to 1, for the both 4α and 4β, the deoxyribose of the abasic site is partially extrahelical (Figure 5.12B and C). The extrahelical conformation predicts that the H2’/2’’ protons of the abasic site move away from A8H8 in both 4α and 4β (red dashed lines, Figure 5.14 B and D) compared to that in B-form DNA (Figure 5.14 A and C). The observed conformation is dictated by and consistent with the weak NOE interactions involving A8H8 and the H2’/2’’ protons of the abasic site (Figure 5.10 B). In addition, in this conformation, the H3’ proton of the abasic site is much closer to A8 H8 (blue dashed lines in Figure B and D) than that in B-form DNA (blue dashed lines in Figure A and C), also consistent with the observed medium NOE signals between A8H8 and H3’ of the abasic site (Figure 5.10 C).
Figure 5.11 NMR structural refinement of 1, 4α, and 4β. A, an overlay of the ten final structures; B, the final averaged structures.
Figure 5.12 The stereo view of the abasic site and the neighboring base pairs in the final structures.

A, 1; B, 4α; C, 4β. The abasic site and the opposite base (A20) are colored in CPK.
Figure 5.13  The conformation of the 4'-oxidized abasic site in 1 compared to that in a theoretical B-form DNA. A, The computer-modeled conformation of the 4'-oxidized abasic site in a B-form DNA. The distance between A8H8 and the H2'/H2” protons are 3.3/2.3 Å, respectively (indicated by the red dashed lines). B, The conformation of the 4'-oxidized abasic site in the final structure. The distance between A8H8 and the H2'/H2” protons are 2.59/4.22 Å, respectively (indicated by the red dashed lines).
Figure 5.14  The conformation of the normal abasic site in 4α and 4β compared to that in a theoretical B-form DNA. The distance between A8H8 and Y-H2'/2'' is colored in red and the distance between A8H8 and Y-H3' is colored in blue (Y, the abasic site). A, The computer-generated conformation of the abasic site (α anomer) in a B-form DNA. The distance between A8H8 and Y-H2'/H2'' is 3.3/2.3 Å. The distance between A8H8 and Y-H3' is 4.4 Å. B, The conformation of the abasic site in the final structure of 4α. The distance between A8H8 and Y-H2'/H2'' is 3.4/3.7 Å. The distance between A8H8 and Y-H3' is 2.7 Å. C, The computer-generated conformation of the abasic site (β anomer) in a B-form DNA. The distances between A8H8 and H2'/H2''/H3' are the same as in A. D, The conformation of the abasic site in the final structure of 4β. The distance between A8H8 and Y-H2'/H2'' protons is 4.2/4.2 Å. The distance between A8H8 and Y-H3' is 2.9 Å.
**Molecular dynamics simulations** To probe the conformation flexibility at each abasic site, the final averaged structure of 1, 4α, and 4β was subjected to molecular dynamics simulations. Water molecules and counter ions (Na⁺) were included in the calculation to better simulate the solution state of DNA molecules. To monitor the time-dependent extrahelical motion, the distances between C1’ of the abasic site and the N1 atom of the opposite base (A20 N1) in 1, 4α, or 4β were monitored during the course of the simulation (Figure 5.15 A-C). In B-form DNA, the C1’-N1 distance, ~6.2 Å, is restrained by Watson-Crick base pairing. As an example, the distance between T24-C1’ and A3-N1 in 1, 4α, or 4β (blue traces in Figure 5.15 A-C) was monitored during the simulation and showed a very small fluctuation (<0.5 Å) from 6 Å. In contrast, for the normal or 4’-oxidized abasic site, because of the lack of the base-pairing interactions, the distance between C1’ of the lesion and N1 of A20 is sensitive to the motion of the abasic site (red traces in Figure 5.15 A-C, see the discussion below). Specifically, the extrahelical motion of the abasic site would likely cause an increase of this distance.

The analysis of the molecular dynamics simulation of 1 reveals that the 4’-oxidized abasic site is mainly intrahelical (the red trace in Figure 5.15 A). The average distance between C1’ of the 4’-oxidized abasic site and N1 of A20 is ~ 6 Å, similar to the value in a normal B-form DNA. However, the extrahelical motion of this site is also present during the simulation indicated by the increase of the distance between C1’ and A20-N1 (Figure 5.15 A). The fluctuation of this distance is less than 2.5 Å. Interestingly, transient H-bond interaction between 1’-OH of the 4’-oxidized abasic site with N1 of A20 is also observed, resulting in the decrease of the distance between C1’ of the lesion and A20-N1 (arrow in Figure 5.15 A). A stereo view of this transient H-bond interaction is presented in Figure 5.15 D.
Figure 5.15 Time-dependent extrahelical motion of the abasic site in 1, 4α, and 4β. The extrahelical motion was monitored by the distance change between Cl' of the abasic site and the N1 atom of the opposite base (A20 N1) (red trace). The blue traces represent the distance between A3N1 and T24Cl'. A, 1. The arrow in A indicates the transient H-bond formation between 1'-OH and A20N1. B, 4α. C, 4β. D, A stereo view of the snap shot of 1 with a transient H-bond between 1'-OH and A20N1 (indicated by the arrow in A).
For both 4α and 4β, the analysis indicates that there is a larger extrahelical motion associated with the abasic site in comparison with 1 as reflected by the large fluctuation of the distance between C1’ of the abasic site and N1 of A20 (> 4 Å) (Figure 5.15 B and C, red trace). In addition, the extrahelial motion occurs approximately every 50 ps (Figure 5.13 B and C), much more frequent than that in 1.

5.4 Discussion

The long-range goal of this structural study is to understand how both ss and ds lesions generated by BLM can be recognized and repaired by DNA repair enzymes. BLM generates two types of ss-DNA lesions: a 3’-PG/5’-P gap and a 4’-oxidized abasic site (Figure 5.1)(6-8). Since the 5’-GTAC-3’ sequence has been identified to be a “hot spot” for BLM-mediated ds cleavage(32), this sequence has been used in structural studies of DNA lesions by our group. A 3’-PG/5’-P gapped lesion in this sequence context (5’-G -P G P-AC-3’) has been synthesized and its structure has been characterized by 2D NMR methods(33). No NOE information was observed between the 3’PG group and adjacent residues, suggesting a flexible conformation of this lesion. In addition, the interaction between 5’-G -P G P-AC-3’ and BLM-Co(III)-OOH studied by 2D NMR offers structural insight into the mechanism of BLM-mediated ds-DNA cleavage(34). Our group has also carried out extensive 2D NMR spectroscopic studies of a normal abasic site in the 5’-GTAC-3’ sequence as a structural analog for the 4’-oxidized abasic site(19). Structural modeling based on NMR-derived distance and dihedral angle restraints suggested that the deoxyribose at the abasic site adopted an extrahelical conformation in both the α and the β anomer(19). The extrahelical conformation of the abasic site in this sequence could be used as the structural element for recognition by Ape1, the major DNA repair enzyme proposed in vivo to repair the abasic sites(17).
In this study, we present the first structure model of duplex DNA containing a 4′-oxidized abasic site (1) using 2D NMR methods. The overall structure of 1 is similar to B-form DNA based on distance and dihedral angle restraints calculated from 2D NMR experiments. The deoxyribose of the lesion adopts an intrahelical conformation, however, its conformation is different from that in B-form DNA. The conformation is largely derived from two unique NOE interactions between H2′/H2″ of the lesion and the base proton 3′ to the lesion. Molecular dynamics simulations were carried out to characterize the conformational flexibility of the 4′-oxidized abasic site. The results indicate that the deoxyribose at the lesion remains largely intrahelical with infrequent extrahelical motion (every ~500 ps).

As a structural comparison, we also determined the structure of a duplex DNA with a normal abasic site (4) in the same sequence context as 1 (5′-CCAAAGYACCGGG-3′). The results indicate that there are two species in solution for 4, which are assigned to the α and β anomers based on the NOE interactions associated with the sugar protons at the abasic site. Structural models of each anomer using these restraints have been determined. In both structures, the deoxyribose of the lesion adopts an extrahelical conformation. This conformation is largely dictated by observed NOE interactions between the sugar protons of the lesion and the base proton 3′ to the lesion. These results are consistent with our previous studies of an abasic site in a very similar sequence context (5′-CCAAAGYACTGGG-3′)(19). Molecular dynamics simulations carried out on each anomer revealed that the normal abasic site is conformationally more flexible than the 4′-oxidized abasic site in this particular sequence. Large extrahelical motions are observed at greatly increased frequency relative to the 4′-oxidized abasic site (every ~50 ps).

The difference in the structure and conformational flexibility between the 4′-oxidized abasic site and the normal abasic site may have important implications for their recognition by DNA repair.
enzymes, such as Ape1. The crystal structure of Ape1 in complex with an abasic site analog (THF) shows that the THF moiety is flipped completely out of helix (Figure 1.10, Chapter 1) and into a specific binding pocket of Ape1(35). A similar extrahelical conformation has been observed in the structures of a number of abasic site-protein complexes, including Endo IV(36), uracil glycosylase(37), and 3-methyl adenine glycosylase(38). However, it remains unclear whether the repair enzyme selectively recognizes the flipped-out abasic site or whether it actively flips the abasic site by recognizing conformational flexibility around the abasic site. Our results for the normal abasic site suggest that in solution, the deoxyribose of the abasic site is flexible and adopts a wide range of extrahelical conformations. Therefore, it is likely that the pre-existing extrahelical conformation of the abasic site, although it might only constitute a small fraction of the conformation space in solution, could be recognized by DNA repair enzymes.

