6-21-2012

Modulation and Recognition of Nucleic Acid Structures

Alexander M. Spring
Georgia State University

Follow this and additional works at: http://scholarworks.gsu.edu/chemistry_diss

Recommended Citation

This Dissertation is brought to you for free and open access by the Department of Chemistry at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Chemistry Dissertations by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.
MODULATION AND RECOGNITION OF NUCLEIC ACID STRUCTURES

by

ALEXANDER M. SPRING

Under the Direction of Markus W. Germann

ABSTRACT

The fidelity of an organism’s genome is central to biology. DNA, however, is constantly being damaged and modified by a variety of sources. As a result of these changes, repair enzymes, polymerases, and other interrogating biomolecules must be able to recognize, repair, and adapt to a multitude of different structures and dynamics presented. Manipulation of natural systems via the development and introduction of novel bases and DNA structures only adds to this complexity. In addition, specific RNA sequences are becoming more prevalent therapeutic and diagnostic targets. These include retroviruses and other viruses that maintain their genome with RNA. Unlike DNA, RNA poses a unique challenge as targets due to their highly diverse secondary and tertiary structures. In this manuscript, three different nucleic acid
systems were chosen to investigate how intramolecular and intermolecular interactions impact their own structure as well as giving further insight into how nucleic acids are recognized and distorted by interrogating damage specific enzymes as well as structure specific proteins.

INDEX WORDS: Nucleic acids, Deoxyrbonucleic acid, DNA, Ribonucleic acid, RNA, Endonuclease IV, Base excision repair pathway, Universal base, Alpha anomer damage, Alpha anomer adenosine, Human immunodeficiency virus, HIV, Rev response element, Zinc fingers, RNA binding zinc fingers, NMR, Fluorescence, Fluorescence resonance energy transfer, FRET
MODULATION AND RECOGNITION OF NUCLEIC ACID STRUCTURES

by

ALEXANDER M. SPRING

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in the College of Arts and Sciences

Georgia State University

2012
MODULATION AND RECOGNITION OF NUCLEIC ACID STRUCTURES

by

ALEXANDER M. SPRING

Committee Chair: Dr. Markus W. Germann

Committee: Dr. Kathryn Grant
Dr. Jerry Smith
Dr. W. David Wilson

Electronic Version Approved:

Office of Graduate Studies
College of Arts and Sciences
Georgia State University
August 2012
DEDICATION

I dedicate this to Derren, my family, and to Mark. Without your support, none of this would be possible. Thank you.
I would like to thank Mark for years of support, patience, tutelage, and friendship. He has proven himself to be not only an invaluable and insightful mentor but also a very close friend. My family has been there every step of the way with their love and support; thank you Dad, Harry, André, Amelia, Conley, Uncle Hugh, Uncle Jack, Judy, and Bill. Derren convinced me to start down this path, stood by me during the highs and lows, listened to me rehearse countless talks, proofread papers in which he had trouble even identifying the parts of speech, and most importantly maintained my sanity throughout this whole process. Thank you. I’d like to thank Dee Daley, Rebecca Peterson, Robin Defoe, David Horvath, Margaret Jones, Jeston and Fay Connell, and David, Evelyn, and Katherine Morris. I’d also like to thank my lab mates and especially Subrata Mishra, Chris Johnson, and Jin Zhang for countless hours of conversation and debate. Additionally I would like to thank our collaborators Dr. Frank Seela his work with the universal base, Dr. Richard Cunningham for his work with endonuclease IV, and Dr. Galen Collier for the extended molecular dynamics studies. Lastly, I’d like to thank the Molecular Basis of Disease program for my funding and insightful meetings and conferences.
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS ........................................................................................................... V

LIST OF TABLES........................................................................................................................ XII

LIST OF FIGURES....................................................................................................................... XIII

1 INTRODUCTION ..................................................................................................................... 1

1.1 Novel and Damaged Bases ................................................................................................. 1

1.2 DNA Damage and Repair ................................................................................................. 2

1.2.1 The BER Enzyme Cascade ......................................................................................... 3

1.2.2 Endonucleases ............................................................................................................. 4

1.3 Viral RNA Targeting with Zinc Fingers .......................................................................... 4

1.4 Chapter Outline ..................................................................................................................... 6

1.5 References: ......................................................................................................................... 7

2 THE N° LINKED 8-aza-7-deazaadenine base: how structure and dynamics allow it to function as a universal base ............................................................................................................ 10

2.1 Preface ................................................................................................................................. 10

2.2 Abstract ............................................................................................................................... 10

2.3 Introduction .......................................................................................................................... 11

2.4 Methods and Materials ....................................................................................................... 13

2.4.1 Oligonucleotides ........................................................................................................... 13
2.4.2 Melting Temperature Studies ................................................................. 13
2.4.3 NMR Sample Conditions ........................................................................ 14
2.4.4 NMR Experiments..................................................................................... 14
2.4.5 Structure Calculations .............................................................................. 15
2.4.6 Computational Methods .......................................................................... 16

2.5 Thermodynamic Results ............................................................................ 18

2.6 Results – NMR Structural Data ................................................................. 19

2.6.1 Glycosidic Bond Orientation .................................................................. 19

2.6.2 Glycosidic Bond Orientation in the Duplex ........................................... 20

2.6.3 Overall Helical Geometry ....................................................................... 20

2.6.4 Base Pairing .......................................................................................... 21

2.6.5 Chemical Shift Perturbations .................................................................. 21

2.6.6 Deoxyribose Conformation ................................................................... 22

2.6.7 Backbone Conformation ......................................................................... 22

2.7 Results – Structure Calculations and Analysis .......................................... 23

2.7.1 A1 Glycosidic Bond Angle ..................................................................... 23

2.7.2 Structure Calculation ............................................................................. 23

2.7.3 Structure Analysis .................................................................................. 24

2.7.4 RDC Implementation .............................................................................. 25
2.7.5 Supercooled Aqueous NMR .................................................................26

2.8 Results – Conformational Exchange, Dynamics, and in-silico Analysis ..........26

2.8.1 “The Case of the Disappearing H2 Proton” ...........................................26

2.8.2 Broadened $^1$H Resonances and ROESY Spectra ....................................27

2.8.3 $^{13}$C T$_1$ Relaxation as a “Probe” for Heightened Dynamics .....................27

2.8.4 Reduced HSQC Peak Intensities...............................................................28

2.8.5 in-silico .....................................................................................................28

2.9 Discussion ....................................................................................................29

2.9.1 Structure ..................................................................................................29

2.9.2 Dynamics ..................................................................................................31

2.9.3 Dynamics and the Disappearing H2 Proton ...............................................32

2.10 Conclusion ..................................................................................................34

2.11 References ..................................................................................................53

2.12 Supplemental Information ............................................................................58

3 ENDONUCLEASE IV AND $\alpha$-ANOMERIC CONTAINING DNA SUBSTRATES: USING FRET TO DETERMINE THE GLOBAL CONFORMATION OF THE DNA SUBSTRATE ...........64

3.1 Introduction .................................................................................................64

3.1.1 APE1 and EndoIV Complex Structures ....................................................65

3.1.2 Endonuclease Activity on $\alpha$ Anomeric Damage .....................................66
3.1.3 Flanking Sequences Impact on Unbound αA Oligonucleotides .................................................67
3.1.4 EndoIV Processing of αA Lesions .............................................................................................68

3.2 Methods and Materials ..................................................................................................................69
3.2.1 Substrate Design, Synthesis, and Purification ............................................................................69
3.2.2 EndoIV Mutant E261Q .................................................................................................................71
3.2.3 FRET ...........................................................................................................................................71
3.2.4 Distance Calculations using Acceptor Emission ...........................................................................73
3.2.5 Fluorescence Emission Measurements .........................................................................................75

3.3 Results and Discussion ....................................................................................................................76
3.3.1 Fluorescein Quantum Yield ..........................................................................................................76
3.3.2 Förster Distance ............................................................................................................................77
3.3.3 Free Substrate FRET Calculations ...............................................................................................77
3.3.4 E261Q – αA Substrate Binding Constants Using Anisotropy .......................................................78
3.3.5 Solution Fluorescence Measurements of the Bound Substrate ...................................................79

3.4 Conclusion ......................................................................................................................................81

3.5 Appendix .........................................................................................................................................82
3.5.1 In-gel FRET Measurements .......................................................................................................82

3.6 References: ....................................................................................................................................98

4 CHARACTERIZATION OF THE RREIIB-TR RNA ........................................................................100
4.1 Preface.........................................................................................................................100

4.2 Overview & Background.................................................................................................100

4.2.1 HIV-1, Rev, and the Rev Response Element (RRE).................................................100

4.2.2 Generation of RRE Binding Peptides .......................................................................101

4.2.3 Using Phage Display to Create a Zinc Finger Protein with Affinity to RREIIB RNA.................................................................................................................................101

4.2.4 The ZF29 RREIIB-Tr Binding Zinc Finger .................................................................102

4.2.5 Isothermal Titration Calorimetry for Thermodynamic Profiling of ZF29...103

4.2.6 Fluorescence and $^{13}$C $^{15}$N NMR Investigation of the RREIIB-Tr RNA........104

4.3 Results and Discussion .................................................................................................106

4.3.1 Solution Fluorescence of 2AP Substitutions of RREIIB-Tr .........................106

4.3.2 NMR of $^{13}$C and $^{15}$N Enriched RREIIB-Tr .....................................................107

4.3.3 Assignment Methodology.......................................................................................109

4.3.4 Testing and Parameterizing Thru-Bond NMR Experiments.........................111

4.4 Conclusion ......................................................................................................................116

4.5 Methods .........................................................................................................................118

4.6 References: ...................................................................................................................140

5 METHODS AND TECHNIQUE DEVELOPMENT ..................................................143

5.1 Supercooled Aqueous NMR Using Agarose Gels ..............................................143
LIST OF TABLES

Table 2-1 Thermodynamic Properties of UB Containing Oligonucleotides ....................... 37
Table 2-2 Summary of NMR Structure Restraints ............................................................. 38
Table 3-1 Free Substrate Distances ................................................................................. 85
Table 3-2 Equilibrium Constants (K_d) for CαAC and CαAG ........................................ 86
Table 3-3 Calculated FRET Distances for the Bound αA Substrates ............................... 87
Table 4-1 Thermodynamic results of single point mutations .......................................... 120
Table 5-1 Optimization Reaction Setup ........................................................................... 159
Table 5-2 Production Reaction Setup ............................................................................... 160
Table 5-3 Software components needed for RESP calculations ..................................... 173
Table 5-4 Boranophosphate charge derivation ................................................................. 174
LIST OF FIGURES

Figure 2-1  A) Stick representations of deoxyadenosine (left) and the UB nucleoside (right). The red numbering on UB nucleoside represents the atomic purine numbering system used in the manuscript. B) “Ball-and-stick” representations of the “anti” and “syn” glycosidic bond orientations of UB nucleoside. Distances between the H7 and H1’ protons are noted. Red = oxygen, blue = nitrogen, cyan = carbon, and white = hydrogen.................................................................39

Figure 2-2  A) The observed NOESY pathways for the UB sequence. B) NOESY spectrum (150 ms mixing time) of the base to H1’ region for the UB structure. Black and red lines represent connectivity for the different strands. Inset shows the unique UB5 H7 to G4 H1’ and G6 H1’ NOE contacts. C) Imino proton spectrum of the UB structure at 280 K. In blue are the resonances for G4 and G6. Note the lack of a T14 imino proton resonance.................................................................40

Figure 2-3  A) The summation of the absolute difference in chemical shifts for the sugar region (|UB – control| = Δ ppm). B) Summation of the absolute difference in chemical shifts for the base region. C) (Top) $^{31}$P spectrum of the UB and control structures at 298 K. Arrows point out the changes in the chemical shifts for G6 and C15. (Bottom) A mapping of the differences in $^{31}$P chemical shifts between the control and UB structures (control – UB). .................................................................41
Figure 2-4  A) The final bundle of structures for the UB structure. The UB 5 residue is in red. The left is a view of the major groove; the right is a view of the minor groove.


Figure 2-5  A) Stick model of the UB structure calculated using both NOESY and RDC restraints. B) Ribbon representation of UB structure (NOESY + RDC restraints). The observed displacement in the backbone is in red and the UB is in red. C) Ribbon representation of the control structure. The same region of the backbone as B) is highlighted in red for comparison. A5 is in blue.

Figure 2-6  A) Base stacking for the UB (left) and control (right) structures. (Left) Blue = G6, red = UB5, green = G6. (Right) Blue = G6, red = A5, green = G4. In gray is the glycosidic sugar connectivity of each base. The bases are viewed down the helix. B) Residues T14 and C15 base stacking for the UB (NOESY + RDC restraints) and control structures. Each of the structures were aligned to C15. The multicolored base is C15 (red = oxygen, blue = nitrogen, cyan = carbon, white = hydrogen); the red base is UB structure T14 base, the blue is the control structure T14 base. The white spheres on C15 represent the H5 and H6 protons.

Figure 2-7  A) The A1, T2, G3, and G4 residues of the UB structure with a ribbon cartoon representing the backbone. Left is with NOESY restraints only; on the right is NOESY + RDC restraints. Red = oxygen, blue = nitrogen, cyan = carbon, and white = hydrogen. G3 and G4 of each structure have been aligned. B) The core G4, UB5, G6, C13, T14, and C15 residues of the UB structure before and after
implementation of RDC restraints. Blue = NOESY only, red = NOESY + RDC restraints.................................................................45

Figure 2-8  The UB structure with only RDC restraints. A) The helix of the RDC only restraints structure viewed from the minor groove. The UB5 residue is in red. B) The predicted base pairing using only RDC restraints. Red = oxygen, blue = nitrogen, cyan = carbon, and white = hydrogen. C) The core G4, UB5, G6, C13, T14, and C15 residues of the UB structure with only RDC restraints viewed from the minor groove. ........................................................................................................46

Figure 2-9  The supercooled aqueous imino proton spectra of the UB structure at -12 °C. The inset is a NOESY spectrum (150 ms mixing time) of the imino – methyl region collected at the same temperature. The red arrow points out the appearance of a third methyl NOE crosspeak at ~14 ppm which is not observed at 280 K. Note the weak NOE crosspeak 13.35 ppm (G4’s imino proton). .................................47

Figure 2-10  HSQC spectra of the base region of A) double stranded UB oligonucleotide at 298 K, B) single stranded UB oligonucleotide at 298 K, and C) double stranded UB oligonucleotide at 330 K. The red arrows point out UB5 C2-H2 resonance. Note its disappearance in the duplex at 298 K (A). .................................................................48

Figure 2-11  A) Radar graphs of the α and ζ backbone torsion angles vs the time course of the MDtar simulation for the UB structure. In red is the measured torsion angle over the simulation and the rings represent the simulation time step. Torsion angles are around the ring counterclockwise with the top center being 0°. B) and C) Histograms for the various helical parameters for the MDtar simulation of the UB
structure; the overall population size for all histograms is 1,000 structure
snapshots.................................................................50

Figure 2-12 The predicted base pairs for the control and UB sequences. The top base
pair is the control A5:T14, the middle is the UB5:T14 base pair with NOESY + RDC
restraints, and the bottom is the UB5:T14 base pair with only RDC restraints. Red =
oxygen, blue = nitrogen, cyan = carbon, gold = phosphorous, and white = hydrogen.
.................................................................51

Figure 2-13 A view of the G4, UB5, and G6 bases from the major groove. The red
spheres represent the O6 carbonyl oxygen of G4 and G6. On the left is the
structure with both NOESY + RDC restraints; on the right is the structure with only
RDC restraints. Note the different orientations of the UB5 H2 proton with respects
to the carbonyls. Red = oxygen, blue = nitrogen, gray = carbon, yellow =
phosphorous, and white = hydrogen.................................................................52

Figure 3-1 Stick diagrams of A) deoxyadenosine and B) αA. Atom position numbering is
noted in A. .................................................................88

Figure 3-2 X-ray crystal structure of endoIV in complex with an abasic containing DNA
substrate.(10) (pdb id: 1QUM) The DNA substrate is in blue with the abasic lesion
in red. The cartoon representation is the endoIV enzyme. The red lines on the DNA
substrate approximate the helical axis on either side of the lesion site, which points
out the induced kink on the substrate.................................................................89
Figure 3-3  A) αA residue (red) manually docked in the endoIV catalytic pocket in place of the abasic residue. (10, 11)(pdb id: 1QUM)  B) APE1 – abasic (red) complex structure focused on the catalytic pocket. (pdb id: 1DE8)........................................90

Figure 3-4  Trends observed in the endoIV substrates.(9) Top from left to right: abasic containing substrate found in complex with endoIV (grey) (pdb id: 1QUM), solution structure of CαAC (red) (pdb id: 1S75), solution structure of CαAG (orange) (pdb id: 2LIB), and a model of canonical B-type CAC (blue). Bottom: observed trends in the structures.................................................................91

Figure 3-5  Denaturing polyacrylamide gel (PAGE) of DNA substrates processed by the wild type endoIV and the mutant E261Q. PAGE conditions are: 15% polyacrylamide (19:1 bis-acrylamide ratio), 8 M urea, and tris-borate-ethylenediaminetetraacetic acid buffer (TBE). The picture is observed colors captured with a portable camera; green is 6-FAM and red is TAMRA. From left to right: DNA only (CαAC), processing of the CαAC substrate with the wild type endoIV (note the processing of the green 6-FAM strand which contains the αA lesion), E261Q with the CαAC substrate, and E261Q with the CαAG substrate (note how neither substrate backbone is cleaved with the E261Q mutant). ........................................................................................................92

Figure 3-6  X-ray crystal structure of the catalytic pocket of the E261Q mutant in complex with an abasic containing substrate. (20) (pdb id: 2NQJ) In grey is the pocket, green is the E261Q mutation, orange are the Zn^{2+} ions, and the ball and stick representation is the abasic lesion site (red = oxygen, cyan = carbon, gold = phosphorous)........................................................................................................93
Figure 3-7  Graph of fluorescence intensity vs. absorbance for 6-FAM and rhodamine B.

The red line is the single labeled 6-FAM CcAC substrate and the blue line is the rhodamine B reference. The trend line is a linear fit to the data. .............................94

Figure 3-8  Emission and absorbance spectra of 6-FAM and TAMRA, respectively. Both spectra have been normalized so the maximum value is equal to 1 for visual comparison only. Red is the 6-FAM fluorescence emission spectrum when excited at 495 nm, in blue is the TAMRA absorption spectrum. The baseline excitation peak has been subtracted from the 6-FAM emission spectrum. .............................95

Figure 3-9  Plotting of anisotropy vs. E261Q concentration. A) 6-FAM  B) TAMRA. Note, both are for the CcAG substrate. ..........................................................96

Figure 3-10  Emission spectra of 6-FAM and TAMRA when directly excited. A) Raw emission spectra from 6-FAM (excitation at 495 nm); lines represent spectra with increasing amounts of E261Q added. B) Raw emission spectra from TAMRA (excitation at 560 nm); lines represent spectra with increasing amounts of E261Q added. Color coding for A) and B) are noted on the spectra. C) Volume adjusted intensities at 519 nm vs. E261Q concentration when excited at 495 nm. D) Volume adjusted intensities at 579 nm vs. E261Q concentration when excited at 560 nm. Note, lines in C) and D) only connect the data points. ..............................................97

Figure 4-1  A) Sequence and secondary structural prediction of the REV Response element. The IIB stem loop is outlined. (2)  B) Sequence and secondary structure of the RREIIB-Tr RNA; upper stem, bulge, and lower stem regions of the RNA are identified.................................................................121
Figure 4-2  Cartoon of the late stage transport of unspliced HIV RNA. Red hexagons represent the Rev protein, the blue box represents the RRE region of the HIV RNA, and the straight black line represents the unspliced HIV RNA. .................................122

Figure 4-3  Cartoon representation of the ZF29 protein. (pdb ID# 2AB3) In red is the alpha helix and in yellow is the beta sheet. The silver sphere represents the Zn\(^{2+}\) and the amino acids side chains with ball and stick representations are the zinc coordinating His 23, His 27, Cys 5, and Cys 10 residues. ........................................123

Figure 4-4  Imino proton spectra of the RREIIB-Tr RNA before (bottom) and after (top) binding with ZF29. Circled in red and shaded in orange point out new peaks formed upon binding. Peak assignments are detailed. The inset is the sequence and secondary structure of the RREIIB-Tr RNA; circled in red are base the observed formed base pairs; the red box outlines the bulge region of the RNA. .......................124

Figure 4-5  A) Cartoon representation of the ZF29 protein with ball and stick representations of the single point mutations made for ITC experiments. Left and right represent different views of the same ZF29. B) A sequence mapping of the ZF29 single point mutations with residue numbering made for ITC experiments. (19) ..................................................................................................................125

Figure 4-6  A) Stick diagram of 2AP nucleoside. Blue = nitrogen, red = oxygen, cyan = carbon, white = hydrogen. B) Sequence and secondary structure of RREIIB-Tr RNA. In red boxes are the substitutions made. C) Fluorescence emission at 380 nm plotted against ZF29 concentration. Red line represents raw data; blue and green
lines represent corrections made for non-specific binding-like curve observed at higher concentrations................................................................. 126

Figure 4-7 $^1$H–$^{15}$N HSQC spectrum of the free RREIIB-Tr RNA. Noted in blue are the typical guanosine (G) and uracil (U) $^{15}$N chemical shift resonances. ..................... 127

Figure 4-8 Overlay of the $^1$H–$^{15}$N HSQC spectra for the RREIIB-Tr bound and unbound with the ZF29 protein. In red is the bound spectrum, in black is the free spectrum. Blown up regions are shown on the right side of the figure. In the top blown up spectrum ($^1$H: 11.6 – 12.5 ppm and $^{15}$N: 145.5 to 150 ppm), the intensity has been increased to visualize the broad new resonances; arrows point out emerging resonances at 11.85 and 12.15 ppm upon ZF29 binding. The $^{15}$N chemical shifts indicate these are guanosine resonances. On the bottom is an expansion of the 12.5 – 13.5 ppm region; the arrow notes the broad resonance corresponding to two peaks at approximately 12.9 ppm. The $^{15}$N chemical shifts indicate these are guanosine resonances, one of which is the G76 imino proton........................................... 128

Figure 4-9 Comparison of Dharmacon and in-house transcribed free and bound RREIIB-Tr RNA imino proton spectra. A) ZF29 bound Dharmacon RNA (20)  B) ZF29 bound in-house transcribed RNA  C) free in-house transcribed RNA  D) free Dharmacon RNA. Chemical shift assignments are noted. (20) .......................................................... 129

Figure 4-10 HNN COSY. On the left are stick representations of the RNA bases with arrows representing magnetization pathways and circles representing observed resonances. On the right is a sample spectrum from the 300 µM RREIIB-Tr RNA. Typical $^{15}$N chemical shift resonances are noted on the left of the spectrum. Dotted
lines connect the hydrogen bonded nitrogen and imino protons. Note the lack of connectivity for G77 and U43, which are base paired to one another and do not involve hydrogen bonding to a nitrogen acceptor.

Figure 4-11 H(N)CO. On the left are stick representations of guanosine and uracil with arrows representing magnetization pathways and circles representing observed resonances. On the right are four spectra of a GMP / UMP sample in DMSO. The four panels represent varying delays of D23, which is equal to \(1/4J_{\text{NCO}}\). Varying this delay allows for some editing to be done to remove resonances (U-C2, G-C2, and G-C6) in case of spectral overlap. Coupling values for U N3-CO4 and U N3-CO2 are as follows: \(U J_{\text{N3CO4}} = 10 \text{ Hz}, U J_{\text{N3CO2}} = 19 \text{ Hz}\). (22)

Figure 4-12 HCP. On the left is a stick representation the adenosine dinucleotide with circles representing observed resonances. Combined with an HCCH-TOCSY, this will give sequential assignments from sequential H1' to H1' resonances.

Figure 4-13 HCCH-TOCSY. On the left is a stick representation the adenosine nucleotide with circles representing observed resonances. On the right are sample spectra from the 300 \(\mu\)M RREIIB-Tr RNA. The top spectrum represents all connected protons and carbons when focused on a specific H1' proton. The spectrum on the bottom represents all connected protons to a C1' resonance of the same H1' proton. The dotted lines connect the proton resonances between the different views.

Figure 4-14 HCN. On the left is a stick representation of the adenosine nucleotide with arrows representing magnetization pathways and circles representing observed
resonances. On the right are two spectra from the 300 µM RREIIB-Tr RNA. The left spectrum is focused on the base proton and the right spectrum is focused on the sugar proton of the same nucleotide. The dotted line connects the N1 connecting the sugar and base spin systems. The horizontal dimension is $^{13}$C, and the vertical is $^{15}$N.

Figure 4-15  ctHMQC. On the top are stick representations of guanosine and cytosine with arrows representing magnetization pathways and circles representing observed resonances. On the bottom is a spectrum from the 300 µM RREIIB-Tr RNA. Green resonances represent the adenosine and guanosine base H8 protons; the red are cytosine and uracil H6 resonances. Green and red are opposite phases.

Figure 4-16  HCCH-COSY. On the left is a stick of adenosine with arrows representing magnetization pathways and circles representing observed resonances. On the right are sample spectra from the 300 µM RREIIB-Tr RNA. Green and red represent opposite phases; the left and right panels are H2 and H8 proton resonances for the same base. Dotted lines connect the C5, C4, and C6 resonances of the base. The F2 dimension (horizontal) is the C2 and C8 resonances (left and right panels respectively).

