Investigation of the Formation of Z-DNA and Other Non-Canonical DNA Conformations Using a Combined Spectroscopic and Biochemical Approach

by

Jameson Ross Bothe

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemistry)
in The University of Michigan
2012

Doctoral Committee:

Professor Hashim M. Al-Hashimi, Chair
Professor Charles L. Brooks III
Professor Jens-Christian D. Meiners
Assistant Professor Kevin J. Kubarych

2012

For Steve.

Acknowledgments

I would like to start by thanking Hashim Al-Hashimi for being my advisor. His constant positive encouragement gave me the freedom and opportunity to scientifically explore basically whatever I found to be interesting as long is it somehow involved nucleic acids. While this sometimes (ok, most of the time) resulted in reaching a dead end or failure, I always learned something. Also, thank you to all past and present members of the Al-Hashimi lab. The bar was most definitely set by all members who joined before me, so thanks to Dr.'s Yan Sun, Alex Hansen, Andrew Stelzer, Max Bailor, Anette Casiano, Liz Dethoff, and Evgenia Nikolova for setting such a high standard to aspire to when starting out. Also thanks to Katie Eichhorn who joined when I did, it was nice to always have someone in lab that could relate to what I was going through during each phase in graduate school. Outside of research I'll miss and look back on fun times with lab members such as walking like an Egyptian or the epic search for a bar that would serve drinks in San Francisco.

I would like to acknowledge and thank my committee members Charles L. Brooks III, Kevin J. Kubarych, and Jens-Christian Meiners for their questions, help along the way, and taking the time to write letters. I'd also like to thank Jens-Christian Meiners for early discussions on supercoiling and minicircles which helped to spark the supercoiling project in the Al-Hashimi lab. Thank you to our collaborators Dr. Ky Lowenhaupt and Professor Alexander Rich for providing us with the Zα protein used in Chapters 3 and 4. Ky always rapidly sent tons of Zα whenever requested and Prof. Rich is truly an inspirational scientist. Finally, a big thank you to Bryan Wang for working with me on the minicircle project, some of his work is found in Chapter 5. This was a very difficult project even for a graduate student let alone an undergrad, so I salute you for your efforts and wouldn't have been able to get nearly as much done without your help.

Also, I could not have completed this major hurdle in life without the inspiration and help from others before entering graduate school. Thanks to my high school teacher Doug Stratton for being the first person to get me excited about science and specifically chemistry. I would like to acknowledge the Onalaska data services crew Kevin Capwell, Lou Kalis, and Sue Ruble for their mentorship, friendship, and support. Working with all of you helped me to develop a strong work ethic and problem solving skills that have been invaluable in grad school. Also, a big thank you to my undergraduate research advisor Dr. Keith Beyer for his mentorship, support, and friendship. I had no idea what I was doing when I first started as a sophomore in the lab. With your office in a separate building it forced me to begin to figure things out independently. I would have totally failed in grad school if you hadn't allowed me to develop as an independent researcher. Also, I wouldn't have known about or considered going to graduate school without your guidance.

Most importantly I thank my family for all of their love and support. To my parents Ken and Linda, who have always supported me without question. I couldn't have asked for better parents. Also, my brother Jared and his wife Julie. Visiting Appleton was always a blast (MC!) and a great break from Ann Arbor. I would also like to thank everyone in Dani's family: the Schultz's, David's, and Nielsen's for their support and always making me feel welcome as a member of the family. Also, Andrew and Nina Stelzer who have been the Ann Arbor Family. You were the local keepers of sanity over these last few years!

Finally, I would like to thank my girlfriend Dani. Without you I would not be writing this thesis. Your unshakable love and support has gotten me through the lowest moments and always made things better. Whenever I was down you were always by my side, so actually it was never bad after all. I could write a thesis on this, but I won't get too sappy so I'll leave it at that. Much love, bun.

Table of Contents

Dedication ii
Acknowledgments iii
List of Figures viii
List of Tables x
List of Appendices xi
Abstract xii
Chapter 1 Introduction 1
1.1 Z-DNA Structure 1
1.2 Structure of Junctions Between B-DNA and Z-DNA 3
1.3 Thermodynamics of Z-DNA Formation 5
1.4 Biological Relevance of Z-DNA 8
1.4.1 Supercoiling Induced Z-DNA Formation 8
1.4.2 Z-DNA and Transcription 9
1.4.3 Z-DNA and Nucleosomes 10
1.4.4 Proteins that Bind Z-DNA 11
1.5 Spectroscopic Methods to Probe DNA Structure and Dynamics 13
1.5.1 CD Spectroscopy 13
1.5.2 Fluorescence Spectroscopy 15
1.5.3 NMR Methods to Characterize the Dynamic Structure Landscape of Nucleic Acids 18
1.6 References 30

Chapter 2 How Fast Can $R_{1\rho}$ Go? Determination of the Timescale Limits for Obtaining Robust Exchange Parameters 39
2.1 Introduction 39

2.2 Methods....40
2.2.1 Grid Searching....40
2.2.2 Bootstrap Analysis....41
2.3 Results and Discussion....42
2.3.1 Exploration of the χ^2 Surface Via Grid Searching....42
2.3.2 Monte Carlo Bootstrap Analysis of Dispersion Data....46
2.3.3 Comparing the limits of $R_{1\rho}$ to CPMG Dispersion....52
2.3.4 Fitting Experimental DNA $R_{1\rho}$ Dispersion Data....54
2.4 Conclusions....55
2.5 References....57

Chapter 3 Sequence-Specific B-DNA Flexibility Modulates Z-DNA Formation....59
3.1 Introduction....59
3.2 Materials and Methods....60
3.2.1 Sample Preparation....60
3.2.2 Spectroscopy....60
3.2.3 Data Analysis....61
3.3 Results and Discussion....61
3.4 Conclusions....69
3.5 References....69

Chapter 4 Incorporation of Non-CG Dinucleotides into Z-DNA: Interplay Between B/Z Junction and Z-DNA Helical Formation....72
4.1 Introduction....72
4.2 Materials and Methods....73
4.2.1 Sample Preparation....73
4.2.2 Fluorescence Spectroscopy....74
4.2.3 CD Spectroscopy....74
4.3 Results and Discussion....75
4.3.1 2-Aminopurine as a Z-DNA Probe....75

4.3.2 CD Spectroscopy as a Probe of B/Z Junction Formation....79
4.3.3 B/Z Junction Mutations Modulate Z-DNA Formation....82
4.3.4 Incorporation of CC Steps into Z-DNA Helices....83
4.3.5 Propagation of Z-DNA Helices with Additional Non-Pyrimidine/Purine Steps....88
4.4 Conclusions....90
4.5 References....90

Chapter 5 Preliminary Studies of Supercoiled DNA: Development and Optimization of a DNA Minicircle System....94
5.1 Introduction....94
5.2 Materials and Methods....95
5.2.1 Synthesis and Purification of DNA Minicircles....95
5.2.2 Analysis of Topoisomer Distributions....97
5.2.3 BAL-31 Nuclease Minicircle Digestion Reactions....98
5.3 Results and Discussion....98
5.3.1 Preparation of DNA Minicircles with Pure Topoisomers....98
5.3.2 Probing the Sequence-Specific Formation of Topoisomers....103
5.3.3 Kinetics of Minicircle Digestion by BAL-31 Nuclease....105
5.4 Conclusions....108
5.5 References....110

Chapter 6 Conclusions and Future Directions....112

Appendices....118

List of Figures

Figure

1.1 Structures of B-DNA (blue) and Z-DNA (black) with some key structural parameters 2
1.2 Crystal structure of a B/Z junction (PDB: 2ACJ) 5
1.3 Transcription induced formation of Z-DNA by negative supercoiling 10
1.4 CD spectra of pure B-DNA and Z-DNA helices 14
1.5 2-Aminopurine as a probe of base stacking interactions 17
1.6 NMR methods for characterizing the DNA/RNA dynamic structure landscape 19
1.7 Characterizing micro-to-millisecond exchange by relaxation dispersion 23
1.8 Graphical representation of $R_{1\rho}$ relaxation dispersion 26
2.1 χ^2surface for slow and fast exchange scenarios 44
2.2 Correlation plot between the $R_{1\rho}$ of positive and negative offsets (Ω) 45
2.3 Example bootstrap fit for a slow-intermediate exchange scenario 48
2.4 Example bootstrap fit for a fast exchange scenario 49
2.5 Bootstrap fits of 1,000 data sets composed of 5% noise corrupted data with varying R_2 51
2.6 Bootstrap fit of 1,000 data sets composed of 5% noise corrupted data 53
3.1 NMR observation of sequence-specific B-DNA flexibility in and around B/Z junctions 62
3.2 Comparison of on-resonance relaxation dispersion data for well resolved resonances in Z-JXN (black) and Z-JXN^{cont} (red) 64
3.3 Comparison of R_2 measured in Z-JXN (black) and Z-JXN^{cont} (red) constructs 65
3.4 CG-induced chemical shift perturbations at terminal residues 66
3.5 Sequence-specific preferences for Z-DNA formation 67
3.6 Sequence-specific preferences for Z-DNA formation at high salt in B/Z junction containing DNA constructs 68

4.1 Combined CD and fluorescence data to probe B/Z junctions 76
4.2 CD and fluorescence spectra of a 2AP labeled $(CA)_6$ 78
4.3 B-DNA base stacking patterns for pyrimidine (Y)/purine (R) repeats 78
4.4 Z-DNA base stacking patterns for pyrimidine (Y)/purine (R) repeats 79
4.5 Probing the incorporation of CC steps into Z-DNA 84
4.6 CD conformational analysis of control constructs in their Zα free state 86
4.7 Probing the structure and stoichiometry of Zα domain/DNA complexes 87
4.8 Z-DNA preferences for non-pyrimidine/purine repeats upon binding the Zα domain 89
5.1 Schematic representation of DNA minicircle synthesis 96
5.2 Denaturing PAGE of typical ligation reaction products with increasing concentrations of chloroquine (0.0 - 1.5 mM) from left to right 101
5.3 Denaturing PAGE of a negatively supercoiled 87 base-paired minicircle purified from ligation reactions 102
5.4 Chloroquine titration of the 84 base-pair minicircle Mini84 103
5.5 Sequence dependence of topoisomer distributions from ligation reactions 104
5.6 BAL-31 Kinetic assay and results for positively supercoiled minicircles 106
5.7 BAL-31 kinetic data for a negatively supercoiled MiniJXN^{cont} minicircle (ΔLk = -0.3) 108

List of Tables

Table

1.1 Experimental and estimated B-to-Z transition energies for each dinucleotide step 6

2.1 95% confidence intervals from bootstrap fitting of synthetic $R_{1\rho}$ data to the Laguerre equation 48

2.2 95% confidence intervals from bootstrap fitting of synthetic $R_{1\rho}$ data to the Laguerre equation with varied noise corruption and R_2 52

2.3 Comparison of $R_{1\rho}$ fits to scenarios in which CPMG fails 54

2.4 95% confidence intervals for $R_{1\rho}$ bootstrap fits of experimental data collected at 26 °C 55

4.1 CD fitting results for each independently 2AP labeled Z-JXN construct bound to the Zα domain 81

4.2 CD fitting results for each independently 2AP labeled Z-CG construct bound to the Zα domain 82

5.1 DNA Minicircle sequences 101

5.2 Rate constants measured for minicircle digestion by BAL-31 nuclease 107

List of Appendices

Appendix

1. Noise Corruption 118

2. Grid Search 119

3. Bootstrap Fitting 121

Abstract

Genomic sequences of DNA not only code for proteins, but also unique structural conformations encompassing a wide variety of helical and base-pairing geometries. Formation of these sequence-specific conformations is driven by a complex network of cellular interactions including ions, proteins, and supercoiling forces that constantly bombard the DNA double helix. One particular conformation Z-DNA, composed of a left-handed helix, is believed to play important roles in gene expression and regulation. Formation of Z-DNA within the genome results in the inversion of right-handed B-DNA and creation of distorted junctions at the intersection of B and Z-form helices. However, much is still unknown about the sequence-dependence of the B-to-Z transition and B/Z junction formation. Here we use a wide range of spectroscopic and biochemical methods to characterize the formation of Z-DNA, B/Z junctions, and other non-canonical DNA conformations.

NMR relaxation dispersion allows for an unprecedented insight into lowly populated conformations. First, we carry out simulations to explore the limits of which systems can be quantitatively characterized by $R_{1\rho}$ relaxation dispersion. Using relaxation dispersion experiments, we show that sites near or at B/Z junctions have a high intrinsic propensity in B-DNA to form non-canonical conformations. By combining NMR dynamics measurements with CD measurements of the B-to-Z transition, we show that mutations that diminish local flexibility at B/Z junctions also reduce the propensity to undergo the B-to-Z transition.

To better characterize the role of B/Z junction formation in the B-to-Z transition, we develop a combined CD and fluorescence spectroscopic approach for quantitatively assessing the formation of B/Z junctions within mixtures of B and Z-DNA. Our studies show that the thermodynamics of B/Z junction formation can significantly influence the B-to-Z transition, allowing for the incorporation of

unfavorable sequences into Z-DNA in order to achieve the most favorable B/Z junction. These new surprising preferences for Z-DNA formation may expand the sequence-space predicted to be available to Z-DNA in genomes. Finally, we have initiated studies on supercoiled DNA using small minicircles and carry out experiments to probe how sequence-specific mutations, shown to have a dramatic effect on relaxed linear DNA, influence the properties of supercoiled DNA.

Chapter 1

Introduction

1.1 Z-DNA Structure

Deoxyribonucleic acid (DNA) is the carrier of genetic information and since Miescher's discovery of nucleic acids in the mid nineteenth century, exhaustive efforts have been carried out to fully understand its function[1]. The key to understanding the biological mechanism of DNA is to have an atomic level understanding of its structure. Up until the early 1950s some insights had been made into the structure of DNA, but an atomic DNA structure had not been proven.

In February of 1953, a DNA structure was proposed by Linus Pauling composed of a triple helix with the phosphates in the core of the helix and bases lining its exterior[2]. This structure was partially based on blurry X-ray diffraction patterns and it quite quickly became apparent that this structure was incorrect particularly because of the unfavorable phosphate negative ion charge repulsion in the helical core. Shortly after in April 1953, Watson and Crick published a DNA structure with the bases in the core of a right-handed double helix which is known as B-DNA (Figure 1.1, B-DNA)[3]. To this day the Watson-Crick B-DNA helix is generally considered and observed to be the most stable DNA structure under most conditions. Even though the DNA double helical structure had been solved, one of the biggest remaining questions was how the two DNA strands would come apart during biological processes. In the fall of 1953 Alexander Rich, a postdoctoral researcher in Pauling's lab, wondered if the double helix might go left-handed as a part of the stand separation process and attempted to build a left-handed helix using atomic models[4]. One day Pauling walked in on Rich's model building and said, "Alex, work on that problem hard. I like the most important discoveries to be made in

Pasadena"[4]. However, Alexander Rich was unable to build a left-handed DNA model and moved on to other projects involving RNA.

Twenty six years later in Cambridge, Massachusetts Alexander Rich and coworkers solved the first high resolution single crystal structure of DNA composed of a $(CG)_3$ duplex[5]. Surprisingly, this $(CG)_3$ crystallized with a left-handed helix held together with Watson-Crick base-pairs (Figure 1.1, Z-DNA). This structure was the opposite of what had been expected where all the bases had literally flipped over from Watson-Crick B-DNA to form a left-handed helix. Instead of all of the bases having a glycosyl *anti* conformation (nucleobases facing away from the sugar) as in B-DNA, the Z-form had bases in alternating *anti*/*syn* conformations (*syn*-nucleobases facing toward the sugar). This alternating structure in Z-DNA causes a zig-zag in the phosphate backbone, hence the name Z-DNA. The flipping over of the bases coupled with the alternating glycosyl bond conformation and a re-puckering of the sugars allows for the bases to maintain Watson-Crick hydrogen bonding when in the Z conformation. Another significant difference between these DNA structures is that Z-DNA is composed of dinucleotide units (two base-pairs) because of the anti/syn glycosyl alternation whereas B-DNA has a mononucleotide unit (one base-pair).