Compared to the normal abasic site, our study suggests that the 4'-oxidized abasic site in 1 is largely intrahelical and is less flexible. Biochemical studies on the repair of the 4'-oxidized abasic site (in a mixture with 3'-PG/5'P gaps generated by BLM) by Ape1 suggest that this lesion is repaired 3 times slower than the normal abasic site in the same sequence context(16). The rate-limiting step of the catalysis by Ape1 remains unclear. Based on the greatly reduced rate on the substrate analog containing a phosphothioate group (39), Strauss et al. proposed that the chemical step of the hydrolysis of the phosphodiester bond (or some proceeding step) is rate-limiting(40). However, the rate of the product release has not been determined and it is possible that the substrate analog changes the rate-limiting step of the catalysis by Ape1. Therefore, it is not clear whether the observed rate reflects the recognition of the lesion.

The 4'-oxidized abasic sites belong to a family of DNA lesions with an oxidized deoxyribose(41). This type of DNA damage is proposed to be repaired via the BER pathway(17).
The structure of another oxidized abasic site lesion, a 2'-deoxyribonolactone (also known as the 1'-oxidized abasic site), has been solved using 2D NMR methods(42). In that study, the 2'-deoxyribonolactone lesion (L) was incorporated into a 11mer duplex DNA, 5'-ACLCA-3'/3'-TGTGT-5'. The structural modeling revealed that the lactone ring at the abasic site adopted an intrahelical conformation based on two medium NOE interactions between H2'/H2'' of L and the 3' base proton. The base (T) opposite L is intrahelical, as suggested by the unperturbed NOE connectivities between the base protons and the H1', H2'/H2'' protons. Interestingly, in contrast to 1 and 4, stable bas-pair interactions were observed on both sides of L as revealed by the observed imino protons. Kinetic analysis of the repair of L in a 30mer duplex (5'-CALAC-3'/3'-GTATG-5') DNA by Ape1(43) showed that a $k_{cat}$ for L (2.3 s$^{-1}$) is similar to that for a normal abasic site (2.4 s$^{-1}$) in the same sequence context with 5 fold higher $K_m$ for L (98 nM compared to 21nM for the abasic site). Since the structural information of L or the abasic site in the 5'-CALAC-3'/3'-GTATG-5' is not available, no conclusions can be drawn about the importance of the conformational flexibility in recognition by Ape1. The structural studies described in this chapter present a starting point to probe the conformational diversity of different abasic site lesions. Further studies on a fast time scale in pre-steady state should yield information that can correlate the difference in structures of abasic sites to the recognition and repair by Ape1.

5.5 References

rate-determining 4'-hydrogen abstraction from poly(dA-dU) associated with the formation of both free base and base propenal. *Biochemistry* 24, 7562-7568.


of the DNA repair enzyme endonuclease IV and its DNA complex: double-nucleotide flipping at abasic sites and three-metal-ion catalysis. *Cell* 98, 397-408.


Chapter 6. Influence of the base (N) opposite an abasic site (X) on the conformation of

d(5’-CCAAAGXACCGGG-3’)/d(3’-GGTTTCNTGGCCC-5’)

6.1 Introduction

The integrity of the genome is under constant threat of endogenous and exogenous DNA damaging agents, which produce a variety of DNA damage products. Our group’s interests in understanding the mechanism of bleomycin (BLM)-mediated DNA damage have led us to the study of the repair of the DNA lesions generated by BLM. Specifically, we are interested in how the lesions generated by BLM are recognized by DNA repair enzymes and whether the difference in the efficiency of removing those lesions would affect BLM’s cytotoxicity among different tumor cell lines. In Chapter 5, we carried out the first structural characterization of duplex DNA containing a 4’-oxidized abasic site and compared its structure to the structure of duplex DNA containing a normal abasic site in the same sequence context. In this chapter, in an effort to understand the influence of the sequence context on the local conformation of a normal abasic site, we describe the first systematic structural characterization of duplex DNA containing an abasic site (X) with four different bases (A, G, T, or C) opposite the lesion in the d(5’-CCA AAG XAC CGG G-3’) sequence.

The normal abasic sites (also called apurinic/apyrimidinic sites, or Ap sites) are common DNA lesions. Lindahl estimated that there is a constant level of ~10,000 abasic sites for a typical human cell(1). Most of the abasic sites are believed to result directly from spontaneous depurination(1), or indirectly from deamination of cytosine to uracil, which is then eliminated by uracil glycolylases(2). Abasic sites also result from hydrolysis of oxidized or alkylated bases by lesion-specific glycosylases(3-5).

The abasic sites in DNA are chemically labile lesions. In solution, the abasic sites exist as a mixture of ring-closed α and β hemiacetals, ring-opened aldehyde, and
aldehyde hydrate(6, 7). The major components of the mixture are α and β hemiacetals, with the aldehyde constituting < 1% of the mixture(6, 7). However, the aldehyde form is susceptible to base-catalyzed β-elimination, which leads to a strand break(8). The half-life of an abasic site in duplex DNA has been measured by our group and Sheppard’s group in 5 different sequence contexts and ranges from 200 to 900 h under physiological conditions(pH 7.5, 150 mM NaCl at 37 °C)(9, 10).

The unrepaired abasic sites are mutagenic(11). Studies on abasic sites as the template for DNA replication suggest that several DNA polymerases (including Pol I and III from E. coli, Pol α from Drosophila, and T4 DNA polymerase) preferentially incorporate dA opposite an abasic site(12-14). The selective incorporation of dA opposite an abasic site is often referred to as the “A rule” (11). In addition, the abasic sites also cause frameshifts in the replication product generated during translesion DNA synthesis catalyzed by Y-family DNA polymerases(15-18).

In vivo, both the abasic sites and 4’-oxidized abasic sites are thought to be repaired via the base excision repair pathway, which is initiated by apurinic/apyrimidinic endonuclease 1 (Ape1) in humans(19, 20). Crystallographic studies from the Tainer group on Ape1 complexed with duplex DNA containing an abasic site analog (tetrahydofuran, or THF) have shed light on the structural features that might govern the specific interactions between this protein and the abasic site lesions (Chapter 1)(21). The unusual structural features for the THF-containing DNA in the DNA-enzyme complexes, i.e., large bending of the DNA helical axis, extrahelical flipping of the abasic site (THF) relative to the phosphodiester backbone, and penetration of protein side chains into the regions adjacent to the abasic site, have been observed in several other structures of
whether the conformation of the abasic site in solution is inherently perturbed and recognized by the repair enzymes or whether the repair enzymes sense flexibility and alter DNA structures in the region of the lesion remains unclear.

Solution structural studies by 2D NMR methods and molecular modeling on oligonucleotides containing site-specifically incorporated abasic sites suggest a different conformation of DNA in solution relative to those of proton-bound in the crystal structure. Recent studies from our group on d(5'-CCAAAGXACTGGG-3', X denotes the abasic site) indicated that the deoxyribose moiety was partially extrahelical and that the overall structure was B-form without significant perturbation(24). Studies of several different sequences containing the abasic site (X) (5'-GAXAC-3’, 5’-AGXCA-3’, and 5’-TTXTT-3’) from the Bolton group suggested that the local conformation around the abasic site is influenced by the identity of the flanking bases and the base opposite the abasic site(25-27). Because the opposite and flanking bases are proposed to participate in the interaction with repair enzymes(21-23), elucidating their roles in determining the local conformation of the abasic site is essential to understanding the recognition of abasic sites by the Ap endonucleases. However, no systematic efforts have been undertaken to investigate the modulation of the abasic site conformation by these adjacent bases.

In the previous chapter, the solution structure of a duplex DNA with an abasic site in the sequence of 5’-CCA AAG XAC CGG G-3’ with an A opposite the abasic site was investigated in detail as a comparison to the structure of a 4’-oxidized abasic site in the same sequence context. In this chapter, duplex DNAs containing a normal abasic site
with the three other bases (G, C, or T, Figure 6.1) opposite the abasic site have been synthesized and characterized by 2D NMR spectroscopy in the same sequence context. Molecular modeling and molecular dynamics calculations have been applied to study the influence of this base on the local conformation of the lesioned DNA. The results, together with the results from Chapter 5, indicate that all four pieces of oligonucleotides adopt overall B-form conformation. The bases opposite the abasic site are all intrahelical. The conformation around the abasic site is more perturbed when the base opposite the lesion is a pyrimidine (C or T). In these cases, the neighboring base pairs (G6-C21 and A8-T19) are closer to each other than those in B-form DNA. Molecular dynamics simulations reveal that the transient H-bond interactions between the pyrimidine (C20 or T20) opposite the abasic site and the base 3' to the abasic site might play a role in perturbing the local conformation. These results provide the first structural insight into the dynamics of abasic sites that are intrinsically modulated by the bases adjacent to the abasic site. The long-range goal of these structural studies is to understand how the conformation of the abasic site in duplex DNA ultimately governs its recognition and the rate of repair by Ape1 and other Ap endonucleases.