Figure 4-17  H(CC)NH. On the left are stick representations of guanosine and uracil with arrows representing magnetization pathways and circles representing observed resonances. On the right is a sample spectrum from the GMP and UMP sample in DMSO. The F2 dimension (horizontal) is the imino proton region, the F1 dimension
(vertical) is the base region. The peak at 11.5 ppm is for U; the peak at 10.8 ppm is for G. ................................................................. 137

Figure 4-18  H6C6/H5C5(C4)NH. On the left is a stick representation of cytosine with arrows representing magnetization pathways and circles representing observed resonances. On the right is a sample spectrum from the UMP and CMP sample in DMSO. The $^1$H dimension is the amino proton region, the $^{13}$C dimension is the C5 (90 – 95 ppm) and C6 (140 – 145 ppm) region. ......................................................... 138

Figure 4-19  Imino proton spectra of 3MU substituted U72 RREIIB-Tr RNA. A) free RNA B) ZF29 bound RNA C) ZF29 bound with non-substituted RNA. Blue diamonds point out the U66 and U45 resonances, red diamonds point out the additional imino protons present when bound to ZF29. (20) ................................................................. 139

Figure 5-1  A) 1% agarose gels with 50 mM sucrose spectra in D$_2$O: top and bottom are phase sensitive NOESY spectra at -7 and +17 °C, respectively (16 scans, 512 increments, 90° shifted sine-squared function). Red and black peaks denote negative and positive peaks respectively. Middle spectrum is $^1$H at +17°C. B) CaAG oligonucleotide (30 μM oligonucleotide, 10 mM sodium phosphate, 100 mM NaCl, pH 6.8) Top panel is the sequence of CaAG oligonucleotide, bottom three spectra are the $^1$H NMR imino proton region (jump and return solvent suppression). [9] The resonance noted with an arrow is the thymine imino proton base paired with the αA ........................................................................................................ 148

Figure 5-2  DNA Templates for RNA Transcription Reactions ........................................... 161

Figure 5-3  RREIIB-Tr DNA template ............................................................................. 162
1 INTRODUCTION

DNA’s structure, although initially solved in 1953 by Watson and Crick, continues to exhibit unique structures. (1) In the PDB database, DNA alone has over 1300 solved structures and there are more than 2300 DNA / protein and 1200 RNA / protein complex structures. (pdb.org as of 6/14/12) Understanding of DNA’s structure, dynamics, and their interactions with interrogating proteins, as well as how it is recognized, processed, and repaired, are critical for our broader understanding of chemical biology.

1.1 Novel and Damaged Bases

The development and introduction of novel DNA bases is a growing field of chemical biology. Indeed, many efforts are underway to design additional “letters” to the DNA genomic code in order to aid in the creation of a “semi-synthetic” organism with additional coding abilities. (2-6) In conjunction with this, several labs are exploring the design and capabilities of universal bases. (7-9) These bases are ambiguous in their pairing and can utilized as quite powerful tools in microbiology by providing unique probe designs, primers, and randomized amplification in PCR and transcription applications. (10) (11-13) Ambiguity in base pairing also occurs in natural systems; for instance inosine incorporation in tRNA allows for broader recognition capabilities with fewer anti-codons. Early attempts at developing universal bases used deoxyinosine as a template. Other universal base designs have been quite varied ranging from systems that lack hydrogen bonding groups, to nitroazole analogs and other canonical base analogs, or as simple as an abasic site. (7, 9, 14) Understanding how these novel bases are accommodated in a duplex environment can aid in their design and further our understanding of how
interrogating biomolecular systems interact with these novel base designs. DNA base modifications also arise from naturally occurring damage.

1.2 DNA Damage and Repair

The fidelity of the genome is critical. The central role of this fidelity is to maintain the proper ordering of only four canonical bases: adenosine, cytosine, guanine, and thymine. However, cellular process must be able to distinguish from a multitude of different types of damage and modifications. These range from the introduction of non-canonical DNA bases through inadvertent but continual damage processes (oxidative damage, single or double strand breaks, and bulky adduct formation) to intentional change of bases by the cell’s own processes (DNA methylation, for instance). If this damage goes uncorrected, the resulting lesion site may cause mutation in the genome, cancer in the host organism, or death. Complex repair pathways exist to address damage to DNA; these include the base excision repair pathway (BER) which recognizes small DNA base damage, the nucleoside excision repair pathway (NER) which recognizes larger, bulkier adducts (but also some small damage as well), mismatch repair, and homologous recombination and nonhomologous end joining repair pathways. (15)

Small base damages in particular are more difficult to recognize. Although the BER is the primary pathway for this repair, the NER pathway has also been implicated in the repair of small damages including abasic sites and ribonucleotide mismatch incorporation. (16, 17) This redundancy in detection of small nucleotide damage underlines the importance their repair as well as the difficulty in their detection. The resilient ability of the BER to process different types of DNA base damage with a relatively small number of enzymes is quite remarkable. (15)
1.2.1 **The BER Enzyme Cascade**

The BER pathway involves a cascade of enzymes, which typically begins with a damage-specific glycosylase that recognizes and cleaves the N-glycosidic bond of the lesion site generating an abasic site. (18) There are two types of glycosylases. Bifunctional glycosylases contain both glycosylase and lyase functionality (e.g. hOGG1) and characteristically utilize a ε-NH₂ for cleaving the N-glycosidic bond, while monofunctional glycosylases are solely glycosylases (hSMUG1 and MutY) and utilize water as the nucleophile. The bifunctional glycosylases further divide into two classes containing β or β, δ lyase functionality. Both cleave the backbone on the 3’ side of the lesion generating an aldehyde adjacent to a 5’ phosphate on the next base. Glycosylases with β, δ functionality go one step further and also cleave the backbone on the 5’ side of the lesion leaving 3’ and 5’ phosphate groups on either side of the removed AP site. The resulting product for both enzyme classes is processed by an endonuclease. For bifunctional glycosylases, the resulting 3’ terminal phosphate or 3’ abasic moiety is processed by an endonuclease producing a single nucleotide gap; for monofunctional glycosylases, the 5’ phosphodiester bond is incised. (18-21) BER continues this repair by either short (1nt) or long patch (2–12nt) repair. Short patch repair is the excision and repair of one base which utilizes polymerase β and ligase III for extension and sealing, respectively. Long patch repair begins in the same manner but polymerases γ or ε are utilized, FEN1 lyase removes the displaced strand, and ligase I seals the nick. (15, 18, 19, 22, 23)
1.2.2 Endonucleases

Rapid recognition and repair of the highly cytotoxic abasic lesion site is required. In humans, APE1 is the endonuclease that recognizes abasic sites and nicks the phosphodiester bond on the 5’ side. APE1 is utilized for 95% of all BER repair activity,\(^{(24, 25)}\) while other organisms have alternate means. The substrate spectrum of this enzyme is quite remarkable, ranging from double stranded to single stranded DNA and even RNA.\(^{(26)}\) APE1 has been recently shown to mediate c-myc RNA degradation and been involved in rRNA maintenance.\(^{(27-29)}\) Endonucleases are evolutionary preserved. The homologs for *E. coli* exonuclease III are APN2 in yeast and APE1 in humans. In addition, *E. coli* contains endoIV which participates in BER but to a lesser extent. Although there is no human analog of this enzyme, it is preserved in yeast as APN1.\(^{(24, 30)}\)

In addition to functioning as an endonuclease, APE1 activity also has 3’ to 5’ exonuclease, 3’ phosphodiesterase, RNaseH activities, and is involved in the regulation of transcription factors.\(^{(26, 31, 32)}\) Intriguingly, APE1 and endoIV, which are both damage general enzymes, may also engage in damage-specific BER activity via recognition and 5’ nicking of α anomeric nucleoside damage.\(^{(33-35)}\)

1.3 Viral RNA Targeting with Zinc Fingers

Unlike most organisms, retroviruses utilize RNA for their genome. These viruses use reverse transcriptases to convert their genomic RNA to DNA and then integrases to incorporate that viral DNA into the host cells genome. This viral genome then exploits the host cell’s sys-
tems to develop new viruses. With the human immunodeficiency virus (HIV), a key part of the lifecycle is the export of unspliced and singly spliced RNA’s for viral packaging. This transport and non-splicing event is dependent on the viral rev protein binding to the rev response element. (36) RNA, either for viral survival or as targets for treatment, represent a promising therapeutic and diagnostic target.

An approach for RNA targeting therapies is the development of RNA binding zinc fingers. Although formed from similar building blocks, DNA and RNA present quite different binding interfaces for zinc finger proteins. DNA is predominately double stranded and tends be regular in structure with a wide and deep major groove. When bound to DNA, CCHH zinc fingers are situated in the major groove making specific contacts between a-helix side chains and typically 3 sequential DNA bases allowing for sequence specific recognition. Further stabilization derives from electrostatic phosphate contacts and non-specific hydrophobic interactions. (37-39) Zinc fingers are quite often found in proteins as tandem repeats that are connected by short linkers, most often of the Krüppel-type TGEKP. The repeats allow for higher sequence specificity while the linkers aid in overall stabilization of the DNA-ZnF complex via proper spacing of the fingers.(38, 40, 41)

In contrast, regular double helical RNA presents a narrow but deep major groove that cannot accommodate an α-helix.(42) However, RNA exhibits a wide variety of secondary and tertiary structures with bulges, loops, extrahelical bases, and junctions that can be interrogated by an α-helix. (43, 44) CCHH zinc finger binding of double stranded helical RNA typically occurs at or near bulges and loops that feature widened major grooves and unique structural features.(42) Electrostatic contacts are generally made with the phosphodiester backbone al-
though some sequence specificity can occur through hydrophobic stacking interactions with extrahelical or bulged bases.\textsuperscript{(45)} The short linkers between tandem ZnF repeats serve a similar purpose as with the DNA binding motifs: proper orientation of the zinc finger.\textsuperscript{(41)} Overall the ZnF-RNA contacts are specific for the spatial presentation of the RNA structure. Because the RNA structure is dependent on the sequence context, this specific spatial recognition corresponds to an indirect sequence readout.

1.4 Chapter Outline

In the following chapters, the structure and dynamics resulting from intramolecular and intermolecular interactions of different nucleic acid systems is explored. Chapter 2 emphasizes how subtle changes to a DNA base impact not only the structure but also the dynamics of the duplex. In Chapter 3, the global conformation of an alpha anomeric containing DNA substrate is investigated in order to determine if its processing by endonuclease IV occurs in a similar manner as its normal abasic target. Chapter 4 details the structural changes occurring in RNA when in complex with an RNA binding zinc finger as well as NMR techniques needed for elucidating an isotopically enriched RNA structure. Techniques developed and other experimental methods are detailed in the final chapter.
1.5 References:


2.1 Preface

Synthesis of the 8-aza-7-deazaadenine (pyrazolo[3,4-d]pyrimidin-4-amine) N^8-(2'-deoxyribonucleoside) deoxyadenosine analog and thermodynamic profiling were performed by our collaborator Dr. Frank Seela at the Universität of Osnabrück, Germany. Extended molecular dynamics and accelerated molecular dynamics simulations were completed by our collaborator Dr. Galen Collier at Clemson University.

2.2 Abstract

Promiscuity in base pairing has the potential to enable novel applications such as PCR primers, randomized sequencing, and probes. Ideally, this universal base should be able to pair indiscriminately to each of the canonical bases with little destabilization of the overall duplex. In reality, most bases probed for this property either greatly destabilize the duplex or do not ambiguously base-pair with each of the natural bases. The novel base 8-aza-7-deazaadenine (pyrazolo[3,4-d]pyrimidin-4-amine) N^8-(2’deoxyribonucleoside) deoxyadenosine analog (UB) has been found through T_m studies to pair with each of the natural DNA bases with little overall destabilization or sequence preference. Our current NMR investigation of a single UB incorporated into a nonamer duplex reveals that the UB participates in base stacking with little overall
perturbation to the global structure, yet participates in an unusual conformational sampling between multiple base pair conformations which causes the UB H2 proton to disappear from all NMR spectra in our sequence environment. This unique sampling may account for the observed thermodynamic stability when opposite the four canonical bases.

2.3 Introduction

The genetic code is dependent on the discriminate pairing of the 4 natural bases. Efforts to expand the “genetic alphabet” with the development of unnatural bases have shown promise in a variety of roles including PCR amplification, RNA transcription, and DNA primers and probes. (1-3) In contrast, a universal base is indiscriminate in its base pairing and interaction with interrogating enzymes. This concept of ambiguous base pairing is already implemented by nature; for example, tRNA’s use of inosine in the wobble position allows broader recognition of codons through ambiguous pairing. Although inosine has been used as the basis for many universal base designs, it does not base pair in an indiscriminate manner. (4, 5) An ideal universal base would base pair with all 4 natural bases uniformly and be incorporated into a DNA duplex environment with little perturbation to the duplex. More so, a truly universal base would be completely promiscuous when encountered by polymerases and allow for randomized transcription and replication. (4) Even if it does not allow for a truly randomized encounter with polymerases, a universal base’s use as probes and transcription applications within a microbiology lab setting would be invaluable. For example, higher RNA yields are generated from T7 RNA polymerase reactions when the transcription region is double stranded (Chapter 5). Inclusion of universal bases in the non-transcribed strand of the DNA template
would allow for different related template sequences with a single opposite strand thus simplifying the template design. The use of a fluorescent probe as a universal base would allow for the monitoring of site-specific structural perturbations. (6) Oligonucleotide microarray chips for the detection of pathogens could also benefit from a universal base, which would allow for a more diverse detection scheme with a single oligonucleotide. (7-9)

The Seela lab has developed a series of N^8 linked 2’ deoxyadenosine analogs that exhibit universal base properties. (10-15) The 8-aza-7-deazaadenine (pyrazolo[3,4-d]pyrimidin-4-amine) N^8-(2’deoxyribonucleoside) deoxyadenosine analog (UB) will base pair indiscriminately with all four natural base pairs despite the unique N^8 glycosidic linkage. * (Figure 2-1 A) (12, 15) The UB opposite each of the four bases in a dodecamer duplex exhibits T_m values within 2 °C of one another. (12) As compared to an A:T base pair, their studies have found that the UB base pairs with each of the four natural bases with only a moderate drop in stability (ΔT_m of -4 to -6 °C), exhibits less fluctuations as compared to mismatches (ΔT_m values of -2 to -12 °C), and is more stable than abasic sites (ΔT_m values of -10 to 17 °C). (12) They have extensively explored the addition of substituents at the 2 and 7 positions in efforts to further stabilize the base and to add probes to the UB. (11, 13-15)

Understanding how a universal base is incorporated into a duplex environment can aid in a refined design and serve as a template to help elucidate how it may interact with polymerases. (16) Here, solution NMR was used to not only determine a high-resolution structure of an oligonucleotide containing a UB opposite a dT but also probe the duplex for regions of heightened motions and chemical exchange. Complemented with extended molecular dynam-

* Note, the purine numbering scheme as outlined in figure 1 is used in this manuscript.
ics, molecular dynamics with time averaged restraints, and chemical shift calculations, the data reveal that the UB is well accommodated in the duplex with little overall impact to global geometry, yet is involved in dynamics on the \( \mu s \) to ms timescale which causes the disappearance of the UB base H2 proton in all NMR spectra under native conditions with little to no impact on any other resonances. The unique base pair conformational sampling of the UB may offer an explanation for the exhibited indiscriminate base pairing.

2.4 Methods and Materials

2.4.1 Oligonucleotides

UB containing oligonucleotides were synthesized and purified as previously described. (10, 12) All other oligonucleotides were purchased from University of Calgary Core DNA Services and purified via ion exchange (Resource 15Q column) and size exclusion liquid chromatography. Purity was assessed via denaturing polyacrylamide gels. Concentrations were calculated using extinction coefficients from the sum of mononucleotides (absorbance at 260 nm, 80 °C). (17) The extinction coefficient for the UB is 6600 M\(^{-1}\)cm\(^{-1}\). (12)

2.4.2 Melting Temperature Studies

Samples for \( T_m \) measurements were in 100 mM NaCl, 10 mM MgCl\(_2\), and 10 mM sodium cacodylate with a pH of 7.0. Total DNA strand concentration was 10 \( \mu \)M. UV absorbance was measured on a Cary-1/1E UV/VIS spectrophotometer. Thermodynamic data were calculated using the MeltWin 3.0 package.
2.4.3 NMR Sample Conditions

NMR samples were prepared in a buffer containing 10 mM sodium phosphate, 100 mM NaCl, with 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) added as an internal reference. The pH* for D$_2$O samples was 6.36; the pH for H$_2$O samples was 6.79. Samples for which imino proton spectra were collected at supercooled conditions were transferred to 1 mm borate glass capillaries with the ends flame sealed; capillaries were bundled in a standard 8 mm NMR tube. (18, 19) Temperatures below -5 °C were reduced at a rate of 1 °C per hour. (20)

2.4.4 NMR Experiments

NMR experiments were conducted on a Bruker Avance 600 MHz NMR equipped with a 5 mm or 8 mm QXI probe, $^1$H{$^{13}$C, $^{15}$N, $^{31}$P}; HSQC and $^{13}$C $T_1$ experiments were conducted using a 500 MHz Avance systems equipped with a 5 mm TXI cryoprobe $^1$H{$^{13}$C, $^{15}$N}. Standard protocols were utilized for the collection of 1D $^1$H and $^{31}$P and 2D COSY, TOCSY, CTNOESY, HPCOR and NOESY spectra as previously described. (21-24) Exceptions and unique details are as follows: for D$_2$O NOESY spectra, mixing times of 75, 150, and 250 ms with an 8 s interscan delay were used; supercooled aqueous imino proton spectra were completed using jump-and-return for water suppression and a NOESY mixing time of 150ms. HSQC and $^{13}$C $T_1$ spectra were collected for both the sugar and base regions and each utilized an interscan delay of 5s. For $^{13}$C $T_1$ relaxation measurement, the Bruker pulse program hsqct1etgpsi was used with delays of 0.01, 0.04, 0.08, 0.120, 0.160, 0.300, and 0.500 s for the base region and 0.01, 0.05, 0.100, 0.150, 0.200, 0.300, and 0.500 s for the sugar region. (25, 26) Experimental peak intensities from $T_1$ spectra
were fit to a standard 1\textsuperscript{st} order exponential decay model in Kaleidagraph and resulting $T_1$ values were calculated. (Formula 2-1)

$$M_z = M_0 e^{(-\tau/T_1)}$$

Formula 2-1

where $M_z$ is the measured peak intensity, $\tau$ is the inversion recovery delay, and $M_0$ is the full intensity of the resonance.

Assignments were completed using Sparky 3.33. (27) Referencing for $^1$H and $^{31}$P was completed using internal DSS and an external capillary containing 85\% H$_3$PO$_4$ respectively. RDC coupling values were determined for the sugar and base regions using f2 coupled HSQC spectra collected in the presence and absence of pf1 phage (ASLA) ($^2$H splitting of 18.0 and 32.2 Hz giving concentrations of 20.3 and 36.3 mg/ml for the UB and control samples respectively). (28)

2.4.5 Structure Calculations

Standard $^1$H, $^{13}$C, and $^{31}$P assignment protocols were utilized as described previously. (22) NOE volumes were calculated in Sparky 3.33 using Gaussian or Sum-over-box methodologies. (27) Structure calculation methods using an iterative RANDMARDI process with MARDI-GRAS, CORMA, and AMBER 9.0 cycles were completed as described previously. (21, 22, 29-34) Total $R^\chi$ values were calculated in CORMA using $\tau_c$ times ranging from 2.5 – 4.0 ns; the $\tau_c$ with the lowest overall $R^\chi$ values was 3.6 ns. Sugar puckering pseudorotation and fraction south calculations using measured 3-bond scalar couplings $^3$J from COSY and $^{31}$P COSY experiments was completed using a graphical method. (35) Epsilon torsion values were calculated from the ratio of peak heights collected from CTNOESY (constant time NOESY) spectra. (24) Standard backbone torsion angle ($\alpha$, $\beta$, $\gamma$, and $\zeta$) and Watson-Crick hydrogen bonding and angle restraints
were included using standard B-type DNA values. (21, 36) The B-type backbone torsion angles (except for the experimentally determined $\epsilon$ torsion angles) and Watson-Crick hydrogen bond restraints were not included for the UB5 and T14 residues. RDC restraints were measured from the base, H1’, H3’, and H4’ regions and implemented as described previously. (22)

The AMBER parm99 force-field was appended to include specific parameters for the UB. (37) Briefly, charge derivations were completed using the R.E.D. script, which uses an iterative process to calculate atomic point charges using Gaussian03 QM (HF 6-31G*) and AMBER RESP calculations. (29, 38-40) The unique C7-N8-N9 bonds were parameterized using the GAFF (General Amber Force-Field) atom definitions. (41) AMBER 9.0’s ANTECHAMBER module was used to build the force-field modification files. (29, 42) (See chapter 5.3 for further details)

Starting structures were canonical B-type DNA with the incorporation of UB at position 5. Fully restrained molecular dynamics were conducted in explicit water (TIP3P) and charges were neutralized with the addition of 16 Na+ counterions; simulation times were 10 ns. The final bundle of structures was selected from the last 10 ps at a rate of 1 structure per ps. Each structure was minimized using all restraints; the structure with the lowest overall AMBER penalties was selected as the final structure. Structural and helical parameters were determined using the Curves+ algorithm. (43)

2.4.6 Computational Methods

Molecular dynamics simulations using time-averaged restraints (MDtar) and AMBER 9.0 with a $\tau = 20$ ps were completed as described previously. (44) Structures were neutralized by the addition of 16 Na+ ions and solvated using a truncated octahedral TIP3P water box that ex-
tended 9 Å from the DNA helix in all directions; MDtar simulations were run for 1 ns. Analysis was conducted using Curves+. (43)

For extended molecular dynamics, the construction and simulation of all molecular systems was accomplished using AMBER 9.0. (29) Starting structures were neutralized by the addition of 16 Na+ counterions proximal to the backbone phosphates, and these systems were then solvated using a truncated octahedral TIP3P water box that extended more than 8 Å from the DNA helix in all directions. (45)

Systems were simulated at a constant pressure of 1.0 bar using 3-D periodic boundary conditions (PBCs) using the explicit-image model. Temperature regulation at 300 K was accomplished using a Langevin dynamics approach with a collision frequency of 2 ps⁻¹. (46) The particle-mesh Ewald (PME) method was used for calculation of long-range electrostatics. (47) Parameterization for the PME method for all simulations included limiting the direct-space summation to 10 Å. (48) 1-4 non-bonded interactions were scaled using a factor of 2.0 and 1-4 electrostatic interactions were scaled using a factor of 1.2. Bond lengths involving heavy atoms and hydrogens (X-H bonds) were held fixed using the SHAKE algorithm, which enabled a 2.0 fs time step to be used for all equilibration and production dynamics. For all simulations, the Cornell et al. 1994 force field was used with the addition of charges for the nonstandard universal base atoms. (49) To avoid end fraying, standard Watson-Crick hydrogen bond and angle restraints were applied to the terminal bases. The production molecular dynamics (MD) simulation of each solvated system were conducted for simulated time period of 1 μs. Each simulation was conducted in triplicate with different random numbers used for seeding the thermostating approach to allow for comparison of structural and temporal features amongst the rep-
licated simulations. Additionally, the replicated MD simulations (3 UB and 3 UB-control simulations) were each conducted using both standard MD and accelerated MD. (50) The accelerated MD simulations were conducted to provide an extended look at the structural features of the 2 helical systems where the temporal limitations of standard MD present constraints. The accelerated MD simulations utilized a potential energy bias applied to dihedral torsions in an effort to enhance conformational sampling. The accelerated MD parameters used included a threshold boost energy (ED) of 625 kcal/mol and a modified potential basin depth tuning parameter (α) of 80 kcal/mol. Systems were analyzed using the AMBER PTRAJ module.

2.5 Thermodynamic Results

Thermodynamic data was collected for oligonucleotides containing the UB opposite each of the four canonical base pairs as well as oligonucleotides containing an A:C mismatch and a canonical A:T base pair, each with the same surrounding oligonucleotide sequence. Thermal stability is similar; $T_m$ and ΔG values for all four UB base pairs are within 1 °C and 0.5 kcal/mol of one another. For all UB containing systems, thermal stability was greater than that of a C:A mismatch. (Table 2-1)
2.6 Results – NMR Structural Data

The DNA duplex chosen for the NMR structural studies is:

\[
\begin{align*}
5' & : \text{ATGGXGCTC} \\
\text{TACCTCGAG} & : 5' \\
18 & : 10
\end{align*}
\]

where X is the UB; for a control sequence, X is dA. NMR assignments for \(^1\)H, \(^{13}\)C, and \(^{31}\)P were completed using standard 1D and 2D NMR method. (21, 22)

2.6.1 Glycosidic Bond Orientation

Because of the unique attachment of the UB, the glycosidic bond orientation must be defined. Adenosine anti and syn conformations place either the H8 or N3 atoms on the endo face of the sugar ring. For the UB, a similar convention is used where the “anti” and “syn” orientations place the H7 or N9 atoms on the endo side of the sugar respectively, which yield quite different H7-H1’ distances of 3.8 Å for “anti” and 2.5 Å for “syn”. This distance was used to characterize the glycosidic bond torsion angle. (Figure 2-1 B)

Energy minimization calculations at the HF 6-31G level (Spartan’04) on a single UB nucleoside indicate two stable conformations representing the “syn” and “anti” glycosidic bond torsion angles with the “syn” conformation most stable. Previous studies of single UB nucleotides reveal that the base is perpendicular to the sugar and in the “syn” type conformation. (12)
2.6.2 **Glycosidic Bond Orientation in the Duplex**

A low mixing time NOESY experiment (75 ms) was used to estimate distance between the UB H7 – H1’ protons. Under low mixing times, spin diffusion is minimized thus allowing for approximate distance calculations. Using the cytosine H5-H6 crosspeak intensity with a distance of 2.46 Å as a reference, a distance of 2.3 Å was calculated for UB5 H7 – H1’ protons. This establishes the base orientation as “syn”. This result was verified with UB5 H7 to H2’, H2”, and H3’ distances. Higher mixing time data in which the MARDIGRAS program was used to resolve spin diffusion effects also supports this conformation. All residues surrounding the UB are in anti conformations.

2.6.3 **Overall Helical Geometry**

NOESY base to base and base to sugar contacts determine the overall helical structure of a DNA duplex. All anticipated NOE base to sugar H1’ contacts were identified, the only exception being a reduction in intensity between the G4 base H8 to UB5 sugar H1’ protons. All base to base, H2’, H2”, and H3’ pathways are all accounted for with the exception of the base to base and base to H3’ pathways for G4 to UB5. (Figure 2-2 A) The NOE intensities are consistent with an overall B type helix. In addition, several unique crosspeaks were observed: G6 sugar H1’ and H3’ to the UB5 base H7 proton as well as UB5 sugar H1’ to G6 sugar H3’ and H4’. (Figures 2-2 A, B) These contacts, with the G4 H8 to UB5 H1’ NOE, confirm an intrahelical orientation of the UB base.
2.6.4 **Base Pairing**

Imino proton NMR was used to detect and assign the base pairs. All imino proton resonances were in anticipated chemical shifts indicative of A:T and G:C Watson-Crick base pairing. The T14 imino proton could not be observed. Of note, the bases flanking the UB5 (G4 and G6) exhibit sharp line shapes. (Figure 2-2 C)

In contrast with adenosines, the UB5 amino group could not be identified in H$_2$O NOESY spectra. All other anticipated imino / amino pathways were detected. A weak NOESY cross-peak between G4’s imino proton and the UB5 base H7 was identified which confirms the UB5 intrahelical orientation.