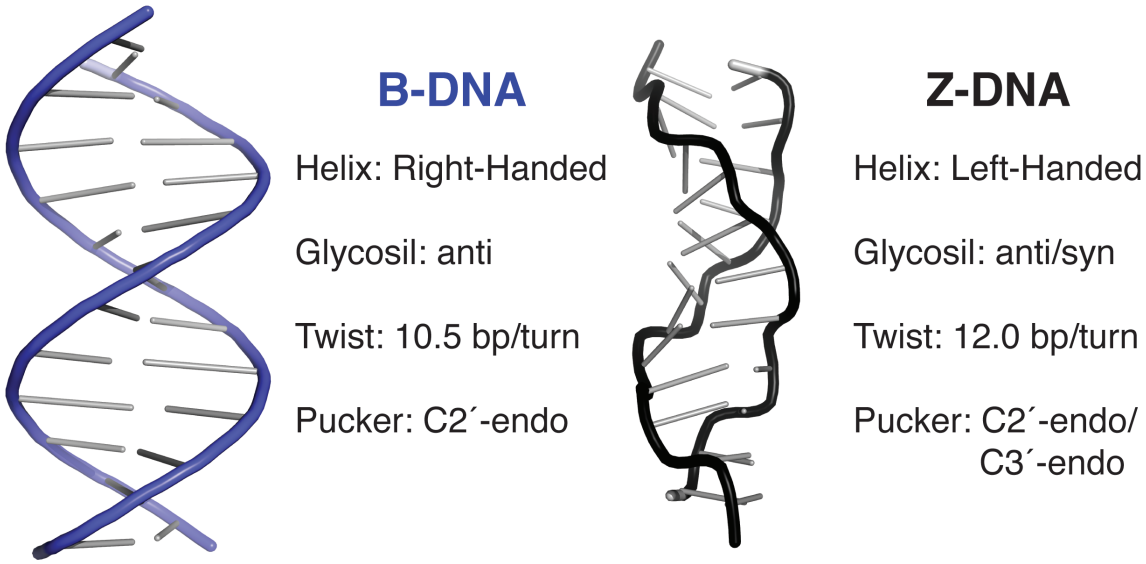

Figure 1.1: Structures of B-DNA (blue) and Z-DNA (black) with some key structural parameters.

This completely surprising X-ray structure was initially met with much skepticism since B-form was the widely accepted DNA conformation. Interestingly, well before the Z-DNA crystal structure of $(CG)_3$ was solved, Pohl and Jovin observed the inversion of the circular dichrosim (CD) spectrum of a poly(dG-dC) when changing the DNA solution's ionic conditions from 0.1 to 4.0 M NaCl[6]. This inversion of the spectrum suggested that the DNA structure had also inverted, but the atomic details of this inverted structure were unknown at the time. In order to decipher if the $(CG)_3$ crystallized in a structure similar to the low or high salt form of poly(dG-dC), Rich and coworkers recorded Raman spectra of the $(CG)_3$ crystals. Raman spectra had previously been collected for the low and high salt forms of poly(dG-dC) and significant spectral differences were observed between each ionic condition[7]. The Raman spectrum of the $(CG)_3$ crystals was nearly identical to the high salt spectrum of poly(dG-dC) indicating that the $(CG)_3$ had crystallized in the high salt conformation[8].

Subsequent studies showed that the formation of Z-DNA is not limited to just CG repeats, rather any sequence of DNA can adopt the Z conformation[9-13]. As discussed in detail in Section 1.3, Z-DNA forms sequence specifically and favors pyrimidine/purine dinucleotide steps[9]. Generally Z-DNA is a higher energy conformation than B-DNA, requiring very specific conditions for its formation such as the presence of high salt, proteins, or negative supercoiling[14]. The ability of DNA to adopt the Z conformation under these very specific conditions leads to the question of its biological relevance. Specifically, another major piece of the genomic Z-DNA puzzle is that if the Z conformation occurred in the genome, it would have to form surrounded by a sea of B-DNA requiring the formation of right/left-handed junctions between B and Z-DNA.

1.2 Structure of Junctions Between B-DNA and Z-DNA

Even in the first report of Z-DNA speculation was made about how B-DNA and Z-DNA might coexist[5]. While it would take another 26 years to solve a high-resolution structure of a B/Z junction[15], multiple experimental studies painted an

accurate, albeit imprecise picture of the B/Z junction. Shortly after the discovery of Z-DNA, nuclease digestion experiments detected significant distortions at the site of B/Z junctions. The nuclease proteins, S1 and BAL-31, used to probe B/Z junctions are sensitive to base-pair opening and single-strand like distortions in the double helix[16-18]. It was observed that both of these nuclease proteins actively cleaved DNA with B/Z junctions indicating a single-strand like feature in the junction. From these studies the size of the junction was estimated to be "on the order of 3 base-pairs"[19], which was eventually observed in the B/Z junction crystal structure[15]. Early CD and Raman spectroscopic studies also indicated that a unique distorted structure existed between the intersection of right and left-handed helices[20,21].

The preceding studies of B/Z junctions were carried out in the context of large plasmids consisting of thousands of base-pairs. Therefore, while a general picture of the B/Z junction structure was emerging, little detail at the atomic level could be determined in such large DNA systems. In 1988, Sheardy synthesized a small 16 nucleotide DNA that formed a B/Z junction under high salt conditions and was able to monitor the B-to-Z transition by CD spectroscopy[22]. This study by Sheardy and continued improvement of methods to synthesize short DNA's opened the door to more detailed spectroscopic investigations that enabled a deeper structural and energetic understanding of junction formation. Shortly after the first CD study of a short B/Z junction duplex a combination of CD, NMR, and Raman spectroscopies were used to generate a more detailed picture of the B/Z junction structure[23,24]. These two independent studies conducted with different DNA sequences both estimated the size of a B/Z junction to be approximately 3 base-pairs.

Finally, in 2005 the first and only crystal structure of a B/Z junction was reported by Ha *et al.* (Figure 1.2)[15]. In this structure, both B and Z-DNA helices are in canonical conformations, with the Z-DNA helix bound to proteins that stabilize its structure (not shown in Figure 1.2). The junction portion of the structure is composed of 3 base-pairs: two link the B and Z-DNA helices together through stacking interactions and one base-pair is extruded (Figure 1.2, pink). As noted by Ha *et al.*, it is remarkable that the handedness of the double helix can be reversed

simply by breaking one base-pair[15]. The B/Z junction resolved in the crystal structure is also consistent with previous observations made on B/Z junctions[16-18,23,24]. For example, the extruded base-pair is likely the explanation for cleavage by S1 and BAL-31 nucleases. Studies following the crystal structure publication using a fluorescent nucleotide sensitive to base stacking further confirmed that a base-pair likely becomes extruded upon formation of a B/Z junction[25].

Over the last 33 years significant strides have been made to characterize the structure of Z-DNA and other biologically relevant structural features including junctions between B-DNA and Z-DNA (B/Z junctions) and junctions between two Z-DNA helices (Z/Z junctions)[26]. However, structure is only part of the story. A full understanding of the impact of Z-DNA must take into consideration the energetics of its formation and its biological significance.

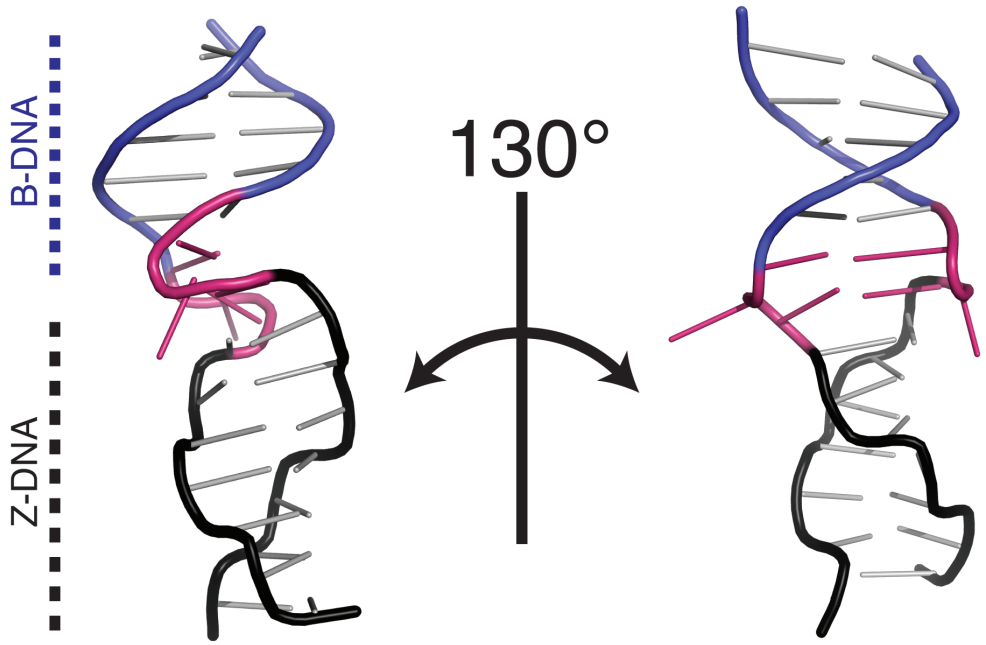

Figure 1.2. Crystal structure of a B/Z junction (PDB: 2ACJ). Structural components are color coded as: Z-DNA (black), B-DNA (blue), B/Z junction (pink). This B/Z junction containing DNA is bound to 4 Zα proteins in the crystal structure (not pictured)[15].

1.3 Thermodynamics of Z-DNA Formation

As observed originally by Pohl and Jovin, very specific sequence and environmental conditions are necessary for the formation of Z-DNA. The first

sequences shown to adopt the Z conformation were composed of CG repeats. Subsequent studies expanded the set of sequences able to adopt Z-DNA to include CA and TG pyrimidine/purine dinucleotide steps[11,27]. Later it was shown that additional dinucleotides, including non-pyrimidine/purine steps, could adopt the Z conformation. The transition free energies of the B-to-Z transition have been experimentally determined for some dinucleotide steps (CG, CA, TA, CC, CT) by inserting sequence repeats into supercoiled plasmids and measuring the transition as a function of supercoiling (Table 1.1)[28-32]. Transition energies for the rest of the DNA dinucleotide steps outside these small few have only been empirically estimated leaving much uncertainty in B-to-Z transition free energies (Table 1.1)[33]. In general, sequences rich in pyrimidine/purine repeats energetically favor Z-DNA formation and in some cases less favorable non-pyrimidine/purine sequences have a several fold higher free energy cost for the transition. For example, the free energy cost for a CG step to undergo the B-to-Z transition is 0.7 kcal/mol, while a CC step has a free energy cost of 2.4 kcal/mol.

Dinucleotide- 5'-3'	Transition Energy (kcal/mol)
CG	0.7*
GC	4.0
CA	1.3*
AC	4.6
TG	1.3*
GT	4.6
TA	2.5*
AT	5.9
CC	2.4
GG	2.4
CT	3.4*
TC	3.4
GA	3.4
AG	3.4*
AA	3.9
TT	3.9

*** B-to-Z transition energy determined experimentally**

Table 1.1: Experimental and estimated B-to-Z transition energies for each dinucleotide step[33]. Dinucleotide steps are listed such that they are in an anti/syn conformation when in Z-form.

Another significant contributor to the thermodynamics of the B-to-Z transition is the formation of B/Z junctions. Very limited sequence-dependent thermodynamic data is available for B/Z junction formation compared to B-to-Z helical conversions. Generally, it is estimated that the free energy cost of B/Z junction formation is ~5.0 kcal/mol [28,29,33]. It has been shown that the free energy of B/Z junction formation is sequence-dependent[34], but systematic studies to characterize every junction sequence have not been carried out. The sequence-dependence of junction formation likely arises from its structural features as revealed in the B/Z junction crystal structure (Figure 1.2)[15]. Here the important sequence-dependent contributors to the B/Z junction formation free energy cost are likely stacking interactions between B-DNA and Z-DNA helices and the extrusion of a base-pair at the junction. Due to the limited available B/Z junction thermodynamic data, theoretical predictions of B-to-Z transition free energies do not account for the sequence-dependence of junction formation and assign the same energy cost (5 kcal/mol) to all B/Z junctions.

Despite the limited amount of thermodynamic data available for the B-to-Z transition, studies have attempted to determine where Z-DNA is likely to form in genomic DNA[35-38]. Predictions of genomic Z-DNA formation have been carried out using a "thermogenomic" approach where Z-DNA hot-spots are pinpointed by calculating the thermodynamic propensity for sequences to form Z-DNA[37]. Typically, these predictions use a statistical mechanics treatment where a two-state zipper model is used to assign Z-DNA propensities to sequences[28,33,36,38]. A sequence's thermodynamic propensity to form Z-DNA is predicted by simultaneously accounting for the free energy difference between the B and Z-helical conformations of each dinucleotide step as well as the cost of B/Z junction formation. In general, computational studies have found that Z-DNA favoring sequences are typically localized near transcription start sites, suggesting that Z-DNA may play a role in transcription[35,36]. However, recent experiments testing these predictions revealed significant disagreement between the models and the observed data[36]. Here Li *et al.* used chemical cross-linking between a Z-DNA binding protein and DNA to probe for the occurrence of Z-DNA in human tumor cells. Strikingly, the authors found that

only 1% of the identified Z-DNA "hot-spots" were localized near transcription start sites[36]. This study provides a powerful demonstration of the gap in our understanding of Z-DNA formation and points towards the need for additional experimental data to improve future theoretical predictions. Also, more experimental methods need to be developed that can consistently and accurately pinpoint extents of Z-DNA formation *in vivo* to test predictions made by these thermogenomic models.

1.4 Biological Relevance of Z-DNA

1.4.1 Supercoiling induced Z-DNA Formation

Shortly after the discovery of Z-DNA the biological relevance, if any, of this alternative DNA structure was a highly sought after question. Early studies showed that the B-to-Z transition in linear DNA could be induced under conditions of extreme ionic strength (> 2.5 M NaCl); however, under physiological ionic conditions B-DNA is the lowest energy conformation[9]. Chemical modifications such as halogenation or methylation of nucleobases were shown to promote the transition at lower ionic strength conditions, but this still did not answer the question if the natural nucleobases could undergo the B-to-Z transition in a biological context[9]. Finally, in 1982 it was shown that CG repeats inserted into negatively supercoiled plasmids could adopt the Z conformation at physiological ionic strengths[13,16]. The discovery of supercoiling induced Z-DNA formation was a major discovery, because it was known that DNA supercoiling occurred in a cellular context.

Supercoiling refers to either an overwinding (positive supercoiling) or underwinding (negative supercoiling) of the DNA double helix in comparison to its canonical, preferred structure. Generally, eukaryotic and most prokaryotic DNA is negatively supercoiled and the level of supercoiling can be dynamically altered through biological processes such as transcription or nucleosome packaging[39]. Over or under-twisting of DNA results in torsional strain that destabilizes the double helix; therefore, promoting the formation of non-canonical conformations that effectively unwind the DNA for negative supercoiling[40,41]. In the case of Z-DNA,

formation of a left-handed helix absorbs unwinding of the right-handed B-DNA helix resulting in an energetically favorable reduction in the level of supercoiling. Additional studies of supercoil-induced Z-DNA formation revealed that Z-DNA could form at levels of supercoiling that occur *in vivo*[42-44]. New insights are still being made into negatively supercoiled induced Z-DNA formation. For example, a recent study by Lee *et al.* using optical tweezers showed that minute levels of negative supercoiling induce the formation of Z-DNA[45]. In summary, negative supercoiling which has been implicated in important biological processes such as transcription and DNA packaging promotes the formation of Z-DNA.

1.4.2 Z-DNA and Transcription

Early studies also began to reveal that Z-DNA was somehow involved with transcription[46-48]. Antibodies that specifically bind the Z conformation were found to localize near actively transcribed sites in cells. However, whether the antibodies were recognizing Z-DNA that was produced by transcription, or instead only inducing Z-DNA in the DNA after it was made accessible by the polymerase was unclear[47]. Later, in 1987 one of the strongest pieces of evidence supporting transcription activated Z-DNA was the transcription model proposed by Liu and Wang where the RNA polymerase complex plows through template DNA generating positive and negative supercoiling in front of and behind the transcription complex, respectively (Figure 1.3)[49]. Negative supercoiling had been shown to be an inducer of Z-DNA, so supercoiling resulting from transcription could be a mechanism for its formation. Multiple studies have found a correlation between Z-DNA formation and upregulated transcription[50-52]. In these studies, the cellular nuclear membrane of mammalian cells was permeabilized to allow for diffusion of foreign molecules into the cellular nucleus. Z-DNA specific antibodies were then introduced into the cells and used to detect formation of Z-DNA during active transcription. Computational studies also support a transcriptional role for Z-DNA by showing that sequences with a high probability to form Z-DNA are localized near transcription start sites in 137 human genes[53].

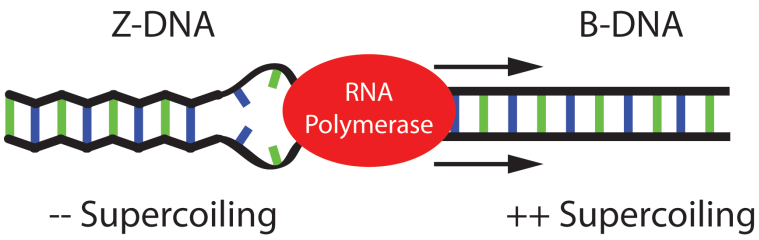

Figure 1.3: Transcription induced formation of Z-DNA by negative supercoiling.