6.2 Experimental Section

*Materials* The oligonucleotides were synthesized on a 10 μmol scale by Invitrogen Inc. *E. coli* uracil-DNA glycosylase (UDG) was purified from expression plasmid pET21a-UDG kindly supplied by Dr. James Stivers, Johns Hopkins University, as previously described. All other chemicals were purchased from Sigma-Aldrich Co.

*Preparation of NMR samples* The single-stranded oligonucleotide containing an abasic
site, d(CCA AAG XAC CGG G) (1), (X denotes the abasic site), was synthesized by treating d(CCA AAG UAC CGG G) with UDG as described in Chapter 5. Compound 1 was annealed in a 1:1 ratio with different templates (2, 3, or 4, Figure 6.1) to generate duplex DNA containing an abasic site with four bases opposite the lesion (referred to as Ab/G, Ab/C, Ab/T, and Ab/A, Figure 6.1) following the same procedure to prepare Ab/A in Chapter 5. The final NMR samples contained Ab/G (2.7 mM), Ab/C (2.5 mM), or Ab/T (2.7 mM) in 325 μL of 10 mM sodium phosphate, 0.2 mM EDTA, pH 6.5 in D$_2$O or 10% D$_2$O/90% H$_2$O (v/v).

Figure 6.1 Sequences of the duplex DNA constructs containing a single abasic site (X) with four different base (in bold). The structure of Ab/A is investigated in Chapter 5.
NMR experiments The NMR experiments were performed on a custom-built 591MHz spectrometer at the Francis Bitter Magnet Laboratory. The raw data were transferred to a Silicon Graphic workstation and were processed using Felix 2001 (Accelrys Inc.). Proton chemical shifts were referenced to an internal standard, sodium 3-(trimethylsilyl)-1-propanesulfonate (TSP) at 0.00 ppm. NOESY (100 ms, 200 ms, and 400 mixing times), TOCSY (60 ms and 110 ms mixing times), and E-COSY were recorded at 25 °C in D2O. For NOESY and TOCSY experiments, datasets of 4096 (t2) x 512 (t1) complex points were acquired with the spectral width of 7022.5 Hz in both dimensions and 32 scans per t1 increment. For the E-COSY experiment, datasets of 4096 (r2) x 1028 (r1) complex points were acquired with 48 scans per r1 increment. A pre-saturation pulse (2 s) was applied during the recycle delay period to suppress solvent signals in all experiments carried out in D2O. Watergate-NOESY experiments (200 ms and 400 ms mixing times) were carried out in 90% H2O/10% D2O at 4 °C. Datasets of 4096 (t2) x 512 (t1) complex points were acquired with spectral widths of 13020.8 Hz in both dimensions and 32 scans per t1 increment. For all the datasets, t1 dimensions were linear-predicted to 4096 points and the data were processed with an exponential window function.

1H-31P HSQC and H3’-selective PH-COSY experiments were performed on the 591 MHz spectrometer at 25 °C in D2O. 1H-31P HSQC data were acquired with a spectrum width of 6009.6 Hz in the 1H dimension and 2000 Hz in the 31P dimension (4096 x 128 complex points). H3’-selective PHCOSY experiments with 6009.6 Hz in 1H dimension and 500 Hz in the 31P dimension (4096 x 256 complex points) were performed to measure coupling constants of the DNA backbone (H3’-C3’-O-P). The H3’ protons are specifically excited by using a band-selective 180° pulse with an EBURP shape(28, 29). In the final
datasets, the $^{31}$P dimension was linear-predicted to 1024 points. The $^{31}$P chemical shifts were referenced to an external trimethyolphosphate standard (0.00 ppm).

*NMR Structure Determination and Refinement* Distance restraints were derived from the volumes of the cross peaks in NOESY spectra (200 ms mixing time). The peak volumes were calculated with the peak picking protocol in the Felix 2000 software package. Volumes of cross peaks between H2' and H2'' (1.772 Å) or between H5 and H6 (2.460 Å) of cytosine were used to calibrate the proton-proton distances. The cross peak intensities were classified as weak, medium, or strong with the distance restraints of 1.5-3.0, 2.0-5.0, or 3.0-6.0 Å respectively. Alternatively, a more rigorous approach of iterative relaxation matrix analysis was applied to estimate the distance restraints using the program MARDIGRAS (30, 31). MARDIGRAS calculations were carried out on the NOESY spectra with a correlation time of 2.0 ns. For each NOESY cross peak, an upper and a lower distance restraint was calculated. The final distance restraints derived from MARDIGRAS calculation agreed well with the manual classification as described above.

Dihedral angle restraints of the deoxyribose moieties were derived from the coupling constants measured in E-COSY experiments. Except for the abasic site, the coupling constants of H1'-H2' and H1'-H2'' were consistent with a Southern (C2'-endo) conformation. The summaries of NMR restraints for Ab/G, Ab/C, and Ab/T are presented in Table 6.1.

All structural calculations were performed on an SGI Altix 3700 LINUX server with 128 Intel Itanium 1.3 GHz processors at the Scripps Research Institute (TSRI) using the SANDER module of AMBER 8.0 (32). The starting model structures were built using the Biopolymer module of Insight 2000 (Accelrys Inc.). Three starting structures were
Table 6.1 Summary of root-mean-square deviation (RMSD), NMR restraints and NMR violations of the final 10 structures of Ab/G, Ab/C, and Ab/T.

<table>
<thead>
<tr>
<th></th>
<th>Ab/G</th>
<th></th>
<th>Ab/C</th>
<th></th>
<th>Ab/T</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>α</td>
<td>β</td>
<td>α</td>
<td>β</td>
<td>α</td>
<td>β</td>
</tr>
<tr>
<td><strong>Pairwise RMSD (Å)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Starting structures</td>
<td>4.54</td>
<td>4.54</td>
<td>4.55</td>
<td>4.55</td>
<td>4.55</td>
<td>4.55</td>
</tr>
<tr>
<td>Final 10 structures</td>
<td>1.55</td>
<td>1.45</td>
<td>1.53</td>
<td>1.41</td>
<td>1.62</td>
<td>1.73</td>
</tr>
<tr>
<td><strong>Number of NMR restraints</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NOE</td>
<td>455</td>
<td>455</td>
<td>478</td>
<td>479</td>
<td>452</td>
<td>453</td>
</tr>
<tr>
<td>Dihedral angle</td>
<td>58</td>
<td>58</td>
<td>63</td>
<td>63</td>
<td>62</td>
<td>61</td>
</tr>
<tr>
<td><strong>Violation of Restraints</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NOE violation (Å)</td>
<td>0.071</td>
<td>0.107</td>
<td>0.112</td>
<td>0.089</td>
<td>0.092</td>
<td>0.110</td>
</tr>
<tr>
<td>Dihedral angle violation (°)</td>
<td>2.27</td>
<td>2.31</td>
<td>2.39</td>
<td>2.38</td>
<td>2.36</td>
<td>2.38</td>
</tr>
</tbody>
</table>
constructed, including an A-form, a B-form and a third form with arbitrary helical parameters. The root-mean square deviation (RMSD) for the three structures is ∼ 4.5 Å. A Generalized-Born implicit solvent model was applied to simulate the solvent dielectric constant. The initial model structures were energy minimized by 500 steps of conjugated gradient minimization followed by 500 steps of steepest descent minimization to remove high energy interactions introduced during the manual model building. The simulated annealing was carried out by heating the structure from 300 K to 900 K in 5 ps followed by cooling back to 300K over the next 15 ps. During the cooling process, weak temperature coupling (slow cooling) was applied for the first 13 ps followed by strong temperature coupling (rapid cooling) for the last 2 ps(33). NMR restraints were enforced during the annealing process with an increasing force constant from 5 to 50 kcal/mol over the first 3 ps and was maintained at 50 kcal/mol for the rest of the calculation(33). The annealed structure was then subjected to a 100-ps constant temperature (300 K) molecular dynamics with NMR restraints. Ten structures for each starting structure (A-, B-, and the third form) were calculated by sampling conformations during the last 10 ps of the molecular dynamics at a 1-ps interval. The resulting 30 structures have an averaged RMSD of <2.0 Å, which indicates a structural convergence after refinement. Ten final structures were chosen from the 30 structures based on the fewest NMR violations to represent the ensemble of the averaged structure. The statistical analysis of the final ten structures is summarized in Table 6.1.

Molecular Dynamics Simulations with an Explicit Solvent Model The molecular dynamics simulations were performed at a constant temperature (300 K) using the software AMBER 8.0 following the protocol by Dupradeau et al.(34). The averaged
structures were first solvated using the TIP3P solvent model with approximately 6,000 water molecules in an octahedral solvent box with a thickness of 10 Å. Twenty four sodium ions were also included in the calculations to neutralize the system. The molecular dynamics calculations were carried out on a 3-ns time scale. The conformation was sampled every 1 ps and was analyzed by the PTRAJ module in Amber 8.0(32).

6.3 Results

Chemical shift Assignments of non-exchangeable protons Standard TOCSY (60 and 110 mixing time) and NOESY (100, 200, and 400 ms mixing time) experiments were performed to assign non-exchangeable protons of DNA(35). The proton chemical shifts associated with the deoxyriboses were assigned using TOCSY experiments. The NOE connectivities between base protons (H6 of pyrimidines, H8 of purines) and deoxyribose protons (H1', H2'/H2'', H3', H4', and most of H5') were established by analyzing the NOESY spectra (200 ms mixing time) to confirm the assignments. NOE connectivities between H5 and H6 protons of cytosines and between methyl and H6 protons of thymidines were also used to assist the assignments of NOE signals in overlapping regions.