2.6.5 **Chemical Shift Perturbations**

Absolute values of differences in chemical shifts ($|UB - control| = \Delta$ ppm) for various sugar and base protons as well as phosphorus resonances were calculated. As shown in Figure 2-3 A and 2-3 B, the largest chemical shift perturbations for both the sugar and base protons were localized to areas around the UB5:T14 base pair. Sugar protons in T14 contained the largest overall chemical shifts ($\Sigma \Delta$ ppm of 0.675) and largest changes in individual proton resonances (H1’ and H2’2 of 0.319 and 0.241 $\Delta$ ppm respectively). (Figure 2-3 A) Base proton chemical shift changes were mainly localized to C15. (Figure 2-3 B)

As compared to the control sequence, $^{31}$P chemical shifts for UB5-P-G6 and T14-P-C15 are shifted 0.52 and -0.56 ppm from their control values. Other perturbations are localized to the sequence around the UB5:T14 base pair. (Figure 2-3 C)
2.6.6 Deoxyribose Conformation

Using a graphical method, fraction south ($f_s$) sugar puckering values were calculated from the individual coupling constants $J_{H1'-H2'}$ and $J_{H1'-H2''}$ as well as sum of couplings $\sum J_{H1'}$, $\sum J_{H3'}$, $\sum J_{H2'}$, and $\sum J_{H2''}$. (35) The results from this analysis place the UB5 sugar in a strong S conformation. All other residues are predominately in an S conformation with the exception of C7 which was ~50% $f_s$. Due to overlap, $f_s$ for G3 was estimated from the $J_{H2'-H3'}$, $J_{H2''-H3'}$, and $J_{H1'-H2'}$ coupling patterns, which estimates it to be predominately S. Due to severe overlap, the conformations of the 3' terminal C9 and T18 sugars could not be determined. For the control sequence, all residues measured were found to be in S conformation.

2.6.7 Backbone Conformation

Experimentally determined epsilon torsion angles from ctNOESY experiments did not reveal a unique perturbation for any of the residues. All residues, including UB5-P-G6 and T14-P-C15, fall within canonical B-type ranges.

$^1H - ^{31}P$ HETCOR experiments interestingly yield a weak intensity for the H3' to $^{31}P$ peak for the UB5-p-G6 sequence. Because of the non-perturbed epsilon values, this is suggestive of dynamics. (Supplemental Information Figure 1 and 2)
2.7 Results – Structure Calculations and Analysis

2.7.1 A1 Glycosidic Bond Angle

The intense crosspeak of A1 H1’ to H8 in both UB and control oligonucleotides indicates the presence of a syn population. An approximate syn population of <25% was calculated from the observed NOE together with theoretical syn vs. anti H1’ to H8 distances. (Supplemental Information Figure 3 and Calculation 1) Because this anomaly was found in both control and UB sequences, restraints were added to the calculations to lock the A1 base into an anti conformation to reflect the major population.

2.7.2 Structure Calculation

The UB containing oligonucleotide is highly restrained containing a total of 440 non-RDC restraints and 47 RDC giving a total of ~27 and ~24 restraints per residue for the NOESY + RDC and non-RDC structures respectively. (Table 2-2) Standard deviations of mass averaged R.M.S.D. values over the entire 10 ns restrained MD trajectories are 0.25 and 0.12 Å for the non-RDC and NOESY + RDC simulations respectively. The average heavy atom R.M.S.D. for the final bundle of structures is 0.37 Å for both the NOESY + RDC and non-RDC structures. (Figure 2-4) The structure with the lowest AMBER restraints violations was chosen as the final structure. Interestingly, CORMA $R^X$ values slightly increased upon the implementation of RDC values, yet both UB structures are still in excellent agreement with NOESY data as exhibited with total CORMA $R^X$ values < 0.06. Total AMBER RDC violations were 5.63 kcal / mol for all restraints. Because of the unanticipated small increase in $R^X$ values upon RDC implementation, a third
structure containing only RDC restraints was calculated. This structure calculation did not use NOESY distance restraints, base pair angle restraint, sugar puckering restraints, or backbone torsion angle restraints. Resulting RDC penalties were minimal with a total AMBER violation of < 2 kcal / mol.

The control structure has slightly fewer restraints than the UB structure; it shows excellent agreement with NOESY data with total CORMA R^x values of 0.051 and 0.050 for the NOESY and NOESY + RDC structures respectively. AMBER RDC violations totaled 3.83 kcal / mol for all restraint. Unlike the UB structure, RDC implementation did not drive up R^x values establishing that the observed increase for the UB structure is a result of the UB modification. An RDC only structure was not calculated for the control. (Table 2-2)

2.7.3 Structure Analysis

For both the non-RDC and NOESY + RDC UB structures, all bases including both the UB5 and T14 bases are intrahelical and the overall global geometry is B-type. (Figure 2-5 A) As expected from the imino proton data, the UB5:T14 base pair does not form a hydrogen bond involving the T14 imino proton. Instead, the amino group of UB5 is hydrogen bonded to T14 O4 carbonyl. (Figure 2-12) This orientation of the UB5:T14 base pair is accommodated in the duplex via backbone perturbations, primarily the torsion angles around the UB (G4 ϖ, UB5 ϖ, G6 ω, and C16 ω). This unusual base pairing results in a heightened stretch and opening helical parameters (1 Å and 27.8 ° respectively) in UB5:T14 base pair as well as an increase of 1.8 Å in the UB5:T14 anomeric carbon distance as compared to the control structure. This results in a displaced backbone, distortions in the helical twist (G4-UB5 30.4°, UB5-G6 29.7°, G6-C7 41.2°),
and a slight bend in the global helical axis (total axis bend of 19 vs. 4° for the UB and control respectively) as compared to the control structure. (Figure 2-5 B and C) As a consequence of the under-twisting and the unique N8 UB base connectivity, the local stacking of G4-UB5-G6 is disrupted with G4 poorly stacked on top of UB5. (Figure 2-6 A)

For completeness, it is noted that the T2-G3 step has an apparent bend in both the UB and control structures. (Figure 2-7A, Supplemental Information Figure 4) This bend is primarily generated by a heightened positive roll of 12.9° in the T2-G3 step. The James group found similar results for T-G / A-C steps with similarly large positive rolls. (59) Intriguingly, when RDC restraints were incorporated, the T2-G3 bend was removed from both the UB and control structures.

2.7.4 RDC Implementation

Although the inclusion of RDC restraints with NOESY data did not result in a major conformational change for the core sequence of the UB structure (G4, UB5, G6, C13, T14, C15 heavy atom R.M.S.D. of 0.63 Å) (Figure 2-7B), the slight increase in total CORMA R^X values for the UB structure and not the control prompted the calculation of an RDC only structure. The resulting structure is similar to the NOESY + RDC results except for the UB5:T14 base pair. Here, the predicted base pair includes a hydrogen bond between the T14 imino proton and UB5 N1. (Figure 2-8 A, B, C)
2.7.5 *Supercooled Aqueous NMR*

Due to the ambiguous orientation of the UB5-T14 base pair, further probing of the imino proton spectra was conducted via supercooled aqueous NMR utilizing 1 mm capillaries. When temperatures were dropped to -10 °C, an additional imino proton peak begins to emerge at 14 ppm, suggestive of the formation of an additional A:T base pair. A NOESY spectrum at -12 °C was used to confirm the identity as T14 through the appearance of a cross peak between the new imino proton and an additional methyl resonance. (Figure 2-9) All observed imino proton resonances shifted as anticipated with decreasing temperatures indicating that an overall global conformation change in the oligonucleotide did not occur.

2.8 Results – Conformational Exchange, Dynamics, and *in-silico* Analysis

2.8.1 “The Case of the Disappearing H2 Proton”

The UB5-H2 proton has a similar chemical environment as compared to an adenosine H2 and is anticipated to have similar NMR and chemical traits. Under denaturing conditions, all 4 H2 protons in the UB sequence are readily identified by both 1D $^1$H NMR and HSQC spectra; comparison to single stranded spectra reveals that the UB5 C2 is slightly shifted from the other A C2 resonances (156 ppm compared to A C2’s of 152 ppm). When annealed, the UB5-H2 proton vanishes; only the 3 A H2 protons can be identified in 1D $^1$H, HSQC, and NOESY spectra. (Figure 2-10 A, B, C) The UB H2 proton reappears upon subsequent denaturing of the duplex, verifying the non-labile nature of the proton, and again vanishes under native conditions. Integration of the $^1$H resonances in the vicinity of the other adenosine H2 peaks does not readily
yield evidence of overlapped H2 peaks, however visual inspection of the spectrum yields an extremely small broad peak in this range suggestive of an additional resonance.

2.8.2 Broadened $^1$H Resonances and ROESY Spectra

NOESY and 1D spectra reveal that the T14 sugar H1’ resonance is broad. Aside from the potential broad peak found in the H2 region as mentioned above, no other protons exhibit broadened peaks.

A ROESY experiment with a 200 ms mixing time was used to detect chemical exchange peaks. A large number of exchange peaks were detected in the base-base and H1’-H1’ / H5-H5 regions. (Supplemental Information Figure 5) Nearly all exchange peaks are localized to residues C13 through T18. No exchange peaks were found in the H2'/H2'', H3'-H3’, and H4'-H4’ regions. Equimolar strand stoichiometry was verified via integration of the base protons at high temperatures (i.e. adenosine H8). No exchange peaks were observed in the control indicating that the anomalous exchange peaks are unique to the UB containing oligonucleotide.

2.8.3 $^{13}$C T$_1$ Relaxation as a “Probe” for Heightened Dynamics

In order to assess the relative mobility of the residues within the oligonucleotide, the $^{13}$C T$_1$ relaxation of the anomic (C1’) and C5, C6, and C8 (pyrimidine and purine respectively) carbons was evaluated at natural abundance. Signal to noise, although somewhat low, allowed for a qualitative assessment of the T$_1$ values with all R values > 0.95. Lower relaxations for residues on the 5’ side of the UB5 (G4 C1’, C15 C5 & C6, C16 C1’ and A17 C1’) were observed. Spectral
overlap prohibited examination of several relaxation rates including G3, G4, and G6 base carbons.

2.8.4 Reduced HSQC Peak Intensities

A number of crosspeaks in the HSQC spectra have lower intensities compared to similar peaks. These include the T14 sugar H1′/C1′, G6 sugar H1′/C1′, and UB5 base H7/C7 crosspeaks. The reduced intensity is not due to unusual $^{1}J_{H-C}$ coupling; all measured couplings were within anticipated ranges with the UB5 $^{1}J_{C7-H7}$ equal to 197 Hz.

Similarly, incomplete relaxation of the resonances cannot account for this observation because $^{13}C$ T$_{1}$ values for all resonances were less than 0.6 s and the HSQC interscan delay was set to 5 s. The measured $^{1}H$ T$_{1}$ value for the UB5 base H7 is 3.2 s, which is lower than the measured adenosine H2 T$_{1}$ values of 5.5 – 6.6 s yet they still exhibit higher peak intensities.

2.8.5 in-silico

A variety of molecular dynamics (MD) simulations were conducted on the UB containing oligonucleotide as well as the control sequence. In order to probe for conformation sampling with the existing empirical data, MDtar was utilized in which restraints are enforced over a period of time as opposed to instantaneous, constant enforcement. Time-averaged restraints were imposed on all distance restraints over a 1 ns trajectory with a 20 ps $\tau$ value; RDC and sugar puckering restraints were not implemented. Using the Curves+ program, helical parame-
ters for the trajectory reveals heightened conformational sampling in the UB5:T14 base pair and on the 5’ side of the UB residue. (Figure 2-11 A, B, C)

Extended MD and accelerated MD (simulation times of 1 μs) were completed on non-restrained UB containing and control sequences. Analysis of the UB5:T14 base pair reveals a reduced involvement of the T14 imino proton in hydrogen bonding as compared to the G4 and G6 imino protons as well as the control structure (UB5:T14 65.7%, A5:T14 71.9%).

2.9 Discussion

2.9.1 Structure

Incorporation of the UB base maintains an overall B-type helical geometry with the UB base intrahelical and involved in base stacking as verified by continuous NOESY base to H1’ and H2’ pathways and the detection of the G4 imino to UB5 base H7 NOESY crosspeak. The unique UB N8 glycosidic connectivity (C1’ to N8) and “syn” orientation of the UB base increases the distance between the UB base H7 proton from the G4 residue yet places the H7 proton in close proximity to the G6 sugar thus explaining the missing pathways between G4 and UB5 and the unanticipated NOE contacts between UB5 base H7 to G6 sugar protons. (Figure 2-2 A, B)

The overall UB oligonucleotide structure is remarkably unaltered compared to the control with all structural perturbations occurring within the core sequence of the duplex (UB5:T14). This is reflected in the observed proton chemical shift differences. Sugar proton chemical shift differences are localized to the nucleotides flanking UB5. In addition, the base proton chemical shift perturbations of C15’s H5 and H6 are the result of aromatic ring field ef-
ffects of the T14 base. When comparing the UB and control structures, the orientation of T14 as compared to C15 is different which exposes the H5 and H6 protons of C15 to different chemical shielding environments. (Figure 2-6 B)

The UB5:T14 base pair for the NOESY + RDC structure is different with respects to regular Watson-Crick base pairing. The unusual UB5 amino - T14 O4 carbonyl hydrogen bond increases base opening and stretch; this combined with the unique UB base N8 connectivity causes an increased anomic carbon distance between UB5 and T14. (Figure 2-12) This enlarged interstrand distance is accommodated within the duplex environment through perturbations in the backbone torsion angles, which for the angles around the UB5-P-G6, is reflected in the 31P data. Distortions in α and ζ torsion angles can perturb 31P chemical shifts from their regular values. As compared to the control structure, the G6 α as well as G4 and UB5 ζ are higher which can be associated with the 31P chemical shift perturbations for UB5-P-G6. The effect of this accommodation is a “shift” in the backbone around UB5 and an under-twisting of the helix (30° as compared to B-type 36°). (Figure 2-5 B)

Intriguingly, RDC implementation with NOESY restraints slightly elevates the total COR-MA R* values; this trend was not observed for the control sequence indicating that the UB perturbation is the cause of this anomaly. Structure calculations with only RDC restraints results in the prediction of a different base pair conformation in which the T14 imino proton is involved in a hydrogen bond. This “reverse Watson-Crick”-like base pair is experimentally supported with supercooled aqueous imino proton spectra where the T14 imino proton is observed in the characteristic Watson-Crick A:T base pairing region; a “trapped” imino proton would appear at further upfield at around 10 – 10.5 ppm. (Figure 2-12)
2.9.2 **Dynamics**

For the UB structure, evidence of dynamic hotspots and multiple conformations were encountered. This was manifested in a missing resonance, broadened and low intensity peaks, and exchange peaks. Supercooled aqueous NMR and $^{13}$C $T_1$ relaxation studies were utilized to probe for dynamics on different timescales. This was complemented by MDtar and extended free MD and accelerated MD to identify areas with increased conformational sampling.

RDC data can be impacted by a large variety of motions and chemical exchange processes ranging from the ps to ms timescales. (51) Structure calculations with only RDC data predicts the formation of a different UB5:T14 base pair conformation as compared to the NOESY based structure. Supercooled conditions will slow base opening rates and dynamics allowing for the potential detection of shorter-lived base pairs involving an imino proton hydrogen bond. Use of this technique resulted in the appearance of an additional broad imino proton resonance at 14 ppm at -10 °C. The resonance has a characteristic chemical shift of an A:T Watson-Crick base pair and was verified as T14’s imino proton from the appearance of an additional methyl crosspeak in NOESY spectra collected at -12 °C. Based on the lifetimes of terminal bases which exhibit shortened base life times and broadened imino proton peaks, a lifetime of < 1ms is estimated for this base pair. (52) (Figure 2-9) This data supports the RDC only structure.

To probe for motions on the ps – ns time frame, which can also impact RDC data, $^{13}$C $T_1$ relaxation rates were measured. The principal relaxation method for $^{13}$C is through dipolar coupling with bound protons. In theory, the $T_1$ relaxation of the similar nuclei within a bio-polymer should be comparable (i.e. C1’ to C1’, …) However, $T_1$ relaxation values are also dependent on correlation time; changes in localized mobility will therefore alter $T_1$ relaxation
pendent on correlation time; changes in localized mobility will therefore alter $T_1$ relaxation times. Another relaxation pathway that may impact T1 relaxation of natural abundance samples is chemical shift anisotropy (CSA). Previous studies of nucleic acids at 11.7 T (500 MHz) have shown that stacking and hydrogen bonding only minimally impact the CSA tensor allowing for direct comparison of the individual bases. (53, 54)

For this size molecule, decreased $T_1$ values equate to increased mobility. Reduced $T_1$ values were encountered on the 5' side of the UB5:T14 base pair, specifically in the G4 sugar and C15 and C16 base carbons, which is indicative of higher mobility in this region. The $^{13}$C $T_1$ relaxation rates did not show a substantial change for UB5 or T14 sugar or base carbons as compared to other residues. This data indicates that the UB5:T14 base pair do not exhibit heightened motions in the ps to ns timescale. This increased motion on the 5' side of the UB is supported with the MDtar data. The sampling of different base pairing by UB5 and T14 is also suggested from the extended MD and accelerated MD data. In these trajectories, the UB5:T14 base pair exhibits a lower population of imino-based base pairing as compared to the control A5:T14 suggesting a lower stability in a base pair involving the T14 imino proton.

### 2.9.3 Dynamics and the Disappearing H2 Proton

A perplexing issue is the disappearance of the UB base H2 proton. The H7 proton of the same UB, surrounding protons, and the nearby T14 methyl protons all remain quite discernable. Under native conditions, the proton cannot be found in any 1D or 2D experiments. However, the proton can be easily identified at high temperatures and in ambient temperatures for
single stranded samples. Whatever causes the UB H2 proton to vanish does not affect any other protons in a similar manner.

The missing UB H2 proton can be explained in context of the surrounding structure and chemical exchange. The “syn” conformation of the UB and the sequence context of G4-UB5-G6 places the H2 proton in the major groove flanked by two carbonyl groups. Carbonyl shielding effects cause quite dramatic changes in the chemical shifts of protons positioned near the plane of the carbon=oxygen bond. (55) For the two conformations of the UB5:T14 base pair, the UB5 H2 proton would experience different environments from the adjacent carbonyls’ shielding. (Figure 2-13) The difference in chemical shifts for the UB H2 resulting from the different carbonyl shielding environments must be relatively large; smaller differences would have just resulted in an observably broad peak.

Broadening of peaks from chemical exchange is dependent on the difference in the chemical shifts and the timescale of the motion. A convenient measure of exchange timescales is the exchange rate at the coalescence point.

\[ k_c = \frac{\pi \Delta \nu}{2^{1/2}} \]  
Equation 2

where \( \Delta \nu \) is the difference in chemical shift (Hz) of the two conformation resonances and \( k_c \) is the exchange rate (s\(^{-1}\)) at the coalescence point. The coalescence point is the exchange rate at which two peaks broaden and coalesce into a single broad peak. Because there are no observable H2 proton resonances, the difference in shielding had to be estimated from model calculations. Due to the complex dependence of not only the impacted proton’s orientation about the carbonyl bond but also the carbonyl’s attached functional groups, chemical shift calculations were completed on the entire G4-UB5-G6 base ring systems. (55, 56) The two conformations of
the UB5:T14 base pair as well as a conformation in which the UB5 was flipped out from the flanking G4 and G6 bases were considered.

The difference in calculated UB H2 chemical shifts from the two base pair conformations was ~450 Hz (600 MHz field strength) giving a exchange lifetime (1/\(k_e\)) of approximately 1 ms. Comparing the completely flipped out model to the base pair increases the difference in chemical shifts to nearly 1.2 ppm (~700 Hz) which reduces the lifetime to approximately 0.6 ms. These theoretical calculations are in agreement with observed spectra. Exchange lifetimes of the base pair conformations on this timescale would result in a broadening of the UB5 base H2 proton under a large portion of the base resonances and would make it difficult to detect.

Further support for this timescale comes from the weakened HSQC intensities of UB5, G6, and T14. Comparison of HSQC peak intensities have been found to give qualitative information of local dynamics on the \(\mu s\) to ms time frame and has been used to characterize broad protein motions and protein-protein interactions. (57, 58)

### 2.10 Conclusion

The solved solution structure of the UB containing duplex gives key insight into its potential use as a molecular probe. The UB is intrahelical, involved in base stacking and base pairing, and does not disrupt the duplex environment. With its unique orientation and accommodation within the duplex, the oligonucleotides groove binding determinants are quite altered as compared to canonical G:C and A:T base pairs. Although it is unknown how major groove binding proteins will interact with this altered landscape, the unusual landscape for groove binding ligands may potentially be exploited with unique ligand designs. For use as an NMR probe, care
needs to be taken with the sequence context; as was encountered with the chosen sequence, the flanking carbonyls caused the H2 proton to vanish. Choosing a sequence without nearby interfering carbonyls (for example A’s or C’s) would avoid this effect and allow the H2 to be used as an NMR probe for groove binding ligands.

The encountered dynamics posed a unique challenge to the NMR analysis. It is highly unusual for a single proton of a DNA base to be overtly impacted by dynamics with little change to any of the surrounding protons or protons of the same base without a groove binding ligand. As an example, the universal bases 5-nitroindole and 5-nitroindol-3-carboxamide were found to be intrahelical, yet similar timescale dynamics resulted in the broadening of several resonances as well as the appearance of NOE exchange peaks. (16) For the UB, a unique combination of flanking bases and motions on the μs to ms timescales generated this vanishing H2 resonance yet did not impact any other resonances.

As observed in the thermal stability, the UB base pairs with each of the canonical bases in an indiscriminate manner. The exhibited dynamics and hydrogen bonding sampling of the UB may help explain this phenomenon. With its unique orientation, UB base stacking will be highly dependent on the identity of the flanking bases and the base-pairing partner. With the observed G4-UBS sequence context, the resulting stacking is stabilizing yet reduced compared to the control G4-A5 step. The UB is involved in hydrogen bonding base pairing schemes and these interactions help further stabilize the orientation of the UB in the duplex. As a consequence however, the interstrand distance is increased and was compensated by backbone dis-
tortions. This strain would only be increased when the UB is opposite a purine and prove to be potentially disruptive to base stacking. The encountered dynamics may result in a stabilizing force by allowing the base to sample multiple hydrogen bonding conformations thus potentially reducing the torsion angle disruption and maintain its stabilizing base stacking. This offers an explanation for the observed enthalpy entropy compensation exhibited in the thermodynamics data.
Table 2-1 Thermodynamic Properties of UB Containing Oligonucleotides

Thermodynamic properties of UB containing oligonucleotides in the same sequence context but with varying base pair partners for the UB. $\Delta G$ was calculated at 294 K using $\Delta G = \Delta H - T\Delta S$. Also included for comparison are the same sequences with an A:T base pair and a A:C mismatch. The X represents the UB. (10 $\mu$M total strand concentration, 100 mM NaCl, 10 mM MgCl$_2$, and 10 mM sodium cacodylate with a pH of 7.0) *Melting curve is not fitable.