1.4.3 Z-DNA and Nucleosomes

The genome of eukaryotic organisms is packaged into cellular nuclei as chromatin, a complex mixture of DNA and packaging proteins[41]. The basic building block of chromatin is nucleosome particles composed of protein-DNA complexes with approximately ~146 base-pairs of DNA. In the nucleosome particle DNA is wrapped around a protein core in a left-handed manner resulting in negative supercoiling[41]. Multiple studies have implicated a role for Z-DNA in nucleosome positioning and remodeling.

Early *in vitro* studies showed that DNA in the Z conformation was not readily incorporated into nucleosomes, suggesting that Z formation *in vivo* would inhibit its incorporation into nucleosomes[54-58]. This led to the hypothesis that sequences localized near transcription start sites in the Z-conformation could make local DNA accessible to transcription factors by reducing nucleosome occupancy[59]. A recent study by Wong *et al.* using yeast demonstrated that Z-DNA could in fact act as a nucleosome boundary element[60]. Sequences favoring Z-DNA formation ($(CG)_9$) were implanted upstream of a TATA box transcription start site and it was observed that nucleosomes had a low occupancy near the $(CG)_9$ region. Further, it was shown that transcription could be upregulated by placing the $(CG)_9$ near the TATA box suggesting that the absence of nucleosomes made the TATA box more accessible to transcription initiation factors. Placement of the $(CG)_9$ at sites more distant from the TATA box showed no enhancement of transcription further supporting this hypothesis.

In addition to the repulsion of Z-DNA incorporation into nucleosomes, sequences incorporated into nucleosomes with a high propensity to form Z-DNA can

play a role in nucleosome regulation. A study by Liu *et al.* on the CSF1 CATA promoter demonstrated that Z-DNA formation could play a role in nucleosome remodeling by absorbing the negative supercoiling resulting from nucleosome removal[61]. Here a chromatin remodeler, BAF, facilitated Z-DNA formation by nucleosome removal of a Z-DNA favoring TG repeat (Table 1.1) located 26 base-pairs upstream of the CSF1 promoter sequence. Z-DNA formation stabilized a nucleosome free state for the surrounding DNA, therefore allowing transcription factors access to the CSF1 CATA promoter sequence.

1.4.4 Proteins that Bind Z-DNA

If Z-DNA had a biological function, then naturally occurring proteins would likely exist that specifically interact with the Z conformation. It took 14 years after the initial discovery of Z-DNA to find such naturally occurring proteins. First, a strategy was developed to detect and characterize proteins that bound Z-DNA. Here, Herbert *et al.* designed a radioactively labeled DNA sequence with bromanated guanines that exists in the Z conformation under physiological ionic conditions[62]. This DNA was then used to detect the binding of proteins specific to the Z conformation using gel band-shifts. Additional studies employing this strategy led to the discovery and sequencing of the double-stranded RNA adenosine deaminase (ADAR1) protein that specifically binds the Z conformation with high affinity[63]. Further studies of ADAR1 pinpointed the Z-DNA binding activity to a 70 amino acid domain named Zα[64]. Spectroscopic and gel shift studies verified that the isolated Zα domain bound the Z conformation with high affinity (low nM) and that it could specifically bind Z-DNA when surrounded by B-DNA[65,66].

The final structural verification that the Zα domain bound Z-DNA was achieved when Kim *et al.* solved the structure of a Zα-$(CG)_3$ complex by X-ray crystallography showing that the $(CG)_3$ DNA was in the same conformation as that observed in the first Z-DNA crystal structure[66]. The crystal structure revealed specific interactions between the Zα domain and the Z form phosphate backbone and nucleotides in the syn conformation. Subsequent X-ray crystallography studies

showed that the Zα domain recognizes conformational features specific to Z-DNA rather than sequence-specific ones[67]. Once the key Z-DNA binding residues were identified in Zα, structures were solved for additional proteins with similar Zα domains including the vaccina virus E3L[68] and the interferon response DAI[69] proteins.

Interestingly, each protein discovered to have a Zα domain has a different function. The ADAR1 protein is an RNA editing protein that deaminates adenosine to yield inosine [70,71]. Since inosines are read as guanines in the ribosome, RNA modifications made by ADAR1 change the amino acid sequence of the coded protein. It has been proposed that the Z-DNA binding affinity might localize ADAR1 to active transcription sites allowing it to modify transcribed RNA[70,72,73]. ADAR1 can also bind RNA in a Z-RNA conformation and it has been shown that RNA's with Z-RNA favoring sequences are more prone to modification[74]. ADAR1 localizes in the nucleus and cytoplasm of the cell and has been implicated as a cellular defense element by targeting and modifying RNA viruses[75-78].

The E3L protein, which contains a Zα domain with a weaker Z-DNA binding affinity than ADAR1, is found in poxviruses. Poxviruses live in the cellular cytoplasm and suppress the cellular interferon defense mechanism. Experimental studies have shown that the Zα domain is essential for the viral pathogenicity of the vaccina poxvirus[79-81]. In one study, mice infected with the natural vaccina virus died within one week, but the mice did not perish upon infection with viruses that had a Zα-deficient E3L protein[79]. Interestingly, mutations also showed that the vaccina virus was always lethal regardless of the identity of the Zα domain attached to E3L whether from E3L, ADAR1, or DAI. Additional studies using vaccina virus infected HeLa cells further confirmed the necessity of the Zα domain in E3L for viral function[80]. Overall, these studies both suggest that the Z-DNA binding component of E3L is crucial for interfering with cellular defense mechanisms.

Another protein, DAI (DNA-dependent activator of IFN) also contains a Zα domain. When the crystal structure of the DAI Zα domain bound to Z-DNA was solved, the function of DAI (DLM-1 at the time) was unknown having initially been found to be upregulated in tumor cells[69,82,83]. Recently, it was found that the DAI

protein acts as a DNA sensor that triggers immune responses upon detection of foreign DNA in the cell's cytosol[84]. Further investigation of the DAI's immune response mechanism revealed that the ADAR1 gene is turned on by the immune response[85,86]. Once transcribed and synthesized, ADAR1 competes with DAI for binding with foreign DNA's and ultimately turns down the immune response. It was also proposed that the E3L protein of the vaccina virus similarly works as a DAI competitor to turn down the cellular immune response[81,86].

In summary, the common thread between the ADAR1, E3L, and DAI proteins is the Zα domain. Each protein has significantly different functions, but all are related in some way to the cellular immune response system. The existence of these proteins revealed that biology has specifically evolved a means to recognize Z-DNA. The discovery of negatively supercoiled induced Z-DNA formation further solidified its biological relevance in processes such as transcription and nucleosome positioning. However, since genomic Z-DNA is only transiently formed, it remains extremely difficult to detect thus leaving open the possibility that Z-DNA may play a more expansive biological role than that has yet to be observed experimentally.

1.5 Spectroscopic Methods to Probe DNA Structure and Dynamics

1.5.1 CD Spectroscopy

Circular dichroism (CD) spectroscopy is one of the most versatile spectroscopic techniques able to probe the polymorphic structure of DNA[87,88]. Here we highlight a few of the advantages of using CD to study structure. First, CD is highly sensitive to DNA structure where each polymorphic structure of DNA including B-DNA, Z-DNA, A-DNA, quadruplex, and others give unique CD spectra (Figure 1.4)[88]. Second, CD can be used to quickly assess DNA structure as a function of many different conditions since CD spectra can be acquired very rapidly and with very little effort in experimental preparation. This is in contrast to more complicated techniques such as NMR that require significant optimization and time consuming assignment experiments. Third, CD requires very little sample (μM

concentrations), allowing for samples to be studied that are cost prohibitive or difficult to prepare. Finally, CD is not limited by the size of the system that can be studied. Unlike NMR, which is limited to a relatively small DNA sample size (< 100 base-pairs), CD can probe DNA's as large as plasmids (1000's of base-pairs) and under conditions of high ionic strength[89,90].

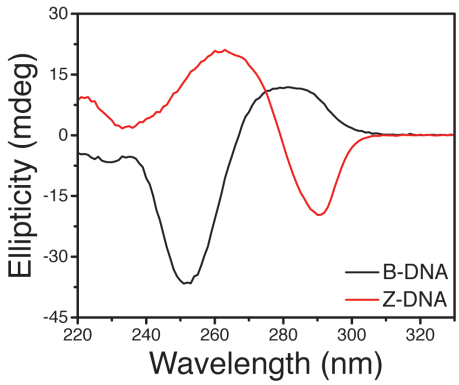

Figure 1.4: CD spectra of pure B-DNA and Z-DNA helices. CD spectra are of a $(CG)_6$ duplex in 0.0 M (black) and 5.4 M (red) NaCl.

The physical basis for CD spectroscopy involves the absorption of circularly polarized light[87]. Here chiral molecules interact uniquely with left and right-handed circularly polarized light and absorb each type to differing extents. Skewed absorption of one type of polarized light, either left or right-handed, gives rise to a CD signal. Achiral molecules do not give a CD signal. For DNA, CD is typically observed for the ultraviolet region spanning ~180-300 nm where the nucleobases interact with light[87,88]. The unit typically used to quantify CD signal is ellipticity, which is expressed in degrees. The ellipticity is the difference in molar extinction coefficients for right and left-handed circularly polarized light multiplied by a scalar factor (3298.2)[87,88]. The theoretical description of the CD signal as a function of wavelength is extremely complicated; therefore, the interpretation of CD spectra of biomolecules is usually strictly empirical[88,91]. Characteristic CD spectra of nucleic acid and protein structural polymorphs are typically assigned by obtaining CD spectra of molecules with known structure, where the structure is then used as the basis for quantifying the structural composition of CD spectra[88,91]. Methods to study

the solution structures of proteins by CD are very well developed with many software packages available for spectral analysis[91]. Generally, these software packages use linear combinations of reference spectra to fit the CD spectra of the target protein. The structural composition of the target protein is then interpolated from the known structures of the reference spectra. Structural interpretation by CD for nucleic acids is not nearly as advanced as it is for proteins. However, the unique CD spectral signatures are known for most DNA structure types[88].

CD spectroscopy is one of the most commonly used methods to study the structure, kinetics, and thermodynamics of Z-DNA. As mentioned previously, the first experimental observation of Z-DNA was made by Pohl and Jovin when the CD spectrum of a poly(dG-dC) inverted upon changing the DNA solution's ionic conditions from 0.1 to 4.0 M NaCl[6]. An example of a salt induced B-to-Z transition for a CG repeat as observed by CD is shown in Figure 1.4 where a right-to-left-handed helical conversion inverts the CD spectrum. Other studies have used CD to rapidly determine the conditions at which Z-DNA will form whether in the presence of specific ionic conditions[9,92,93], proteins that bind Z-DNA[65,66], or Z-DNA stabilizing inorganic complexes[94-96]. Many studies have also been carried out to probe the mechanism of the B-to-Z transition by rapidly inducing the transition and monitoring it by CD spectroscopy[97,98]. Other studies have quantitatively analyzed CD spectra of Z-DNA as a function of different ionic conditions to reveal intermediates along the B-to-Z transition pathway[34,99]. In summary, CD spectroscopy is a powerful technique to characterize the structure of biomolecules and has proven to be a versatile method to probe the formation of Z-DNA.

1.5.2 Fluorescence Spectroscopy

Another highly sensitive probe of DNA structure is fluorescence spectroscopy[100-102]. Fluorescent nucleotide analogues have been used to study a wide range processes associated with nucleic acid structural transitions such as folding, dynamics, hybridization, and protein-nucleic acid interactions. While the natural nucleobases are not fluorescent, small chemical modifications can be made to make them fluorescent. Chemical approaches to generating fluorescent

nucleobase analogues include tethering fluorophores to nucleotides, expanded nucleobases, and nucleotides that have been substituted with fluorescent aromatic groups[102]. The goal for all of these chemical modifications is to limit energetic and structural perturbations that arise from changing the nucleotide's chemical composition. An ideal fluorescent nucleotide maintains similar hydrogen bonding patterns and molecular size upon substitution into a DNA sequence. Each nucleotide has various fluorescent analogues and the development of new improved analogues is an active area of research[102].

One of the earliest developed and still most popular nucleotide analogues is 2-aminopurine (2AP), which can be used as a purine mimic to probe base stacking interactions[103,104]. 2AP differs from adenine only by the position of its amino group, which is moved from the 6 to 2 position (Figure 1.5a). This slight modification allows for 2AP to remain in a Watson-Crick like conformation while maintaining two hydrogen bonds when base-paired with thymine. This similarity to adenine makes 2AP an ideal fluorescent analogue and its substitution has been shown to result in minimal structural perturbations[105,106]. While 2AP minimally perturbs DNA structure, its substitution does result in some local energetic destabilization[105-107]. 2AP can also be substituted for guanines, but with a larger energetic perturbation compared to adenine as a 2AP-C base pair can only form 2 hydrogen bonds instead of the 3 found in a G-C pair. Furthermore, 2AP has the advantage of being widely commercially available.

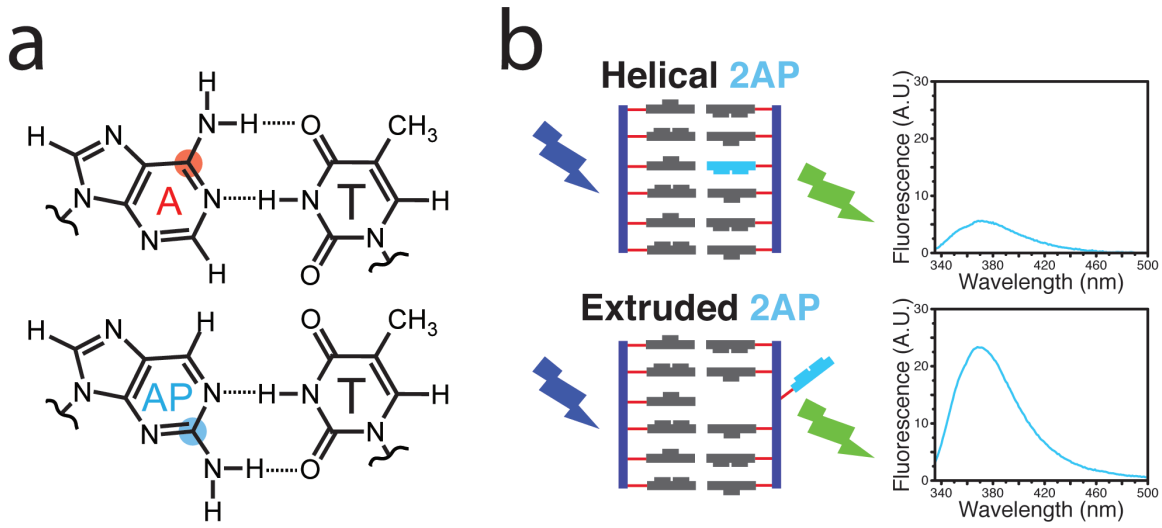

Figure 1.5: 2-Aminopurine as a probe of base stacking interactions. a) Comparison of adenine (red) and 2AP (blue) where the amino group is highlighted. b) The fluorescence of a helical 2AP (top, light blue) is quenched significantly compared to an extruded 2AP without base stacking interactions (bottom, light blue).

2AP is highly sensitive to base stacking interactions as its fluorescence is significantly quenched upon stacking with other nucleobases[103,104]. Any change in stacking or extrusion of the 2AP base can result in up to a 100 fold increase in fluorescence intensity (Figure 1.5b). While 2AP is highly sensitive to any change in stacking interactions, structural interpretation of changes in 2AP steady-state fluorescence have proven to be elusive in many systems[108-110]. A change in fluorescence intensity absolutely indicates a structural change has occurred, but the identity of the structural change cannot be determined based purely on steady-state fluorescence. For example, large enhancements in fluorescence could arise due to the extrusion of a helical base or helical kinking where stacking interactions are lost. Therefore, when interpreting 2AP data, all possible structural changes must be considered when trying to understand a change in fluorescence intensity.

2AP has recently been utilized as a probe of Z-DNA structure[25,111,112]. The first application involved the creation of a molecular thermometer based on the B-to-Z transition where the thermometer's fluorescence reported on temperature[111,112]. The thermometer utilized the entropy of the B-to-Z transition

where Z-DNA has lower entropy than B-DNA. Thus, as the system is heated the amount of B-DNA increases resulting in a reduction in fluorescence intensity. Temperature dependent changes in fluorescence could be attributed to differences in base stacking interactions between B-DNA and Z-DNA. When the 2AP is in the B conformation, it stacks with both it's 3' and 5' neighboring bases. However while in the Z conformation it only has a strong stacking interaction with its 5' neighboring base. This loss of stacking interactions, when in Z-form, results in a large increase in fluorescence intensity compared to B-DNA. 2AP is also an ideal probe of B/Z junctions because bases at the site of the junction become extruded and unstacked[113]. A study by Kim *et al.* provided additional evidence supporting the extrusion of bases reported in the B/Z junction crystal structure and showed that base extrusion likely occurs regardless of if the system is in solution or crystal form[25,113]. 2AP nucleotides have also been used to study the formation of Z/Z, and A/Z junctions in both DNA and RNA[25]. In Summary, 2AP fluorescence based methods are powerful for characterizing changes in DNA base stacking interactions. 2AP has been shown to be sensitive to the B-to-Z transition, but caution must be made in the structural interpretation of changes in 2AP fluorescence.