For Ab/G, the NOE connectivities in the base-H1' region are straightforward due to dispersion of signals (shown in Figure 6.2 for Ab/G). On the strand of Ab/G containing the abasic site, the NOE connectivity is disrupted between G6 and A8 as indicated by × (Figure 6.2), consistent with the presence of the abasic site. NOE connectivities are observed on the opposite strand to the abasic site between T19 and C21, suggesting that the base opposite to the abasic site (G20) adopts an intrahelical conformation as in
regular B-form DNA. Similar results were obtained for Ab/A (Figure 5.4, Chapter 5).

For Ab/C, the NOE connectivities between the H1' protons and the base protons are shown in Figure 6.3. Surprisingly, weak NOE interactions are detected between G6 H1' and A8 H8 (the dashed box in Figure 6.3), in contrast to the absence of NOE connectivities for Ab/G and Ab/A. The base-H1' NOE connectivities for the residues on the strand opposite the abasic site (C20 to T24) were difficult to assign due to overlapping signals. To confirm the base-H1' connectivities for these residues, the NOESY spectrum shown in Figure 6.4A was used. This region shows the connectivities between the base proton and the methyl protons of thymidines. The chemical shift of the base proton (H6) of C21 was assigned by its connectivity with T22 methyl protons (Figure 6.4A). In addition, the TOCSY spectrum showing the through-bond connectivities between H5 and H6 of cytosines is presented in Figure 6.4B to confirm the assignments of base protons associated cytosines.

For Ab/T, the base-H1' NOE connectivities are shown in Figure 6.5. Weak NOE interactions were detected between G6 H1' and A8 H8 (indicated by the arrow in Figure 6.3), similar to those observed for Ab/C. The base-H1' NOE connectivities for the residues on the strand opposite the abasic site (G18 to T24) were more difficult to establish due to the overlapping signals associated with T19, T20, T22, T23, and T24. For these residues, the NOE connectivities are also analyzed in the region between the base proton and the methyl protons of thymidines (Figure 6.6, solid lines). In this region, the NOE connectivities were readily established from G18 to T20 and from C21 to T24 via the well-resolved signals of the methyl protons of thymidines.
Figure 6.2 The NOESY spectrum (200 ms mixing time) of Ab/G showing the NOE connectivities between base (H8/H6) protons and H1'. The cross peak between A8H8 and C9H5 is indicated by an arrow. "x" denotes disruption of the connectivity between G6 and A8 by the abasic site.
Figure 6.3 The NOESY spectrum (200 ms mixing time) of Ab/C showing the NOE connectivity between base (H8/H6) protons and H1'. Two cross peaks (of α and β anomers) between A8H8 and G6H1' (in the dashed box) are shown at a lower contour level. The cross peaks between A8H8 and C9H5 are also indicated by arrows.
Figure 6.4 Chemical shift assignments of base protons of C20, C21, T22, T23, and T24 for Ab/C. A: The NOESY (200 ms mixing time) spectrum showing the connectivities associated with T methyl protons (solid lines). B: The TOCSY (60 ms mixing time) spectrum showing the cross peaks between H5 and H6 protons of C. The dashed lines indicated the chemical shift assignments of C21 (associated with the α and β anomers of the abasic site), which are consistent with the observed connectivities with T22 methyl protons.
Figure 6.5 The NOESY spectrum (200 ms mixing time) of Ab/T showing the NOE connectivity between base (H8/H6) protons and H1' protons. The cross peaks between A8H8 and G6H1' and between A8H8 and C9H5 are indicated by the arrow and the label.
Figure 6.6 The NOESY (200 ms mixing time) spectrum showing the connectivities associated with T methyl protons (solid lines). The NOE interactions involving the T methyl protons confirm the chemical shift assignments of base protons of G18-T20 and C21-T24.

Chemical shift Assignments of exchangeable protons The exchangeable protons (imino protons of guanines and amino protons of cytosines) of each DNA with an abasic site are indicative of Watson-Crick base pair interactions and global conformations. These exchangeable protons were assigned by Watergate-NOESY experiments at 4 °C in 10% D₂O/H₂O (v/v). Figure 6.7 (A-D) shows the imino proton region (12-15 ppm) with the overlaid 1D spectrum of each oligonucleotide. This region shows the cross peaks of T-imino to A-H2 protons and G-imino to C-amino protons. For Ab/A, 8 out of total 12 imino protons were detected by the Watergate-NOESY experiment (Figure 6.7A). The imino protons associated with terminal base pairs (C1-G26 and G13-C14) and base pairs adjacent to the abasic site (G6-C21 and A8-T19) could not be detected in either 1D or 2D spectra. The absence of these imino proton signals suggests weak base pair interactions.
and fast exchange with solvent.

For Ab/G, besides some 8 imino signals detected in Ab/A, a weak signal is detected in the 1D spectrum (Figure 6.7B, indicated by a red arrow). This signal is assigned as the T19 imino protons based on a weak cross peak with the A8H2 proton. For Ab/C, a similar weak, broad feature is observed, which might also be associated with T19 (indicated by the red arrow in Figure 6.7 C). Since there was no detectable cross peak associated with this signal due to its low intensity, the assignment of this signal to T19 is tentative.

For Ab/T, in addition to the 8 imino proton signals that were detected in other constructs, a weak signal appears as a doublet upfield from the imino protons of guanines (red arrow, Figure 6.7 D). This signal has been assigned to the imino protons of G6 based on their cross peaks with the amino protons of C21 (solid red line, Figure 6.7 D). The doublet feature is associated with the two anomeric states of the abasic site in Ab/T (α and β, see the discussion below). Another interesting feature observed for the imino protons in Ab/T is the signal associated with G18 (blue lines in Figure 6.7 D). This signal is the downfield-shifted compared to those in other three constructs (blue lines in Figure 6.7 A-C). The detection of the G6 imino proton and the downfield-shifted G18 imino protons (G6 and G18 residues are directly adjacent to or near the abasic site in Ab/T) suggest that the base-pair interactions around the abasic site in Ab/T are different from those in other constructs.
Figure 6.7 Watergate-NOESY spectra (200 ms mixing time) of the four duplex DNAs containing an abasic site showing the cross peaks involving T-imino to A-H2 and G-imino to C-amino protons. The 1D spectra containing the imino proton signals are overlaid on the top of the 2D spectra. The NOE signals associated with the G18 imino proton are indicated by the blue solid line. The NOE signals associated with the G6 imino proton in D (Ab/T) are indicated by the red solid line.
In contrast to the imino protons, amino protons of each cytosine (including the amino protons of terminal C1 and C14, as well as C21 adjacent to the abasic site) have been assigned in the base-H1’ region where the amino protons show strong cross peaks with both C-H5 and C-H6. An example of the strategy to assign the chemical shifts of the cytosine amino protons is shown in Figure 6.8 for Ab/G. The NOE connectivities between each C-H5 proton and C-H6/ C-amino protons are indicated by red solid lines. Strong cross peaks between the two amino protons are indicated by red dashed lines. One of the two amino protons is shifted downfield relative to the other due to the H-bonding interaction with the G-imino proton. The intensity of cross peaks involving the amino protons at C21 was much weaker than other C-amino protons (Figure 6.8, Cross peak H and its connected signals), suggesting that the H-bonding interactions of this G6-C21 base pair, which is 3’ to the abasic site, are weaker than other base pairs. The same strategy has been used to assign the exchangeable amino protons of cytosines in the other constructs. The chemical shift assignments of both non-exchangeable and exchangeable protons are summarized in Table 6.2 (Ab/G), Table 6.3 (Ab/C), Table 6.4 (Ab/T), and Table 5.4 in Chapter 5 (Ab/A).
Figure 6.8 The Watergate-NOESY spectrum (200 ms mixing time) of Ab/G as an example to illustrate the strategy to assign the chemical shifts of 2 amino protons of C. The NOE connectivities between C-H6 protons and C-H5/C-amino (H41/H42) protons are indicated by the solid red lines. The NOE connectivities between the two amino protons of C are indicated by the dashed red lines. Cross peaks between the two amino protons of C are assigned as A (C15), B (C16), C (C2), D (C10), E (C9), F (C1), G (C14), and H (C21).
Table 6.2 Proton chemical shift assignments of Ab/G.