<table>
<thead>
<tr>
<th>Duplex</th>
<th>$T_m$ (°C)</th>
<th>$\Delta H$ (kcal mol$^{-1}$)</th>
<th>$\Delta S$ (cal mol$^{-1}$ K$^{-1}$)</th>
<th>$\Delta G_{293}$ (kcal mol$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5’ A-T-G-G-A-G-C-T-C</td>
<td>42.8</td>
<td>-77.0</td>
<td>-218.1</td>
<td>-12.9</td>
</tr>
<tr>
<td>3’ T-A-C-C-T-C-G-A-G</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5’ A-T-G-G-A-G-C-T-C</td>
<td>&lt;25$^a$</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3’ T-A-C-C-T-C-G-A-G</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5’ A-T-G-G-X-G-C-T-C</td>
<td>34.0</td>
<td>-67.1</td>
<td>-192.8</td>
<td>-10.4</td>
</tr>
<tr>
<td>3’ T-A-C-C-T-C-G-A-G</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5’ A-T-G-G-X-G-C-T-C</td>
<td>33.2</td>
<td>-79.1</td>
<td>-232.7</td>
<td>-10.7</td>
</tr>
<tr>
<td>3’ T-A-C-C-A-C-G-A-G</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5’ A-T-G-G-X-G-C-T-C</td>
<td>33.5</td>
<td>-64.4</td>
<td>-184.3</td>
<td>-10.2</td>
</tr>
<tr>
<td>3’ T-A-C-C-G-C-G-A-G</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5’ A-T-G-G-X-G-C-T-C</td>
<td>33.9</td>
<td>-62.0</td>
<td>-176.3</td>
<td>-10.2</td>
</tr>
<tr>
<td>3’ T-A-C-C-G-C-G-A-G</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 2-2 Summary of NMR Structure Restraints

Summary of NMR restraints used for the structure development of the UB and control structures. Qualitative distance restraints involving exchangeable protons were obtained from a H₂O NOESY spectrum with a mixing time of 150ms. The final structure rMD calculations were run for 10 ns; charges were neutralized with 16 Na⁺ ions and simulations were conducted in explicit TIP3P water. Final structures underwent fully restrained energy minimization (2000 steps steepest decent). R.M.S.D. values were calculated in VMD version 1.9.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>UB structure</th>
<th>Control structure</th>
<th>Force Constant(k) a</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance Restraints</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-exchangeable (RANDMARDI)</td>
<td>210</td>
<td>146</td>
<td></td>
</tr>
<tr>
<td>intra residue</td>
<td>139</td>
<td>78</td>
<td>20 or 30 b</td>
</tr>
<tr>
<td>inter residue (sequential)</td>
<td>71</td>
<td>68</td>
<td>30</td>
</tr>
<tr>
<td>Average well width (Å)</td>
<td>0.69</td>
<td>0.79</td>
<td></td>
</tr>
<tr>
<td>Qualitative exchangeable distance restraints</td>
<td>34</td>
<td>0</td>
<td>30</td>
</tr>
<tr>
<td>Qualitative A1 anti-glycosidic angle restraints</td>
<td>2</td>
<td>2</td>
<td>30</td>
</tr>
<tr>
<td>Watson-Crick Base Pair Restraints</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Watson-Crick distance restraints</td>
<td>21</td>
<td>23</td>
<td>25</td>
</tr>
<tr>
<td>Watson-Crick flat angle restraints</td>
<td>21</td>
<td>23</td>
<td>10</td>
</tr>
<tr>
<td>Backbone Torsion Angle Restraints</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Broad angle DNA restraints (α, β, γ)</td>
<td>61</td>
<td>66</td>
<td>50</td>
</tr>
<tr>
<td>y (CTNDOESY) torsion angle restraints</td>
<td>16</td>
<td>16</td>
<td>50</td>
</tr>
<tr>
<td>Endocyclic torsion angle restraints</td>
<td>75</td>
<td>90</td>
<td>50</td>
</tr>
<tr>
<td>Total Restraints</td>
<td>440</td>
<td>366</td>
<td></td>
</tr>
<tr>
<td>Average Restraints / Residue</td>
<td>24.4</td>
<td>20.3</td>
<td></td>
</tr>
<tr>
<td>RDC Restraints</td>
<td>47</td>
<td>42</td>
<td>1.0 (dwt)</td>
</tr>
<tr>
<td>Total Restraints with RDC</td>
<td>487</td>
<td>408</td>
<td></td>
</tr>
<tr>
<td>Average Restraints / Residue</td>
<td>27.1</td>
<td>22.7</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>UB structure</th>
<th>Control structure</th>
<th>Force Constant(k) a</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOESY only structure CORMA R⁺</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tₘ (ms)</td>
<td>intra</td>
<td>inter</td>
<td>total</td>
</tr>
<tr>
<td>75 ms</td>
<td>5.14</td>
<td>5.75</td>
<td>11.34</td>
</tr>
<tr>
<td>150 ms</td>
<td>4.88</td>
<td>5.58</td>
<td>10.46</td>
</tr>
<tr>
<td>250 ms</td>
<td>4.14</td>
<td>5.46</td>
<td>9.60</td>
</tr>
<tr>
<td>NOESY + RDC structure CORMA R⁺</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tₘ (ms)</td>
<td>intra</td>
<td>inter</td>
<td>total</td>
</tr>
<tr>
<td>75 ms</td>
<td>5.57</td>
<td>6.64</td>
<td>12.21</td>
</tr>
<tr>
<td>150 ms</td>
<td>5.19</td>
<td>5.82</td>
<td>11.01</td>
</tr>
<tr>
<td>250 ms</td>
<td>4.49</td>
<td>5.18</td>
<td>9.67</td>
</tr>
<tr>
<td>Average Heavy Atom R.M.S.D. (Å)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final Bundle of Structures (NOESY only)</td>
<td>0.37</td>
<td>0.40</td>
<td></td>
</tr>
<tr>
<td>Final Bundle of Structures (NOESY + RDC)</td>
<td>0.37</td>
<td>0.40</td>
<td></td>
</tr>
<tr>
<td>Final AMBER Penalties(NOESY only)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMBER distance penalty (kcal/mol)</td>
<td>37.5</td>
<td>15.5</td>
<td></td>
</tr>
<tr>
<td>AMBER angle penalty (kcal/mol)</td>
<td>0.1</td>
<td>0.3</td>
<td></td>
</tr>
<tr>
<td>AMBER Torsion angle penalty (kcal/mol)</td>
<td>4.0</td>
<td>11.2</td>
<td></td>
</tr>
<tr>
<td>Final AMBER Penalties(NOESY + RDC)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMBER distance penalty (kcal/mol)</td>
<td>45.9</td>
<td>17.0</td>
<td></td>
</tr>
<tr>
<td>AMBER angle penalty (kcal/mol)</td>
<td>0.2</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>AMBER Torsion angle penalty (kcal/mol)</td>
<td>4.8</td>
<td>12.4</td>
<td></td>
</tr>
<tr>
<td>Total RDC alignment constraint (kcal/mol)</td>
<td>5.6</td>
<td>3.8</td>
<td></td>
</tr>
</tbody>
</table>

* (kcal/mol * unit of violation)
* 20 kcal/mol for intraresidue sugar restraints, 30 kcal / mol for all others
Figure 2-1  A) Stick representations of deoxyadenosine (left) and the UB nucleoside (right). The red numbering on UB nucleoside represents the atomic purine numbering system used in the manuscript. B) “Ball-and-stick” representations of the “anti” and “syn” glycosidic bond orientations of UB nucleoside. Distances between the H7 and H1’ protons are noted. Red = oxygen, blue = nitrogen, cyan = carbon, and white = hydrogen.
Figure 2-2  A) The observed NOESY pathways for the UB sequence.  B) NOESY spectrum (150 ms mixing time) of the base to H1' region for the UB structure. Black and red lines represent connectivity for the different strands. Inset shows the unique UB5 H7 to G4 H1' and G6 H1' NOE contacts.  C) Imino proton spectrum of the UB structure at 280 K. In blue are the resonances for G4 and G6. Note the lack of a T14 imino proton resonance.
Figure 2-3  A) The summation of the absolute difference in chemical shifts for the sugar region (|UB – control| = Δ ppm). B) Summation of the absolute difference in chemical shifts for the base region. C) (Top) $^{31}$P spectrum of the UB and control structures at 298 K. Arrows point out the changes in the chemical shifts for G6 and C15. (Bottom) A mapping of the differences in $^{31}$P chemical shifts between the control and UB structures (control – UB).
Figure 2-4  A) The final bundle of structures for the UB structure. The UB 5 residue is in red. The left is a view of the major groove; the right is a view of the minor groove. B) Final bundle of structures for the UB5:T14 base pair. Red = oxygen, blue = nitrogen, cyan = carbon, and white = hydrogen.
Figure 2-5  A) Stick model of the UB structure calculated using both NOESY and RDC restraints. B) Ribbon representation of UB structure (NOESY + RDC restraints). The observed displacement in the backbone is in red and the UB is in red. C) Ribbon representation of the control structure. The same region of the backbone as B) is highlighted in red for comparison. A5 is in blue.
Figure 2-6 A) Base stacking for the UB (left) and control (right) structures. (Left) Blue = G6, red = UB5, green = G6. (Right) Blue = G6, red = A5, green = G4. In gray is the glycosidic sugar connectivity of each base. The bases are viewed down the helix. B) Residues T14 and C15 base stacking for the UB (NOESY + RDC restraints) and control structures. Each of the structures were aligned to C15. The multicolored base is C15 (red = oxygen, blue = nitrogen, cyan = carbon, white = hydrogen); the red base is UB structure T14 base, the blue is the control structure T14 base. The white spheres on C15 represent the H5 and H6 protons.
Figure 2-7  A) The A1, T2, G3, and G4 residues of the UB structure with a ribbon cartoon representing the backbone. Left is with NOESY restraints only; on the right is NOESY + RDC restraints. Red = oxygen, blue = nitrogen, cyan = carbon, and white = hydrogen. G3 and G4 of each structure have been aligned. B) The core G4, UB5, G6, C13, T14, and C15 residues of the UB structure before and after implementation of RDC restraints. Blue = NOESY only, red = NOESY + RDC restraints.
Figure 2-8  The UB structure with only RDC restraints.  A) The helix of the RDC only restraints structure viewed from the minor groove. The UB5 residue is in red.  B) The predicted base pairing using only RDC restraints. Red = oxygen, blue = nitrogen, cyan = carbon, and white = hydrogen.  C) The core G4, UB5, G6, C13, T14, and C15 residues of the UB structure with only RDC restraints viewed from the minor groove.
Figure 2-9  The supercooled aqueous imino proton spectra of the UB structure at -12 °C. The inset is a NOESY spectrum (150 ms mixing time) of the imino – methyl region collected at the same temperature. The red arrow points out the appearance of a third methyl NOE cross-peak at ~14 ppm which is not observed at 280 K. Note the weak NOE crosspeak 13.35 ppm (G4’s imino proton).
Figure 2-10 HSQC spectra of the base region of A) double stranded UB oligonucleotide at 298 K, B) single stranded UB oligonucleotide at 298 K, and C) double stranded UB oligonucleotide at 330 K. The red arrows point out UB5 C2-H2 resonance. Note its disappearance in the duplex at 298 K (A).
Figure 2-11 A) Radar graphs of the $\alpha$ and $\zeta$ backbone torsion angles vs the time course of the MDTar simulation for the UB structure. In red is the measured torsion angle over the simulation and the rings represent the simulation time step. Torsion angles are around the ring counterclockwise with the top center being 0°. B) and C) Histograms for the various helical parameters for the MDTar simulation of the UB structure; the overall population size for all histograms is 1,000 structure snapshots.
Figure 2-12 The predicted base pairs for the control and UB sequences. The top base pair is the control A5:T14, the middle is the UB5:T14 base pair with NOESY + RDC restraints, and the bottom is the UB5:T14 base pair with only RDC restraints. Red = oxygen, blue = nitrogen, cyan = carbon, gold = phosphorous, and white = hydrogen.
Figure 2-13 A view of the G4, UB5, and G6 bases from the major groove. The red spheres represent the O6 carbonyl oxygen of G4 and G6. On the left is the structure with both NOESY + RDC restraints; on the right is the structure with only RDC restraints. Note the different orientations of the UB5 H2 proton with respect to the carbonyls. Red = oxygen, blue = nitrogen, gray = carbon, yellow = phosphorous, and white = hydrogen.
2.11 References


function is regulated via a flexible pocket and novel phosphorylation-dependent mechanism, *Mol Cell Biol* 29, 1306-1320.


2.12 Supplemental Information

Supplemental Information Figure 1  Epsilon Torsion Angles

Calculated epsilon torsion angles from the ctNOESY data. Red is the universal containing sequence, blue is the control sequence.
Supplemental Information Figure 2  HPCOR Spectrum.

This spectrum correlates H3’n – P – H4’n+1. Circled in red are the crosspeaks for UB5H3’ – P and P – G6H4’. Note the weakened intensity for the UB5H3’ – P crosspeak.
**Calculation of syn Population of the A1 Glycosidic Torsion Angle**

**Given:**
cytosine H5-H6 distance is 2.46 Å  
experimental H5-H6 NOE integration is 2.02 x 10⁸ 

**and:**
experimental A1 H1’ – H8 NOE integration is 1.07 x 10⁸  
adenosine syn glycosidic bond H1’ – H8 distance is 2.2 Å  
adenosine anti glycosidic bond H1’ – H8 distance is 3.7 Å 

**Formula 1:**  
\[
\text{NOE A} / \text{NOE B} = \frac{r_B^6}{r_A^6}
\]

**Using formula 1:**
Anticipated NOE anti conformation integration 3.95 x 10⁸  
Anticipated NOE syn conformation integration 1.74 x 10⁷  

**Assuming two populations of syn and anti and no spin diffusion:**
\[
X(\text{anti NOE integration}) + (1-X)(\text{syn NOE integration}) = \text{actual NOE integration}
\]

\[
X(3.95 \times 10^8) + (1-X)(1.74 \times 10^7) = 1.07 \times 10^8
\]

\[
X = 23.7\%
\]

The population of A1 is <25% syn

---

**Supplemental Information Calculation 1  Calculation of the syn glycosidic bond population for the A1 base proton.**

NOESY integrations were taken from spectra with a 75 ms mixing time to minimize spin diffusion effects. Theoretical values for adenosine syn and anti H8 to H1’ distances were taken from: Wüthrich, K., *NMR of Proteins and Nucleic Acids*, John Wiley & Sons, Inc., New York, NY, 1986, page 236.
Supplemental Information Figure 3  NOESY Integration Values of the Base to H1’ Crosspeaks.

A) UB sequence, B) control sequence. NOESY spectra were collected at 298 K with a 75 ms mixing time. Noted in red are the A1 H8-H1’ and A1 H1’-H8 crosspeak integrations.
Supplemental Information Figure 4  NOESY Integration Values of the Base to H2’2 Crosspeaks

A) UB sequence, B) control sequence. NOESY spectra were collected at 298 K with a 150 ms mixing time. Noted in red are the G3H8 – T2H2’2 and A17H8 – C16H2’2 cross peaks. The shorter distances for the H2’2 – H8 distances observed in the T2-G3 kink is reflected in smaller integrations values.
Supplemental Information Figure 5  ROESY Spectrum.

ROESY spectrum with a 250 ms mixing time at 298 K focused on the H1’ to H1’ region. Red and blue represent opposite phasing with exchange peaks occurring in blue and NOESY peaks in red. On the right is a mapping of the observed chemical exchange peaks (noted with boxes) on the DNA sequence.
3 ENDONUCLEASE IV AND α-ANOMERIC CONTAINING DNA SUBSTRATES: USING FRET TO DETERMINE THE GLOBAL CONFORMATION OF THE DNA SUBSTRATE

3.1 Introduction

As part of the repair process of the base excision repair pathway (BER), glycosylases recognize and convert non-bulky DNA base lesions (oxidative damage and deoxyuracil incorporation for instance) into a highly cytotoxic abasic site via cleavage of the glycosidic bond. In addition, abasic sites can result from spontaneous depurination. DNA damage is estimated to occur as many as 10,000 times per cell per day. (1) Abasic sites are recognized and processed by endonucleases, which nick the backbone on the 5’ side of the lesion and are a critical component of the BER. In bacteria, the enzymes exonuclease III and endonuclease IV (endoIV) are the abasic processing enzymes in the BER pathway; they also repair abasic sites formed by depurination or other sources. (2) In humans there is no endoIV homolog, however APE1 (a homolog of the bacteria exonuclease III) is the primary endonuclease in the BER pathway as well as abasic detection and processing. Although APE1 has been implicated in several other processes and the detection of multiple types of abasic sites in both ds and ss DNA as well as RNA, their enzyme activity is primarily limited to the detection of different types of abasic lesion sites. (3, 4)

Surprisingly, both endoIV and APE1 can detect and process alpha anomeric damage. (5-7) Alpha anomeric damage is the inversion of chirality at the anomeric position of the sugar,
which places the base on the opposite face of the sugar ring. (Figure 3-1) EndoIV enzyme activity of α-anomeric lesion is similar to an abasic site; the backbone is cut on the 5’ side of the lesion. (5-7) In addition, it has been found that the endoIV enzyme activity with α adenosine (αA) containing oligonucleotides is modulated based on the identity of the lesion’s flanking bases. (8, 9) To gain insight into the unbound substrates, solution NMR structures were solved for both the C-αA-C and C-αA-G sequences. They each reveal a slight kink and widening of the major groove with the C-αA-C sequence context more pronounced than the C-αA-G. These characteristics are believed to drive detection of the αA lesion site. (8, 9) Although these NMR structures provided insight into the substrate quality of the unbound oligonucleotides, the structural consequences of enzyme binding and catalysis was not addressed. In this chapter, efforts to characterize the DNA substrate (both C-αA-C and C-αA-G sequence contexts) in complex with a non-enzymatically active endoIV mutant using fluorescence energy transfer (FRET) are explored.

### 3.1.1 APE1 and EndoIV Complex Structures

The X-ray crystal structures of endoIV and APE1 in complex with abasic containing DNA substrates give valuable insight into the effects of binding on the oligonucleotide substrates. (PDB #’s 1DE8 and 1QUM for the APE1 and endoIV structures, respectively) (10, 11) (Figure 3-2) Upon binding to the DNA substrate, APE1 induces a 35° kink in the DNA while endoIV induces a more severe kink of 67°. Both enzymes interact with the DNA via the minor groove, which is significantly enlarged. Recognition occurs via loops inserted into the major and minor
grooves, which destabilize the DNA in search for flexible regions. Once localized, the abasic residue is flipped out and strand cleavage ensues. Interestingly, APE1 undergoes little structural rearrangement upon binding the substrate, which forces the DNA to conform to the protein binding site. (11, 12) In contrast, for endoIV both the substrate and enzyme undergo structural changes. (10) Bending of intact DNA would be energetically unfavorable, in addition a regular β-anomeric nucleotide would not fit into the well-defined active binding pockets of APE1 or endoIV and would not result in an enzymatically competent complex, thus giving the endonucleases specificity for abasic lesion sites. (4, 12)

3.1.2 Endonuclease Activity on α Anomeric Damage

In addition to abasic residues, it was found that endoIV can also process α anomeric nucleoside lesion site. (5, 13) The human APE1 endonuclease also exhibits similar recognition capabilities. (6, 7) In α-anomeric lesions, the stereochemistry of C1’ is inverted placing the base on the opposite side of the deoxyribose, producing a radically different structure and altered base pairing. (5) (Figure 3-1) This lesion is produced under anaerobic conditions by hydroxyl radicals and is a major lesion produced by γ-irradiation under anoxic conditions is an α-adenosine (αA) lesion. (5) The altered nucleotide is mutagenic because it directs the incorporation of C, A, and T in vitro and generates deletions in vivo. (13) Although α-anomeric lesions in mammals have yet to be identified, such damage in hypoxic tumor cells exposed to radiation cannot be ruled out. (14) The observation that α-anomeric nucleoside lesion repair capability is preserved may suggest a biological relevance.
To gain more insight, the abasic residue of the DNA in complex with either the endoIV or APE1 crystal structures was replaced with an αA moiety. (Figure 3-3) The endoIV structure reveals a cavity that readily accommodates an αA residue. (10) In the case of APE1, it is clear that small conformational changes of the αA residue and/or of the binding pockered would be required for a successful fit. In either case it is readily apparent that a regular β-anomeric residue could not fit for either endoIV or APE1.

### 3.1.3 Flanking Sequences Impact on Unbound αA Oligonucleotides

Thermodynamic profiling revealed only a slightly lower stability of αA containing oligonucleotides as compared with an unmodified control. (8, 9) The NMR solution structure of the C-αA-C containing sequence is a B type decamer duplex and reveals that the αA lesion site is indeed intrahelical and involved in reverse Watson–Crick base orientation. The stacking of the αA lesion is achieved at the expense of only one backbone torsion angle, ζ, causing a concomitant increase in the minor groove width downstream (3’ of the lesion) and a significant increase in roll and decrease in twist for the bases flanking the lesion. Together, these structural features result in an overall kink of 18°, which is supported by numerous local restraints at the site of the lesion. (Figure 3-4) With exception of the enhanced sugar puckering for C6, the decamer is quite stable and static. Specifically, there was no evidence from NMR data that would indicate the presence of either multiple structures or significant dynamics. (8) With these results, coupled with the recognition pattern of the endonucleases with abasic lesions, it is believed that the initial local distortion caused by the αA lesion (a widened minor groove and kink) triggers
the endonuclease base-flipping activity. Because of the altered stacking pattern, the energetic cost of driving the nucleic acid into its final conformation is reduced. Enzyme studies of endoIV acting on the αA lesion found that activity is modulated by the flanking sequences. Examination of the C-αA-G NMR structure reveals a similar distortion as the C-αA-C, yet dampened with less of a distortion in base stacking resulting in a more normal B-type minor groove and smaller kink. (9) (Figure 3-4) This supports the concept that variations in the αA base stacking environment change the minor groove topology and the energetics needed to further deform the DNA and flip out the base. (4, 8, 9)

3.1.4 EndoIV Processing of αA Lesions

The NMR structures of αA containing oligonucleotides have given valuable insight into nature of the unbound endoIV substrate and a potential explanation of the observed modulation of enzyme activity. To further understand endoIV’s activity with αA containing substrates, structural knowledge of the αA substrate in complex with the endoIV enzyme is needed. As exhibited by the crystal structure of the endoIV – abasic complex, the enzyme imposes a kink of 67° on the substrate bringing the 5′ ends to a measured distance of 34.2 Å (from a theoretical B-type distance of 40.8 Å), which is in line with the imposed kink.(10) (Figure 3-2) (pdb 1QUM) If the αA containing substrate is processed in a similar manner, it is anticipated that the DNA oligonucleotide will have a similarly induced kink.

Fluorescence resonance energy transfer (FRET) is useful in detecting distances from 10 to 100 Å. It has been used quite extensively to study the global bending and kinking in DNA oli-
gonucleotides for both free DNA as well as DNA–protein complexes. FRET with 5′ labeled fluorophores has been used for numerous studies of bending and kinking effects from bound proteins; examples include: a 95° bend in DNA/HMG-D protein (15), an increase in DNA length upon binding to the DNA adenine methyltransferase M.EcoRI (16), and a 45° bend in DNA when bound to the HU protein. (17) Using distances calculated from FRET measurements of αA containing substrates with 5′ attached fluorophores, an induced kink in the DNA upon binding with endoIV can be deduced.

3.2 Methods and Materials

3.2.1 Substrate Design, Synthesis, and Purification

Oligonucleotides of 15 bp tagged with 5′ dyes were used as the substrate. (Scheme 1) This size oligonucleotide is slightly longer than the endoIV–abasic substrate crystal structure (13 bp) in order to avoid potential interference between the fluorophores and enzyme. The dye pair fluorescein (6-FAM) and tetramethyl rhodamine (TAMRA) using C6 amino linkers were chosen as the donor and acceptor dyes respectively because of their well characterized use with 5′ labeled oligonucleotides. Excitation and emission for the two dyes are: 6-FAM excitation at 494 nm, emission at 521 nm and TAMRA excitation at 556 nm and emission at 580 nm (excitation and emission wavelengths provided by Glen Research).
Scheme 1. Substrate Sequences and Labeling Scheme

C-αA-C duplex  5’ (6-FAM)-GCTATCCXCGACGTC
CGATAGGTGCTGCAG-(TAMRA) 5’

C-αA-G duplex  5’ (6-FAM)-GCTATCCXGGACGTC
CGATAGGTCTGCAG-(TAMRA) 5’

X=αA; 6-FAM and TAMRA are connected to the duplex with C6 amino linkers

The αA containing oligonucleotides were synthesized on an Applied Biosystems solid-phase DNA synthesizer and deprotected. Standard phosphoramidites and the DMT 6-FAM phosphoramidite were purchased from Glen Research; the DMT αA phosphoramidite was purchased from Chem Genes Corp. All oligonucleotides were synthesized with trityl-on and subsequently purified using the Glen-Pak™ purification cartridges with established manufacturer protocols; with this method, near 100% 6-FAM labeling is ensured. (18) The opposite strand oligonucleotides were purchased from Integrated DNA Technologies. Sequences to be TAMRA conjugated included the 5’ C6 amino linker and were mixed with a 6-fold excess of TAMRA NHS ester (Glen Research) in a sodium carbonate buffer (pH 9.0) for 2 h at 37 °C per manufacturer protocols. Purification was completed via size exclusion liquid chromatography and was verified with denaturing polyacrylamide gels; purity of unlabeled strands was verified by denaturing polyacrylamide gels stained with SYBR Green II. Concentrations were calculated using extinction coefficients from the sum of mononucleotides (absorbance at 260 nm, 80 °C); for labeled strands, 6-FAM and TAMRA extinction coefficients at 260 nm, provided by Glen Research, were included (20,900 and 32,300 M⁻¹cm⁻¹ for 6-FAM and TAMRA respectively). (19)
3.2.2 *EndoIV Mutant E261Q*

A mutation at position 261 from glutamate to glutamine was designed by the Cunningham lab; the E261Q mutant will bind an abasic containing substrate in a similar manner as the wild type endoIV yet will not cleave the backbone of the DNA. (20) The 261 glutamate is positioned in the active site and is key to the enzymatic function of the enzyme. Comparisons between the wild type and E261Q mutant crystal structures display few differences with an R.M.S.D. of 0.25 Å for the Cα atoms. (20) Visualization of the active site of the E261Q mutant shows a large pocket capable of containing an αA residue similar to the wild type endoIV. (Figure 3-6) The E261Q mutant was expressed and purified by the Cunningham lab as previously described. (20) Enzyme inactivity was verified by incubating the αA substrates with the E261Q and wild type endoIV and observing the substrate products on denaturing gels. The E261Q was not active. (Figure 3-5)

3.2.3 *FRET*

FRET, the energy transfer between two fluorophores, allows for distance calculations ranging from 10 to 100 Å. (21) The transfer does not occur via emitted photons, rather by the direct energy transfer from a donor fluorophore to an acceptor. If both species are fluorescent, the result of this transfer is a reduction of intensity of the donor fluorescence emission and an increase in intensity of the acceptor. The rate of energy transfer (κ) is dependent on not only the chosen dye pair but also the distance between the fluorophores, refractive index of the medium, relative orientation and mobility of the fluorophores (κ²), the quantum yield of the donor dye (Φ₀), and the spectral overlap of the two fluorophores J(λ).


\[ k_t = \frac{9\ln(10)\Phi D^2 f(\lambda)}{128\pi^5 N_A \eta^4 \tau_D} \]  
\text{Formula 3-1}

\[ J(\lambda) = \frac{\int_0^\infty (\varepsilon^A(\lambda)F(\lambda)^4 d\lambda}{\int_0^\infty F(\lambda) d\lambda} \approx \frac{\sum(e^A(\lambda)F(\lambda)^4 \Delta \lambda}{\sum F(\lambda) \Delta \lambda} \]  
\text{Formula 3-2}

\[ E = \frac{k_t}{\tau_D^{-1} + k_t} \]  
\text{Formula 3-3}

where, \( k_t \) is the energy transfer rate, \( \tau_D \) is the lifetime of the donor in the absence of the acceptor, \( N_A \) is Avogadro’s number, \( r \) is the distance between the two fluorophores, \( J(\lambda) \) is the spectral overlap, \( \varepsilon \) is the extinction coefficient of the acceptor, \( F \) is the normalized fluorescence emission of the donor, and \( E \) is the energy transfer efficiency. The characteristic Förster distance \( (R_0) \) is the distance at which the energy transfer efficiency is 50% (when \( k_t = 1/\tau_D \)). This distance is unique to each individual dye pair, the medium, and the spectral overlap of the fluorophores.

\[ R_0^6 = \frac{9\ln(10)\Phi D^2 f(\lambda)}{128\pi^5 N_A \eta^4} \]  
\text{Formula 3-4}

The spectral overlap of the dye pair \( (J) \) is the major determinant of the Förster distance with the term \( \lambda^4 \) of the overlap integral being the dominant influence. Distances can be calculated from the observed transfer efficiency \( (E) \) using:

\[ E = \frac{R_0^6}{R_0^6 + r^6} \]  
\text{Formula 3-5}

As seen, the efficiency is directly correlated to the distance between the dye pair. The \( 6^{th} \) root dependence between efficiency and distance is a result of the dipole-dipole interaction. (22)
3.2.4 Distance Calculations using Acceptor Emission

FRET efficiencies can be determined by either tracking changes in acceptor or donor fluorescence emissions of the dye pairs used. Efficiencies were calculated based on the measured increase in emission from the acceptor dye. The method being used is a slightly modified version of the “Ratio A” method. (23, 24) Advantages include:

- Single sample measurements (2 scans per sample).
- The method is not concentration dependent, so samples can be easily compared.
- Decrease in donor emission may be a result of other processes (ie unanticipated quenching).
- Increase in acceptor emission from energy transfer can only be from FRET and other processes that can be accounted and corrected for.