1.5.3 NMR Methods to Characterize the Dynamic Structure Landscape of Nucleic Acids

A complete and predictive understanding of DNA behavior ultimately will require atomic characterization to go beyond static structures towards a description of the free energy landscape. The free energy landscape[114] provides a unified and complete description regarding the dynamic properties of nucleic acids that are relevant for understanding their function (Figure 1.6, landscape). Here, the free energy is specified for every conformation that can be adopted by the DNA/RNA. The fractional population of a given conformer depends on its relative free energy, whereas the rate with which two conformers interconvert depends on the free energy barrier that separates them. Cellular cues perturb the free energy landscape, diminishing barriers and/or stabilizing conformers that are otherwise unfavorable, and thereby redistribute the conformer populations to effect specific biological outcomes. While rich in information, the free energy landscape is very complex, and

cannot be fully measured experimentally. Fortunately, important insights into biological function can be obtained by focusing on a sub-set of conformers that populate minima along the energy landscape. Studies increasingly show that such 'low-hanging fruit' conformers are often the ones that are stabilized by cellular cues to carry out biological function[115-117]. These more appreciably populated conformers are also more amenable to experimental characterization using spectroscopic techniques that probe dynamic fluctuations in structure along the energy landscape. We will refer to this partial energy landscape as the 'dynamic structure landscape'.

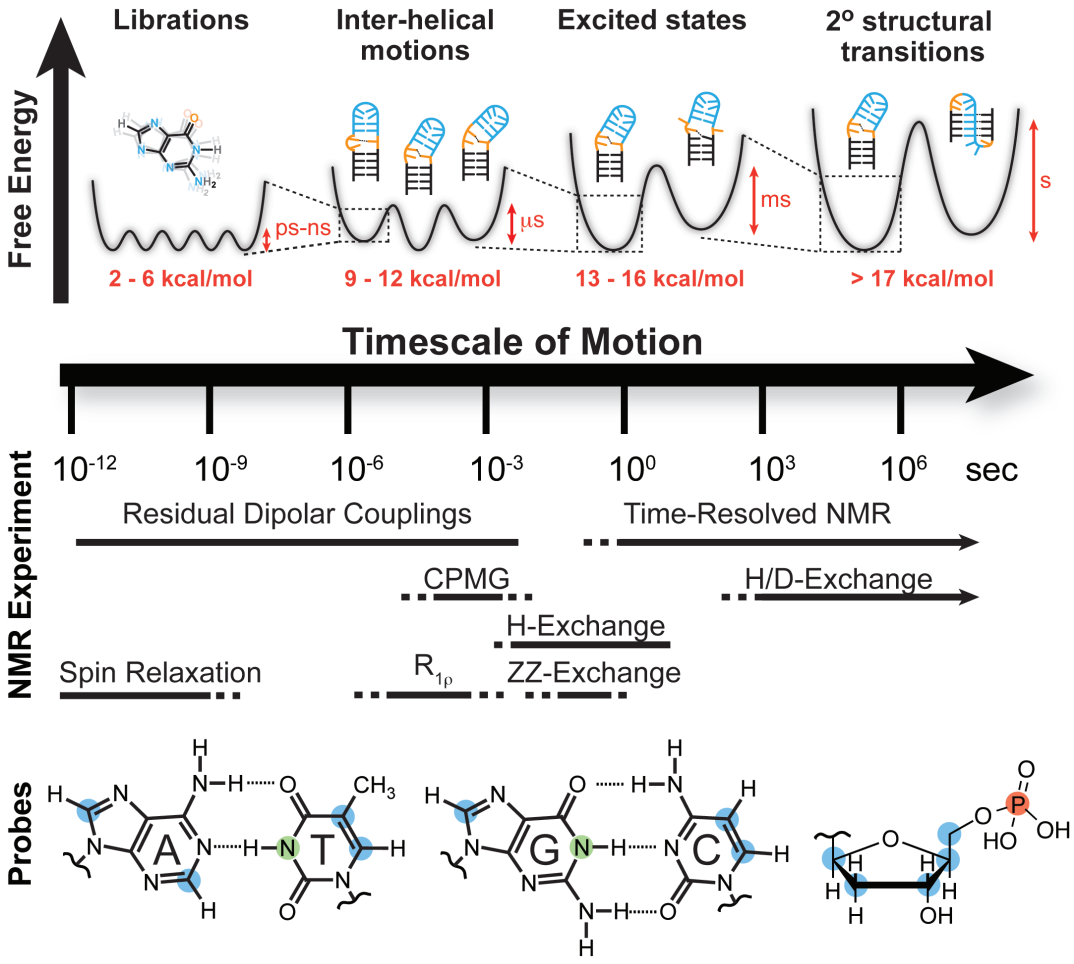

Figure 1.6: NMR methods for characterizing the DNA/RNA dynamic structure landscape. ***Dynamic Structure Landscape:*** Transition free energies (red, kcal/mol) corresponding to typical timescales of interconversion were estimated using transition state theory at 25 ºC. ***NMR Experiments:*** Solid lines indicate the timescales at which each NMR experiment is optimally suited for, while the dotted lines indicate timescales that are difficult to probe. ***Probes:*** Nuclei most commonly

used for RNA dynamics measurements: protonated carbons (blue), imino protons (green), backbone phosphorus (red).

The importance of dynamics to protein function has been appreciated ever since X-ray structures of myoglobin revealed no open passageway between the heme iron and external solution back in 1966[118]. Among many techniques that are now being developed to study DNA/RNA dynamics at atomic resolution, solution-state NMR spectroscopy, which has contributed significantly toward the characterization of protein dynamics[119,120], has a unique role to play (Figure 1.6, NMR). First, NMR can be used to measure dynamics at atomic resolution, comprehensively for sugar, base, and backbone moieties across different residues (Figure 1.6, probes). Second, multiple interactions can be measured at a given site to deduce structural, kinetic, and thermodynamic characteristics of not one, but many motional modes occurring at different timescales. Third, NMR enjoys broad sensitivity to motions spanning picosecond to second and longer timescales (Figure 1.6, NMR) and can be used to characterize very subtle changes in conformation, including those involving minutely populated conformers (on the order of $\sim 10^{-7}$ %) that have exceptionally short lifetimes (on the order of nanoseconds). Last but not least, NMR provides a powerful approach for exploring how the dynamic structure landscape is modulated by cellular cues and time-resolved methods can be used to follow these perturbations in real time.

As shown in Figure 1.6, there are a multitude of NMR experiments that can be used to characterize structural dynamics spanning picosecond-minute timescales. Each NMR technique has a timescale sweet spot for which it is applicable. Fast motions occurring at ps-ns timescales can be quantified using spin relaxation experiments where the amplitude and timescale of motions can be determined. Dynamical transitions occurring over ps-ms timescales can be characterized using residual dipolar couplings. Here the amplitude of motions can be determined, but unfortunately the absolute timescale of the motions is unknown. Experiments that probe exchange events occurring on the μs timescale or slower commonly take advantage of the difference in Larmor frequencies between

exchanging states. Here, structural transitions result in inequivalent electronic environments of nuclei making their corresponding chemical shifts different. Relaxation dispersion experiments are useful for quantifying processes that occur on μs-ms timescales where the thermodynamics and kinetics of the exchange process can be determined. Transitions occurring on the ms-s timescale can be quantified using ZZ Exchange experiments where again the thermodynamics and kinetics of the exchange process can be determined. Conformational transitions that occur at second or slower timescales can be quantified using time-resolved experiments. Here, a system in equilibrium is typically chemically perturbed and the race to the new equilibrium state is monitored in real time.

For the purposes of this study we focus primarily on relaxation dispersion techniques to probe lowly populated conformational states. Sometimes biomolecules transition into conformational sub-states, often referred to as 'excited states', that populate different local minima along the energy landscape, and that require loss of key stabilizing interactions that are partially restored by formation of new ones (Figure 1.6, landscape). These excited states are not to be confused with electronic excited states, rather they are conformational states higher in free energy than the ground state conformation. NMR relaxation dispersion techniques provide a rare opportunity to characterize the population, lifetime, and conformation of these lowly populated (as low as 1%) and transient (lifetimes < 1 ms) excited states[121,122].

To understand relaxation dispersion, consider a nucleus that exchanges between a major ground (G) and minor (E) excited state with chemical shifts, ω_G and ω_E, respectively. In the absence of exchange, two NMR peaks are observed with chemical shifts ω_G and ω_E with integrated volumes that reflect their relative populations (Figure 1.7a). Now consider what happens when G and E states exchange at a rate ($k_{ex} = k_{GE} + k_{EG}$) comparable to their NMR frequency difference ($\Delta\omega = \omega_G - \omega_E$). In this case, the frequency of a given nucleus stochastically fluctuates between ω_G and ω_E. Because nuclei in different molecules spend varying amounts of time in the G and E state, their magnetization no longer precess in synchrony

(Figure 1.7b). The resulting 'fanning out' of the magnetization leads to a decay in the net oscillating signal and an additional exchange contribution (R_{ex}) to transverse relaxation (R_2) (Figure 1.7c). The ensuing line broadening of signals can result in the total disappearance of the E signal (Figure 1.7a). Relaxation dispersion experiments probe the invisible E state by measuring the exchange broadening contribution to the visible G state signal following the application of a series of RF pulses in the Carr-Purcell-Meiboom-Gill (CPMG) experiment, or constant RF field in the $R_{1\rho}$ experiment (Figure 1.7b). For example, in the CPMG experiment, a series of 180º pulses reverse the precession of magnetization at constant time intervals, τ_{cp}; as a result, any dephasing that accrues prior to the 180° pulse is partly refocused in the following period, with the extent of refocusing increasing with shorter τ_{cp} delays (Figure 1.7b). The exchange contribution is measured as a function of τ_{cp} in CPMG and the power/frequency of RF field in $R_{1\rho}$. The resulting relaxation dispersion curve (Figure 1.7c) is typically fit to a two-state model. For slow ($k_{ex} << |\Delta\omega|$) to intermediate ($k_{ex} \sim |\Delta\omega|$) exchange, this yields the population (p_E), lifetime ($\tau_{ex} = 1/k_{ex}$), and chemical shift (ω_E) of the excited state, the latter carrying structural information, whereas for fast exchange ($k_{ex} >> |\Delta\omega|$), only τ_{ex} and $\Phi=p_G p_E \Delta\omega^2$ can be determined and additional experiments are needed to resolve $\Delta\omega$ and p_E. The CPMG experiment only yields the absolute difference in chemical shift $|\omega_G - \omega_E|$, and additional experiments[123,124] are typically needed to determine ω_E.

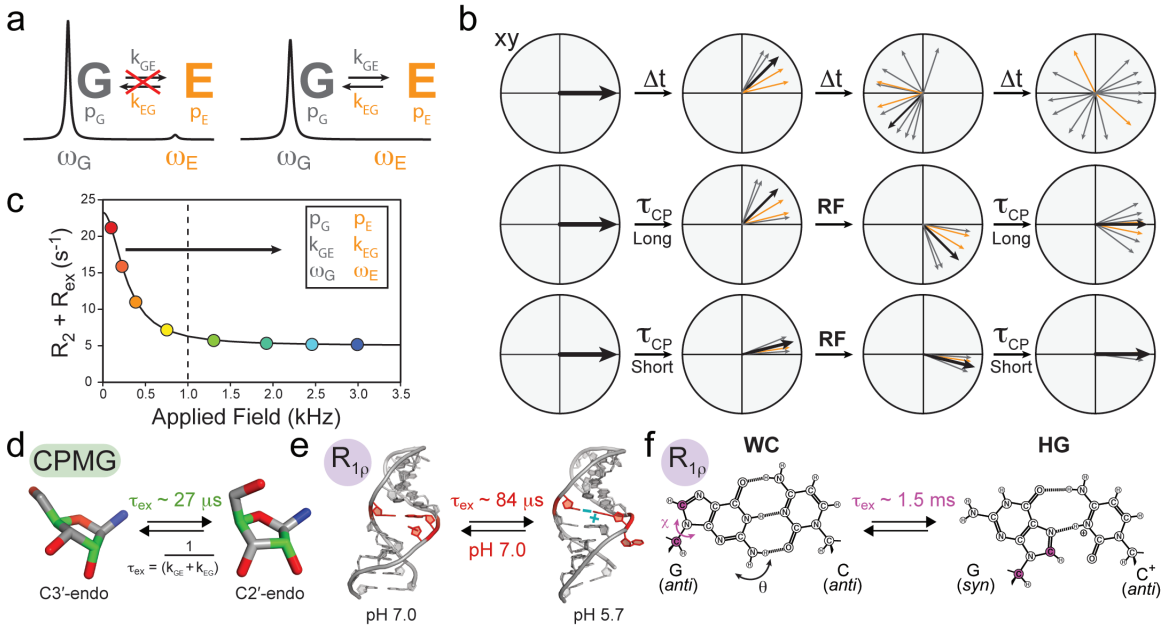

Figure 1.7: Characterizing micro-to-millisecond exchange by relaxation dispersion. (a) Exchange between ground (G) and excited (E) states leads to broadening of the G signal and disappearance of the E signal. (b) Fanning out of bulk magnetization due to exchange can be suppressed by application of RF fields. (c) Characteristic relaxation dispersion curve showing the power dependence of R_{ex} which can be used to extract in favorable cases the populations (p), exchange rates (k) and chemical shifts (ω) of G and E. (d-f) Examples of conformational exchange characterized by CPMG and $R_{1\rho}$ carbon relaxation dispersion. (d) Sugar repuckering in a GCAA tetraloop using selectively labeled C2' and C4' (green) probes[125]. (e) Transition toward an excited state structure in the U6 RNA involving a CA^+ base-pair and looping out of a uridine bulge[126]. (f) Transition to Hoogsteen (HG) GC^+ base-pairs in canonical duplex DNA (Adapted from ref 144).

CPMG relaxation dispersion can be used to probe exchange processes occurring at microsecond to tens of millisecond timescales[121,122]. The CPMG experiment proves difficult to apply to nucleic acids because ^{13}C experiments are complicated by extensive ^{13}C-^{13}C scalar couplings that are difficult to suppress due to challenges in achieving selective carbon excitation with hard 180° pulses[125,127]. This problem was recently addressed by preparing RNA samples that are specifically ^{13}C labeled at the C2' and C4' sugar positions[125] (Figure 1.7d, green). CPMG experiments on these samples led to the observation of excited state sugar conformations in the GCAA tetraloop with populations of 15-30% and lifetimes of

30-42 microseconds that most likely correspond to exchange between C3'-endo and C2'-endo sugar pucker conformations[125] (Figure 1.7d).

The $R_{1\rho}$ experiment has been more widely used than the CPMG experiment in studies of nucleic acids because spin lock fields can be used more selectively to reduce or eliminate unwanted ^{13}C-^{13}C interactions. Additional benefits include the ability to directly determine ω_E at a single magnetic field strength without the need for additional experiments[128]. The upper limit for motional timescale is comparable to, if not greater than that of the CPMG experiment (tens to hundreds of microseconds) and is limited by the amount of RF power that can be dissipated into the probe. While technical considerations have traditionally limited the lowest RF spin lock strength to ~1,000 Hz (Figure 1.7c) and the timescales to exchange processes faster than ~500 microseconds, recent advances permit use of much lower spin lock fields[129-131] (on the order of 100 Hz) extending the timescale sensitivity to tens of milliseconds.

The $R_{1\rho}$ relaxation dispersion experiment measures line broadening as a function of the power and frequency of a constant RF spin lock field[132]. The $R_{1\rho}$ relaxation constant is determined by measuring the decay of the projection of the magnetization about the applied magnetic field (ω_{eff}) as a function of time (Figure 1.8a). In an $R_{1\rho}$ experiment, the magnetization is aligned about an applied RF field ω_{eff} that is determined by the amplitude of the applied pulse (ω_1) and the frequency offset at which the pulse is applied (Ω). Here,

$$R_{1\rho} = R_1 \cos^2\theta + R_2 \sin^2\theta \tag{1}$$

with,

$$\theta = \tan^{-1}\left(\frac{\omega_1}{\Delta\Omega}\right), \tag{2}$$

$$\Delta\Omega = \overline{\Omega} - \omega_{rf}, \tag{3}$$

$$\overline{\Omega} = p_G\Omega_G + p_E\Omega_E, \tag{4}$$

where ω_{rf} is the frequency of the applied field, Ω_G is the resonance frequency of the ground state, Ω_E is the resonance frequency of the excited state, and R_1 and R_2 are the longitudinal and transverse relaxation constants, respectively. In the limit at

which the applied magnetic field is way off resonance ($\Delta\Omega >> 0$) from the ground state $R_{1\rho}$ is essentially a measure of R_1 whereas if the magnetic field is applied on resonance ($\Delta\Omega = 0$) with the ground state, $R_{1\rho}$ is effectively an R_2 measurement.