<table>
<thead>
<tr>
<th></th>
<th>H8/H6</th>
<th>Me/H5</th>
<th>H1'</th>
<th>H2'</th>
<th>H2''</th>
<th>H3'</th>
<th>H4'</th>
<th>H5''</th>
<th>-NH2, -NH,</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>7.766</td>
<td>5.984</td>
<td>5.989</td>
<td>2.067</td>
<td>2.507</td>
<td>4.680</td>
<td>n.d.</td>
<td>3.740/3.774</td>
<td>8.018/7.100</td>
</tr>
<tr>
<td>A3</td>
<td>8.253</td>
<td>5.781</td>
<td>2.760</td>
<td>2.844</td>
<td>5.044</td>
<td>4.368</td>
<td>117/4.004</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A4</td>
<td>8.106</td>
<td>5.801</td>
<td>2.600</td>
<td>2.785</td>
<td>5.044</td>
<td>4.412</td>
<td>4.208</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A5</td>
<td>7.959</td>
<td>5.963</td>
<td>2.516</td>
<td>2.782</td>
<td>4.992</td>
<td>4.399</td>
<td>4.208</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ab-α</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ab-β</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G12</td>
<td>7.742</td>
<td>5.791</td>
<td>2.600</td>
<td>2.742</td>
<td>4.992</td>
<td>4.312</td>
<td>4.156/4.117</td>
<td>13.229</td>
<td></td>
</tr>
<tr>
<td>C14</td>
<td>7.820</td>
<td>5.999</td>
<td>6.014</td>
<td>2.164</td>
<td>2.563</td>
<td>4.689</td>
<td>n.d.</td>
<td>3.822/3.774</td>
<td>8.016/7.134</td>
</tr>
<tr>
<td>G17</td>
<td>7.866</td>
<td>5.677</td>
<td>2.703</td>
<td>2.747</td>
<td>5.010</td>
<td>4.364</td>
<td>4.121</td>
<td>13.138</td>
<td></td>
</tr>
<tr>
<td>G18</td>
<td>7.688</td>
<td>5.928</td>
<td>2.491</td>
<td>2.696</td>
<td>4.914</td>
<td>4.373</td>
<td>4.204</td>
<td>12.866</td>
<td></td>
</tr>
</tbody>
</table>

("n.d., not determined due to overlapping signals.

253
Table 6.3 Proton chemical shift assignments of Ab/C.

<table>
<thead>
<tr>
<th></th>
<th>H8/H6</th>
<th>Me/H5</th>
<th>H1'</th>
<th>H2'</th>
<th>H2''</th>
<th>H3'</th>
<th>H4'</th>
<th>H5''/H2'</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>7.815</td>
<td>5.999</td>
<td>6.003</td>
<td>n.d.</td>
<td>n.d.</td>
<td>4.681</td>
<td>n.d.</td>
<td>3.762/3.778</td>
</tr>
<tr>
<td>A3</td>
<td>8.248</td>
<td>5.808</td>
<td>2.757</td>
<td>2.837</td>
<td>5.045</td>
<td>4.368</td>
<td>4.112</td>
<td>4.129</td>
</tr>
<tr>
<td>A4</td>
<td>8.078</td>
<td></td>
<td>2.557</td>
<td>2.744</td>
<td>5.014</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A5-α</td>
<td>7.892</td>
<td>5.851</td>
<td>2.350</td>
<td>2.601</td>
<td>4.956</td>
<td>4.333</td>
<td>4.204</td>
<td></td>
</tr>
<tr>
<td>A5-β</td>
<td>7.875</td>
<td>5.881</td>
<td>2.645</td>
<td>4.954</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G6-α</td>
<td>7.504</td>
<td>5.689</td>
<td></td>
<td>4.979</td>
<td></td>
<td></td>
<td></td>
<td>4.502</td>
</tr>
<tr>
<td>G6-β</td>
<td>7.557</td>
<td>5.704</td>
<td></td>
<td>4.983</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ab-α</td>
<td>5.468</td>
<td>5.472</td>
<td>2.221</td>
<td>1.950</td>
<td>4.515</td>
<td></td>
<td>n.d.</td>
<td>4.105/4.138</td>
</tr>
<tr>
<td>Ab-β</td>
<td></td>
<td></td>
<td>2.030</td>
<td>2.1122</td>
<td>4.439</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C9-β</td>
<td>7.345</td>
<td></td>
<td>4.380</td>
<td></td>
<td></td>
<td></td>
<td>4.130</td>
<td>4.198</td>
</tr>
<tr>
<td>C16</td>
<td>7.430</td>
<td>5.666</td>
<td>5.611</td>
<td>2.033</td>
<td>2.384</td>
<td>4.850</td>
<td>n.d.</td>
<td>4.084/4.117</td>
</tr>
<tr>
<td>G18-α</td>
<td>7.655</td>
<td>5.915</td>
<td>4.157</td>
<td>2.659</td>
<td>4.152</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G18-β</td>
<td>5.942</td>
<td>2.514</td>
<td>2.531</td>
<td>4.200</td>
<td>4.194</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C20-β</td>
<td>7.779</td>
<td>5.835</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C21-β</td>
<td>7.832</td>
<td>5.904</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T24</td>
<td>7.357</td>
<td>1.760</td>
<td>5.721</td>
<td>1.971</td>
<td>2.299</td>
<td>4.861</td>
<td>n.d.</td>
<td>4.184</td>
</tr>
</tbody>
</table>

(a n.d., not determined due to overlapping signals).
Table 6.4  Proton chemical shift assignments of Ab/T.

<table>
<thead>
<tr>
<th></th>
<th>H8/H6</th>
<th>Me/H5</th>
<th>H1'</th>
<th>H2'</th>
<th>H2''</th>
<th>H3'</th>
<th>H4'</th>
<th>H5''/''</th>
<th>-NH2, -NH, or H2</th>
</tr>
</thead>
<tbody>
<tr>
<td>A3</td>
<td>8.252</td>
<td>5.789</td>
<td>2.761</td>
<td>2.845</td>
<td>5.048</td>
<td>4.367</td>
<td>4.121/4.000</td>
<td>3.748/3.778</td>
<td></td>
</tr>
<tr>
<td>A5</td>
<td>7.923</td>
<td>5.896</td>
<td>2.433</td>
<td>2.716</td>
<td>4.997</td>
<td>4.361</td>
<td>4.188</td>
<td>4.118</td>
<td></td>
</tr>
<tr>
<td>Ab-α</td>
<td>3.748</td>
<td>7.092/7.958</td>
<td>3.778</td>
<td>3.748</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ab-β</td>
<td>5.982</td>
<td>5.982</td>
<td>2.072</td>
<td>2.220</td>
<td>4.680</td>
<td>4.118</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C14</td>
<td>7.821</td>
<td>6.000</td>
<td>5.991</td>
<td>2.056</td>
<td>2.165</td>
<td>4.695</td>
<td>4.151</td>
<td>3.810/3.834</td>
<td>7.147/7.999</td>
</tr>
<tr>
<td>G17</td>
<td>7.864</td>
<td>5.749</td>
<td>2.700</td>
<td>2.781</td>
<td>5.005</td>
<td>4.373</td>
<td>4.118/4.047</td>
<td>13.128</td>
<td></td>
</tr>
<tr>
<td>G18-α</td>
<td>7.710</td>
<td>5.989</td>
<td>2.546</td>
<td>2.753</td>
<td>4.970</td>
<td>4.392</td>
<td>4.211</td>
<td>13.090</td>
<td></td>
</tr>
<tr>
<td>G18-β</td>
<td>7.706</td>
<td>5.989</td>
<td>2.546</td>
<td>2.753</td>
<td>4.970</td>
<td>4.392</td>
<td>4.211</td>
<td>13.090</td>
<td></td>
</tr>
<tr>
<td>T20-β</td>
<td>7.547</td>
<td>1.558</td>
<td>5.998</td>
<td>2.072</td>
<td>2.525</td>
<td>4.813</td>
<td>n.d.</td>
<td>n.d.</td>
<td></td>
</tr>
<tr>
<td>G26</td>
<td>7.817</td>
<td>6.143</td>
<td>2.519</td>
<td>2.357</td>
<td>4.644</td>
<td>4.146</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(a n.d., not determined due to overlapping signals).
Chemical shift assignments of the protons associated with the abasic site deoxyribose

Assignments of the protons associated with the abasic-site deoxyribose are key to distinguishing the α and β anomers and to identifying the NOE interactions between the abasic site and neighboring bases. The H1’, H2’/H2”, and H3’ protons associated with the abasic site deoxyribose were assigned using the 60 ms TOCSY spectra (Figure 6.9 and for Ab/A, see Chapter 5, Figure 5.6). In all of the oligonucleotides, the chemical shifts of the protons associated with the abasic site deoxyribose are shifted to some extent upfield (H1’, 5.0-5.5 ppm; H2’/2’, ~1.8/2.0 ppm; H3’, 4.5-4.6 ppm) from the generic deoxyribose protons (H1’, 5.3-6.5 ppm; H2’/2’, ~2.0-3.0 ppm; H3’, 4.7-5.0 ppm). The base opposite the abasic site modulates the chemical shift of the H1’. When it is a purine, the chemical shift of H1’ is between 5.2-5.3 ppm (Figure 6.9 A). When it is a pyrimidine, the chemical shift of H1’ is between 5.5 and 5.6 ppm in the downfield-shifted region of the chemical shifts of most H1’ deoxyribose protons (Figure 6.9 B and C). The H4’ and H5’/5” protons could not be assigned due to spectral crowding in those regions.