The non-modified ratio A calculation is as follows:

\[
\text{RatioA} = \frac{F^A(ex = D, em = A)}{F^A(ex = A, em = A)}
\]

\[
\text{RatioA} = \left[ E d^+ \frac{\varepsilon^D(D, \text{max})}{\varepsilon^A(A, \text{max})} + \frac{\varepsilon^A(D, \text{max})}{\varepsilon^A(A, \text{max})} \right] \frac{\phi^A(\text{em}–dl–acc)}{\phi^A(\text{em}–sl–acc)}
\]

where \(F^A\) is the measured acceptor fluorescence emission of a double labeled at the indicated excitation (ex) and emission (em) wavelengths for donor (D) and acceptor (A) maximum wavelengths, \(E\) is the transfer efficiency, \(d^+\) is the labeling efficiency of the donor dye, \(\varepsilon^D\) and \(\varepsilon^A\) are the extinction coefficients at the indicated wavelengths in the parenthesis for the donor and acceptor respectively, and \(\phi^A\) is the quantum yield of the acceptor dye under double labeled
(numerator) and single labeled (denominator) conditions. By setting the equations equal to each other, E can be solved for.

The observed $F^A$ is made up of three components: enhanced emission from FRET, fluorescence emission from the donor “tail” (T), and direct excitation of the acceptor dye (DIRECT). To compensate for the donor tail, a ratio of emission at the donor and acceptor’s wavelengths is used ($^{SD}f$):

$$^{SD}f = \frac{^{SD}F(ex = D, em = D)}{^{SD}F(ex = D, em = A)}$$  \hspace{1cm} \text{Formula 3-8}$$

$$T = \frac{^{DL}F(ex = D, em = D)}{^{SD}f}$$  \hspace{1cm} \text{Formula 3-9}$$

where $^{SD}F$ is the fluorescence emission of the fluorescein only labeled sample, and $^{DL}F$ is the emission of the double labeled sample.

Although the term $\varepsilon^A(D, \text{max}) / \varepsilon^A(A, \text{max})$ in formula 3-7 adjusts for direct excitation of the acceptor, an alternative method uses the ratio of direct measurements of acceptor fluorescence emissions ($^{SA}f$) in a similar manner as the tail contribution (Toth, 1998):

$$^{SA}f = \frac{^{SA}F(ex = A, em = A)}{^{SA}F(ex = D, em = A)}$$  \hspace{1cm} \text{Formula 3-10}$$

$$\text{Direct} = \frac{^{DL}F(ex = A, em = A)}{^{SA}f}$$  \hspace{1cm} \text{Formula 3-11}$$

where $^{SA}F$ is the fluorescence emission of the acceptor only labeled sample and D is the direct excitation of the acceptor dye when excited at the donor’s excitation wavelength maximum.

This reduces the classical ratio A method for calculating E to:

$$\frac{\langle P^A(ex=D,em=A) - T - \text{Direct} \rangle}{P^A(ex=A,em=A)} = Ed^* \left[ \frac{\varepsilon^D(D,\text{max})}{\varepsilon^A(A,\text{max})} \right]$$  \hspace{1cm} \text{Formula 3-12}$$
In formula 3-7, the term $\phi^A / \phi^A$ compensates for change in quantum yield resulting from a shift in the acceptor emission wavelength. In the system used, such a shift was not observed and therefore this term can be set equal to 1. Distances can be calculated using $E$ from formulas 3-12 and 3-5.

### 3.2.5 Fluorescence Emission Measurements

The DNA concentration for all solution fluorescence measurements was 165 nM (duplex concentration). Unbound substrate samples were in 10 mM sodium phosphate (NaP), pH 7.5 and 100 mM NaCl. The buffer used for the bound FRET analysis consisted of 10 mM NaP at pH 7.5, 100 mM NaCl, and 1 μM β-mercaptoethanol. Because FRET measurements are dependent on the ratio of the acceptor fluorescence intensities, any excess single strand TAMRA labeled DNA would result in errors in the FRET calculations; for the systems studied, all analyzed duplexes contain an excess of 6-FAM labeled strand.

Solution fluorescence measurements were made on Perkin Elmer LS55 fluorescence spectrometer using excitation and emission slit widths of 5 nm and 3 nm respectively with a scanning rate of 150 nm/min averaged over three scans. Anisotropy measurements were completed on the same instrument. Anisotropy was calculated using:

$$\text{anisotropy} = \frac{I_\parallel - GI_\perp}{I_\parallel + 2GI_\perp} \quad \text{Formula 3-13}$$

where $I_\parallel$ and $I_\perp$ are fluorescence intensity with polarizers set parallel and perpendicular to one another respectively, and G is the correction for the sensitivity of the polarizers. It is calculated as $G = I_{HV} / I_{HH}$ where $I_{HV}$ is the intensity when the excitation polarizer is set to horizontal and the
emission polarizer set to vertical and $I_{HH}$ is the intensity when both are set to horizontal. The G-factor was measured for each wavelength and slit combination used.

Binding constants ($K_d$) were calculated using a 1:1 binding model, which was fit to measured anisotropy data in Kaleidagraph 4.1.3. The concentration of the bound species was calculated using:

$$[PD] = \frac{([P_{total}] + [D_{total}) + K_d) \pm \sqrt{([P_{total}] + [D_{total}) + K_d]^2 - 4[P_{total}][D_{total}]}}{2}$$

Formula 3-14

where P and D represent E261Q and DNA substrate concentrations, respectively and PD represents the complex concentration. Fraction bound was determined from the concentration of the bound species.

### 3.3 Results and Discussion

#### 3.3.1 Fluorescein Quantum Yield

The DNA sequence chosen for the studies contained a 5’ terminal guanosine (G), which quenches fluorescein. Due to the partial dependence of the Förster distance on the quantum yield of the donor dye, the effective fluorescein quantum yield for the $\alpha$A substrate was calculated. By comparing the absorbance and emission of an unknown dye to a reference fluorophore at multiple concentrations, the quantum yield of the unknown can be calculated using:

$$Q_{unk} = \frac{Q_{std} (F_{unk}/A_{unk})}{[(F_{std}/A_{std}) (n_{unk}/n_{std})]^{2}}$$

Formula 3-15
where $Q$ is the quantum yield, $F$ is the fluorescence intensity, $A$ is the absorbance, $n$ is the refractive index, unk is the unknown dye, and std is the standard. The chosen standard was rhodamine B in water; the single labeled 6-FAM duplex was used as the unknown. Rhodamine B has an excitation maximum of 514 nm and a quantum yield of 0.31. (26) Because rhodamine B’s standard quantum yield was determined in water, as is the aA substrate, the term $(n_{unk} / n_{std})^2$ in formula 3-15 can be disregarded. Further, concentrations for fluorescence emission measurements were kept below 0.1 OD to avoid possible inner filter effects of the dyes. The quantum yield for fluorescein bound to the aA substrate is 0.61 +/- 0.06. (Figure 3-7) which is lower than anticipated quantum yield of 0.9 indicating that the adjacent guanosine does indeed quench the fluorophore.

3.3.2 Förster Distance

Emission and absorbance spectra of single labeled 6-FAM and TAMRA, respectively, were measured giving a spectral overlap of $2.89 \times 10^{15} \text{ M}^{-1}\text{cm}^{-1}\text{nm}^4$. (Figure 3-8) Using a $k^2$ of 2/3, $f_0$ of 0.61, and a refractive index of 1.33, a Förster distance of $56 \pm 1\text{Å}$ was calculated using formula 3-4 (the variance arises from the error range of the quantum yield).

3.3.3 Free Substrate FRET Calculations

Contributions of T and DIRECT were determined using single labeled duplexes; the resulting ratios were 13.5 and 8.0 respectively. Calculated FRET distances for the C-αA-C, C-αA-G, and C-A-C sequences are detailed in Table 1 along with 5’-5’ distances determined from NMR
solution structures (C-αA-C and C-αA-G) as well as a canonical B-type helix for the C-A-C duplex. An effective linker length of approximately 14 Å adequately fits each FRET measurement and will be used as the linker length for endoIV bound substrate. In all cases, the FRET determined distances are in agreement with measured distances on the NMR structures as well as the idealized B-type CAC structure.

Of note, placement of the dyes on the 5’ ends relative to each other is different for the solved αA containing NMR structures and the theoretical B-type control structure. When treated as a two dimensional object, the difference in distances with an 18 degree kink is approximately 1 Å. The slight difference in placement of the dyes, however, offsets the differences in distances between the oligonucleotides.

3.3.4 E261Q – αA Substrate Binding Constants Using Anisotropy

The 1:1 binding disassociation constants (K_d) were calculated using the measured anisotropy changes of 6-FAM for both the CαAC and CαAG substrates when bound to E261Q. Due to high standard deviations observed in the TAMRA anisotropy values combined with little overall change in magnitude, K_d values could not be calculated from the TAMRA data. Addition of E261Q results in the increase of the 6-FAM anisotropy values (0.06 to 0.17). (Figure 3-9) These data give K_d values of 60 ± 4 nM and 78 ± 6 nM for the CαAC and CαAG substrates respectively. (Table 3-2) E261Q’s higher affinity for the CαAC substrate is in agreement with the results from the free substrate NMR analysis in which the CαAG flanking context was found to be more B-type and thought to partially conceal the αA lesion site from enzyme recognition. (9)
3.3.5 Solution Fluorescence Measurements of the Bound Substrate

Upon addition of E261Q, the observed fluorescence emission of both 6-FAM and TAMRA increased when the dyes were directly excited; this is in contrast to the anticipated results: a decrease in 6-FAM and unchanged TAMRA emissions. (Figure 3-10) To address possible causes, the 6-FAM and TAMRA volume adjusted fluorescence intensities were plotted against E261Q concentration. (Figure 3-10) For 6-FAM, a specific binding curve is observed; calculations of $K_d$ from these curves (Table 3-2) suggest the change in fluorescence intensity is a result of E261Q specific binding to the lesion site. Similar plots for TAMRA do not yield a specific binding curve and are instead suggestive of a weak affinity association. As discussed previously, the substrate sequence context contains a guanosine adjacent to 6-FAM that partially quenches the fluorophore. Based on the observation that increased fluorescence emission is correlated to E261Q binding, it is theorized that upon binding with E261Q, the guanosine and 6-FAM stacking is decreased thus lowering guanosine’s quenching effect. Using a 6-FAM single labeled CαAC duplex, the quantum yield increase for the donor is estimated to be 1.3 fold higher than the free substrate.

In an effort to compensate for the observed changes, the Förster distance was recalculated for the substrate when bound to E261Q. The Förster distance is dependent on the donor quantum yield as well as the dye pair’s orientation factor $\kappa^2$. The quantum yield for 6-FAM in the bound substrate increases to 0.79 (free substrate quantum yield was 0.61). For $\kappa^2$, a value of 2/3 represents a randomized orientation with respects to the dye pair. For 6-FAM, the measured anisotropy increases from 0.06 and 0.17; these values are relatively low and are indicative of free rotation. The TAMRA anisotropy changes from approximately 0.15 to 0.25 is
consistent with a stronger interaction between the dye and the duplex end. However, this value is less than that of other studies, which found TAMRA to experiences enough mobility for random orientation between the two dyes. (17, 23, 27) In addition to these changes, a red-shift in the donor emission of 2 nm was observed; this phenomenon has been observed with other bent DNA and its impact to the Förster distance is nominal at < 0.5 Å. (17) The revised calculated Förster distance for the 6-FAM – TAMRA dye pair for the bound substrate using the increased quantum yield is 59.1 Å.

In order to avoid the impacts of low affinity binding between TAMRA and E261Q, fluorescence measurements at low protein concentrations (< 70% binding) were used. To compensate for incomplete binding, the fraction bound (f_{bound}) was calculated from K_d. Using a two state system (free and bound substrate), the effective transfer efficiency for the bound system (E_{bound}) can be calculated from the observed efficiency (E_{observed}) and efficiency of the free substrate (E_{free}) using:

$$E_{\text{observed}} = (1-f_{\text{bound}})E_{\text{free}} + f_{\text{bound}}E_{\text{bound}}$$

Formula 3-16

With E_{bound}, the FRET distance for the bound substrate can be calculated normally using formula 3-5. The calculated distances for both substrates are outlined in Table 3-3.

The bound distances calculated for each substrate are similar in both magnitude and error. Of note, at higher protein concentrations (binding > 70%), the calculation compensation efforts were lost and calculated distances changed (all distance were > 60 Å). Using a two dimensional model, a theoretical dye to dye distance of 54 Å is anticipated for a kink of 67°. The observed distances are 55 ± 2 Å and 56 ± 2 Å for the CαAC and CαAG bound substrates respectively. These distances correlate to a calculated kink angle of 68° ± 6°.
3.4 Conclusion

For the free substrate, the FRET calculations were consistent with measured distances from the NMR structure PDB files as well as the theoretical B-type CAC control sequence. Interestingly, for the substrates bound to E261Q, the calculated $K_d$ values match the anticipated trend from the NMR analysis with the endoIV mutant showing a higher affinity towards the CαAC substrate as compared to the CaAG substrate. Unexpectedly, the addition of E261Q to the αA substrates resulted in an increase in fluorescence emission for both directly excited 6-FAM and TAMRA. The enzyme is known to participate in end-stacking with DNA at high protein concentrations. (R. Cunningham, personal communication) This end stacking may be the cause of the increased TAMRA emission; observation of the increase in TAMRA emission is suggestive of weak affinity binding with the increase more significant when the 1:1 E261Q binding was near completion. The increase in 6-FAM intensity correlates with E261Q binding and is thought to be the result of a decrease in stacking between 6-FAM and the adjacent guanosine.

The calculated FRET distances for the two substrates in complex with endoIV were calculated over several measurements (>9) and are in relatively close agreement with one another (average calculated distance is within 1 Å of one another). The anomalous effects of binding on the fluorescence emission were compensated for at low concentrations. The resulting distances and kink angle are in range with the anticipated values observed from the X-ray crystal structure. Unfortunately, access to an abasic containing oligonucleotide serving as a control was not available. However, the data collected to date suggests that endoIV processes αA substrates in a similar manner as abasic containing substrates. The identification of other similarly
structured lesion sites may lead to other DNA damage sites that are processed in a similar manner as abasic and αA damage.

3.5 Appendix

An alternative method for FRET calculations was investigated (in-gel FRET). However, due to the unanticipated increase in donor fluorescence, it could not be utilized. This appendix details the method.

3.5.1 In-gel FRET Measurements

Measurement of fluorescence in polyacrylamide gels (PAGE) offers an alternative approach to FRET measurements (in-gel FRET). (28) The advantages of this method include the significantly reduced amount of materials needed for measurements and the separation of bound and unbound substrate. The limitations are increased errors in readings and the required gel shift to occur. (28) There must be an adequate mobility difference between the bound and unbound species in order to discretely measure the intensities of the bands separately, yet too high of a disparity in mobilities may result in a difference in the band diffusion resulting in difficulties in measuring intensities or an inability to capture both bands on the gel.

In-gel FRET measurements utilize the decrease in donor fluorescence to determine the transfer efficiency. Due to instrument limitations, comparison of the acceptor emission from direct excitation and donor excitation is not possible (different lasers are utilized). The com-
parison of donor fluorescence emission from the free species and bound species allows for a
calculation of energy transfer efficiency as follows:

Efficiency calculated from donor decreased emission is:

\[ E = 1 - \left( \frac{l_{\text{complex, dl}}}{l_{\text{complex, sl}}} \right) \]  \hspace{1cm} \text{Formula 3-17}

where \( E \) is energy transfer efficiency, \( l_{\text{complex}} \) is the emission intensity of the donor, \( \text{dl} \) is double
labeled with donor and acceptor, and \( \text{sl} \) is single labeled with donor only. Given total DNA pop-
ulation (\( N \)) is related to free and bound states with:

\[ N_{\text{DNA}} = N_{\text{complex}} + N_{\text{free}} \]  \hspace{1cm} \text{Formula 3-18}

then the total intensities observed in both free and bound states can be related as:

\[ I_0 = I_{\text{free}} + I_{\text{complex, sl}} \]  \hspace{1cm} \text{Formula 3-19}

where \( I_0 \) is the total fluorescence intensity and \( I_{\text{free}} \), and \( I_{\text{complex, sl}} \) are the intensities of the un-
bound substrate and single labeled bound substrate (i.e. where FRET can not occur). Rearrang-
ing and substituting in \( E \) from formula 3-17:

\[ I_0 = I_{\text{free}} + \left( \frac{l_{\text{complex, dl}}}{1-E} \right) \]  \hspace{1cm} \text{Formula 3-20}

\[ l_{\text{complex, dl}} = \left( E-1 \right)I_{\text{free}} + \left( 1-E \right)I_0 \]  \hspace{1cm} \text{Formula 3-21}

As seen with formula 3-21, plotting observed donor emission intensities of the bound
state vs the unbound state will yield a linear line with a slope equal to \( E-1 \).

In order to utilize this method, however, impacts of the PAGE environment on the donor
fluorophore properties must be taken into account. Fortunately, PAGE effects on fluorescein
have been studied. The quantum yield of fluorescein in PAGE is reduced by 35% as compared
to solution measurements; this quenching is independent of PAGE concentration. Anisotropy of
fluorescein is not significantly impacted thus allowing for the same \( \kappa^2 \) value of 2/3. (28) This
allows for the recalculation of an effective in-gel Förster distance from solution-based measurements. With the calculated E from formula 3-21 and the PAGE corrected R obed distance, the bound dye pair distance can be calculated using formula 3-5.

Of note, this method will calculate a difference in E between the free and bound states present in the gel. If energy transfer between the dye pairs is already present in the free state, then the E of the free species must be determined with traditional solution-based FRET measurements and added to the observed E for the in-gel bound state. In addition, care must be taken to adjust for the quantum yield reduction of fluorescein in PAGE; converting the solution determined E of the free state to a distance, then converting back to an effective in-gel E using the recalculated in-gel Förster distance will give a corrected in-gel E for the free state.
Table 3-1 Free Substrate Distances

\(^A\) FRET distances represent 6-FAM to TAMRA distances determined using formulas 3-12 and 3-5; 
\(^B\) C-A-C 5′-5′ distance is from a canonical B-type DNA structure; \(^C\) PDB 5′-5′ distances are measured distances from O5 to O5 for the PDBs 1S75 and 2LIB, \(^D\) Linker distances are based on the difference between FRET distance and PDB distances.

<table>
<thead>
<tr>
<th></th>
<th>C-A-C (Å)</th>
<th>C-αA-C (Å)</th>
<th>C-αA-G (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Determined FRET Distance(^A)</td>
<td>65 ± 1</td>
<td>66 ± 1</td>
<td>66 ± 1</td>
</tr>
<tr>
<td>PDB 5′-5′ Distance</td>
<td>51(^B)</td>
<td>51(^C)</td>
<td>51(^C)</td>
</tr>
<tr>
<td>Linker Distance(^D)</td>
<td>14</td>
<td>14</td>
<td>14</td>
</tr>
</tbody>
</table>
Table 3-2  Equilibrium Constants ($K_d$) for CαAC and CαAG

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Anisotropy (nM)</th>
<th>6-FAM Emission (nM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C-αA-C</td>
<td>60 ± 4</td>
<td>63 ± 13</td>
</tr>
<tr>
<td>C-αA-G</td>
<td>78 ± 6</td>
<td>114 ± 16</td>
</tr>
<tr>
<td>Sequence</td>
<td>average $E_{\text{bound}}$</td>
<td>average R (Å)</td>
</tr>
<tr>
<td>------------</td>
<td>-----------------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>C-αA-C</td>
<td>0.61 ± 0.05</td>
<td>55 ± 2</td>
</tr>
<tr>
<td>C-αA-G</td>
<td>0.59 ± 0.04</td>
<td>56 ± 2</td>
</tr>
</tbody>
</table>

Table 3-3  Calculated FRET Distances for the Bound αA Substrates
Figure 3-1 Stick diagrams of A) deoxyadenosine and B) $\alpha$A. Atom position numbering is noted in A.
Figure 3-2  X-ray crystal structure of endoIV in complex with an abasic containing DNA substrate. (pdb id: 1QUM) The DNA substrate is in blue with the abasic lesion in red. The cartoon representation is the endoIV enzyme. The red lines on the DNA substrate approximate the helical axis on either side of the lesion site, which points out the induced kink on the substrate.
Figure 3-3  A) αA residue (red) manually docked in the endoIV catalytic pocket in place of the abasic residue. (10, 11)(pdb id: 1QUM)  B) APE1 – abasic (red) complex structure focused on the catalytic pocket. (pdb id: 1DE8)
Figure 3-4  Trends observed in the endoIV substrates.\(^9\) Top from left to right: abasic containing substrate found in complex with endoIV (grey) (pdb id: 1QUM), solution structure of CαAC (red) (pdb id: 1S75), solution structure of CαAG (orange) (pdb id: 2LIB), and a model of canonical B-type CAC (blue). Bottom: observed trends in the structures.
Figure 3-5  Denaturing polyacrylamide gel (PAGE) of DNA substrates processed by the wild type endoIV and the mutant E261Q. PAGE conditions are: 15% polyacrylamide (19:1 bis-acrylamide ratio), 8 M urea, and tris-borate-ethylenediaminetetraacetic acid buffer (TBE). The picture is observed colors captured with a portable camera; green is 6-FAM and red is TAMRA. From left to right: DNA only (CαAC), processing of the CαAC substrate with the wild type endoIV (note the processing of the green 6-FAM strand which contains the αA lesion), E261Q with the CαAC substrate, and E261Q with the CαAG substrate (note how neither substrate backbone is cleaved with the E261Q mutant).
Figure 3-6 X-ray crystal structure of the catalytic pocket of the E261Q mutant in complex with an abasic containing substrate. (20) (pdb id: 2NQJ) In grey is the pocket, green is the E261Q mutation, orange are the Zn$^{2+}$ ions, and the ball and stick representation is the abasic lesion site (red = oxygen, cyan = carbon, gold = phosphorous).
Figure 3-7 Graph of fluorescence intensity vs. absorbance for 6-FAM and rhodamine B. The red line is the single labeled 6-FAM CαAC substrate and the blue line is the rhodamine B reference. The trend line is a linear fit to the data.
Figure 3-8 Emission and absorbance spectra of 6-FAM and TAMRA, respectively. Both spectra have been normalized so the maximum value is equal to 1 for visual comparison only. Red is the 6-FAM fluorescence emission spectrum when excited at 495 nm, in blue is the TAMRA absorption spectrum. The baseline excitation peak has been subtracted from the 6-FAM emission spectrum.
Figure 3-9  Plotting of anisotropy vs. E261Q concentration. A) 6-FAM  B) TAMRA. Note, both are for the CαAG substrate.
Figure 3-10  Emission spectra of 6-FAM and TAMRA when directly excited.  A) Raw emission spectra from 6-FAM (excitation at 495 nm); lines represent spectra with increasing amounts of E261Q added.  B) Raw emission spectra from TAMRA (excitation at 560 nm); lines represent spectra with increasing amounts of E261Q added.  Color coding for A) and B) are not noted on the spectra.  C) Volume adjusted intensities at 519 nm vs. E261Q concentration when excited at 495 nm.  D) Volume adjusted intensities at 579 nm vs. E261Q concentration when excited at 560 nm.  Note, lines in C) and D) only connect the data points.
3.6 References:

4 CHARACTERIZATION OF THE RREIIB-TR RNA

4.1 Preface

The zinc finger single point mutations, expression, and purification were completed by Subrata Mishra as well as the natural abundance NMR characterization of the RREIIB-Tr and the solution structures of the free and bound zinc finger proteins. The isothermal titration calorimetry (ITC) work was completed in collaboration with Subrata Mishra.

4.2 Overview & Background

4.2.1 HIV-1, Rev, and the Rev Response Element (RRE)

The Human Immunodeficiency Virus-1 (HIV-1) rev protein binds to a highly conserved specific RNA structure called the rev responsive element (RRE) in HIV introns.\(^1\),\(^2\) Rev binding to RRE is absolutely required for the production of virions. In early stages of HIV mRNA expression, the transcribed viral RNA undergoes multiple splice events before nucleocytoplasmic transport. This RNA encodes for various HIV-1 regulatory proteins, including rev, which predominantly localizes in the nucleus. In later stages of mRNA expression, as rev accumulates, it binds as a monomer to the IIB loop of RRE (RREIIB). (Figure 4-1 A) Subsequent multimerization binding to the RRE inhibits splicing and signals for cytoplasm transport. The unspliced and singly spliced viral RNA encodes the viral structural proteins as well as provides unspliced genomic RNA for virion packaging. This event ushers in the late phase of the cellular HIV infection: virion assembly and budding. (Figure 4-2)\(^3\)-\(^7\) The importance of the rev-RRE interaction for HIV-1 replication makes it an obvious target for therapeutic intervention by molecular means.\(^8\)
4.2.2 Generation of RRE Binding Peptides

The RNA binding region of rev has been isolated to an arginine rich sequence of the protein with a high specificity for the purine rich bulge in RREIIB (K_0 of 1 to 3 nM).(5) Peptides based on this sequence express the same specificity and have been shown to adopt an α-helical conformation when bound. Solution NMR structures of the rev peptide have been solved with various versions of the RREIIB RNA construct including a truncated version, RREIIB-Tr.(4, 9) (Figure 4-1 B) Complex formation between rev and RREIIB-Tr causes the formation of two GC and two purine-purine (GG, GA) base pairs in the bulge, which are critical to rev binding (3). Subsequent studies of the rev peptide have identified key side-chain interactions required for binding. In particular, N40 forms hydrogen bonds to the major groove face of the GA base pair in the bulge region.(10, 11) Other studies showed that a higher α-helical predisposition of the peptide improved binding.(12) Expanding on this, the Frankel lab has shown that inclusion of a modified version of the RRE binding rev peptide sequence into a zinc finger motif maintains a ββα fold and exhibits RREIIB specificity.(13) An alternative approach is to employ phage display techniques to optimize the proteins sequence of zinc fingers to effect specific binding to RREIIB.(8, 14)

4.2.3 Using Phage Display to Create a Zinc Finger Protein with Affinity to RREIIB RNA

The Darby lab utilized phage display to develop a zinc finger with RREIIB specificity. (14) In addition to being highly stable, the zinc finger motif has been well studied in protein engineering and biocombinatorial schemes. (14, 15) With phage display, multiple amino acids of a target protein can be changed simultaneously and are expressed on the surface of the bacte-
riophage. Selection of high affinity proteins is accomplished via binding of the protein to a desired target; subsequent washing will eliminate non-binding and weakly binding phage (i.e. phage with changed proteins that do not bind). Phage expressing stronger binding proteins are subsequently eluted, amplified in bacteria, isolated, and put through additional cycles of phage display as needed. Sequencing the phage DNA allows for the sequence determination of members that exhibit the desired properties. (14, 16)

The starting construct for the RREIIB binding zinc finger phage display was a tandem zinc finger design using the 4th zinc finger of the TFIIB protein as a template. The amino acids on the alpha helix in positions -2 to 10 were subjected to changes leaving the core zinc binding residues as well as the beta sheet unaltered. The resulting product, RR1, was a tandem zinc finger with a high affinity and specificity to the RREIIB RNA exhibiting a $K_d$ of 7.9 nM. This is affinity is similar to the monomeric rev protein specificity to RREIIB. (14) Subsequently, it was found that only one of the zinc fingers (RR1-zf2) was primarily responsible for binding; removal of the second acidic zinc finger (RR1-zf1) resulted in a $K_d$ of 26 nM indicating that the majority of the binding determinants are present on this zinc finger. (17)

4.2.4 **The ZF29 RREIIB-Tr Binding Zinc Finger**

The small size and stability of a single zinc finger is an attractive approach as a therapeutic or diagnostic agent. With the quite promising results from the RR1-zf1 zinc finger, further enhancements to the design were pursued. The RR1-zf2 zinc finger served as the precursor for the smaller ZF29 protein developed by the Germann lab. (Figure 4-3) The ZF29 and ZF29G29R proteins were found to exhibit high affinity and specificity for the RREIIB-Tr RNA with $K_d$ values
of 149 +/- 21 and 57 +/- 6 nM for ZF29 and ZF29G29R respectively. Further, the ZF29 and ZF29G29R proteins were found to bind the same bulge region of RREIIB-Tr as the rev peptide.\textsuperscript{(18, 19)} The solution structures of ZF29 and ZF29G29R (gly to arg mutation at position 29) confirm the $\beta\beta\alpha$ fold of the CCHH type zinc finger. Both zinc fingers are dependent on this $\beta\beta\alpha$ fold for high affinity binding. This is unlike the rev incorporated zinc finger, which is still capable of binding to the RREIIB target without zinc and the proper $\beta\beta\alpha$ fold (3 fold decrease in affinity).\textsuperscript{(13)} The ZF29 and ZF29G29R show no binding affinity when unfolded.