When conformational exchange occurs and a nucleus' Larmor frequency changes during the spin lock period, another component, R_{ex}, must be accounted for in $R_{1\rho}$. Here $R_{1\rho}$ becomes,

$$R_{1\rho} = R_1 \cos^2\theta + (R_2 + R_{ex})\sin^2\theta \tag{5}$$

to account for exchange. In the $R_{1\rho}$ experiment, the magnetization of the ground and excited states is initially aligned with their respective effective fields. During the duration of the spin lock period, some molecules in the ground state convert to the excited state and some in the excited state convert to the ground state. Once a spin has switched conformational states, its magnetization begins to precess about its new corresponding effective field (ω_E or ω_G). As these states exchange back and forth, the overall projection of the magnetization about the effective field is reduced due to the dephasing of the magnetization as a result of exchange. Shown in Figure 1.8 is an example of an exchanging system under the influence of two spin locks with different amplitudes (ω_1). Application of a weak spin lock (ω_1 = 100 Hz) results in significant dephasing of the locked magnetization due to the differences of the exchanging effective fields. For weak spin locks there is a large angular difference between the G and E effective fields and their corresponding frequencies of rotation. Application of a stronger spin lock (ω_1 = 3000 Hz) results in less dephasing, and correspondingly the effects from conformational exchange are suppressed in $R_{1\rho}$ (Figure 1.8b,c).

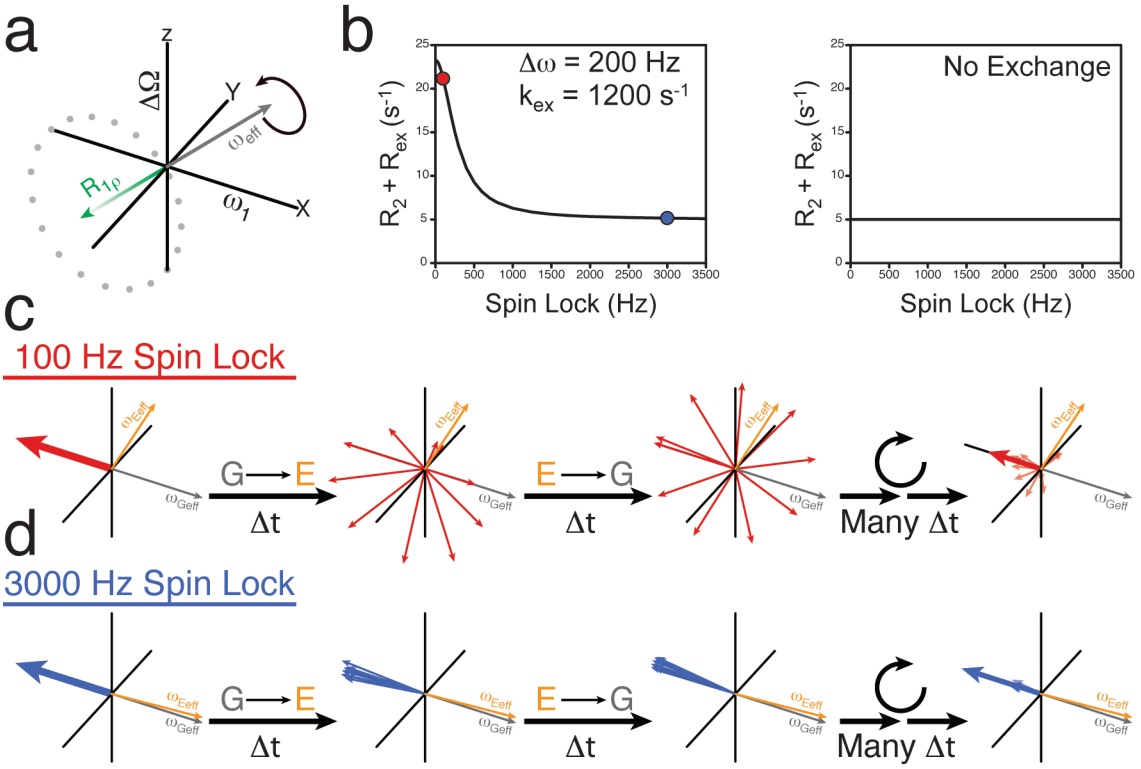

Figure 1.8: Graphical representation of $R_{1\rho}$ relaxation dispersion. a) $R_{1\rho}$ relaxation is the rate of relaxation of the magnetization (green) aligned about an effective field ω_{eff}. Any magnetization that is not aligned about ω_{eff} will precess about the applied field ω_{eff} (indicated with circles). b) $R_{1\rho}$ dispersion profiles for a system undergoing millisecond conformational exchange (left) and no conformational exchange (right). c) Enhancement of the magnetization decay from conformational exchange under the influence of a weak (100 Hz) spin lock field. d) Suppression of exchange effects on the magnetization decay from conformational exchange under the influence of a stronger (3000 Hz) spin lock field.

The evolution of the magnetization under the influence of a spin lock can be described by the Bloch-McConnell equations[128,133,134]. This set of differential equations describes the time evolution for the ground (M_{gx}, M_{gy}, M_{gz}) and excited state (M_{ex}, M_{ey}, M_{ez}) magnetization components as a function of time,

$$\frac{d}{dt}\begin{pmatrix} M_{Gx} \\ M_{Ex} \\ M_{Gy} \\ M_{Ey} \\ M_{Gz} \\ M_{Ez} \end{pmatrix} = \begin{pmatrix} -k_{GE}-R_2 & k_{EG} & -\delta_G & 0 & 0 & 0 \\ k_{GE} & -k_{EG}-R_2 & 0 & -\delta_E & 0 & 0 \\ \delta_G & 0 & -k_{GE}-R_2 & k_{EG} & -\omega_1 & 0 \\ 0 & \delta_E & k_{GE} & -k_{EG}-R_2 & 0 & -\omega_1 \\ 0 & 0 & \omega_1 & 0 & -k_{GE}-R_1 & k_{EG} \\ 0 & 0 & 0 & \omega_1 & k_{GE} & -k_{EG}-R_1 \end{pmatrix} \begin{pmatrix} M_{Gx} \\ M_{Ex} \\ M_{Gy} \\ M_{Ey} \\ M_{Gz} \\ M_{Ez} \end{pmatrix} + R_1 \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ M_{G0} \\ M_{E0} \end{pmatrix} \quad (6)$$

with,

$$\delta_G = \Omega_G - \omega_{rf}, \tag{7}$$

$$\delta_E = \Omega_E - \omega_{rf}, \tag{8}$$

, k_{GE} is the rate constant for ground to excited state transitions, k_{EG} is the rate constant for excited to ground state transitions, ω_1 is the amplitude of the applied field (s^{-1}), and ω_{rf} is the frequency (s^{-1}) at which the spin lock is applied.

The Bloch-McConnell equations fully describe the evolution of the magnetization and the resulting $R_{1\rho}$, however they are not practical for determining exchange parameters (k_{ex}, p_E, $\Delta\omega$) from experimental data due to the computational cost of solving Equation 6. Multiple simple algebraic equations which are more amenable for fitting experimental data have been derived that accurately predict $R_{1\rho}$ under certain conditions[132]. The first such equation derived was the fast exchange equation[135],

$$R_{1\rho} = R_1 \cos^2\theta + R_2 \sin^2 + \sin^2\left(\frac{\Phi k_{ex}}{\omega_{eff}^2 + k_{ex}^2}\right), \tag{9}$$

with,

$$\Phi = p_G p_E \Delta\omega^2, \tag{10}$$

$$\Delta\omega = \Omega_G - \Omega_E. \tag{11}$$

Here the timescale of exchange, k_{ex}, can be elucidated but the highly desirable excited state parameters of chemical shift and population cannot be determined due to their coupling in Φ without additional knowledge of the system. Therefore, this equation is useful for estimating the rate of exchange when the system is in fast exchange ($k_{ex} >> \Delta\omega$).

Palmer and coworkers have derived a number of more complicated algebraic equations that are accurate for exchange scenarios in the intermediate and slow exchange regimes ($k_{ex} \leq \Delta\omega$)[128,136-138]. The most versatile equation was derived by determining the characteristic sixth-order polynomial of the Bloch-McConnell equations and then using Laguerre's method for polynomial root finding[138]. Here,

$$R_{1\rho} = R_1 \cos^2\theta + R_2 \sin^2\theta + \frac{\sin^2\theta p_G p_E \Delta\omega^2 k_{ex}}{\omega_G^2\omega_E^2/\omega_{eff}^2 + k_{ex}^2 - \sin^2\theta p_G p_E \Delta\omega^2\left(1 + \frac{2k_{ex}^2(p_G\omega_G^2 + p_E\omega_E^2)}{\omega_G^2\omega_E^2 + \omega_{eff}^2 k_{ex}^2}\right)} \quad (12)$$

with,

$$\omega_{eff}^2 = \Delta\Omega^2 + \omega_1^2, \quad (13)$$

$$\omega_G^2 = (\Omega_G - \omega_{rf})^2 + \omega_1^2, \quad (14)$$

$$\omega_E^2 = (\Omega_E - \omega_{rf})^2 + \omega_1^2. \quad (15)$$

Equation 12, which we will refer to as the Laguerre equation accurately calculates $R_{1\rho}$ consistent with the Bloch-McConnell equations when $p_E \leq 30\%$[138]. If $p_E \geq 30\%$, caution should be used when interpreting data as the accuracy of the Laguerre is reduced. With these simple algebraic expressions in hand, experimental data can be rapidly fit to determine exchange parameters.

Some of the earliest experimental $R_{1\rho}$ studies on nucleic acids revealed microsecond timescale exchange processes involving an excited state CA^+ base-pair within the active site of the lead-dependent ribozyme that may be important for catalysis[139]. More recent studies suggest that excited states that are coupled to protonation of nucleobases may be quite common. For example, Blad *et al.* uncovered a microsecond timescale exchange processes directed towards an excited state structure involving a CA^+ base-pair within a functionally important region of the U6 RNA element from the spliceosome which is accompanied by the looping out of a uridine bulge[126,140] (Figure 1.7e), whereas Nikolova *et al.* uncovered an even slower millisecond exchange process directed towards excited state GC^+ Hoogsteen base-pairs in canonical duplex DNA[141] (Figure 1.7f). The above two studies provide rare examples in which the structure of the excited state could be determined. In both cases, this was accomplished by trapping the excited state, either by lowering the pH or by introducing chemical modifications. With these structures in hand, the authors were able to gain insights into the conformational pathway linking the ground and excited states using various computational methods[141,142].

Because the relationship between chemical shift and nucleic acid structure is not fully developed, it is generally not feasible to determine nucleic acid excited state structures *de novo* by relaxation dispersion. While this has made it difficult to directly link observed micro- to millisecond motions to function, potential roles have been proposed. For example, Zhang *et al.* observed a concerted exchange process in the $PreQ_1$-binding riboswitch involving three contiguous residues that form a sharp turn that may provide a dynamic lid for entry of the $PreQ_1$ ligand into the deep binding pocket of the aptamer domain[143]. Micro- to millisecond motions were also observed in two universally conserved ribosomal A-site internal loop adenines that were interpreted to potentially reflect the looping in and out motions needed to proof read the mRNA message during translation[131]. Taken together, these studies establish excited states as a ubiquitous feature of the nucleic acid dynamic structure landscape. Further studies are however required to structurally characterize these states and more firmly establish their link to function.

One of the goals of this dissertation is to utilize relaxation dispersion experiments to identify dynamic sites in DNA that have a high propensity to form non-canonical conformations. First, we carry out simulations to determine conditions that are necessary to obtain a full quantitative characterization of the populations and chemical shifts of excited states from $R_{1\rho}$ relaxation dispersion experiments. We then apply dispersion experiments to identify dynamic sites in DNA with a high propensity to form non-canonical conformations. With a residue level understanding of sites with high conformational flexibility in hand, we test how a change in local stability influences the formation of highly distorted conformations, specifically B/Z junctions. We also develop a combined CD and fluorescence spectroscopic approach to further understand the formation of Z-DNA when coupled to the creation of B-Z junctions. Finally, studies of supercoiled DNA are initiated and experiments are carried out to begin to understand how sequence specific mutations, shown to have a dramatic effect on relaxed linear DNA, influence the properties of supercoiled DNA.

Some of the NMR portion of this chapter was published in a Nature Methods Review[144].

1.6 References

1. R. Dahm, Discovering DNA: Friedrich Miescher and the early years of nucleic acid research. *Hum Genet* **122**, 565 (2008).
2. L. Pauling, R. B. Corey, A Proposed Structure For The Nucleic Acids. *Proc Natl Acad Sci U S A* **39**, 84 (1953).
3. J. D. Watson, F. H. Crick, Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**, 737 (1953).
4. "At Caltech with Linus Pauling", Cold Spring Harbor Laboratory Oral History Collection, Cold Springs Harbor, http://library.cshl.edu/oralhistory/interview/scientific-experience/advice-young-scientists/caltech-linus-pauling/, 2012.
5. A. H. Wang *et al.*, Molecular structure of a left-handed double helical DNA fragment at atomic resolution. *Nature* **282**, 680 (1979).
6. F. M. Pohl, T. M. Jovin, Salt-induced co-operative conformational change of a synthetic DNA: equilibrium and kinetic studies with poly (dG-dC). *J Mol Biol* **67**, 375 (1972).
7. F. M. Pohl, A. Ranade, Stockbur.M, Laser Raman-Scattering of 2 Double-Helical Forms of Poly(Dg-Dc). *Biochim Biophys Acta* **335**, 85 (1974).
8. T. J. Thamann, R. C. Lord, A. H. Wang, A. Rich, The high salt form of poly(dG-dC).poly(dG-dC) is left-handed Z-DNA: Raman spectra of crystals and solutions. *Nucleic Acids Res* **9**, 5443 (1981).
9. A. Rich, A. Nordheim, A. H. J. Wang, The Chemistry and Biology of Left-Handed Z-DNA. *Annu Rev Biochem* **53**, 791 (1984).
10. A. Nordheim *et al.*, Negatively supercoiled plasmids contain left-handed Z-DNA segments as detected by specific antibody binding. *Cell* **31**, 309 (1982).
11. D. B. Haniford, D. E. Pulleyblank, Facile transition of poly[d(TG) x d(CA)] into a left-handed helix in physiological conditions. *Nature* **302**, 632 (1983).
12. J. Feigon, A. H. Wang, G. A. van der Marel, J. H. van Boom, A. Rich, Z-DNA forms without an alternating purine-pyrimidine sequence in solution. *Science* **230**, 82 (1985).
13. L. J. Peck, A. Nordheim, A. Rich, J. C. Wang, Flipping of cloned d(pCpG)n.d(pCpG)n DNA sequences from right- to left-handed helical structure by salt, Co(III), or negative supercoiling. *Proc Natl Acad Sci U S A* **79**, 4560 (1982).
14. A. Rich, S. Zhang, Timeline: Z-DNA: the long road to biological function. *Nat Rev Genet* **4**, 566 (2003).
15. S. C. Ha, K. Lowenhaupt, A. Rich, Y. G. Kim, K. K. Kim, Crystal structure of a junction between B-DNA and Z-DNA reveals two extruded bases. *Nature* **437**, 1183 (2005).
16. C. K. Singleton, J. Klysik, S. M. Stirdivant, R. D. Wells, Left-handed Z-DNA is induced by supercoiling in physiological ionic conditions. *Nature* **299**, 312 (1982).