The sugar protons at each abasic site exist as two sets of signals (Figure 6.9, solid red lines), consistent with the α and β anomic forms. The assignment of each anomer was achieved using NOESY experiments (200 ms mixing time) based on the relative intensities of the NOE signals between H1’ and H2’/2”. For the α anomer, the H1’ is closer to H2’, producing a more intense cross peak than to H2”. The intensity pattern is reversed for the β anomer. The NOESY spectra containing the abasic site proton regions are shown in Figure 6.10 (Ab/G), Figure 6.11 (Ab/C) and Figure 6.12 (Ab/T) with the connectivity of H2’ and H2” indicated by solid red lines. The ratio of α:β in each case is approximately 60:40 based on relative signal intensities associated with each anomer.
Figure 6.9 Chemical shift assignments of protons associated with the abasic site deoxyribose by TOCSY experiments (60 ms mixing time). For each oligonucleotide, the through-bond connectivities between H1' and H2'/2'' and between H3' and H2'/2'' are indicated by the solid red lines. The dashed red lines indicate the connectivities between H1' and H3' through the same H2'/2'' proton.
Figure 6.10 The NOESY spectrum (200 ms mixing time) of Ab/G showing the NOE connectivities between abase protons (H1' or H3' to H2'/H2''). Chemical shifts of the abasic protons are labeled.
4.3

.. ".. .

•
c-O .. Ji

4.5

'P

ex.

~

•

4.7

,.•

- 4.9
E

Co

.,g.

-

:I:

~H3'
aH3'

/

5.1
5.3
Hl'
~Hl'

5.5

,

5.7
2.9

2.7

1.7

1.5

Figure 6.11 The NOESY spectrum (200ms mixing time) of Able showing the NOE
connectivity between abase protons (H l' or H3' to H2' 1H2"). Chemical shifts of the
abasic protons are labeled.

259


Figure 6.12 The NOESY spectrum (200 ms mixing time) of Ab/T showing the NOE connectivity between abase protons (H1' or H3' to H2'/H2''). Chemical shifts of the abasic protons are labeled.
Sugar conformations

To obtain insight into the sugar pucker of DNA, the coupling constants of the deoxyribose protons have been measured using E-COSY experiments. For all the residues in each duplex (except the abasic site), the values of $J_{H1',H2'}$ are between 8 and 10 Hz and $J_{H1',H2''}$ are between 5 and 6.5 Hz. This range of coupling constants is characteristic for a 2'-endo sugar pucker (S-type) for B-form DNA. The coupling constants have also been measured for the abasic site in Ab/C and Ab/A (Figure 6.13). The observed coupling constants are consistent with a near-N-type (O1'-endo) conformation for α anomer and a near-S-type (C1'-exo) conformation for β anomer for both Ab/A and Ab/C. For Ab/G and Ab/T, coupling constant information at the abasic site was not available due to the low cross peak intensities. Therefore, the sugar conformations were not restrained during the modeling in Ab/G or Ab/T.

Phosphodiester backbone conformations

In order to characterize the conformation of the phosphodiester backbone, the dihedral angles of H3'-C3'-O3'-P have been determined by measuring the $J_{H3',P}$ coupling constants using H3'-selective $^3$P-$^1$H COSY experiments. Due to the small dispersion of H3' and $^3$P chemical shifts, not all cross peaks were well resolved. For Ab/G and Ab/A, all $^3$P chemical shifts were between -3.8 and -5.0 ppm and all the $J_{H3',P}$ values were within 6-8 Hz, consistent with those expected for a generic B-form DNA with regular backbone conformation.

For Ab/C, the majority of the $^3$P chemical shifts were between -3.8 and -5.0 ppm and the $J_{H3',P}$ values were within 6-8 Hz, consistent with B-form DNA. In addition, three $^3$P signals were shifted downfield between -3.1 ppm and -3.4 ppm (cross peak C, D, and E in Figure 6.14). These signals were assigned to be associated with the bases adjacent to the abasic sites (G6, A8, and T19) based on the proton chemical shifts (H3' and H4') with
these signals in $^{31}\text{P}-^1\text{H}$ HSQC experiments. Although the coupling constants associated with these peaks were within the normal range (6-8 Hz), the perturbed $^{31}\text{P}$ chemical shifts suggest that the backbone conformations at these sites are different from B-form DNA.

Figure 6.13 Coupling constants between H1' and H2'/H2'' measured by E-COSY experiments for the abasic sites. A, Ab/A; B, Ab/C. The chemical shift assignments of these abasic protons are also shown. Due to overlapping signals, the coupling constant between H1' and H2'' for the $\alpha$ anomer in Ab/C cannot be accurately determined.
Figure 6.14 H3'-selective PH-COSY experiment to measure the backbone H3'-C3'-O3'-P coupling constants of Ab/C. The phase-sensitive cross peaks are assigned as A: Ab7-α, B, Ab7-β, C: A8, D: T19, and E: G6.

For Ab/T, most of the 31P chemical shifts were between -3.8 and -5.0 ppm and the $J_{H3'}$ values were within 6-8 Hz, consistent with B-form DNA. In addition, two 31P signals are shifted downfield between -3.2 ppm and -3.5 ppm (cross peak C and D in Figure 6.15). These two signals are assigned to be associated with G6 and T19 that are adjacent to the abasic sites based on the proton chemical shifts (H3' and H4') with these signals in 31P-1H HSQC experiments. The coupling constants associated with these two 31P signals are within the normal range (6-8 Hz). As discussed above for Ab/C, the perturbed 31P chemical shifts suggest that the backbone conformations at those sites are different from B-form DNA.
Figure 6.15 H3'-selective PH-COSY experiment to measure the backbone H3'-C3'-O3'-P coupling constants of Ab/T. The phase-sensitive cross peaks are assigned as A: Ab7-α, B, Ab7-β, C: G6, D: T19.

The Karplus equation ($^3J$(HCOP) = 15.3 cos$^2\phi$ – 6.1 cos$\phi$ +1.6) was used to derive dihedral angle constraints for H3'-C3'-O3'-P from the experimentally determined coupling constants ($J_{H3',P}$). As noted above, the $^{31}$P chemical shifts and $J_{H3',P}$ associated with most residues in all the oligonucleotides are within the normal envelop expected for B-form DNA. Therefore, the H3'-C3'-O3'-P dihedral angles for these residues were constrained to be $-40^\circ \pm 25^\circ$ as in B-form DNA. For G6, A8 and T19 in Ab/C and G6 and T19 in Ab/T, the unusually downfield shifted $^{31}$P chemical shifts may be indicative of possibly perturbed dihedral angles. For these residues in Ab/C and Ab/T, all four possible solutions to the Karplus equation ($\pm 40^\circ$ and $\pm 115^\circ$) were tested in separate structural calculations. Only the solution that consistently gave the lowest energy was
used in the final structural calculations. The final H3'-C3'-O3'-P dihedral angle constraints for Ab/C were set to 40° ± 25° (G6), -40° ± 25° (A8), and -40° ± 25° (T19); for Ab/T, the H3'-C3'-O3'-P dihedral angle constraints were set to 40° ± 25° (G6) and -40° ± 25° (T19).

**NOE interactions involving the protons associated with the abasic site and adjacent bases** A number of NOE interactions associated with the abasic sites and neighboring bases have been identified and are summarized in Table 6.5. These NOE interactions, along with the coupling constants, are essential in defining the conformation of the abasic site. The NOE interactions for Ab/A were presented in Chapter 5. For each anomer of Ab/G, three weak NOE interactions were observed between the sugar protons (H2'/H2'', and H3') of the abasic site and the base proton (H8) of A8 that is 3' to the abasic site. For Ab/C, medium NOE interactions were observed between A8H8 and H3' of the abasic site, as well as weak NOE interactions between A8H8 and H2'/H2'' of the abasic site. For Ab/T, medium NOE interactions were observed between A8H8 and H3' of the abasic site for each anomer and a weak NOE interaction between A8H8 and H2' of the abasic site in the α anomer.

Based on previous studies(24) and the results in Chapter 5 for Ab/A, a partial extrahelical conformation of the abasic site is usually indicated by this unique NOE pattern: medium NOE interactions between the H3' proton of the abasic site and A8-H8 and weak NOE interactions between the H2'/2'' protons and A8-H8. The NOE pattern for Ab/C and Ab/T therefore predicts that the abasic site deoxyribose in these two constructs most likely adopts a partial extrahelical conformation.

For Ab/C and Ab/T, several unusual NOE interactions in the abasic site region were
also observed between G6 and A8, the two bases adjacent to the abasic site. For Ab/C, a weak NOE signal was observed between A8-H8 and G6-H1’ for both anomers (indicated by the dashed box in Figure 6.3). In addition, a weak NOE signal between A8-H8 and G6-H2” was also observed. These NOE interactions suggest that A8 and G6 are in close proximity in this construct relative to those in B-form DNA. These NOE interactions, together with the downfield-shifted 31P signals associated with the G6 and A8 residues, indicate that the abasic site region of Ab/C adopts a different conformation from that of Ab/A and Ab/G.