Upon ZF29 binding to RREIIB-Tr, additional imino proton peaks are observed. (Figure 4-4) These imino protons suggest the formation of additional base pairs within the RNA, however the possibility of ZF related peaks could not be ruled out. Earlier work completed on a $^{13}$C, $^{15}$N labeled ZF sample revealed that of the new peaks formed, 2 are a result of his residues (His 23 and His 27).\textsuperscript{(20)} The remaining peaks must be a result of RREIIB-Tr related protons. The imino proton spectra of the ZF29 bound species, as compared to rev protein binding the RREIIB-Tr RNA, are different suggesting different mode of binding for ZF29 and ZF29G29R proteins.

\textbf{4.2.5 \textit{Isothermal Titration Calorimetry for Thermodynamic Profiling of ZF29}}

A study using isothermal titration calorimetry (ITC) was used to characterize the driving forces behind binding and further enhance the design of the zinc finger.\textsuperscript{(19)} Single point mutations were made on the alpha helix, tip, and beta sheets of ZF29 (Figure 4-5 A and B); by varying salt, buffer, and pH conditions with these various mutants, binding dependence on ion displacement and proton uptake can be elucidated. Beta sheet mutations at positions H6 and R12 were found important in proton uptake and ion displacement, respectively. The data indicate a
binding induced change in the pKa of H6 (from 6.7 to 8.3 for the unbound and bound states respectively) and an electrostatic contact between the RNA phosphate backbone and R12. It is theorized that these residues, which flank either side of ZF29, stabilize the zinc finger in the widened bulge of the RNA. Residues at the tip of ZF29 were also found to be critical for binding; alanine mutations made to N15 and D16 resulted in 2 and 3 fold reductions in binding affinity and disrupt the imino proton spectra of RREIIB-Tr indicating that the tip is critical for the structure of the resulting complex. (Figure 4-5, Table 4-1) Positions N21 and G29 of the alpha helix were also probed. The N21 position was investigated to see if it serves a similar role as N40 of the rev peptide which participates in hydrogen bonds with the major groove face of the G:A base pair formed upon binding. (10) Mutations to alanine and glutamine had opposite effects where alanine reduced binding and glutamine enhanced binding affinity (-1.6 and +1.9 fold changes). (Figure 4-5, Table 4-1) Because the N21A mutation results in a similar imino proton spectrum as the control, it does not serve a similar role as the rev peptide’s N40. The G29 mutation to arginine results in an additional electrostatic contact and a higher binding affinity as compared to ZF29. (19) The data combined allowed for the design of a double mutant where the 29 and 21 positions were mutated to arginine and glutamine respectively, which exhibits a $K_d$ of 35 +/- 10 nM. (19) (Table 4-1)

4.2.6 Fluorescence and $^{13}$C$^{15}$N NMR Investigation of the RREIIB-Tr RNA

To gain further structural specificities of ZF29 / RREIIB-Tr binding, solution fluorescence of 2-aminopurine (2AP) substituted RREIIB-Tr was conducted. The fluorescent 2AP is similar in structure to guanosine yet lacks the C6 carbonyl, which results in destabilized base pairing as
compared to guanosine. (Figure 4-6A) Also, NMR of $^{13}$C and $^{15}$N enriched RREIIB-Tr was used to identify the type of imino protons detected during complex formation. Ultimately, a high-resolution solution structure of the ZF29 / RREIIB-Tr complex structure is desired in order to give insight into the exact nature of the specific contacts made. These contacts may provide critical insight in the creation of a novel HIV RRE transport inhibitor. However, because of the size of the complex formed (29 aa, 34 bases), traditional unlabeled NMR methods for determining solution structures are not possible due to resonance overlap. Previous work by Subrata Mishra provided the solution structure as well as the relative orientation of the bound ZF29G29R protein. (20) In order to determine the structure of the bound RREIIB-Tr RNA and the overall complex structure, $^{13}$C and $^{15}$N enriched RNA is needed for heteronuclear NMR. Protocols and methods for transcribing and purifying RNA with T7 RNA polymerase were developed. (Chapter 5.2) The method produces relatively high RNA yields (approximately 8 ODs per 500 µL reaction for the RREIIB-Tr RNA) with no observed N+1 transcription products. In addition to optimizing the yields, minimization of the number of steps involved in purification was pursued in order to avoid inadvertent RNase contamination; the purification protocols developed using ion exchange and size exclusion liquid chromatography results in a 2-step purification scheme with yields pure enough for NMR and ITC analysis.

By utilizing isotopically enriched NTPs for the RNA transcription, 2D and 3D heteronuclear NMR experiments are possible, but many of these experiments have not been used in the Germann lab. Various through-bond 2D and 3D NMR experiments were tested, adapted, and parameterized for the RREIIB-Tr system. In addition, a methodology for elucidating the identity
of the resonances and bases was developed. These experiments and methodology will be needed to develop a solution structure of the RREIIB-Tr / ZF29G29R complex.

4.3 Results and Discussion

4.3.1 Solution Fluorescence of 2AP Substitutions of RREIIB-Tr

Several key guanosine bases located in the vicinity of the bulge region anticipated to be critical to the complex structure were substituted with 2AP. (Figure – 4-6B) These substitutions continued to exhibit 1:1 binding stoichiometry with ZF29G29R, yet result in reduced binding affinities (ranging from 210 to 670 nM $K_d$ values for the various 2AP mutations) indicating that any destabilization to the bulge region is detrimental to ZF binding. (19) In addition to observing 2AP’s impact on binding affinity, the fluorescence of 2AP can also be utilized to monitor base stacking changes upon binding. The fluorescence intensity of 2AP is dependent on base stacking: increases in stacking will reduce the intensity while decreased stacking will increase fluorescence intensities. (21) Fluorescence spectra were measured for the various 2AP substitutions in an effort to give potential insight into the stacking changes taking place upon binding. The G50 2AP spectrum exhibits decreased fluorescence intensity indicating enhanced base stacking. In contrast, G47 and G70 2AP spectra have increased intensities suggesting a decrease in base stacking. The double mutant, G48-G71 2AP, exhibited little change in fluorescence suggestive of either no change in stacking or compensation changes in stacking. (Figure 4-6C) Because the 2AP substitution itself disrupts base pairing due to the elimination of the carbonyl group, a direct correlation to changes occurring in the non-mutated RREIIB-Tr cannot
be made. This analysis does give additional evidence that the RREIIB-Tr’s bulge region undergoes a conformational change upon binding to accommodate the ZF.

4.3.2 NMR of $^{13}$C and $^{15}$N Enriched RREIIB-Tr

NMR of $^{13}$C and $^{15}$N isotopically enriched nucleic acids gives the ability to use heteronuclear NMR experiments to determine the chemical shifts and coupling constants of these nuclei. The chemical shifts and coupling constants for $^1$H, $^{13}$C, and $^{15}$N nuclei are well characterized for nucleic acids. (22) Upon ZF29 binding with RREIIB-Tr, several additional imino proton resonances are observed. (Figure 4-4) The unambiguous identity of these protons had not yet been determined using non-labeled RNA. Inspection of the bulge region indicates that the imino protons should belong to guanosines, yet the potential involvement of the flipped out U42 cannot be ruled out. Because the ZF29 binding method is believed to differ from the rev peptide, the rev – RREIIB-Tr complex structure cannot be used as an absolute predictor for the base pairs formed.

The $^{15}$N-$^1$H HSQC (heteronuclear single quantum coherence spectroscopy) spectrum correlates protons bonded to nitrogens. This NMR experiment in the imino proton region can distinguish G’s imino protons from U’s using the $^{15}$N chemical shifts. (Figure 4-7) (22) The imino proton spectrum of the free RREIIB-Tr had already been assigned using NOESY spectra. (20) The HSQC spectrum of the free RREIIB-Tr collected supports the earlier assignment. (Figure 4-7) Of note, the loop G55 imino proton does not appear in spectra at 298 K but does at 280 K indicative of a trapped imino proton rather than involvement in a hydrogen bond, which is also consistent with the chemical shift of the imino proton (10.5 ppm).
When ZF29 is bound, the changes in the imino proton spectra can be tracked in the HSQC spectra. (Figure 4-8) Previous work on assigning the bound imino proton spectrum was conducted using natural abundance NOESY spectra. The identity of the additional imino protons could not be elucidated with these spectra, but modifications to the RREIIB-Tr RNA with 2AP and H3 methyl substituted uracil (3MU) offered a hypothesis for their identities. In particular, the 3MU substitution for U72 resulted in the loss of imino proton resonances; it was theorized that RREIIB-Tr’s U72 stacks in the bulge upon binding and may contribute the observed imino proton at 11.85 ppm. (20) In contrast, the HSQC spectra collected demonstrate that the identity of the additional base pairs formed involve only G’s, not the U72. In addition, the spectra obtained from the in-house transcribed RNA exhibit slightly sharper line shapes when bound to ZF29 as compared to the Dharmacon RNAs. (Figure 4-9) With this, the 1D imino proton and HSQC spectra reveal that potentially 4 additional imino protons arise upon binding including a previously unobserved new resonance near G76 at 12.85 ppm which corresponds to a broadened HSQC peak. (Figure 4-8) The additional peaks near 11.8, 12.6, and 12.7 ppm were all identified as G’s from their $^{15}$N chemical shifts. With 5 potential G’s in the bulge and stem region and only 4 new imino protons resonances, additional experiments are needed to identify which residues these imino protons belong to. Further, HSQC spectra cannot reveal the identity of base pair partner or the conformation of the base pair; for these, additional heteronuclear NMR experiments are required.

Of interest, a unique observation was made in the $^1$H-$^{13}$C HSQC spectra; complex formation anomalous peaks were found in the base region. After comparing the free RNA and complex spectra, it was found that several base peaks shift upon complex formation. The ano-
lous peaks were a result of incomplete complex formation; both free and bound RNA peaks were observed. This supplies an alternative method for determining complete complex formation. This is especially useful for D$_2$O based samples, since previous methods relied on imino proton observations.

### 4.3.3 Assignment Methodology

By utilizing $^{13}$C and $^{15}$N isotopically enriched RNA samples, several avenues of experiments exist in order to elucidate the type and identity of the imino protons and other key nuclei as well as gathering data for an overall structure of the RNA-ZF complex. Indeed, with all of the potential experiments available to NMR spectroscopy, it is theoretically possible to identify nearly every resonance of a nucleic acid. (22) This, however, would result in an extended use of precious NMR time. Although precautions are taken upfront to reduce the likelihood and impact of RNase contamination, any unanticipated contamination would further limit the time available for study. As such, it is critical to have an overall assignment strategy which will give the essential data needed for unambiguous assignment without poorly utilizing the time involved in gathering data for “un-needed” resonances.

In order to effectively identify and assign base and sugar spin systems along with sequential residue assignments, a systematic approach with alternative assignment schemes must be adopted. “Entry points” must be first identified and tied to their spin systems. These base and sugar spin systems must then be correlated to one another and then overall residue spin systems can be sequentially correlated. This allows for multiple methods for sequential assignment. For example, imino protons are easily identified in 1D imino proton spectra and can
be identified as G’s or C’s from HSQC spectra. Sequence identity of these can often be derived from NOESY spectra. However, gaps in the imino proton NOESY connections will undoubtedly be encountered in the bulge region thus not allowing for a NOESY sequential assignment of the imino protons in this region. It would prove useful to identify the base spin system associated with each imino proton. With this correlation, standard NOESY base to H1’ pathways could then be used to determine the sequential assignment of the correlated imino protons. In areas of extreme pathway interruption where both H1’ to base and imino NOESY pathways are discontinuous, through bond correlation of phosphorous to sugar spin systems utilizing the phosphodiester linkage can establish unambiguous sequential assignments. This would aid in the identification of any ambiguous or missing NOESY pathway assignments. For the RREIIB-Tr system, it is anticipated that discontinuous pathways will be encountered in the bulge region as well as the upper stem region due to the flipped out A68 base. These methods will prove to be invaluable.

Along with sequential assignments, identifying the base pair partners and orientation is critical. 1D and HSQC imino proton spectra can identify which bases have base pairs involving an imino proton, but they do not indicate the type of base pair formed or the identification of the base pair partner or the hydrogen bonding acceptor (ie nitrogen or carbonyl). In addition, these spectra cannot distinguish a hydrogen bond between an imino proton and another base or a protein sidechain. Also, determination of the other nuclei involved in hydrogen bonds (amino protons and carbonyls, for example) can clarify the orientation of the base pairs. Unusual base pair formation with amino protons is possible. For example, in the bulge region of RREIIB-Tr, one possible base pair formation is G48:G70. In the rev peptide complex structure,
this base pair is formed, and involves the hydrogen bonding of each G’s imino protons to carbonyls; this results in an intense crosspeak in the NOESY imino-imino region. (4) In our system, such a crosspeak is not observed. (20) If a G48:G70 base pair is formed, another potential conformation is an N1-carbonyl, N7-amino base pair, which only involves the imino proton of one of the G’s. When involved in a hydrogen bond, the amino protons are no longer degenerate and result in the appearance of two proton resonances for the amino group. Correlation of the amino protons to bases would, in this case, prove useful. Similar arguments can be made for base pairs involving imino protons and carbonyls; upon involvement in a hydrogen bond, the carbonyl chemical shift is changed. By mapping these changes, it is possible to define the conformations of the various new basepairs formed upon ZF binding.

4.3.4 Testing and Parameterizing Thru-Bond NMR Experiments

Many of the different nucleic acid pulse programs were evaluated and tested. A 300 μM $^{13}$C and $^{15}$N enriched RREIIB-Tr sample was used in conjunction with 10 mM AMP ($^{13}$C and $^{15}$N) sample in D$_2$O as well as a 3.75 mM GMP / UMP / CMP ($^{13}$C and $^{15}$N) sample in dry DMSO in order to observe exchangeable protons. Below are the specific experiments tested separated by category.

Identification of base pairing: In smaller systems, quite often a well resolved NOESY spectrum will help identify imino protons through imino-imino and imino-base contacts. Although these techniques will be utilized, through-bond experiments will help resolve ambiguous assignments due to overlap as well as determination of base pair partners.
• **HSQC:** Standard HSQC spectra can identify the nitrogen chemical shift of the imino proton involved in the base pair. With this, G resonances can be distinguished from U resonances. This technique was used with the free and bound RREIIB-Tr sample. (Figures 4-7 and 4-8) This does not, however, give the identity of the opposite base (C or A). In addition, multiple version of the pulse program were evaluated to test for effectiveness with the free vs. bound RREIIB-Tr; these included normal hsqc, echo-anti-echo, signal enhanced, and TROSY (Transverse Relaxation Optimized SpectroscopY) version.

• **HNN-COSY:** This experiment correlates the imino proton to the base pairing nitrogen across the hydrogen bond utilizing $^{2}J_{NN}$ couplings. (23) Based on chemical shift analysis, the types of base pairings can be determined (G-C vs A-U). This experiment was tested on the free RREIIB-Tr sample with each of the imino proton base pairs identifiable. Of note, G:U base pairs involving both imino protons can not be observed in this spectrum because of the lack of involvement of a hydrogen bond accepting nitrogen. Also, any hydrogen bond that is formed between the imino proton and a carbonyl will not be observed. (Figure 4-10)

• **H(N)CO:** Correlates imino protons to intra-nucleoside carbonyl carbons in guanosine and uracil (i.e. imino proton to the same base’s carbonyl). (24-26) Carbonyl chemical shifts are shifted if involved hydrogen base pairing; carbonyl “finger-printing” before and after binding can help identify any unique base pairing that may be occurring within the bulge region. This experiment was tested with a sample containing GMP and CMP in DMSO; well resolved carbonyl resonances for GMP’s C2 and C6 as well as UMP’s C2 and C4 were observed. Interestingly, varying the delay $1/4J_{NCO}$ resulted in an ability to “fine-tune” the spectra, which may prove useful when collecting the data. (Figure 4-11)
Residue Sequential Assignment and Sugar Proton Spin System Correlation: In RNA, the H2’ and H3’ sugars are quite often overlapped; expansion into the carbon dimension may allow for assignment of some of these resonances. Identification of the sugar spin systems not only allows for the sequential assignment of RNA through standard base to sugar NOESY pathways, but also can help identify any unique sugar contacts between the RREIIB-Tr and ZnF upon complex formation as well as the potential of defining some sugar puckerings.

- **HCP-HCCH TOCSY:** Gives sequential assignments by following pathways from $1'_n$ to $1'_{n+1}$ via $1’-3’-P-4’-1’$ pathways. (27) H1’ chemical shifts are isolated from other ribose protons. This can be used in conjunction with typical NOESY base to sugar ‘walks’. Unfortunately, highly concentrated samples of labeled DNA or RNA oligonucleotides are required for this experiment. The 300 μM RREIIB-Tr sample did not generate a interpretable spectrum. An alternative to this experiment is the simplified version, HCP, which correlates 3’-P-4’ pathways. In conjunction with the HCCH-TOCSY, as described below, 3’ to 1’ correlation is a possibility. (Figure 4-12)

- **HCCH-TOCSY:** This experiment correlates the sugar spin system. (28) The pulse program was successfully tested with the 300 μM RREIIB-Tr sample. Although overlap was encountered in several areas, many of the individual spin systems could be identified. This experiment will prove valuable in identifying any unusual base to sugar contacts or the establishment of sugar puckering for any well-separated spin systems. (Figure 4-13)
Correlating base spin systems to sugar spin systems: In the event that base to H1’ NOESY resonances are overlapped or non-contiguous, through-bond correlation of base and sugar spin systems will be critical.

• **HCN:** This experiment correlates the H1’ to N1/N9 and H6/H8 to N1/N9. (29) Combined, these will allow for base to sugar spin system correlations. The pulse program was tested with the 300 μM RREIIB-Tr sample as well as the monophosphate samples. Many of the RREIIB-Tr sugar and base spin systems were successfully correlated using the HCN spectra, however signal overlap in all dimensions was encountered in some areas thus potentially limiting the use of this pulse program on fully labeled samples. (Figure 4-14)

Base Spin System Correlation: In the absence of available NOESY data from noncontiguous pathways, through bond correlations will be critical to for assignment of isolated protons. In addition, the constant time HMQC experiment will simplify the differentiation of H6 and H8 protons.

• **Constant Time $^1$H-$^{13}$C HMQC (ctHMQC):** Although similar to the HSQC experiment, the constant time option removes f2 dimension homonuclear couplings (ie $^1J_{CC}$). (30) A side benefit of the experiment is the ability to distinguish different categories of carbons via phasing. When a constant time of $1/J_{CC}$ is chosen, carbons with zero or two carbon neighbors will exhibit an opposite phase as those with one or three neighbors thus giving the ability to distinguish H8/C8 resonances (i.e. a sequence of N9-C8-N7) from H6/C6 resonances (a se-
sequence of N1-C6-C5), which both have similar carbon and proton chemical shifts. As seen in Figure 4-15, the opposite phasing is quite valuable in assignment of the base region. In addition, lack of f2 coupling simplifies the interpretation of the spectra. (Figure 4-15)

- **HCCH COSY:** This experiment correlates H2 to H8 in adenosine allowing for H8 to H1’ through bond correlation when the data is combined with the HCN experiment. (31) The experiment was tested with the 300 µM RREIIIB-Tr sample. The H8 and H2 protons are correlated through C4 and C5. Interpretation can be somewhat challenging as the phasing of the various resonances is dependent on which resonance is observed (ie C2 or C8) (Figure 4-16)

- **H(CC)NH-TOCSY:** correlates imino protons to H8 in guanosine as well as uracil imino protons to H6; this is critical experiment for identification of the new base pairs in the absence of traditional NOESY methods. (32) The experiment was successfully tested on the GMP/UMP sample in DMSO; the 300 µM RREIIIB-Tr sample was not sufficiently concentrated for the pulse program to be successful. (Figure 4-17)

- **H6/H5(C4N)H:** This experiment correlates the amino protons to H6 in cytosine. (33) The experiment was tested with a GMP / CMP sample in DMSO with successful identification of the CMP amino protons. (Figure 4-18)

Although many other pulse programs are available, the selection was winnowed to identify the specific interactions anticipated in the RREIIIB-Tr sample.
4.4 Conclusion

The data collected gives additional insight into the nature of the bases involved in ZF binding. The 2AP study reveals a strong dependence on proper base pairing and ZF29 binding indicating that any alterations to the bulge structure inhibit ZF binding as is anticipated with RNA binding zinc fingers. The observed fluorescence changes of the 2AP substitutions may be the result of either the normal reorganization of the bulge region upon binding or the rearrangement of an altered structure to better match the needed ZF binding structure. The previously solved solution structure of the bound ZF29G29R as compared to the unbound solution structure indicate that very few changes are occurring within the ZF’s tertiary structure or hydrophobic core. (18, 20) Regardless of the causes of the 2AP fluorescence changes, the data provides additional evidence that the RREIIB-Tr changes its conformation to accommodate the binding of the ZF29G29R ligand with no significant reorganization of the zinc finger.

The heteronuclear NMR data collected thus far gives key details on the identity of the additional imino protons observed on binding. Sequence analysis of the bulge region implicated G’s as the predominate source of the additional imino proton signals, yet previous 3MU substitution experiments suggested U72 as a source of one of the imino proton resonances. Although the HSQC data rules this out as an option, it does not negate U72’s involvement in binding. The observed ZF bound imino proton spectrum of the U72-3MU substituted RREIIB-Tr is broadened considerably compared to the non-mutated RREIIB-Tr. (Figure 4-19) U72’s role in binding is currently unknown but clearly necessary for stable binding. In addition, the use of the $^{13}$C chemical shifts as an indicator for complex formation will eliminate the need for unnec-
ecessary solvent changes to H$_2$O thus reducing the potential for inadvertent introduction of RNase contamination.

The protocols and methodology developed will be critical for the elucidation of an NMR solution structure of the RREIIB-Tr / ZF29G29R complex. NMR spectra of in-house transcribed RNA exhibits sharper line shapes upon binding to the protein as compared to the Dharmacon RNA. (Figure 4-9) The only differences between the two sources of RNA are the in-house RNA’s inclusion of a 5’ terminal phosphate group and the lack of protection groups that must be removed. The sharper peaks will be of value when analyzing the data. When testing the NMR experiments, utilizing a low concentration RREIIB-Tr sample in conjunction with labeled monophosphates, multiple versions of existing experiments could be evaluated and parameterized as needed. This procedure could easily be implemented for other instances where pulse program testing and adaptation is needed when the target molecule supply is limited. The vast majority of the experiments were successfully tested. Several of the experiments on the low concentration RREIIB-Tr sample were completed with encouraging results (HNN-COSY, HCN, CT-HSQC, and HCCH-TOCSY). Many of the experiments involving imino protons required higher concentrations, utilizing monophosphates in dry DMSO allowed for evaluation with quite promising results. Unfortunately, highly concentrated isotopically enriched RNA or DNA was not available for the evaluation of the $^{31}$P sequential experiments. For the complex structure, however, the concentration of the isotopically enriched RNA will be > 1 mM.

Although through bond correlations are needed in order to solve the solution structure of the RREIIB-Tr / ZnF complex, the possibility of signal overlap in a 2nd or 3rd dimension can not be ruled out. For example, in the data collected thus far, the nitrogen, carbon, and proton re-
regions of the HCN spectra are overlapped in a few areas, which makes unambiguous assignments impossible in some cases. To circumvent this issue, it may be necessary to strategically label certain categories of residues (i.e., G’s or C’s only). This will help reduce the amount of overlap in crowded regions of the spectra and simplify the identification of resonances. If ambiguity persists in proton rich regions, perdeuterated RNAs are an option in that they will reduce the number of protons in the spectra. (34) With parameterization of the NMR experiments completed upfront, a strategy in place for the NMR collection and analysis, and optimization of RNA transcription and purification protocols, data collection on prepared RREIIB-Tr/ZF29G29R samples can be expedited.