17. C. K. Singleton, J. Klysik, R. D. Wells, Conformational Flexibility of Junctions between Contiguous B-Dnas and Z-Dnas in Supercoiled Plasmids. *P Natl Acad Sci-Biol* **80**, 2447 (1983).
18. M. W. Kilpatrick, C. F. Wei, H. B. Gray, Jr., R. D. Wells, BAL 31 nuclease as a probe in concentrated salt for the B-Z DNA junction. *Nucleic Acids Res* **11**, 3811 (1983).
19. R. D. Wells *et al.*, Left-handed DNA helices, supercoiling, and the B-Z junction. *Cold Spring Harb Symp Quant Biol* **47 Pt 1**, 77 (1983).
20. J. Klysik, S. M. Stirdivant, J. E. Larson, P. A. Hart, R. D. Wells, Left-Handed DNA in Restriction Fragments and a Recombinant Plasmid. *Nature* **290**, 672 (1981).
21. R. M. Wartell, J. Klysik, W. Hillen, R. D. Wells, Junction between Z-Conformation and B-Conformation in a DNA Restriction Fragment - Evaluation by Raman-Spectroscopy. *P Natl Acad Sci-Biol* **79**, 2549 (1982).
22. R. D. Sheardy, Preliminary Spectroscopic Characterization of a Synthetic DNA Oligomer Containing a B-Z Junction at High Salt. *Nucleic Acids Res* **16**, 1153 (1988).
23. R. D. Sheardy, S. A. Winkle, Temperature-Dependent Cd and Nmr-Studies on a Synthetic Oligonucleotide Containing a B-Z Junction at High Salt. *Biochemistry-Us* **28**, 720 (1989).
24. Z. Dai, G. A. Thomas, E. Evertsz, W. L. Peticolas, The Length of a Junction between the B-Conformations and Z-Conformations in DNA Is 3 Base-Pairs or Less. *Biochemistry-Us* **28**, 6991 (1989).
25. D. Kim *et al.*, Base extrusion is found at helical junctions between right- and left-handed forms of DNA and RNA. *Nucleic Acids Res* **37**, 4353 (2009).
26. M. de Rosa *et al.*, Crystal structure of a junction between two Z-DNA helices. *Proc Natl Acad Sci U S A* **107**, 9088 (2010).
27. A. Nordheim, A. Rich, The sequence (dC-dA)n X (dG-dT)n forms left-handed Z-DNA in negatively supercoiled plasmids. *Proc Natl Acad Sci U S A* **80**, 1821 (1983).
28. L. J. Peck, J. C. Wang, Energetics of B-to-Z transition in DNA. *Proc Natl Acad Sci U S A* **80**, 6206 (1983).
29. M. D. Frank-Kamenetskii, A. V. Vologodskii, Thermodynamics of the B-Z transition in superhelical DNA. *Nature* **307**, 481 (1984).
30. A. V. Vologodskii, M. D. Frank-Kamenetskii, Left-handed Z form in superhelical DNA: a theoretical study. *J Biomol Struct Dyn* **1**, 1325 (1984).
31. M. J. Ellison *et al.*, An assessment of the Z-DNA forming potential of alternating dA-dT stretches in supercoiled plasmids. *Biochemistry-Us* **25**, 3648 (1986).
32. M. J. Ellison, R. J. Kelleher, 3rd, A. H. Wang, J. F. Habener, A. Rich, Sequence-dependent energetics of the B-Z transition in supercoiled DNA containing nonalternating purine-pyrimidine sequences. *Proc Natl Acad Sci U S A* **82**, 8320 (1985).
33. P. S. Ho, M. J. Ellison, G. J. Quigley, A. Rich, A computer aided thermodynamic approach for predicting the formation of Z-DNA in naturally occurring sequences. *The EMBO journal* **5**, 2737 (1986).

34. R. D. Sheardy *et al.*, Sequence Dependence of the Free-Energy of B-Z Junction Formation in Deoxyoligonucleotides. *J Mol Biol* **231**, 475 (1993).
35. G. P. Schroth, P. J. Chou, P. S. Ho, Mapping Z-DNA in the Human Genome - Computer-Aided Mapping Reveals a Nonrandom Distribution of Potential Z-DNA-Forming Sequences in Human Genes. *J Biol Chem* **267**, 11846 (1992).
36. H. Li *et al.*, Human genomic Z-DNA segments probed by the Z domain of ADAR1. *Nucleic Acids Res* **37**, 2737 (2009).
37. P. S. Ho, Thermogenomics: thermodynamic-based approaches to genomic analyses of DNA structure. *Methods* **47**, 159 (2009).
38. D. Zhabinskaya, C. J. Benham, Theoretical Analysis of the Stress Induced B-Z Transition in Superhelical DNA. *Plos Comput Biol* **7**, (2011).
39. F. Kouzine, D. Levens, Supercoil-driven DNA structures regulate genetic transactions. *Front Biosci* **12**, 4409 (2007).
40. A. V. Vologodski*i, *Topology and physics of circular DNA*. (CRC Press, Boca Raton, 1992), pp. 179 p.
41. A. D. Bates, A. Maxwell, *DNA topology*. (Oxford University Press, Oxford ; New York, ed. 2nd, 2005), pp. xviii, 198 p.
42. D. B. Haniford, D. E. Pulleyblank, The in-vivo occurrence of Z DNA. *J Biomol Struct Dyn* **1**, 593 (1983).
43. A. R. Rahmouni, R. D. Wells, Stabilization of Z DNA in vivo by localized supercoiling. *Science* **246**, 358 (1989).
44. B. Wittig, T. Dorbic, A. Rich, The level of Z-DNA in metabolically active, permeabilized mammalian cell nuclei is regulated by torsional strain. *J Cell Biol* **108**, 755 (1989).
45. M. Lee, S. H. Kim, S. C. Hong, Minute negative superhelicity is sufficient to induce the B-Z transition in the presence of low tension. *Proc Natl Acad Sci U S A* **107**, 4985 (2010).
46. A. Nordheim *et al.*, Antibodies to left-handed Z-DNA bind to interband regions of Drosophila polytene chromosomes. *Nature* **294**, 417 (1981).
47. H. J. Lipps *et al.*, Antibodies against Z DNA react with the macronucleus but not the micronucleus of the hypotrichous ciliate stylonychia mytilus. *Cell* **32**, 435 (1983).
48. F. Lancillotti, M. C. Lopez, P. Arias, C. Alonso, Z-DNA in transcriptionally active chromosomes. *Proc Natl Acad Sci U S A* **84**, 1560 (1987).
49. L. F. Liu, J. C. Wang, Supercoiling of the DNA template during transcription. *Proc Natl Acad Sci U S A* **84**, 7024 (1987).
50. B. Wittig, T. Dorbic, A. Rich, Transcription is associated with Z-DNA formation in metabolically active permeabilized mammalian cell nuclei. *Proc Natl Acad Sci U S A* **88**, 2259 (1991).
51. B. Wittig, S. Wolfl, T. Dorbic, W. Vahrson, A. Rich, Transcription of human c-myc in permeabilized nuclei is associated with formation of Z-DNA in three discrete regions of the gene. *The EMBO journal* **11**, 4653 (1992).
52. V. Muller, M. Takeya, S. Brendel, B. Wittig, A. Rich, Z-DNA-forming sites within the human beta-globin gene cluster. *Proc Natl Acad Sci U S A* **93**, 780 (1996).
53. G. P. Schroth, P. J. Chou, P. S. Ho, Mapping Z-DNA in the human genome. Computer-aided mapping reveals a nonrandom distribution of potential Z-

DNA-forming sequences in human genes. *The Journal of biological chemistry* **267**, 11846 (1992).
54. J. Nickol, M. Behe, G. Felsenfeld, Effect of the B--Z transition in poly(dG-m5dC) . poly(dG-m5dC) on nucleosome formation. *Proc Natl Acad Sci U S A* **79**, 1771 (1982).
55. J. Ausio, G. Zhou, K. van Holde, A reexamination of the reported B----Z DNA transition in nucleosomes reconstituted with poly(dG-m5dC).poly(dG-m5dC). *Biochemistry-Us* **26**, 5595 (1987).
56. J. M. Casasnovas, F. Azorin, Supercoiled induced transition to the Z-DNA conformation affects the ability of a d(CG/GC)12 sequence to be organized into nucleosome-cores. *Nucleic Acids Res* **15**, 8899 (1987).
57. M. M. Garner, G. Felsenfeld, Effect of Z-DNA on nucleosome placement. *J Mol Biol* **196**, 581 (1987).
58. C. Davey, S. Pennings, J. Allan, CpG methylation remodels chromatin structure in vitro. *J Mol Biol* **267**, 276 (1997).
59. K. van Holde, J. Zlatanova, Unusual DNA structures, chromatin and transcription. *Bioessays* **16**, 59 (1994).
60. B. Wong, S. Chen, J. A. Kwon, A. Rich, Characterization of Z-DNA as a nucleosome-boundary element in yeast Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* **104**, 2229 (2007).
61. R. Liu *et al.*, Regulation of CSF1 promoter by the SWI/SNF-like BAF complex. *Cell* **106**, 309 (2001).
62. A. G. Herbert, A. Rich, A method to identify and characterize Z-DNA binding proteins using a linear oligodeoxynucleotide. *Nucleic Acids Res* **21**, 2669 (1993).
63. A. Herbert, K. Lowenhaupt, J. Spitzner, A. Rich, Chicken double-stranded RNA adenosine deaminase has apparent specificity for Z-DNA. *Proc Natl Acad Sci U S A* **92**, 7550 (1995).
64. A. Herbert *et al.*, A Z-DNA binding domain present in the human editing enzyme, double-stranded RNA adenosine deaminase. *Proc Natl Acad Sci U S A* **94**, 8421 (1997).
65. I. Berger *et al.*, Spectroscopic characterization of a DNA-binding domain, Z alpha from the editing enzyme, dsRNA adenosine deaminase: Evidence for left-handed Z-DNA in the Z alpha-DNA complex. *Biochemistry-Us* **37**, 13313 (1998).
66. Y. G. Kim *et al.*, The Zab domain of the human RNA editing enzyme ADAR1 recognizes Z-DNA when surrounded by B-DNA. *J Biol Chem* **275**, 26828 (2000).
67. S. C. Ha *et al.*, The structures of non-CG-repeat Z-DNAs co-crystallized with the Z-DNA-binding domain, hZ alpha(ADAR1). *Nucleic Acids Res* **37**, 629 (2009).
68. S. C. Ha *et al.*, A poxvirus protein forms a complex with left-handed Z-DNA: crystal structure of a Yatapoxvirus Zalpha bound to DNA. *Proc Natl Acad Sci U S A* **101**, 14367 (2004).

69. T. Schwartz, J. Behlke, K. Lowenhaupt, U. Heinemann, A. Rich, Structure of the DLM-1-Z-DNA complex reveals a conserved family of Z-DNA-binding proteins. *Nat Struct Biol* **8**, 761 (2001).
70. P. Barraud, F. H. Allain, ADAR proteins: double-stranded RNA and Z-DNA binding domains. *Curr Top Microbiol Immunol* **353**, 35 (2012).
71. C. E. Samuel, Adenosine deaminases acting on RNA (ADARs) are both antiviral and proviral. *Virology* **411**, 180 (2011).
72. D. Placido, B. A. Brown, 2nd, K. Lowenhaupt, A. Rich, A. Athanasiadis, A left-handed RNA double helix bound by the Z alpha domain of the RNA-editing enzyme ADAR1. *Structure* **15**, 395 (2007).
73. M. Koeris, L. Funke, J. Shrestha, A. Rich, S. Maas, Modulation of ADAR1 editing activity by Z-RNA in vitro. *Nucleic Acids Res* **33**, 5362 (2005).
74. B. A. Brown, 2nd, K. Lowenhaupt, C. M. Wilbert, E. B. Hanlon, A. Rich, The zalpha domain of the editing enzyme dsRNA adenosine deaminase binds left-handed Z-RNA as well as Z-DNA. *Proc Natl Acad Sci U S A* **97**, 13532 (2000).
75. R. Cattaneo, Biased (A-->I) hypermutation of animal RNA virus genomes. *Curr Opin Genet Dev* **4**, 895 (1994).
76. S. M. Horikami, S. A. Moyer, Double-stranded RNA adenosine deaminase activity during measles virus infection. *Virus Res* **36**, 87 (1995).
77. A. G. Polson, B. L. Bass, J. L. Casey, RNA editing of hepatitis delta virus antigenome by dsRNA-adenosine deaminase. *Nature* **380**, 454 (1996).
78. D. R. Taylor, M. Puig, M. E. R. Darnell, K. Mihalik, S. M. Feinstone, New antiviral pathway that mediates hepatitis C virus replicon interferon sensitivity through ADAR1. *J Virol* **79**, 6291 (2005).
79. Y. G. Kim *et al.*, A role for Z-DNA binding in vaccinia virus pathogenesis. *Proc Natl Acad Sci U S A* **100**, 6974 (2003).
80. J. A. Kwon, A. Rich, Biological function of the vaccinia virus Z-DNA-binding protein E3L: gene transactivation and antiapoptotic activity in HeLa cells. *Proc Natl Acad Sci U S A* **102**, 12759 (2005).
81. J. O. Langland *et al.*, Suppression of proinflammatory signal transduction and gene expression by the dual nucleic acid binding domains of the vaccinia virus E3L proteins. *J Virol* **80**, 10083 (2006).
82. Y. Fu *et al.*, Cloning of DLM-1, a novel gene that is up-regulated in activated macrophages, using RNA differential display. *Gene* **240**, 157 (1999).
83. S. C. Ha *et al.*, Biochemical characterization and preliminary X-ray crystallographic study of the domains of human ZBP1 bound to left-handed Z-DNA. *Biochim Biophys Acta* **1764**, 320 (2006).
84. A. Takaoka *et al.*, DAI (DLM-1/ZBP1) is a cytosolic DNA sensor and an activator of innate immune response. *Nature* **448**, 501 (2007).
85. K. Nishikura, Editor meets silencer: crosstalk between RNA editing and RNA interference. *Nat Rev Mol Cell Bio* **7**, 919 (2006).
86. Z. Wang *et al.*, Regulation of innate immune responses by DAI (DLM-1/ZBP1) and other DNA-sensing molecules. *Proc Natl Acad Sci U S A* **105**, 5477 (2008).
87. G. D. Fasman, *Circular dichroism and the conformational analysis of biomolecules*. The language of science (Plenum Press, New York, 1996), pp. ix, 738 p.

88. J. Kypr, I. Kejnovska, D. Renciuk, M. Vorlickova, Circular dichroism and conformational polymorphism of DNA. *Nucleic Acids Res* **37**, 1713 (2009).
89. M. F. Maestre, J. C. Wang, Circular dichroism of superhelical DNA. *Biopolymers* **10**, 1021 (1971).
90. S. Brahms, S. Nakasu, A. Kikuchi, J. G. Brahms, Structural-Changes in Positively and Negatively Supercoiled DNA. *Eur J Biochem* **184**, 297 (1989).
91. N. J. Greenfield, Using circular dichroism spectra to estimate protein secondary structure. *Nat Protoc* **1**, 2876 (2006).
92. M. Behe, G. Felsenfeld, Effects of Methylation on a Synthetic Polynucleotide - the B-Z Transition in Poly(Dg-M5dc).Poly(Dg-M5dc). *P Natl Acad Sci-Biol* **78**, 1619 (1981).
93. J. H. Vandesande, L. P. Mcintosh, T. M. Jovin, Mn-2+ and Other Transition-Metals at Low Concentration Induce the Right-to-Left Helical Transformation of Poly[D(G-C)]. *Embo J* **1**, 777 (1982).
94. A. Johnson, Y. Qu, B. Vanhouten, N. Farrell, B-]Z DNA Conformational-Changes Induced by a Family of Dinuclear Bis(Platinum) Complexes. *Nucleic Acids Res* **20**, 1697 (1992).
95. B. Norden, P. Lincoln, B. Akerman, E. Tuite, DNA interactions with substitution-inert transition metal ion complexes. *Met Ions Biol Syst* **33**, 177 (1996).
96. Z. G. Wu *et al.*, Formation of Sequence-Independent Z-DNA Induced by a Ruthenium Complex at Low Salt Concentrations. *Angew Chem Int Edit* **50**, 11962 (2011).
97. M. A. Fuertes, V. Cepeda, C. Alonso, J. M. Perez, Molecular mechanisms for the B-Z transition in the example of poly[d(G-C) x d(G-C)] polymers. A critical review. *Chem Rev* **106**, 2045 (2006).
98. V. M. Gonzalez, M. A. Fuertes, J. M. Perez, C. Alonso, Kinetics of the salt-induced B- to Z-DNA transition. *Eur Biophys J Biophy* **27**, 417 (1998).
99. V. Ivanov, K. Grzeskowiak, G. Zocchi, Evidence for an intermediate state in the B-to-Z transition of DNA. *J Phys Chem B* **107**, 12847 (2003).
100. D. P. Millar, Fluorescence studies of DNA and RNA structure and dynamics. *Curr Opin Struc Biol* **6**, 322 (1996).
101. M. J. Rist, J. P. Marino, Fluorescent nucleotide base analogs as probes of nucleic acid structure, dynamics and interactions. *Curr Org Chem* **6**, 775 (2002).
102. R. W. Sinkeldam, N. J. Greco, Y. Tor, Fluorescent analogs of biomolecular building blocks: design, properties, and applications. *Chem Rev* **110**, 2579 (2010).
103. D. C. Ward, E. Reich, L. Stryer, Fluorescence Studies of Nucleotides and Polynucleotides .I. Formycin 2-Aminopurine Riboside 2,6-Diaminopurine Riboside and Their Derivatives. *J Biol Chem* **244**, 1228 (1969).
104. H. W. Lee, K. T. Briggs, J. P. Marino, Dissecting structural transitions in the HIV-1 dimerization initiation site RNA using 2-aminopurine fluorescence. *Methods* **49**, 118 (2009).