**Table 6.5** Summary of the NOE interactions between protons around the abasic sites.

<table>
<thead>
<tr>
<th>Ab/A(α)</th>
<th>Ab/A (β)</th>
<th>Ab/G (α)</th>
<th>Ab/G(β)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A8H8-AbH3’ m</td>
<td>A8H8-AbH3’ m</td>
<td>A8H8-AbH3’ w</td>
<td>A8H8-AbH3’ w</td>
</tr>
<tr>
<td>A8H8-AbH2’ w</td>
<td>A8H8-AbH2’ w</td>
<td>A8H8-AbH2’ w</td>
<td>A8H8-AbH2’ w</td>
</tr>
<tr>
<td>A8H8-AbH2” w</td>
<td>A8H8-AbH2” w</td>
<td>A8H8-AbH2” w</td>
<td>A8H8-AbH2” w</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ab/T (α)</th>
<th>Ab/T (β)</th>
<th>Ab/C (α)</th>
<th>Ab/C(β)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A8H8-AbH3’ m</td>
<td>A8H8-AbH3’ m</td>
<td>A8H8-AbH3’ m</td>
<td>A8H8-AbH3’ m</td>
</tr>
<tr>
<td>A8H8-AbH2’ w</td>
<td>A8H8-AbH2’ w</td>
<td>A8H8-AbH2’ w</td>
<td>A8H8-AbH2’ w</td>
</tr>
<tr>
<td>A8H8-G6H1’ w</td>
<td>A8H8-G6H1’ w</td>
<td>A8H8-G6H1’ w</td>
<td>A8H8-G6H1’ w</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A8H8-G6H2” w</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>G6H4’-AbH2’ w</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>G6H4’-AbH2” w</td>
<td></td>
</tr>
</tbody>
</table>

For Ab/T, a weak NOE signal was observed between A8-H8 and G6-H1’ for each anomer (arrow in Figure 6.4), similar to the observation for Ab/C. This NOE interaction suggests that A8 and G6 in Ab/T are also in closer proximity compared to those in B-form DNA. No NOE signals were observed between A8H8 and G6-H2”, suggesting subtle structural differences between Ab/C and Ab/T in the abasic site region.

266
Averaged structures of Ab/G, Ab/C, and Ab/T: Ten structures of the α and β anomers of Ab/G, Ab/C, and Ab/T (also Ab/A in Chapter 5) were calculated from 3 starting structures using the experimentally determined distance and dihedral angle constraints. None of the ten structures contained NOE violations > 0.5 Å or dihedral angle violations > 5°. The RMSD’s of the final ten structures are between 1.5 and 1.7 Å (Table 6.1). An overlay of the ten structures for each construct is shown in Figure 6.16. The results reveal that the overall structures are B-form with the bases (colored in blue) opposite the abasic site stacked intrahelically between flanking base pairs. Compared to the base opposite to the lesion, different conformations are observed for the abasic site deoxyribose (colored in red, Figure 6.16) and bases adjacent to it (G6 and A8). The discussion below will focus on the abasic site conformation in the averaged structure of each construct (shown in Figure 6.17-6.19).
Figure 6.16 The overlay of ten final structures of Ab/A, Ab/G, Ab/C, and Ab/T. The abasic site is colored in red and the base opposite the lesion is colored in blue. A, Structures with the α anomer at the abasic site; B, Structures with the β anomer at the abasic site.
A (α anomer)

B (β anomer)

Figure 6.17 A stereo view of the conformation of the abasic site region in the averaged structure of Ab/G. A, α anomer; B, β anomer. In B, only the abasic site and the opposite bases are colored for clarity.
A stereo view of the conformation of the abasic site region in the averaged structure of Ab/C. A, α anomer; B, β anomer. The abasic site deoxyribose and the base opposite the lesion are colored.
Figure 6.19 The stereo view of the conformation of the abasic site region in the averaged structure of Ab/T. A, α anomer; B, β anomer. The abasic site deoxyribose and the base opposite the lesion are colored.
Abasic site conformation The averaged structures of Ab/G (Figure 6.17) show that in the α anomer, the abasic site adopts an intrahelical conformation and in the β anomer, the abasic site is partially extrahelical. The NOE interactions associated with the abasic site protons in Ab/G are weak (Table 6.5) and accommodate a variety of conformations. In the final ten structures for the α anomer, two of them show that the 1'-OH group is within the H-bonding distance from the imino group of the opposite G20. However, no imino proton signals were detected for G20, implying that the H-bonding interaction, if exists, is transient. The partially extrahelical conformation for the Ab/G β anomer is very similar to that of Ab/A determined in Chapter 5.

Figure 6.18 and 6.19 show the abasic site conformations of Ab/C and Ab/T. In all structures, the abasic site is extrahelical. In addition, the backbone conformation around the abasic site is perturbed so that G6 and A8 are closer to each other than in a normal B-form DNA. These results are consistent with and defined by the NOE interactions between the base proton of A8 and several sugar protons of G6 in both Ab/C and Ab/T (Table 6.5).

Molecular dynamics simulations of abasic site-containing duplex DNA Molecular dynamics simulations were carried out to characterize the conformational flexibility in the abasic site region of all constructs. For Ab/G, the conformational flexibility is dominated by the motion of the abasic site moiety. Figure 6.20 shows an overlay of 10 snapshots representing 20 ps of molecular dynamics simulations for the α and β anomer of Ab/G. The results indicate that the abasic site deoxyribose adopts a range of conformations from intrahelical to extrahelical (Figure 6.20 A and B). The extrahelical motion of the abasic site deoxyribose occurs without significant distortion of
conformations of adjacent residues, which is illustrated by the snapshot of the \( \alpha \) anomer of Ab/G with an extrahelical abasic site (Figure 6.20 C). A similar extrahelical motion of the abasic site has also been observed with Ab/A during the simulation (Chapter 5).

Figure 6.20 Molecular dynamics simulation of Ab/G showing the conformational flexibility at the abasic site. A, \((\alpha \) anomer) An overlay of ten snapshots representing 20 ps molecular dynamics simulations to show the conformational flexibility of the abasic site (colored in CPK); B, \((\beta \) anomer) An overlay of ten snapshots representing 20 ps molecular dynamics simulations to show the conformational flexibility of the abasic site (colored in CPK). C: A stereo view of one of the ten snapshots in A \((\alpha \) anomer) showing the extrahelical conformation of the abasic site (Ab).
The molecular dynamics simulations of Ab/C reveal that the abasic site deoxyribose remains partially extrahelical without significant conformational change. The flexibility of the abasic site region is largely reflected by the motion associated with A8, 3’ to the abasic site. Figure 6.21A shows that A8 (in stick) adopts a range of conformations (indicated by the double-headed arrows) that are capable of forming H-bonding interactions with both T19 (its expected base-pair partner) and C20 (the base opposite the abasic site). Figure 6.21 B shows the snapshot with the transient H-bonding interaction between A8 and C20 (viewed from the top along the helical axis). Quantitative analysis of the distance (red trace, Figure 6.22) and angle (blue trace, Figure 6.22) between the H-bond donor of C20 and the accepter of A8 during simulation suggests that the optimal H-bond interaction occurs every ~15 ps between C20 and A8 during the simulation (black, triangular dot, Figure 6.22).

The molecular dynamics simulations of Ab/T reveal that the abasic site deoxyribose remains partially extrahelical, similar to Ab/C. The flexibility of the abasic site region of Ab/T is also dominated by the motion associated with A8 3’ to the abasic site. Figure 6.23A shows that A8 (in stick) adopts a range of conformations (indicated by the double-headed arrows) that are capable of forming H-bonding interactions with both T19 (its expected base-pair partner) and T20 (the base opposite the abasic site). Figure 6.23 B shows the snapshot with the transient H-bonding interaction between A8 and T20 (viewed from the top along the helical axis). Quantitative analysis suggests that the optimal H-bond interaction occurs every 8 ps between T20 and A8, more frequent than that in Ab/C (Figure 6.24). This observation can be explained by the stronger interaction (2 H-bonds) between T20 and A8 than that in Ab/C (1 H-bond between C20 and A8).
Figure 6.21 Molecular dynamics of Ab/C (α anomer) suggesting transient H-bonding interactions between C20 and A8 formation. A, a stereo view of a snapshot of the conformation at the abasic site. For the conformation of A8 (in stick), 5 snapshots representing 10 ps molecular dynamics simulations are overlaid to show the conformational flexibility of A8 (indicated by the double-headed arrow). B, a snapshot of the conformation showing the transient H-bonding interactions (red dashed line) between A8 and C20. The conformation of T19, the expected base-pair partner of A8, is also shown.
Figure 6.22 Quantitative analysis of the transient H-bond formation between A8 and C20 in Ab/C in molecular dynamics simulation. Red trace, the distance between the proposed H-bond donor (C20 amino-N) and the accepter (A8-N1); blue trace, the angle of donor-H-accepter. The black triangle indicates the optimum H-bonding interaction (the distance < 3 Å and the angle > 160°).
Figure 6.23  Molecular dynamics of Ab/T (α anomer) suggesting transient H-bonding interactions between T20 and A8. A, a stereo view of a snapshot of the conformation at the abasic site. For the conformation of A8 (in stick), 5 snapshots representing 10 ps molecular dynamics simulations are overlaid to show the conformational flexibility of A8 (indicated by the double-headed arrow). B, a snapshot of the conformation showing the transient H-bonding interactions (red dashed lines) between A8 and T20. The conformation of T19, the expected base-pair partner of A8, is also shown.
Figure 6.24 Quantitative analysis of the transient H-bond formation between A8 and T20 in Ab/T in molecular dynamics simulation. Red trace, the distance between the proposed H-bond donor (T20 imino-N) and the accepter (A8-N1); blue trace, the angle of donor-H-accepter. The black triangle indicates the optimum H-bonding interaction (the distance < 3 Å and the angle > 160°).
6.4 Discussion

Very few studies have been reported on the influence of the sequence context on the structure and conformational flexibility of the abasic site lesion. Most studies have been focused on abasic site analogs (36-38). The Bolton group has reported structural modeling of duplex DNA containing normal abasic sites using 2D NOE spectroscopy (25-27, 39). The discussion will be focused on the results that give insight into the role of the base opposite the abasic site in the structure of the lesion. The structures of the abasic site (X) in the sequence of 5'-GAXAC-3' with A or C opposite the lesion have been investigated (25, 26). For both oligonucleotides, the molecule adopts an overall B-form DNA conformation with the base opposite the lesion stacking intrahelically between adjacent base pairs. For the abasic site with A opposite the lesion, a nearly equal population of α and β anomers were observed with slightly more α anomer. However, conflicting results were reported for the structural modeling of this construct. In one study, Goljer et al. showed that for the α anomer, the abasic site deoxyribose was extrahelical and for the β anomer, the abasic site was intrahelical (26). In a later study of the abasic site with A opposite the lesion in the exact same sequence, Beger et al. showed that that the abasic site deoxyribose in both anomers were intrahelical (25). Because no NOE interactions were observed between the abasic site protons and adjacent bases, the significance of the conformation of the abasic sites in the modeled structure is unclear.