4.5 Methods

Isotopically enriched RNA was transcribed and purified as discussed in chapter 5.2. ZF29 and ZF29G29R were expressed and purified as previously described. (18-20) NMR buffers for the RREIIB-Tr RNA consisted of 10 mM sodium phosphate and 50 to 100 mM NaCl. The $^{13}\text{C}^{15}\text{N}$ isotopically enriched monophosphates studied in D$_2$O were in 10 mM sodium phosphate and 100 mM NaCl. Monophosphates studied in dimethyl sulfoxide-d6 (DMSO) were lyophilized twice in order to remove any remaining H$_2$O. NMR experiments were conducted on a Bruker Avance 600 MHz NMR equipped with a 5 mm QXI probe, $^1\text{H}\{^{13}\text{C},^{15}\text{N},^{31}\text{P}\};$ HSQC and $^{13}\text{C}$ T$_1$ experiments were conducted on 500 MHz Avance systems equipped with a 5 mm TXI cryoprobe $^1\text{H}\{^{13}\text{C},^{15}\text{N}\}$.

Steady state solution fluorescence was completed in a buffer consisting of 10 mM sodium phosphate, 100 mM NaCl and 200 µM beta-mercaptoethanol. Excitation and emission for
the 2AP substitutions were 310 and 380 nm respectively. High ZF29 concentrations resulted in a non-specific binding-type curve in the emission spectra; this was adjusted for by fitting a linear line to the non-specific binding region.
Table 4-1  Thermodynamic results of single point mutations.

Thermodynamic data was obtained by curve fitting the ITC binding isotherm in Origin 7.0 software. Binding affinities are expressed as dissociation constants (K_d). The errors reported are also from curve fitting to a 1:1 site binding model. The binding free energies were calculated from ΔG=-RT ln (K_d) and the entropies from ΔS=(ΔH − ΔG)/T. Error in ΔG and ΔH were 0.02 kcal mol⁻¹ and 0.04 kcal mol⁻¹, respectively. (19)

<table>
<thead>
<tr>
<th>Protein</th>
<th>n</th>
<th>K_d (nM)</th>
<th>ΔG (kcal mol⁻¹)</th>
<th>ΔH (kcal mol⁻¹)</th>
<th>ΔS (cal mol⁻¹ K⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZF29</td>
<td>1.02</td>
<td>149 ± 21</td>
<td>-9.3 ± 0.1</td>
<td>-7.2 ± 0.1</td>
<td>7.2</td>
</tr>
<tr>
<td>ZF29H6A</td>
<td>0.94</td>
<td>207 ± 25</td>
<td>-9.1 ± 0.1</td>
<td>-5.3 ± 0.1</td>
<td>13</td>
</tr>
<tr>
<td>ZF29H6K</td>
<td>1.01</td>
<td>95 ± 29</td>
<td>-9.6 ± 0.2</td>
<td>-4.3 ± 0.1</td>
<td>17.7</td>
</tr>
<tr>
<td>ZF29R12A</td>
<td>0.98</td>
<td>266 ± 10</td>
<td>-9.0 ± 0.0⁺</td>
<td>-5.7 ± 0.0⁺</td>
<td>11</td>
</tr>
<tr>
<td>ZF29N15A</td>
<td>1.09</td>
<td>322 ± 60</td>
<td>-8.9 ± 0.1</td>
<td>-8.2 ± 0.2</td>
<td>2.3</td>
</tr>
<tr>
<td>ZF29D16A</td>
<td>1.00</td>
<td>450 ± 45</td>
<td>-8.7 ± 0.1</td>
<td>-8.0 ± 0.2</td>
<td>2.4</td>
</tr>
<tr>
<td>ZF29N21A</td>
<td>0.94</td>
<td>232 ± 40</td>
<td>-9.0 ± 0.1</td>
<td>-6.7 ± 0.2</td>
<td>8.1</td>
</tr>
<tr>
<td>ZF29N21Q</td>
<td>0.94</td>
<td>80 ± 31</td>
<td>-9.7 ± 0.2</td>
<td>-7.8 ± 0.2</td>
<td>6.3</td>
</tr>
<tr>
<td>ZF29G29R</td>
<td>0.99</td>
<td>57 ± 6</td>
<td>-9.9 ± 0.1</td>
<td>-5.5 ± 0.1</td>
<td>14.6</td>
</tr>
<tr>
<td>ZF29N21QG29R</td>
<td>1.06</td>
<td>35 ± 10</td>
<td>-10.2 ± 0.2</td>
<td>-7.6 ± 0.2</td>
<td>8.6</td>
</tr>
</tbody>
</table>
Figure 4-1  A) Sequence and secondary structural prediction of the REV Response element. The IIB stem loop is outlined. (2)  B) Sequence and secondary structure of the RREIIB-Tr RNA; upper stem, bulge, and lower stem regions of the RNA are identified.
Figure 4-2 Cartoon of the late stage transport of unspliced HIV RNA. Red hexagons represent the Rev protein, the blue box represents the RRE region of the HIV RNA, and the straight black line represents the unspliced HIV RNA.
Figure 4-3 Cartoon representation of the ZF29 protein. (pdb ID# 2AB3) In red is the alpha helix and in yellow is the beta sheet. The silver sphere represents the Zn$^{2+}$ and the amino acids side chains with ball and stick representations are the zinc coordinating His 23, His 27, Cys 5, and Cys 10 residues.
Figure 4-4  Imino proton spectra of the RREIIB-Tr RNA before (bottom) and after (top) binding with ZF29. Circled in red and shaded in orange point out new peaks formed upon binding. Peak assignments are detailed. The inset is the sequence and secondary structure of the RREIIB-Tr RNA; circled in red are base the observed formed base pairs; the red box outlines the bulge region of the RNA.
The interaction between Rev and RREIIB has been characterized by biochemical, mutational and structural methods. NMR structural studies have utilized an RREIIB analog RREIIBTR and the Rev peptide, a 17 amino acid arginine-rich motif (Rev34–50), to demonstrate that binding occurs in the major groove of the RNA.

Rev peptide–RREIIBTR interaction induces the formation of two new purine-purine base pairs (G-G and G-A). These base pairs are formed in the bulge in the stem–loop IIB and are critical to Rev binding.

An extensive list of strategies for inhibiting the pathogenesis of the virus has been documented. To interfere with the essential RRE–Rev interaction, these approaches use RNA-based strategies such as anti-sense RNA, RNA decoys, RNA aptamers, ribozymal siRNA, protein-based strategies that involve transdominant negative proteins, chimeric nucleases, intracellular antibodies, peptides, and small organic compounds.

Earlier, we demonstrated that C2H2 zinc finger proteins (znfs) ZNF29 and ZNF29G29R designed by phage display bind the same RNA bulge that Rev utilizes with nanomolar affinity.

The major groove is narrow in a regular RNA helix, while the RREIIB major groove is opened at the bulge containing the purine-purine mismatches. This facilitates access of α-helical recognition elements of Rev as well as znfs. Our current studies focus on understanding the energetics of the underlying molecular processes that facilitate RNA–znf recognition. We have studied the energetic perturbations to the znf–RNA system resulting from mutations on ZNF29 and the RNA target RREIIBTR by isothermal titration calorimetry (ITC). Effects of salt, pH, and temperature on the binding and the role of solvent were investigated. Our findings suggest strategies for enhancing the binding affinity of zinc finger proteins to the target RNA.

**Results and Discussion**

**Design of zinc finger and RNA mutants**

The sequence of the zinc finger has been optimized by phage display, which produced a

---

**Figure 4-5** A) Cartoon representation of the ZF29 protein with ball and stick representations of the single point mutations made for ITC experiments. Left and right represent different views of the same ZF29. B) A sequence mapping of the ZF29 single point mutations with residue numbering made for ITC experiments. (19)
Figure 4-6  A) Stick diagram of 2AP nucleoside. Blue = nitrogen, red = oxygen, cyan = carbon, white = hydrogen. B) Sequence and secondary structure of RREIIB-Tr RNA. In red boxes are the substitutions made. C) Fluorescence emission at 380 nm plotted against ZF29 concentration. Red line represents raw data; blue and green lines represent corrections made for non-specific binding-like curve observed at higher concentrations.
Figure 4-7 $^1$H-$^{15}$N HSQC spectrum of the free RREIIB-Tr RNA. Noted in blue are the typical guanosine (G) and uracil (U) $^{15}$N chemical shift resonances.
Figure 4-8 Overlay of the $^1$H-$^{15}$N HSQC spectra for the RREIIB-Tr bound and unbound with the ZF29 protein. In red is the bound spectrum, in black is the free spectrum. Blown up regions are shown on the right side of the figure. In the top blown up spectrum ($^1$H: 11.6 – 12.5 ppm and $^{15}$N: 145.5 to 150 ppm), the intensity has been increased to visualize the broad new resonances; arrows point out emerging resonances at 11.85 and 12.15 ppm upon ZF29 binding. The $^{15}$N chemical shifts indicate these are guanosine resonances. On the bottom is an expansion of the 12.5 – 13.5 ppm region; the arrow notes the broad resonance corresponding to two peaks at approximately 12.9 ppm. The $^{15}$N chemical shifts indicate these are guanosine resonances, one of which is the G76 imino proton.
Figure 4-9  Comparison of Dharmacon and in-house transcribed free and bound RREIIB-Tr RNA imino proton spectra.  A) ZF29 bound Dharmacon RNA (20)  B) ZF29 bound in-house transcribed RNA  C) free in-house transcribed RNA  D) free Dharmacon RNA. Chemical shift assignments are noted. (20)
Figure 4-10 HNN COSY. On the left are stick representations of the RNA bases with arrows representing magnetization pathways and circles representing observed resonances. On the right is a sample spectrum from the 300 µM RREIIB-Tr RNA. Typical $^{15}$N chemical shift resonances are noted on the left of the spectrum. Dotted lines connect the hydrogen bonded nitrogen and imino protons. Note the lack of connectivity for G77 and U43, which are base paired to one another and do not involve hydrogen bonding to a nitrogen acceptor.
Figure 4-11  H(N)CO. On the left are stick representations of guanosine and uracil with arrows representing magnetization pathways and circles representing observed resonances. On the right are four spectra of a GMP / UMP sample in DMSO. The four panels represent varying delays of D23, which is equal to $1/4J_{\text{NCO}}$. Varying this delay allows for some editing to be done to remove resonances (U-C2, G-C2, and G-C6) in case of spectral overlap. Coupling values for U N3-CO4 and U N3-CO2 are as follows: $U J_{\text{N3CO4}} = 10$ Hz, $U J_{\text{N3CO2}} = 19$ Hz. (22)
Figure 4-12  HCP. On the left is a stick representation the adenosine dinucleotide with circles representing observed resonances. Combined with an HCCH-TOCSY, this will give sequential assignments from sequential H1’ to H1’ resonances
Figure 4-13 HCCH-TOCSY. On the left is a stick representation the adenosine nucleotide with circles representing observed resonances. On the right are sample spectra from the 300 µM RREIIB-Tr RNA. The top spectrum represents all connected protons and carbons when focused on a specific H1’ proton. The spectrum on the bottom represents all connected protons to a C1’ resonance of the same H1’ proton. The dotted lines connect the proton resonances between the different views.
Figure 4-14 HCN. On the left is a stick representation of the adenosine nucleotide with arrows representing magnetization pathways and circles representing observed resonances. On the right are two spectra from the 300 µM RREIIB-Tr RNA. The left spectrum is focused on the base proton and the right spectrum is focused on the sugar proton of the same nucleotide. The dotted line connects the N1 connecting the sugar and base spin systems. The horizontal dimension is $^{13}$C, and the vertical is $^{15}$N.
Figure 4-15 cTHMQC. On the top are stick representations of guanosine and cytosine with arrows representing magnetization pathways and circles representing observed resonances. On the bottom is a spectrum from the 300 µM RREIIB-Tr RNA. Green resonances represent the adenosine and guanosine base H8 protons; the red are cytosine and uracil H6 resonances. Green and red are opposite phases.
Figure 4-16 HCCH-COSY. On the left is a stick of adenosine with arrows representing magnetization pathways and circles representing observed resonances. On the right are sample spectra from the 300 μM RREIIB-Tr RNA. Green and red represent opposite phases; the left and right panels are H2 and H8 proton resonances for the same base. Dotted lines connect the C5, C4, and C6 resonances of the base. The F2 dimension (horizontal) is the C2 and C8 resonances (left and right panels respectively).
Figure 4-17  H(CC)NH. On the left are stick representations of guanosine and uracil with arrows representing magnetization pathways and circles representing observed resonances. On the right is a sample spectrum from the GMP and UMP sample in DMSO. The F2 dimension (horizontal) is the imino proton region, the F1 dimension (vertical) is the base region. The peak at 11.5 ppm is for U; the peak at 10.8 ppm is for G.
Figure 4-18 H6C6/H5C5(C4)NH. On the left is a stick representation of cytosine with arrows representing magnetization pathways and circles representing observed resonances. On the right is a sample spectrum from the UMP and CMP sample in DMSO. The $^1$H dimension is the amino proton region, the $^{13}$C dimension is the C5 (90 – 95 ppm) and C6 (140 – 145 ppm) region.
Figure 4-19 Imino proton spectra of 3MU substituted U72 RREiIB-Tr RNA. A) free RNA B) ZF29 bound RNA C) ZF29 bound with non-substituted RNA. Blue diamonds point out the U66 and U45 resonances, red diamonds point out the additional imino protons present when bound to ZF29. (20)
4.6 References:


5 METHODS AND TECHNIQUE DEVELOPMENT

5.1 Supercooled Aqueous NMR Using Agarose Gels

Supercooled aqueous NMR can provide essential data for biomolecules, yet such information is cumbersome to obtain via capillaries. Agarose gels allow for NMR measurements under supercooled conditions; the gels are simple to prepare, invisible to NMR, and non-interactive with most biomolecules and organics. Here we demonstrate their use with nucleic acids, small organic molecules, and peptides with both 1D and 2D homonuclear and heteronuclear applications.

NMR spectroscopy is often limited by the physical constraints imposed by the solvent. Changing the solvent is typically not an option because this may change the structure or solubility of the solute. Studies of biological molecules and drug compounds often encounter this predicament since they are generally limited to an aqueous environment. However, low temperature NMR studies could provide important insights. At low temperature, the dynamics of biomolecules are slowed which can often reveal conformational equilibria. In addition, it can allow the detection of protons that are in exchange with the solvent (i.e. water). For small molecules, NOESY analysis is often hampered because of an unfortunate combination of field strength and correlation time which may result in either weak negative crosspeaks (as compared to the diagonal) or even null crosspeaks making structure determination of small molecules difficult. Macromolecules have much larger correlation times resulting in positive crosspeaks even at lower field strengths. For small molecules, moving to higher field strengths often circumvents these issues. Alternatively, the higher viscosity of water at lower tempera-
tures increases the correlation time for small molecules. This approach may be an alternative to moving to higher field strengths for NOESY analysis.

Supercooled aqueous NMR, with liquid temperatures significantly below 0°C can extend these phenomena beyond the usual constraints of water. This state can be achieved using small diameter capillaries for which temperatures as low -31°C have been reported. [1] Using a bundle of 1 mm glass capillaries placed inside a standard NMR tube, temperatures as low as -18°C can be reached. [2,3] Although capillaries can achieve supercooled conditions, their implementation can be cumbersome. In addition, the overall sample volume is smaller which results necessarily in lower sensitivity. To overcome this drawback higher sample concentrations are required which may then cause problems with aggregation and solubility. As a result, supercooled aqueous NMR using capillaries, although attractive, has not been widely adopted. [4]

An alternative technique to capillaries uses filtration and exhaustive centrifugation of a sample to remove free particulates in order to reach temperatures as low as -6°C in a standard NMR tube. [5] This technique is relatively labor intensive and requires considerable preparation of the sample.

In contrast, agarose gels are relatively simple to prepare and require little preparation of the sample. As reported by the Temussi lab, agarose gels allow for supercooled aqueous solution NMR and are invisible in NMR. They found that the aqueous environment in the agarose matrix is not crowded and yields similar results to free solution NMR. [6] Agarose is noninteractive with most biomolecules and organics; also, DNA, RNA, and protein sample recovery from agarose gels is well documented. [7,8] Despite this rather easy implementation, literature searches do not yield evidence that the method has been utilized. Here we expand on the Te-
mussi study to report the application of this technique with nucleic acids, small organic molecules, and small peptides, all in 1% agarose gels in both H2O and D2O. We demonstrate that this method promises a straightforward approach to supercooled aqueous NMR for a variety of systems.

Samples were mixed with agarose powder and heated in a water bath at 85 – 90°C to melt the agarose. The samples were mixed well, transferred to an NMR tube, and allowed to gel at ambient temperatures. Sample preparation was found to be critical in order to obtain sharp lines. For example, inhomogeneous samples resulting from partially unmelted agarose or incorporation of air bubbles will result in broadened NMR peaks and an inability to shim properly. If this happens, the sample can be re-melted in the NMR tube, mixed thoroughly, and re-gelled at ambient temperatures. All samples were prepared in 1% agarose unless otherwise noted in the text. Spectra were collected on an Avance 600MHz NMR with an 8 mm or 5 mm QXI probe.

Using water based, 1% agarose gels, temperatures as low as -13°C were routinely achieved before freezing; freezing points for 100% D2O samples, which has a melting point of +3.8°C, were approximately -9°C in 1% agarose gels (10 mM sodium phosphate, 100 mM NaCl, 1% agarose). All samples tested could reproducibly be measured at -10°C before freezing.

Changing the percentage of agarose to 2% had a negligible impact on the freezing point of the sample. Freezing the sample in agarose does not break the NMR tube, which allows for probing of freezing conditions without fear of either damaging or contaminating the probehead or loosening the sample. These are important concerns considering precious biomolecular samples and modern high sensitivity probeheads. Unlike capillaries, agarose gel temperatures could be re-
duced relatively quickly to -5°C and then dropped approximately 1°C per 10 minutes. In
comparison standard protocols for capillary temperature reduction is approximately 1°C per hour
when reaching temperatures below -5°C. [4]

Despite the agarose matrix, line shapes similar to free solution NMR were obtained as
evidenced by a 50 mM sucrose sample. As anticipated, line shapes broadened at lower tem-
peratures. (Figure 5-1A, middle spectrum). In addition, standard ¹H-¹³C natural abundance
HSQC experiment could be readily recorded at -7°C without interference from the agarose ma-
trix. The impact on the rotational correlation times of small molecules was explored via NOESY
spectra of 50 mM sucrose in D₂O suspended in 1% agarose. NOESY spectra collected at 17°C
resulted in negative crosspeaks, however, spectra collected at -7°C yielded positive crosspeaks
indicating that the W₀ relaxation pathway is dominant (Figure 5-1A). This simplifies the quanti-
tative use of NOESY data for structure determination of small molecules.

The effects of supercooled temperatures on conformational dynamics were probed to
show the existence of a weak base pair in an alpha anomic adenosine (αA) containing DNA
oligonucleotide (Figure 5-1B, top). In this duplex the thymine opposite the αA presents a broad
imino proton resonance which may escape detection. [9] As temperatures are lowered to -8°C,
base opening rates and solvent exchange are reduced resulting in a dramatic sharpening of the
thymine imino proton resonance (Figure 5-1B). This demonstrates the existence of a base pair
and also permits the determination of its connectivity via NOE experiments. This sample was
also used to test feasibility for low concentration samples in an agarose matrix. The DNA con-
centration was 30 μM; using jump and return solvent suppression, well-resolved spectra at -8°C
were obtained with 16k scans (Figure 5-1B). [10]
Reduced solvent exchange rates with labile protons were also tested using the tripeptide GGR, a small, disordered peptide. At ambient temperatures, the only amide proton resonance detected is that of Arg3; when temperatures were reduced to –8°C, Gly2 and Arg3 amide protons as well as Arg NH protons could be resolved at –8°C in both 1H and TOCSY spectra.

Our experiments were conducted with an 8 mm probe for convenience and comparison with capillaries. The larger sample volume provides an advantage for less soluble samples and to improve signal to noise. On the other hand the performance of larger diameter probes are inherently more sensitive to salt. The much more widely available 5 mm probes can also be used with this approach. One consideration that must be taken when utilizing this system is heat sensitive biomolecules. Samples must be stable to approximately 87°C to melt regular agarose; in low melting agarose, this is reduced to 65°C. Even at 65°C, use of agarose with some proteins or temperature sensitive systems is problematic. But, for systems that can tolerate these temperatures (i.e. small nucleic acids, organics), agarose suspension can be easily implemented and extend data to supercooled temperatures.

As demonstrated, suspending aqueous samples in 1% agarose is a simple and convenient method for reaching temperatures of -8 to -13°C. Sample shimming is straightforward yielding line shapes comparative to regular solution NMR, sensitivity is not impacted, the agarose matrix is invisible in both $^1$H and $^{13}$C spectra, standard NMR experiments can be employed, inadvertent sample freezing does not risk breaking the NMR tube, and sample extraction from agarose is well documented in molecular biology protocols. [7,8] For many biomolecules, just a few degrees of supercooled conditions can provide much needed data.
Figure 5-1  A) 1% agarose gels with 50 mM sucrose spectra in D$_2$O: top and bottom are phase sensitive NOESY spectra at -7 and +17 °C, respectively (16 scans, 512 increments, 90° shifted sine-squared function). Red and black peaks denote negative and positive peaks respectively. Middle spectrum is $^1$H at +17°C. B) CαAG oligonucleotide (30 μM oligonucleotide, 10 mM sodium phosphate, 100 mM NaCl, pH 6.8) Top panel is the sequence of CαAG oligonucleotide, bottom three spectra are the $^1$H NMR imino proton region (jump and return solvent suppression). [9] The resonance noted with an arrow is the thymine imino proton base paired with the αA.
5.1.1 References

5.2 RNA Transcription Protocols Utilizing T7 RNA Polymerase

Note, all amounts and target concentrations were optimized for the transcription of the 34 nt RREIIB-Tr RNA sequence. Sequence lengths as well as content will require optimization of MgCl₂ concentration.

5.2.1 Critical Components of the Reaction

T7 RNA Polymerase: Stock concentration should be approximately 0.7 – 1 OD/ml. Polymerase should be in the short-term use buffer. Optimization steps demonstrated that overall T7 polymerase reaction concentration had little effect on yields. It should be noted, however, that T7 polymerase reaction concentrations too high will result in some reagents crashing out of solution.

DNA template: Synthesized oligonucleotides must contain the T7 recognition site: 5’TAA TAC GAC TCA CTA TA3’. The transcription sequence should be on the 3’ side of the recognition site on the opposite strand. The transcription sequence is read 3’ to 5’ and the RNA is transcribed 5’ to 3’. Literature suggests that targeted RNA sequences should, if possible, initiate with a G (on the 5’ end of the RNA). A closed hairpin was found to be by far the optimum template geometry followed by an open hairpin (3:1 difference in yields). The double stranded template was found to have 20 fold less yield than the closed hairpin.
**NTP concentration:** Reaction yields were found to be nearly linear with NTP concentrations with a maximum of approximately 15 – 20 mM total NTP concentration.

**MgCl$_2$ concentration:** Reaction yields are highly dependent on MgCl$_2$ concentrations. MgCl$_2$ concentrations are given as a ratio to NTP concentration. Further, MgCl$_2$ concentrations should be optimized for each unique RNA target transcription. For the RREIIIB-Tr sequence, it was found that the maximum occurred around 3:1 (MgCl$_2$:NTP) with a steep decline in RNA yields occurring above that.

**Time:** Reaction yields plateau at 2.5 hours regardless of all reaction conditions.

Only MgCl$_2$ concentration needs to be optimized; all other reaction components (aside from NTP concentration) were found to have minimal impact on yields.

### 5.2.2 Precautions and Advice

- First and foremost: **AVOID RNase CONTAMINATION!** Wear gloves, clean ALL glassware with RNase Zap, utilize RNase free pipette tips and Eppendorf tubes, do NOT use “lab general” buffers and solutions, clean the pH meter well with ElectroZap, clean the FPLC thoroughly with RNase Zap before loading the RNA reaction, if possible work in an isolated area (ie fume hood).
• When combining reagents, all components (except for the BSA and T7 RNA polymerase) should be at room temperature; failure to do so can result in reagents crashing out of solution.

• The order in which reaction components are added should be followed; again, deviation will result in reagents (DNA template, some NTP’s, and BSA) to crash out of solution (a result of high concentrations during the initial addition steps).

• Once all components are added, the mixture should be mixed well; T7 RNA polymerase is in glycerol and will “settle” if not mixed with other components.

• All pipetters need to be calibrated in order to avoid reagent amounts being “out of balance” with one another.

• DTT will in time oxidize and has approximately a 6 month shelf life once put into solution. If it is found that reactions do not scale up in a linear manner, DTT deterioration may have occurred.

• The 10X buffer should be stored in the refrigerator.

• BSA and T7 RNA polymerase should remain on ice until added to the reaction.

• Salt to NTP ratio imbalance or inadvertent addition of other salts will have a tremendously negative impact on reaction yields.

• The transcription reaction can be initiated with GMP instead of GTP, thereby eliminating the triphosphate on the 5′ end of the RNA. Use an equimolar amount of GMP as GTP.

• It was found that the reaction is product inhibited (RNA, not magnesium phosphate precipitate or pyrophosphate buildup).
• Some RNase contamination is a by-product of the T7 RNA polymerase production. Although RNase Secure is added to the reaction to help minimize the impact, the reaction temperature is far below the optimum activity of RNase Secure. As such, initial ion-exchange purification (or alternate purification) must be completed the same day as the transcription reaction. If left in the reaction mixture, the RNA will degrade.

• The white precipitate during the reaction is magnesium phosphate. Addition of inorganic pyrophosphatase does eliminate the precipitate, however yields are not improved. The precipitate gives a visual cue that the reaction was successful.

• Note, if utilizing gel filtration in place of ion exchange for the initial RNA purification, gel filtration utilizing DI H₂O does not work; rNTP’s aggregate and elute with the RNA (interestingly this is not observed with dNTP’s and DNA). If choosing to use gel filtration, a phosphate buffer with NaCl should be used. The RNA elute can subsequently be desalted normally.

• During ion exchange purification, glycerol contained within the reaction will significantly increase the back pressure on the FPLC; the flow rate should be lowered in order to keep pressure below 2 MPa. The pressure typically falls just before NTP elution. If pressures are too high, consider utilizing gel filtration or ethanol precipitation (both alternative protocols are outlined below).

• Be sure to spin-down the sample before loading onto FPLC in order to avoid loading the precipitate onto the column.
• Although not observed in high quantities for the RREIIB-Tr transcript, n+1 and abandoned G-run transcripts are potential side products from the reaction. RNA yields should be run on a denaturing PAGE to ensure purity.