105. T. M. Nordlund *et al.*, Structure and Dynamics of a Fluorescent DNA Oligomer Containing the Ecori Recognition Sequence - Fluorescence, Molecular-Dynamics, and Nmr-Studies. *Biochemistry-Us* **28**, 9095 (1989).
106. A. Dallmann *et al.*, 2-Aminopurine Incorporation Perturbs the Dynamics and Structure of DNA. *Angew Chem Int Edit* **49**, 5989 (2010).
107. P. O. Lycksell *et al.*, Base pair opening dynamics of a 2-aminopurine substituted Eco RI restriction sequence and its unsubstituted counterpart in oligonucleotides. *Nucleic Acids Res* **15**, 9011 (1987).
108. Y. V. Reddy, D. N. Rao, Binding of EcoP15I DNA methyltransferase to DNA reveals a large structural distortion within the recognition sequence. *J Mol Biol* **298**, 597 (2000).
109. H. Gowher, A. Jeltsch, Molecular enzymology of the EcoRV DNA-(Adenine-N (6))-methyltransferase: kinetics of DNA binding and bending, kinetic mechanism and linear diffusion of the enzyme on DNA. *J Mol Biol* **303**, 93 (2000).
110. C. Beck, A. Jeltsch, Probing the DNA interface of the EcoRV DNA-(adenine-N6)-methyltransferase by site-directed mutagenesis, fluorescence spectroscopy, and UV cross-linking. *Biochemistry-Us* **41**, 14103 (2002).
111. R. Tashiro, H. Sugiyama, A nanothermometer based on the different pi stackings of B- and Z-DNA. *Angew Chem Int Ed Engl* **42**, 6018 (2003).
112. R. Tashiro, H. Sugiyama, Biomolecule-based switching devices that respond inversely to thermal stimuli. *J Am Chem Soc* **127**, 2094 (2005).
113. S. C. Ha, K. Lowenhaupt, A. Rich, Y. G. Kim, K. K. Kim, Crystal structure of a junction between B-DNA and Z-DNA reveals two extruded bases. *Nature* **437**, 1183 (2005).
114. H. Frauenfelder, S. G. Sligar, P. G. Wolynes, The Energy Landscapes and Motions of Proteins. *Science* **254**, 1598 (1991).
115. D. D. Boehr, R. Nussinov, P. E. Wright, The role of dynamic conformational ensembles in biomolecular recognition. *Nat Chem Biol* **5**, 789 (2009).
116. N. Leulliot, G. Varani, Current topics in RNA-protein recognition: Control of specificity and biological function through induced fit and conformational capture. *Biochemistry-Us* **40**, 7947 (2001).
117. H. M. Al-Hashimi, N. G. Walter, RNA dynamics: it is about time. *Curr Opin Struc Biol* **18**, 321 (2008).
118. M. F. Perutz, F. S. Mathews, An x-ray study of azide methaemoglobin. *J Mol Biol* **21**, 199 (1966).
119. A. G. Palmer, NMR characterization of the dynamics of biomacromolecules. *Chem Rev* **104**, 3623 (2004).
120. A. Mittermaier, L. E. Kay, Review - New tools provide new insights in NMR studies of protein dynamics. *Science* **312**, 224 (2006).
121. A. G. Palmer, C. D. Kroenke, J. P. Loria, Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Method Enzymol* **339**, 204 (2001).
122. D. M. Korzhnev, L. E. Kay, Probing invisible, low-populated states of protein molecules by relaxation dispersion NMR spectroscopy: An application to protein folding. *Accounts Chem Res* **41**, 442 (2008).

123. N. R. Skrynnikov, F. W. Dahlquist, L. E. Kay, Reconstructing NMR spectra of "invisible" excited protein states using HSQC and HMQC experiments. *J Am Chem Soc* **124**, 12352 (2002).
124. R. Auer *et al.*, Measurement of signs of chemical shift differences between ground and excited protein states: a comparison between H(S/M)QC and R-1 rho methods. *J Biomol NMR* **46**, 205 (2010).
125. J. E. Johnson, Jr., C. G. Hoogstraten, Extensive backbone dynamics in the GCAA RNA tetraloop analyzed using 13C NMR spin relaxation and specific isotope labeling. *J Am Chem Soc* **130**, 16757 (2008).
126. H. Blad, N. J. Reiter, F. Abildgaard, J. L. Markley, S. E. Butcher, Dynamics and metal ion binding in the U6 RNA intramolecular stem-loop as analyzed by NMR. *J Mol Biol* **353**, 540 (2005).
127. T. Yamazaki, R. Muhandiram, L. E. Kay, Nmr Experiments for the Measurement of Carbon Relaxation Properties in Highly Enriched, Uniformly C-13,N-15-Labeled Proteins - Application to C-13(Alpha) Carbons. *J Am Chem Soc* **116**, 8266 (1994).
128. O. Trott, A. G. Palmer, R-1 rho relaxation outside of the fast-exchange limit. *J Magn Reson* **154**, 157 (2002).
129. F. Massi, E. Johnson, C. Y. Wang, M. Rance, A. G. Palmer, NMR R-1 rho rotating-frame relaxation with weak radio frequency fields. *J Am Chem Soc* **126**, 2247 (2004).
130. D. M. Korzhnev, V. Y. Orekhov, L. E. Kay, Off-resonance R1(p) NMR studies of exchange dynamics in proteins with low spin-lock fields: An application to a fyn SH3 domain. *J Am Chem Soc* **127**, 713 (2005).
131. A. L. Hansen, E. N. Nikolova, A. Casiano-Negroni, H. M. Al-Hashimi, Extending the range of microsecond-to-millisecond chemical exchange detected in labeled and unlabeled nucleic acids by selective carbon R(1rho) NMR spectroscopy. *J Am Chem Soc* **131**, 3818 (2009).
132. A. G. Palmer, F. Massi, Characterization of the dynamics of biomacromolecules using rotating-frame spin relaxation NMR spectroscopy. *Chem Rev* **106**, 1700 (2006).
133. H. M. Mcconnell, Reaction Rates by Nuclear Magnetic Resonance. *J Chem Phys* **28**, 430 (1958).
134. J. Cavanagh, *Protein NMR spectroscopy : principles and practice*. (Academic Press, Amsterdam ; Boston, ed. 2nd, 2007), pp. xxv, 885 p.
135. D. G. Davis, M. E. Perlman, R. E. London, Direct measurements of the dissociation-rate constant for inhibitor-enzyme complexes via the T1 rho and T2 (CPMG) methods. *J Magn Reson B* **104**, 266 (1994).
136. O. Trott, D. Abergel, A. G. Palmer, An average-magnetization analysis of R-1 rho relaxation outside of the fast exchange. *Mol Phys* **101**, 753 (2003).
137. D. Abergel, A. G. Palmer, On the use of the stochastic Liouville equation in nuclear magnetic resonance: Application to R-1 rho relaxation in the presence of exchange. *Concept Magn Reson A* **19A**, 134 (2003).
138. V. Z. Miloushev, A. G. Palmer, R(1p) relaxation for two-site chemical exchange: General approximations and some exact solutions. *J Magn Reson* **177**, 221 (2005).

139. C. G. Hoogstraten, J. R. Wank, A. Pardi, Active site dynamics in the lead-dependent ribozyme. *Biochemistry-Us* **39**, 9951 (2000).
140. N. J. Reiter, H. Blad, F. Abildgaard, S. E. Butcher, Dynamics in the U6 RNA intramolecular stem-loop: a base flipping conformational change. *Biochemistry-Us* **43**, 13739 (2004).
141. E. N. Nikolova *et al.*, Transient Hoogsteen base pairs in canonical duplex DNA. *Nature* **470**, 498 (2011).
142. V. Venditti, L. Clos, N. Niccolai, S. E. Butcher, Minimum-Energy Path for a U6 RNA Conformational Change Involving Protonation, Base-Pair Rearrangement and Base Flipping. *J Mol Biol* **391**, 894 (2009).
143. Q. Zhang, M. Kang, R. D. Peterson, J. Feigon, Comparison of solution and crystal structures of preQ1 riboswitch reveals calcium-induced changes in conformation and dynamics. *J Am Chem Soc* **133**, 5190 (2011).
144. J. R. Bothe *et al.*, Characterizing RNA dynamics at atomic resolution using solution-state NMR spectroscopy. *Nat Methods* **8**, 919 (2011).

Chapter 2

How Fast can $R_{1\rho}$ Go? Determination of the Timescale Limits for Obtaining Robust Exchange Parameters

2.1 Introduction

NMR spectroscopy is a powerful technique for characterizing the kinetics, thermodynamics, and structures of complex biomolecules at atomic resolution. One particular NMR technique, relaxation dispersion, allows for the characterization of lowly populated short-lived conformational states, existing as low as one percent of the time for 10^{-3}-10^{-5} seconds, that are in a dynamic equilibrium with a dominant ground state[1,2]. These relaxation dispersion methods are particularly powerful for characterizing systems undergoing chemical/conformational exchange at rates occurring on the μs-ms timescale and allow for the determination of the kinetics, thermodynamics, and chemical shifts of lowly populated (excited) states. Over the last decade relaxation dispersion methods have evolved significantly in the field of biomolecular dynamics, and recently achieved the pinnacle of solving the structures of conformational states populated less than five percent of the time[3-7]. These studies relied on analysis of populations, rate constants, and chemical shifts obtained from fitting experimental relaxation dispersion data to simple models.

Much effort has been directed towards establishing criteria for obtaining reliable populations, chemical shifts, and rate constants from data fitting both by improving experimental data collection and by testing the limits of theoretical models used for fitting[8-13]. Previous theoretical efforts have primarily focused on Carr-Purcell-Meiboom-Gill (CPMG) based methods. Statistical analysis of CPMG data has resulted in improved experimental strategies for minimizing data collection times and has roughly defined the limits (populations, kinetics, and chemical shifts) of what can be accurately quantified experimentally[10,13]. While CPMG relaxation

dispersion is the preferred technique for proteins, $R_{1\rho}$ methods are much more easily applied to nucleic acids[14]. Although a multitude of algebraic equations have been derived to accurately model $R_{1\rho}$ relaxation dispersion data as a function of applied power (ω_1) and offset (Ω), an analysis of the exchange regimes in which robust fitting parameters can be obtained has not been undertaken[15]. In this chapter we carry out simulations of $R_{1\rho}$ relaxation dispersion data to probe the limits of which exchange scenarios allow for the determination of robust fitting parameters, which are necessary for excited state structural characterization. Also, we highlight some experimental pitfalls that might bias curve fitting which could result in incorrect fitted parameters.

2.2 Methods

2.2.1 Grid Searching

Synthetic $R_{1\rho}$ relaxation dispersion curves were generated using the Laguerre equation (Chapter 1, Equation 12) and were made to emulate typical experimental nucleic acid data sets consisting of 150 data points with varying powers (ω_1) and offsets (Ω)[7]. These data sets were then uniformly noise corrupted with 5% error by randomly selecting a value from a normal distribution centered at the Laguerre $R_{1\rho}$ with a standard deviation of 5% (Appendix 1). The exchange parameters are defined identical to Chapter 1 where $\Delta\omega = \omega_G - \omega_E$, p_E is the excited state population, and k_{ex} is the sum of the rate constants for conversions between ground and excited states $k_{ex} = k_{GE} + k_{EG}$. χ^2 grids consisting of approximately 100,000 points were generated by scanning $0 < p_E < 0.5$ in increments of .005, -3 ppm $< \Delta\omega <$ 3 ppm generally in increments less than 0.1 ppm and varying ranges of k_{ex} depending on the motional timescale. Each calculation used a ^{13}C frequency of 150 MHz that is achieved with a magnetic field strength of 14.1 T. All of the calculations in this chapter were carried out using Mathematica 8.0 (Wolfram Research) and the corresponding code is included in the Appendix when noted. We also note that the grid search code was written to accommodate data collected at multiple fields and that the code takes advantage of the parallel computing packages included in Mathematica 8.0

(Appendix 2). A grid consisting of 100,000 grid points required approximately 10 minutes to calculate using a 2.0 GHz Intel Core i7 processor with 4 cores. For this study, grids were generated with a $\Delta\omega$ of 2.12 ppm (2,000 s^{-1}) for highly asymmetric (5%) and asymmetric (30%) excited state populations (p_E) with transverse (R_2) and longitudinal (R_1) relaxation constants of 11.0 s^{-1} and 1.5 s^{-1}, respectively.

2.2.2 Bootstrap Analysis

To estimate the robustness of data fitting, we carried out a bootstrap procedure where 1,000 new data sets were generated from a parent data set similar to a recent bootstrap analysis reported for CPMG[13,16,17]. Data sets for bootstrap analysis were created as follows. First, each $R_{1\rho}$ data point from a parent set of j points was assigned a number between 1 and j. Then each bootstrap data set was generated by randomly selecting a data point between 1 and j from the parent data set and inserting it into a new bootstrap data set j times. Here the bootstrap data sets are of the same size (j) of the original data set and each data point from the parent data set can be represented multiple times or completely excluded. The parent data sets used for bootstrapping were composed of the same powers and offsets as the grid search data (150 total points), noise corrupted 5%, and unless noted otherwise, R_2 and R_1 were 11 s^{-1} and 1.5 s^{-1}, respectively.

The bootstrap data sets were fit using the "Minimize" function built into Mathematica 8.0. According to the Mathematica manual, this function is best suited for finding a global minimum solution as opposed to other built in functions that employ computationally less costly algorithms. These less costly methods such as the Levenberg-Marquardt method can easily become trapped in local minima. The Minimize function was used to fit all five parameters in the Laguerre expression (Chapter 1, Equation 12): R_1, R_2, p_E, k_{ex}, $\Delta\omega$. In order to diminish any bias in curve fitting, we setup the following constraints for fitting parameters: $0.001 < p_E < 0.999$, $-7 \text{ ppm} < \Delta\omega < 7 \text{ ppm}$, $100 \text{ s}^{-1} < k_{ex} < 50{,}000 \text{ s}^{-1}$, $0.5 \text{ s}^{-1} < R_1 < 4 \text{ s}^{-1}$, and $R_{2actual} - 10 \text{ s}^{-1} < R_2 < R_{2actual} + 10 \text{ s}^{-1}$, unless noted otherwise. It was found that for cases with slow-intermediate exchange, the Minimize function did not always find the global

minimum χ^2 when fitting 5 parameters. However, the Minimize function always converged to a global minimum χ^2 when fitting 3 parameters (with a fixed R_1 and R_2). The Mathematica code for the bootstrap fitting procedure can be found in Appendix 3 and fits consisting of 1,000 bootstraps required approximately 10-20 minutes for completion using 4 Intel Core i7 cores on the Flux cluster (University of Michigan). The bootstrap Mathematica code was written to accommodate data collected at multiple fields.

2.3 Results and Discussion

2.3.1 Exploration of the χ^2 Surface Via Grid Searching

Data fitting for $R_{1\rho}$ relaxation dispersion data is typically carried out by minimizing a χ^2 function,

$$\chi^2 = \sum_{i=1}^{N} \left(\frac{R_{1\rho(i)}^{Calc} - R_{1\rho(i)}^{Exp}}{\sigma_{Exp(i)}} \right)^2 \quad (1)$$

where σ is the experimental error, $R_{1\rho(i)}^{Calc}$ is the value of $R_{1\rho}$ predicted by a model, $R_{1\rho(i)}^{Exp}$ is the experimental $R_{1\rho}$, and N is the number of points being fit. The model that matches best with the observed experimental data results in the lowest χ^2 value. Software packages such as Origin (OriginLab Corporation) are typically used to fit experimental data where minimization algorithms such as the Levenberg-Marquardt method are employed to vary model parameters in an effort to minimize χ^2. One pitfall of these algorithms is that the search can get caught in a local minimum of the χ^2 surface resulting in a bad fit, where the "correct" global minimum is missed. While more computationally costly than search algorithms, unbiased grid searches reveal all minima of the χ^2 surface allowing for a definitive hunt for the global minimum.

The first step in determining the experimental limits of $R_{1\rho}$ relaxation dispersion is to systematically investigate its χ^2 surface as a function of different conformational exchange regimes. Accurate parameters such as the population of the excited state (p_E), timescale of exchange (k_{ex}), and excited state chemical shift

($\Delta\omega$) can only be reliably determined if there is a clear global minimum in the χ^2 surface. An understanding of how the χ^2 surface varies for different exchange regimes can act as a guide to recognize where and when potential local minima occur and how difficult it is to discern them from the global minimum χ^2.