The structure of the abasic site with C opposite the lesion was also examined by Beger et al. in the same sequence context (5'-GAXAC-3') in an effort to understand the influence of the base opposite to the abasic site (25). Their modeling suggested that the overall structure is B-form with the abasic site and C opposite the lesion adopting an
intrahelical conformation. Unlike the abasic site with A opposite the lesion, Beger et al. observed two downfield-shifted $^{31}$P signals that were assigned to the residues adjacent to the abasic site(25). This observation suggested that the backbone conformation was perturbed from generic B-form DNA in the abasic site with C opposite the lesion. However, no dihedral angle constraints were derived for the structural calculation. More interestingly, their results suggested that there was only one anomer for the abasic site with C opposite the lesion. Based on the relative NOE intensities associated with abasic site protons, this species was assigned as the $\beta$ anomer. The absence of the $\alpha$ anomer in this sequence is inconsistent with their previous studies that showed a mixture of $\alpha$ and $\beta$ anomers for the this site with four different bases opposite this lesion using a duplex DNA containing a 1'-[13C] labeled abasic site(6).

In summary, the modeling studies from the Bolton group described above suggested that the base opposite the abasic site could influence the conformation of the abasic site. However, their studies suffered from the lack of NOE interactions between the abasic site and the adjacent bases. Furthermore, in one study, the H5 proton of C 3’ to the abasic site was mistakenly assigned as the H1’ proton of the abasic site(25), which compromised their assignments of the anomeric state of the abasic site. Because of these limitations, the relevance of the structure modeling from these studies remains uncertain.

Compared to the previous studies, our studies on duplex DNA containing the abasic site with all four bases opposite the lesion have provided the first glimpse of such structures in the 5’-G\textbf{X}AC-3’ sequence context. Similar to previous studies(24-27, 39), the structural modeling based on distance and dihedral angle constraints from 2D NMR analysis suggests that the DNA is overall B-form for all lesions. Furthermore, in all cases,
the bases opposite the lesion are all intrahelical. The chemical shifts of the abasic site protons (H1’, H2’, H2″, and H3’) were assigned for all lesions and in all cases, a mixture of α and β anomers were observed. The conformation of the abasic site was calculated based on the NOE interactions with adjacent bases and dihedral angles associated with the abasic site.

For the abasic site with A opposite the lesion, studies suggest that the abasic site deoxyribose for both anomers are partially extrahelical (Chapter 5). The extrahelical conformation is consistent with the observed NOE interaction pattern between the abasic site protons (H1’ and H2’/2″) and the base proton 3’ to the lesion (Figure 5.14, Chapter 5). This result is also consistent with the previous study from our group on the abasic site in a very similar sequence context (5’-CCAAAGXACTGGG-3’/3’-GGTTTCATGACCC-5’)(24). For the corresponding lesion with G opposite the abasic site, weak NOE interactions were also observed between the abasic site protons (H1’ and H2’/2″) and the base proton 3’ to the lesion. These weak interactions are not sufficient to define a unique conformation for the abasic site deoxyribose. Molecular modeling suggests that the abasic site in the α anomer is intrahelical and that the 1’-OH group of the abasic site has the tendency to form H-bonding interaction with the imino group of G opposite the lesion. The abasic site in the β anomer adopts an extrahelical conformation that is similar to the abasic site with A opposite the lesion.

For the abasic site with pyrimidines (C or T) opposite the lesion, the deoxyriboses are partially extrahelical with both α and β anomers. These results are defined by and consistent with a medium NOE interaction between H3’ of the abasic site and the base proton 3’ to the lesion. The most interesting structural feature for the abasic site with a
pyrimidine opposite the lesion is the unusually close proximity between bases adjacent to the abasic site (G6 and A8). This conformation is supported by NOE interactions between these two bases and unusual backbone conformations determined by the coupling constants of H3'-C3'-O3'-'P.

At present, it is not clear whether this structural feature exists for abasic sites with pyrimidines opposite the lesion in any sequence context. Previous NMR characterization on an abasic site analog (THF, or F) in the 5'-CGTGFGTGC-3' sequence with either C or T opposite the lesion also revealed NOE interactions between the two bases adjacent to THF, similar to that observed in this study(37). Although those studies did not lead to final refined structures, the NOE interaction patterns suggested that the two bases adjacent to THF were closer to each other than those in B-form DNA. Further studies of the abasic site in more sequence contexts are required to draw a more general conclusion on the role of the pyrimidine opposite the abasic site in defining the conformation of the lesion.

Molecular dynamics calculations have been carried out to simulate the conformational flexibility of the abasic site region with four bases opposite the lesion. For the abasic site with a purine opposite the lesion, the conformational change is largely due to the extrahelical movement of the abasic site sugar. For the abasic site with a pyrimidine opposite the lesion, the simulation revealed much smaller motion associated with the abasic site deoxyribose. Instead, we observed transient H-bonding interactions between A8 3' to the abasic site and the pyrimidine opposite the lesion. This transient H-bonding interaction is not observed for the abasic site with a purine opposite the lesion. We propose that this H-bonding interaction is due to the lack of the base stacking
interaction between the unpaired pyrimidine with adjacent 3' bases.

Examples of transient H-bonding interactions between an unpaired base and an adjacent base normally paired with another base have been documented. Computer modeling of an abasic site analog (THF, or F) with C opposite the THF in the 5'-TCFTA-3' sequence without NMR restraints revealed a similar transient H-bonding interaction between T 3' to THF and C opposite the lesion during simulation(40). Such transient H-bonding interactions were not observed during the simulation of THF with G opposite the lesion in the same sequence context(40).

Are the transient H-bonding interaction observed between the base 3’ to the abasic site (A8) and the pyrimidine (C20 or T20) opposite to the lesion biologically interesting? Some insight has been provided by the structural studies of Y-family DNA polymerases. The transient H-bonding interactions between the incoming nucleotide and bases adjacent to the abasic sites have been proposed to be responsible for frameshift mutations in translesion DNA replication by Y-family DNA polymerases(15-18). Several crystal structures of Dpo4 (a Y-family DNA polymerase from S. solfataricus) have been solved in a complex with a primer/THF-containing-template(41). In one structure, the incoming nucleotide (dCTP, the would-be base opposite THF) formed base pair interactions with G 5’ to THF, which would lead to a –1 frameshift. In the structure, the THF moiety was flipped partially extrahelical to accommodate the structural requirement of the base pair interaction between the incoming nucleotide and the base 5’ to THF. Although such conformation was observed in a complex with the DNA polymerase, our results suggest that the inherent structural flexibility of the abasic site could cause the transient base pair interaction even without DNA polymerases.
Finally, we are interested in how the conformation of the abasic site governs its recognition and ultimately, repair by DNA repair enzymes. The distinct conformations modulated by the base opposite the lesion suggest that the interaction with repair enzymes is sequence dependent. Takeuchi and Demple et al. have examined the effect of the base opposite the abasic site on Ap endonuclease (E. coli endonuclease IV and exonuclease III) activity(42). They observed similar $k_{\text{cat}}$ and $K_m$ among different sequences. Since it remains unclear which step is rate limiting during the catalysis by Ap endonucleases, the steady-state kinetic constants may not be indicative of the interaction between the enzyme and the substrate DNA containing abasic sites. Further pre-steady state kinetic analysis on the repair of the abasic site in a large sequence space, combined with structural analysis of DNA lesions, would lead to the mechanism of sequence-dependent recognition of the abasic site lesions in vivo.

6.5 References


(37) Cuniasse, P., Fazakerley, G. V., Guschlbauer, W., Kaplan, B. E., and Sowers, L.


Education

**Ph. D.**
Massachusetts Institute of Technology
Biological Chemistry,
Thesis Advisor: Prof. JoAnne Stubbe
2006

**B. S.**
Nanjing University, Nanjing
Biochemistry
1999

Research Experience

Massachusetts Institute of Technology
Graduate Research Assistant
Cambridge, MA
2000-2006

Nanjing University
Undergraduate Research Assistant
Nanjing, China
1996-1999

Publications and Presentations

J. Chen, & J. Stubbe, Boston DNA Repair and Mutagenesis Seminar, November 2003, Cambridge, MA.

Honors and Awards

Sigma Xi 2004-
The People’s Fellowship 1996-1999
Hollison Scholar 1995-1996
The President’s Special Award 1995
Gold Medalist, 27th International Chemistry Olympiad 1995