Reagents needed:

10X reaction buffer: 400 mM Tris, 60 mM MgCl₂, 100 mM DTT, 20 mM spermidine, pH = 8.1 (store in 4 °C)

DNA Template should be ~3 μM

5.2.3 Optimization Protocol

20 μL reactions are sufficient for MgCl₂ optimization. Premix an equimolar amount of rNTP’s and rGMP; each reaction uses a total of 5 μL NTP/GMP mixture (assuming 100 mM concentration on each). The general reaction make-up is outlined in table 1 (be sure to add components in the order listed):

Reactions should be run 3 hours at a constant 37 °C. Quantify reaction yields via SYBR Green II fluorescence on a 15% denaturing PAGE (or other % gel depending on RNA length). No reaction material purification is needed; 2 μL of the reaction mixture should be sufficient for quantification.

Do to the steep drop-off in RNA yields with high salt concentrations and the relatively low improvement within a relatively large salt concentration range, the lower end in salt concentrations should be utilized.
5.2.4 Production Protocols

500 µL reaction sizes give high yields, require few reactions for NMR quantities, yet minimize potential RNase contamination impacts on RNA yields. RNA yield in a 500 µL reaction range from 6 OD’s to 8 OD’s (for RREIIB-Tr production). Again, transcription length will have a tremendous impact on overall yields. Literature suggests that longer transcripts will result in higher yields and efficiency; binding and initiation are the rate determining steps.

The general reaction makeup for the RREIIB-Tr transcription reaction is outlined in table 2 (be sure to add components in the order listed). If deviation from general reaction makeup is needed due to volume constraints, consider increasing the stock concentration of BSA and RNase Secure, lowering the T7 RNA polymerase amount slightly, and lowering the DNA template amount slightly; increasing the stock concentration of MgCl₂ should be avoided due to accidental overage in overall MgCl₂ reaction concentrations. As an example, for ¹³C, ¹⁵N labelled RNA production, rNTP concentrations ranged from 55 mM to 100 mM; in order to compensate for volume constraints, the BSA and RNase Secure stock concentrations were doubled, T7 RNA polymerase addition was lowered to 150 µL, and the DNA template addition was lowered to 45 µL. Yields for the reaction averaged 7 OD / 500 µL reaction.

5.2.5 Purification

Purification must happen on the same day as the transcription reaction. Briefly, initial purification is completed on a Mono Q ion exchange column utilizing a tris buffer with a NaCl gradient. This method allows for separation of NTP’s and RNA from other reaction components. The RNA elute can then be desalted utilizing a normal HiTrap desalting column. MAKE
SURE ALL COLUMNS AND THE FPLC ARE SANITIZED WITH RNase ZAP BEFORE LOADING THE SAMPLE! Utilizing this methodology reduces the number of steps involved during purification and thereby reducing the potential for RNase contamination. Listed below are the main purification protocols as well as alternative methods.

**Mono Q:**

Loading buffer: 20 mM tris, pH = 9.0

Elution buffer: 20 mM tris, 1 M NaCl, pH = 9.0.

The gradients of interest are: 20 – 24% NTP’s are eluted, 48 – 53% RNA is eluted.

Flow rate: up to 1 ml/min, but watch pressure!

Dilute the sample 3 fold with DI H₂O, mix well, and spin down the precipitate before loading. Due to pressure buildup, load only 0.7 mL of the diluted reaction mixture at a time. Pressure will increase when the sample is injected (possibly up to 3X increase); reduce the flow rate to maintain a pressure below 2 MPa.

**Desalting:**

Desalt utilizing the 4 chamber HiTrap desalting column (each chamber has a 5 mL V₀) in DI H₂O.

Run a test injection, but optimum amounts for the RREIIB-Tr production run were 2 mL load amounts and a flow rate of 1.5 mL/min.
When completed, lyophilize and bring up in solvent of choice. Verify purity on a 15% denaturing PAGE.

5.2.6 Alternative Purification Methods

Superdex 75 column: Size exclusion. Utilize a 10 mM NaP, 100 mM NaCl, pH = 7.0 solution for purification. As stated before, utilizing DI H₂O results in aggregation of the rNTP’s and elution of RNA and NTP at once. Flow rate should be 0.4 mL/min, elution of RNA occurs around 11 to 12 mL. Cons: time consuming and high volume elution.

Methods to isolate RNA from much of the reaction components prior to FPLC purification methods are as follows:

Ethanol Precipitation: A 3:1 ratio of ethanol to reaction volume was utilized. The mixture should be held at -20 °C for a minimum of 30 minutes and centrifuged for 15 minutes at 12,000 x G. The supernatant was removed and the pellet was rehydrated with H₂O.

Phenol / chloroform extraction: (Phenol is highly toxic; it should be handled with nitrile gloves in a fume hood) A 5:1 mixture of phenol and chloroform should be saturated in an acetate buffer at pH 4.64. Due to potential damage to the pH meter, the pH is measured by mixing 200 µl of the phenol chloroform mixture with 800 µl of methanol and 1ml of water. For RNA extraction, equal amounts of the phenol / chloroform mixture and chloroform are added to the reaction mixture. The mixture is mixed well and allowed to separate. In an acidified phenol /
chloroform environment, the RNA will remain in the aqueous layer while the DNA and proteins will migrate to the organic layer. The aqueous layer containing the RNA was then removed and desalted to remove the remaining phenol.

These methods (ethanol precipitation and phenol / chloroform extraction) did not improve yields for the RREIIB-Tr reactions and proved to be bothersome when loading onto the Superdex column due to phenol contamination. For the RREIIB-Tr RNA product, the previously discussed purification scheme with ion exchange followed by size exclusion liquid chromatography proved sufficiently pure for ITC and NMR studies and were comparable to 3rd party purchased RNA.
Table 5-1  Optimization Reaction Setup

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>10X buffer</td>
<td>2 µL</td>
</tr>
<tr>
<td>rGMP/rNTP mixture</td>
<td>5 µL</td>
</tr>
<tr>
<td>MgCl₂ (300 mM)</td>
<td>varied from 0 to 4.2 µL</td>
</tr>
<tr>
<td>DI H₂O</td>
<td>varied per MgCl₂ addition</td>
</tr>
<tr>
<td>BSA (1 mg/mL)</td>
<td>1 µL</td>
</tr>
<tr>
<td>RNase Secure</td>
<td>0.8 µL</td>
</tr>
<tr>
<td>T7 RNA Polymerase</td>
<td>5 µL</td>
</tr>
<tr>
<td>DNA</td>
<td>2 µL</td>
</tr>
<tr>
<td>Total reaction volume</td>
<td>20 µL</td>
</tr>
</tbody>
</table>

Reactions should be run 3 hours at a constant 37 °C.
Table 5-2  Production Reaction Setup

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>10X buffer</td>
<td>50 μL</td>
</tr>
<tr>
<td>DI H₂O</td>
<td>0 μL (or as determined from optimization)</td>
</tr>
<tr>
<td>rATP (100 mM)</td>
<td>25 μL</td>
</tr>
<tr>
<td>rGTP (100 mM)</td>
<td>25 μL</td>
</tr>
<tr>
<td>rCTP (100 mM)</td>
<td>25 μL</td>
</tr>
<tr>
<td>rUTP (100 mM)</td>
<td>25 μL</td>
</tr>
<tr>
<td>rGMP (100 mM)</td>
<td>25 μL</td>
</tr>
<tr>
<td>MgCl₂ (300 mM)</td>
<td>55 μL (or as determined from optimization)</td>
</tr>
<tr>
<td>BSA (1 mg/mL)</td>
<td>25 μL</td>
</tr>
<tr>
<td>RNase Secure</td>
<td>20 μL</td>
</tr>
<tr>
<td>T7 RNA polymerase</td>
<td>175 μL</td>
</tr>
<tr>
<td>DNA template</td>
<td>50 μL</td>
</tr>
<tr>
<td>Total reaction volume</td>
<td>500 μL</td>
</tr>
</tbody>
</table>

Reactions should be run 3 hours at a constant 37 °C.
Figure 5-2  DNA Templates for RNA Transcription Reactions
Figure 5-3  RREIIIB-Tr DNA template
5.2.7 References

5.3 AMBER Parameterization for Novel DNA Residues

AMBER [Cornell, et al] calculations of biomolecular systems are completed through the use of the AMBER force-field, which is a molecular mechanics systems based on empirical observations (Formula 5-1). For systems composed of the canonical residues (for nucleic acids: adenosine, cytosine, guanosine, and thymine / uracil), the PARM99 force-field has been shown to consistently reproduce close to experimental results. Although the parm99 force-field is effective for many biological-based system calculations, it is limited to canonical residues. For systems containing novel residues and components, the AMBER force-field must be appended. Specifically, the point charges and force constants for each of the new force-field components must be supplied. Derivation of these should be completed at a consistent level as the rest of the force-field parameters; under-calculation of novel residue parameters would result in a poorly described system and over-calculation would be a waste of resources as the system would be defined by the lesser developed components (ie the rest of the force field).

Formula 5-1. AMBER Force Field:

\[ U(r) = \sum K_1(r-r_{eq})^2 + \sum K_2(\theta-\theta_{eq})^2 + \sum \frac{V_n}{2} (1+\cos[n\varphi-\gamma]) + \sum \frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \sum \frac{q_i q_j}{\varepsilon R_{ij}} \]

In an effort to augment the AMBER force field, Wang, J. et al. (2004), developed the General AMBER Force Field (GAFF). The GAFF force-field contains many unique bonds, angles, and torsion angles to address the behaviour of small organic molecules and ligand-biomolecular interactions. To somewhat automate the process of atom type and other parameter defini-
tions, the ANTECHAMBER [Wang, J. et al. (2005)] module of AMBER was developed. This module “scans” the molecule of interest, analyses atom and bond properties (i.e. aliphatic vs. aromatic) and generates a listing of GAFF or AMBER parameters for addition to the force field. In theory, these parameters can be simply appended to the normal force field via two files known as the .frcmod and .off files. In practice, however, over reliance on the ANTECHAMBER module can prove to be extremely problematic.

In addition to the atom, bond, angle, and torsion parameters, the atomic point charges must be calculated. Within the AMBER force-field, point charges were calculated using the HF 6-31G* restrained electrostatic potential (RESP) charge model. For GAFF however, many of the parameters were developed using the AM1-bond charge correction (BCC) model in which Mulliken charges are calculated at the AM1 semi-empirical level, and then generates RESP-like charges through bond charge distributions. This latter method (AM1-BCC) is much less costly in computational time as compared to HF 6-31G* RESP calculations (minutes vs. hours or days) and has been shown to produce similar results as full RESP calculations. For efforts that involve a large number of unique molecules, the AM1-BCC calculation is ideal, but for solitary novel residues, the RESP charge derivation is preferred in that it better matches the rest of the force-field. Further, with the universal base residue, there was a fairly large deviation between the RESP calculation and the AM1-BCC calculation.

This document outlines the software components and procedures for development of AMBER parameters for two novel nucleic acid residues: the borano-phosphate linkage and the universal base.
5.3.1 **Charge Derivation:**

If time permits and the parm99 force field is being used, charges should be complete using the RESP charge model. Although implementation can be somewhat “tricky”, matching to the rest of the parm99 force-field (especially for components of biopolymers) is ideal. If not possible, then AM1-BCC charges can be used.

**RESP charge derivation: software components and overview:**

RESP charges are derived using a QM based calculation program (Gaussian03) and AMBER’s RESP module. The starting structure should be well minimized using the HF 6-31G* level of theory. This structure is then supplied to Gaussian where MEP charge calculations are completed at the same level of theory. These charges and the structure are then fed to AMBER’s RESP module where RESP charges are calculated. It has been shown that simple 3-D reorientation of the molecule will actually generate somewhat different charges. Because highly reproducible results are required, the molecule should be reoriented multiple times with QM MEP charges and RESP charges calculated for each reorientation.

Each program, unfortunately, uses a different file format, so inter-conversion between the programs can be tedious. The R.E.D. script somewhat automates the charge calculation, RESP calculation, and reorientation process. Table 5-3 details the software components needed for RESP calculations. The R.E.D. script takes the minimized structure with the number and types of reorientations desired and automates the QM and RESP calculations, and in the end generates an ending mol2 file which contains the final RESP charges for the system. This file can then be fed into ANTECHAMBER for atom, bond, angle, and torsion type parameterization.
It is important to note that all components (with the possible exception of BABEL and VMD) must be installed on the same machine; RED calls on both AMBER and Gaussian03. Further, it should be noted that RED error prompts are a bit cryptic; individual components should be tested separately in order to determine their functionality.

The R.E.D. ver 3.0 script is free for educational use. Their website is: [http://q4md-forcefieldtools.org/RED/](http://q4md-forcefieldtools.org/RED/). The installation and supporting documentation is well written; utilization of their tutorial is quite effective in learning the overall software control.

OPENBABEL 2.2.0 is freeware for the sole purpose of file conversion between different visualization file formats. The website to download OPENBABLE is: [http://openbable.org](http://openbable.org). Installation is simple and well documented. Only one file conversion is needed for RED usage: conversion from a MOL2 to a PDB file format.

**RED usage (borano-phosphate linkage as an example):**

It is important to choose a molecular model that adequately represents your target system but is not overly complex. The borano-phosphate linkage is a prime example of this concept. Early calculations for this linkage included the linkage attached to a sugar, the linkage attached to a sugar with a base, and the linkage attached to two sugars with simplified bases on each. This approach proved to be very problematic. Aside from being extremely computationally expensive, minimized structures were far from canonical representations and charges were not what were anticipated. When restraints were added to the structure other parts did not minimize well and unusual conformations were observed; ending charge derivation results were not reproducible due to the complex nature of the starting structures. Further, computa-
tion times for minimization began to reach multi-day calculations and several calculation efforts failed all together.

It was originally anticipated that the charge distribution for the borano-phosphate linkage would not overly impact the charges outside of the phosphate group. This was observed in several of the earlier calculations. To this end, the system was modeled as a borano-phosphate dimethyl group [figure 3], which yielded acceptable and reproducible results.

The starting borano-phosphate dimethyl system was gradually minimized until the HF 6-31G* level of theory was used (MMFF → ST0 → HF 3-31G → HF 6-31G*). The B-P distance was restrained to the crystal structure value of 1.91 Å. [Summers, J.S., et al. (1997)]. Once completed, a PDB file is needed for input into Ante-R.E.D. Within Gaussian03, the structure is saved as a MOL2 file. OPENBABEL is then used to convert this file to a PDB. Ante-RED is then used to generate the needed P2N file for RED input. The P2N file was edited to include the proper -1 charge, multiplicity of 0, and 14 reorientations. It is imperative that the overall charge be entered. The altered p2n and Gaussian03 log files were used as inputs for the RED script. The calculations were completed within approximately 30 minutes.

The ending values were within remarkable agreement with the test dimethyl phosphate calculation with a slightly different overall distribution around the phosphate group. What is encouraging about this calculation specifically is the agreement in charges between the 05’ and 03’ atoms between the dimythyl phosphate and the borano dimethyl phosphate. This indicates that charge distribution was limited to the borano, phosphate, and lone oxygen groups. To this end, the charges were “balanced” according to the charges used in the normal DA (deoxy-adenosine within a sequence) parm99 parameters. Specifically, the overall charge for the
phosphate group in DA was used as the overall charge in the borano-phosphate group. The O5’ and O3’ atoms were both approximately 0.01 units below the calculated dimethyl phosphate values; to this end, the DA values adjusted by 0.01. The remaining charge was then distributed between the B, BH’s, P, and =O atoms weighted according the borano dimethyl phosphate calculation. This ensures the same charge distribution as predicted by the RESP calculation, but with easy incorporation into the AMBER force field parameters for the canonical bases. For example, the P atom has a weight of approximately 2.8691: 0.9973/(.9973-.0722-.1925-.1925-.19325). The calculated value was then reduced by the weight times the difference in overall charge between the calculated phosphate group charge and the DA phosphate group charge: 0.9973 – (-1.4271 - (-1.4049))*2.869. (Table 5-4) Final values for AMBER charges were taken from “scheme 2” as outlined in Table 5-4.

**AM1-BCC charge derivation:**

Charges can also be calculated using the AM1-BCC method. The process is quite simple. ANTECHAMBER is executed using a starting PDB file: antechamber –i filename.pdb –fi pdb –o filenameout.mol2 –fo mol2 –c bcc. It should be noted that results for the universal base showed significant variances between key components of the glycosidic linkage atoms.

**5.3.2 Atom, Bond, Angle, and Torsion Parameterization**

ANTECHAMBER, at first glance, appears to automate the entire process for residue atom selection. The results, however, are not always desirable. The universal base N8-8-aza-7-deaza-2’-deoxyadenosine had to be parameterized for AMBER use during NMR structure calcu-
lations. The structure is an adenosine analog with the main exception being the N8 linkage between the sugar and the base and the placement of the nitrogens in the 5 member ring. The atom connections and torsion angles for this linkage were ill-defined within the parm99 force-field.

The residue was parameterized using the nucleoside analog (ie without the phosphate group) of the universal base. The resulting values and parameters were then appended to the DA definitions to create the UB.frcmod and UB.off files. Charges were derived using the RESP charge model as described previously. It should be noted that significant variations occurred between AM1-BCC charges and RESP charges in the N8 and C1’ linkage atoms. RESP charges were used for the force-field modification.

ANTECHAMBER was first used to generate a GAFF based force-field definitions. This technique worked quite well noting very few exceptions to the GAFF defined parameters. However, when this structure was then built to standard AMBER atom definitions with GAFF supplements, atom type definitions were inaccurate and the resulting structure, when minimized in AMBER, ignored sp² geometry in the nitrogens. Upon inspection, the atom types selected by ANTECHAMBER ignored the aliphaptic and aromatic ring definitions and chose the ring atoms as aliphatic atoms. Further, ANTECHAMBER inaccurately defined the C7 atom and the N8 atom. These atoms had to be manually defined as: N8 = N8 and C7 = CV. Attempts to minimize the structure still resulted in sp² like behavior regardless of the atom type definitions. It should be a good practice to review all atom types defined by ANTECHAMBER and ensure that they accurately the atoms within the molecule.
Inspection of the .frcmod file showed that all atoms and atom interactions within the molecule were being redefined utilizing GAFF standards. Because the structure is an adenosine analog, such definition is unnecessary. To this end, each definition of the within the .frcmod file was compared to definitions within the parm99 force-field for equivalent values and all definitions that appeared in the parm99 force-field were deleted. Still, minimization of the structure with this revised .frcmod file resulted in ignored sp² geometry. Further inspection of the .frcmod file revealed the use of improper torsion angles. These values were ignored during the first culling. Improper torsion angles define branched atom groups and impact, among other things, pucker and out-of-plane “bowing” of atom groups. It was found that improper torsion angles are not required in the .frcmod file; removal of these improper torsion angles resulted in proper sp² geometry behavior.

The above .frcmod and .off files were modified for use in a non-terminal nucleotide version of the universal base. Specifically, the universal base was built and minimized in Gaussian03 with the appropriate phosphate group. RESP charges were added to the mol2 file and a prep file was created within ANTECHAMBER. The only charges that were adjusted were the values surrounding the glycosidic linkage atoms. Care was taken to assure that the overall charge value was balanced with the DA charge value (ie to avoid a charge different than -1). Further, atom types were verified with the previously created universal base nucleoside. This prep file along with the .frcmod file from the nucleoside were read into xLEaP and an .off file was created. See attached for printouts of the .off and .frcmod files.

When the universal base was incorporated into an oligonucleotide context, it was found that the orientation of the phosphate group in the .off file was “backwards” which resulted in
non B-type conformation of the backbone torsion angles. The .off file was rebuilt and the phosphate group was reversed using the DA structure as a model. The resulting sequence build contained correct B-type backbone geometry.
Table 5-3  Software components needed for RESP calculations

<table>
<thead>
<tr>
<th>Software package</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian03</td>
<td>QM calculations</td>
</tr>
<tr>
<td>AMBER 9</td>
<td>RESP calculations</td>
</tr>
<tr>
<td>R.E.D. ver. 3</td>
<td>Automation script</td>
</tr>
<tr>
<td>OPENBABEL 2.2.0</td>
<td>MOL2 → pdb file conversion</td>
</tr>
<tr>
<td>VMD</td>
<td>molecular viewer</td>
</tr>
</tbody>
</table>
Table 5-4 Boranophosphate charge derivation

*Note, for DA and Dimethylphosphate, these values represent the charge of O for the P=O bond.

<table>
<thead>
<tr>
<th></th>
<th>dm Phos.</th>
<th>dA</th>
<th>calc'd</th>
<th>calc'd</th>
<th>accepted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Earlier Study</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>1.1659</td>
<td>1.1659</td>
<td>1.2019</td>
<td>0.9973</td>
<td>1.0610</td>
</tr>
<tr>
<td>B</td>
<td>-0.1352</td>
<td>-0.7761*</td>
<td>-0.7927*</td>
<td>-0.0722</td>
<td>-0.0768</td>
</tr>
<tr>
<td>H (BH3)</td>
<td>-0.2136</td>
<td></td>
<td></td>
<td>-0.1925</td>
<td>-0.2048</td>
</tr>
<tr>
<td>O (P=O)</td>
<td>-0.7761</td>
<td>-0.7761</td>
<td>-0.7927</td>
<td>-0.7748</td>
<td>-0.7761</td>
</tr>
<tr>
<td>O5'</td>
<td>-0.4954</td>
<td>-0.4954</td>
<td>-0.4741</td>
<td>-0.4676</td>
<td>-0.4854</td>
</tr>
<tr>
<td>O3'</td>
<td>-0.5232</td>
<td>-0.5232</td>
<td>-0.4741</td>
<td>-0.461</td>
<td>-0.5132</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>phosphodiester</td>
<td>-1.4048</td>
<td>-1.4049</td>
<td>-1.3317</td>
<td>-1.3558</td>
<td>-1.4049</td>
</tr>
<tr>
<td>O5' weight</td>
<td>0.486</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O3' weight</td>
<td>0.514</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P weight</td>
<td>2.869</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B weight</td>
<td>-0.208</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H (BH3) weight</td>
<td>-0.554</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
5.3.3 References


Collier, G. Fall 2005 Semester Progress Report, Germann lab.


*R.E.D. III.x development was initiated at TSRI in Professor D. A. Case’s laboratory & finalized at the "Faculté de Pharmacie" in Amiens by N. Grivel, P. Cieplak, and F.-Y. Dupradeau*


5.4 Molecular Dynamics Utilizing Time-Averaged Restraints and Analysis with PDQPRO

5.4.1 Introduction

Structure development using solution NMR techniques requires the use of empirically determined restraints. Fine-level structure restraints include 2 atom distances, 3 atom angles, and 4 atom torsion angles. Global restraints include residual dipolar restraints (RDC) and residual chemical shift anisotropy (RCSA) allowing for orientation of atoms about a global alignment vector (dipolar bond vectors and chemical shift anisotropy principal axis system for RDC and RCSA respectively). (1, 2) For structure determination, these restraints are generally implemented using restrained molecular dynamics (rMD) in which instantaneous restraints are applied at all times with energy penalties imposed when deviations occur.

An alternative calculation process involves the use of time averaged restraints (MDtar) in which the restraints are enforced over a period of time rather than instantaneous and constant enforcement. The theory and background are well covered in existing literature, yet their implementation within AMBER is less detailed. (3, 4) Also, a number of scripts were written to implement the PDQPRO algorithm which sorts through and scores the MDtar structures as related to the NOESY intensity data. (5) The following details their implementation.

5.4.2 Implementation of MDtar

AMBER allows for time-averaging for distance, angle, and torsion restraints. For AMBER 9.0, implementation of this methodology is straightforward with the use of the DISAVE, ANGAVE, TORAVE and DISAVI, ANGAVI, and TORAVI commands. These commands are imple-
mented in the AMBER input file (.*in) under the weight change section and apply to all re-
straints in that class (i.e. all distance restraints). As an example of syntax, the following line
employs the XXXAVE command for time-averaged distance restraints in a 1 ns MD simulation
(AMBER 9.0):

```
&wt  type='DISAVE', istep1=0, istep2=1000000, value1=20, value2=3, /
```

In this syntax, istep1 and istep2 are the beginning and ending steps for weighting, value
1 is equal to tau (the time for exponential decay) and is in the units ps (for this example, 20 ps),
and value2 is equal to the power used in averaging. For nucleic acids, it has been found that a
tau of approximately 20 ps or less and a power of 3 for distances are sufficient. However, prob-
ing of various tau values should be employed. If time-averaging is implemented on angles or
torsions, the values are linear and value2 should be set to -1 (i.e. value2=-1) which imple-
ments linear averaging in AMBER.

The additional XXXAVI commands further defines how time-averaging will be imple-
mented and how often data is recorded. The following is an example of syntax for distance
time-averaging:

```
&wt  type='DISAVI', istep2=1000, IINC=1, /
```

...  
DUMPAVE=avgout
where istep2 is how often data is written to the DUMPAVE file (avgout in this case) and IINC=1 sets calculations using the pseudoforce option which is required if broad conformation sampling is desired. If this is not desired, do not include the IINC=1 command. Test MD runs should be conducted where the IINC toggle not used to determine if pseudoforces need to be implemented. In the case of the universal base structure, use of exact forces (i.e. where IINC=1 was not used) resulted in instability and drastically increased penalties.

Starting structures for MDtar should be refined via traditional rMD methodologies. (3) Overall MDtar times need to be adequately long enough for averaging to be effective. It has been found that a minimum of 10 times tau is sufficient to allow conformational sampling. (4) In practice, however, because of increased computing capabilities, larger multiples can be employed. For example, with the universal base MDtar examples, an overall time of 1 ns was used (tau values ranged from 5 ps to 50 ps).

### 5.4.3 Implementation of the PDQPRO Algorithm

MDtar simulations can result in an extremely large number of potential structures which must be ranked as possibilities in a unbiased method. For example, the universal base MDtar simulations resulted in structure being written every 1 ps yielding 1000 potential structures per MDtar simulation; with multiple MDtar simulations, several thousand structures had to be analyzed. The PDQPRO algorithm developed by Dr. Ulyanov was utilized for this analysis. (5)

Because NOE intensities are not linear to calculated distances, the PDQPRO algorithm compares the dipolar relaxation rates of theoretical structures determined via MDtar simula-
tions and experimental NOESY data and then calculates the probability of each potential structure with regards to agreement to the experimental data.

Experimental dipolar relaxation rates are calculated in the MARDIGRAS program and theoretical dipolar relaxation rates for each proposed structure are calculated in the CORMA program. Because thousands of files must be analyzed via CORMA and each program utilizes different PDB file formats, scripts were written to automate much of the process. PDQPRO output is quite simple, a list of the top structures with probabilities, thus winnowing the potential structures from thousands to few.
5.4.4 References