It is well known that $R_{1\rho}$ relaxation dispersion measurements that utilize low power spin lock fields (ω_1 < 1,000 Hz) are excellent for quantifying lowly populated excited states when exchange occurs on slow to intermediate chemical shift timescales ($k_{ex}/\Delta\omega \leq 1$)[15]. Indeed, the χ^2 grid for a slow-intermediate exchanging system with an excited state population of 5% reveals a very distinct global χ^2 minimum (Figure 2.1a). This is partially due to significant asymmetry of the $R_{1\rho}$ curve as a function of offset (Ω), where the exchange contribution to $R_{1\rho}$ is greater when the spin-lock frequency (ω_{rf}) is nearly on or on resonance with the minor state (Figure 2.2a). Similar grids with very distinct global minima were observed in all cases when $k_{ex} \leq \Delta\omega$ for excited state populations of 5% and 30%.

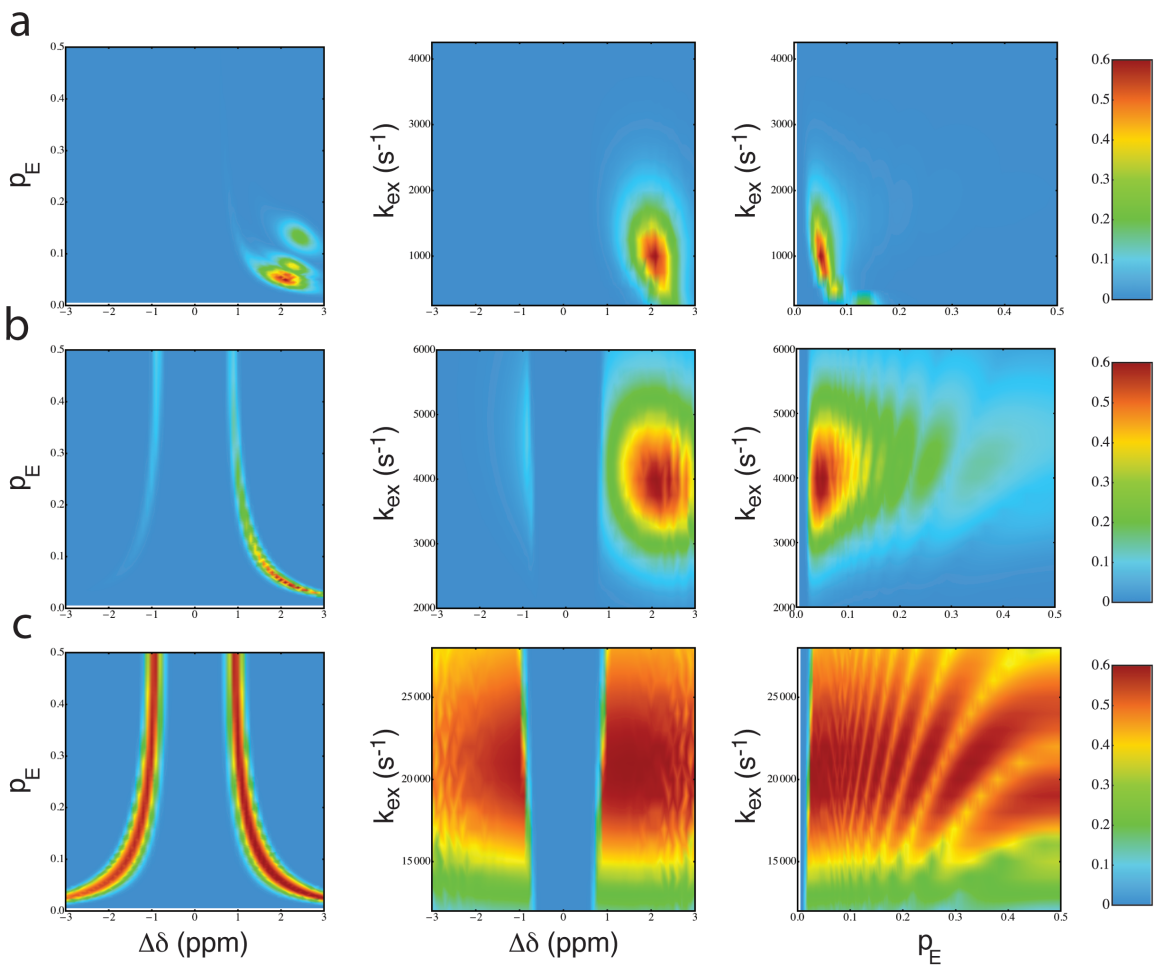

Figure 2.1: χ^2surface for slow and fast exchange scenarios. Here, $\exp\left(-\frac{\chi^2}{2\chi^2_{min}}\right)$ has been plotted with, a) p_E = 5%, k_{ex} = 1,000 s^{-1}, and Δω = 2,000 s^{-1}, b) p_E = 5%, k_{ex} = 4,000 s^{-1}, and Δω = 2,000 s^{-1}, c) p_E = 5%, k_{ex} = 20,000 s^{-1}, and Δω = 2,000 s^{-1}.

Next we explored what happens to the χ^2 grid when the exchanging system goes into the fast exchange regime ($k_{ex} > \Delta\omega$). In the limit of fast exchange, the fast exchange equation predicts that $R_{1\rho}$ is symmetric as a function of offset where the chemical exchange contribution to relaxation (R_{ex}) is a function of $\Delta\omega^2$ coupled to the excited state population (Chapter 1, Equation 9). The grid search of a system in moderately fast exchange ($k_{ex}/\Delta\omega = 2$) begins to reveal this symmetry (Figure 2.2b), where there is clearly a coupling between Δω and p_E and local χ^2 minima appear at the opposite Δω of the correct solution (Figure 2.1b). However, even though local

minima begin to appear at the opposite chemical shift, the global minimum is still very distinct from the incorrect local minima (Figure 2.1b). The sign of the excited state chemical shift is also easily discernible from this χ^2 grid, but the uncertainty in the excited state population and chemical shift becomes larger compared to slow-intermediate exchange scenarios. As $k_{ex}/\Delta\omega$ becomes even larger (10X), the χ^2 grid becomes nearly perfectly symmetric therefore making it almost impossible to reliably determine either the chemical shift or population of the excited state (Figure 2.1c and 2.2c).

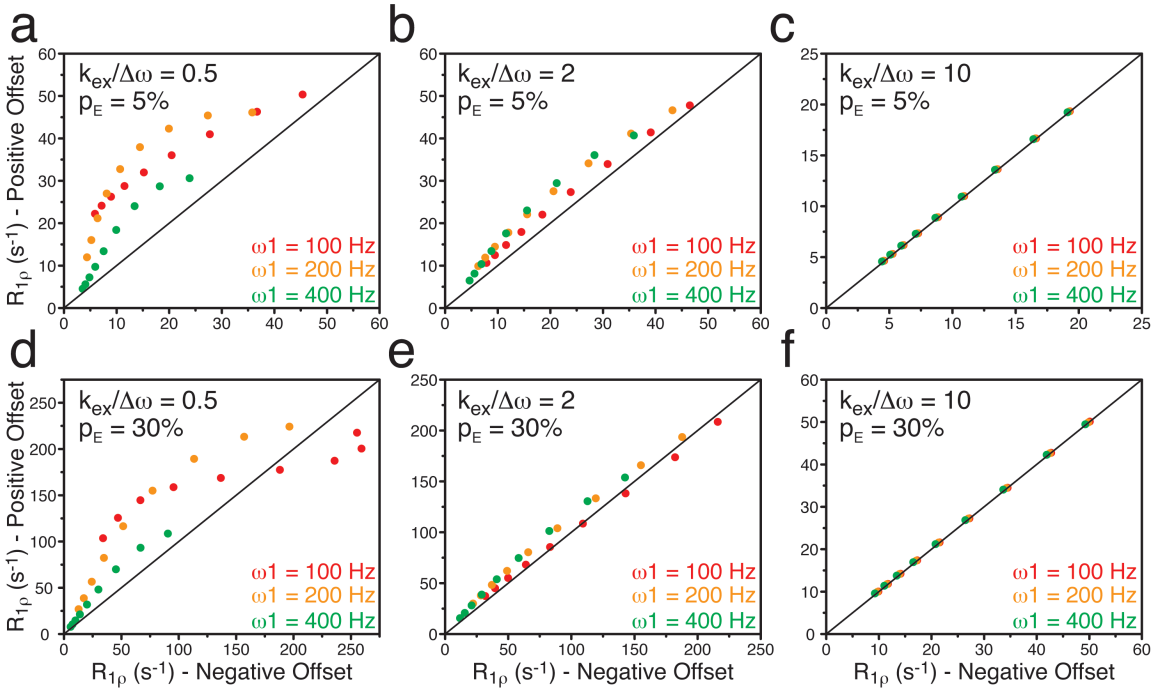

Figure 2.2: Correlation plot between the $R_{1\rho}$ of positive and negative offsets (Ω). For a) k_{ex} = 1,000 s^{-1}, p_E = 5% b) k_{ex} = 4,000 s^{-1}, p_E = 5% c) k_{ex} = 20,000 s^{-1}, p_E = 5% d) k_{ex} =1,000 s^{-1}, p_E = 30% e) k_{ex} = 4,000 s^{-1}, p_E = 30% f) k_{ex} = 20,000 s^{-1}, p_E = 30%. Plotted spin lock powers are ω_1 = 100 Hz (red), ω_1 = 200 Hz (orange), ω_1 = 400 Hz (green).

These grids for fast exchange scenarios reveal pitfalls that could occur when trying to fit dispersion data when the system is in fast exchange. As the ratio of $k_{ex}/\Delta\omega$ exceeds ~5X, the difference between the global and local minima becomes very small. Therefore, depending on the quality of the search algorithm, data fitting might result in a solution that is trapped in a local minimum. For example, fitting

solutions could be highly dependent on initial model parameters that might ultimately trap the fit in a local minimum. The small difference between the χ^2's of positive and negative excited state chemical shifts when $k_{ex}/\Delta\omega > \sim 5$ necessitates the collection of data at very high quality in order to assure that the correct model parameters are determined. Finally, we note that it is crucial that data collection is symmetric as a function of offset due to the symmetry about positive and negative $\Delta\omega$. Collection of additional data points at either positive or negative offsets could bias the χ^2 function where the quality of their fit would add additional weight to a particular sign of $\Delta\omega$.

2.3.2 Monte Carlo Bootstrap Analysis of Dispersion Data

With an understanding of χ^2 surfaces as a function of exchange regime in hand, we sought to test the limits of which exchange scenarios can be reliably characterized by $R_{1\rho}$ relaxation dispersion. Not much previous work has focused on determining the limits of $R_{1\rho}$ relaxation dispersion; rather, previous studies have primarily explored the limits of CPMG relaxation dispersion. For example, a recent study by Vallurupalli *et al.* characterized the limits for CPMG dispersion for systems with asymmetrically populated excited states ($p_E < 5\%$) and found that when $k_{ex}/\Delta\omega$ exceeded ~2, data fitting did not converge to a common solution[13]. Here the authors employed a Monte Carlo bootstrap analysis to determine the uncertainty in the parameters resulting from fitting experimental data[16,17]. We carried out a very similar analysis with synthetic and experimental data sets to determine the scenarios in which reliable parameters can be determined from fitting $R_{1\rho}$ dispersion data.

In our bootstrap procedure, we initially limited the p_E, $\Delta\omega$, and k_{ex} parameters to the following ranges $0 < p_E < 0.5$, -4 ppm $< \Delta\omega <$ 4 ppm, and 100 $s^{-1} < k_{ex} <$ 50,000 s^{-1} in order to minimize bias in curve fitting. For each search, the Mathematica Minimize function chooses random seeds within the specified ranges to begin global minimization. Upon carrying out the bootstrap procedure with these search restrictions we found that in many cases the fitting did not converge to a

single solution, even for some slow-intermediate exchange scenarios with clear global minima. When the fitting did not converge to the correct solution, the minimization became stuck at a p_E of 0.5 with a $\Delta\omega$ of the incorrect sign, indicating that the search became caught in a minimum associated with the major state population for the p_E parameter. In order to fix this convergence problem we expanded the population search range to include $0.001 \leq p_E \leq .999$ and $-7 \leq \Delta\omega \leq 7$. When fits resulted in $p_E > 0.5$, the excited state population was adjusted to be $(1-p_E)$ and the sign of $\Delta\omega$ from the fits was flipped. By expanding the search space to include all populations we found that the fits had significantly improved convergence to the global minimum χ^2.

Grid search analysis revealed that in cases of slow-intermediate exchange, there exists a very distinct global minimum in the χ^2 surface. As expected, bootstrap analysis of 5% noise corrupted slow-intermediate exchange data ($k_{ex}/\Delta\omega=1$) resulted in fits that converge to the correct solution where there is little deviation between the 1,000 bootstrap fits for p_E, $\Delta\omega$, and k_{ex} (Figure 2.3). For slow exchange scenarios, most of the 1,000 bootstrap fits resulted in obtaining a global minimum with a χ^2 of ~1. However, we observed that a small fraction of the bootstraps for slow exchange ($k_{ex} < \Delta\omega$) clearly did not converge having a final $\chi^2 > 20$. The failure to find the global minimum in these cases might be because the global minimum is very narrow in the χ^2 surface, making it similar to finding a needle in a haystack. We varied the allowed ranges for each fitting parameter and found that convergence of the fitting was dependent on k_{ex}. Decreasing the allowable range for k_{ex} resulted in convergence to the global minimum for every slow exchange bootstrap fit. We also note that when fitting only 3 parameters (p_E, k_{ex}, $\Delta\omega$) and keeping R_2 and R_1 constant, we always observed convergence to a χ^2 of ~1 with large ranges allowed for k_{ex} ($100\ s^{-1} < k_{ex} < 50{,}000\ s^{-1}$). The 95% bootstrap confidence intervals for 3 and 5 parameter fitting were nearly identical when the 5 parameter fits converged to the global minimum χ^2. 95% confidence intervals for bootstrap fits including excited state populations of 5 and 30% are reported in Table 2.1.

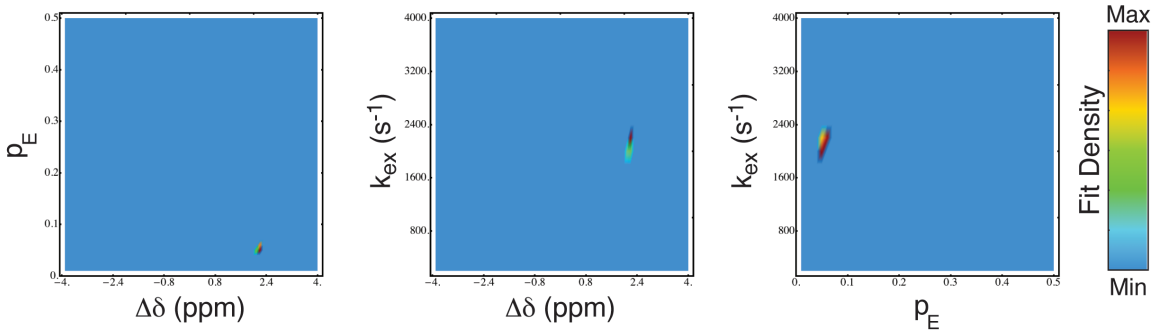

Figure 2.3: Example bootstrap fit for a slow-intermediate exchange scenario. Plotted are results for the fitting of 1,000 data sets composed of 5% noise corrupted data with p_E = 5%, k_{ex} = 2,000 s^{-1}, Δω = 2,000 s^{-1}. The color key for fit density is indicated on the right.

p_E (%)	k_{ex} (s^{-1})	Δω (ppm)	Actual p_E (%)	Actual k_{ex} (s^{-1})	Actual Δω (ppm)	$k_{ex}/\Delta\omega$
3.9-6.1*	394-696*	2.07-2.18*	5	500	2.12	0.3
4.8-5.5	879-1071	2.02-2.15	5	1000	2.12	0.5
4.8-5.2*	1939-2084*	2.06-2.18*	5	2000	2.12	1.0
4.0-5.3	3727-4079	2.03-2.38	5	4000	2.12	2.0
3.4-6.2	7661-8502	1.89-2.64	5	8000	2.12	4.0
1.9-8.9	11625-23011	1.41-3.04	5	16000	2.12	8.0
2.4-23.1	18567-24461	1.14-3.06	5	20000	2.12	10.0
2.0-47.7	21282-37113	-2.56-2.93	5	30000	2.12	15.0
28.9-30.6*	492-535*	2.07-2.15*	30	500	2.12	0.3
28.8-30.3	973-1061	2.09-2.15	30	1000	2.12	0.5
29.1-30.9	1889-2003	2.09-2.16	30	2000	2.12	1.0
29.2-33.6	3848-4025	2.05-2.14	30	4000	2.12	2.0
23.9-31.7	7717-8187	2.09-2.29	30	8000	2.12	4.0
20.2-35.0	15438-17028	2.04-2.44	30	16000	2.12	8.0
24.4-45.2	19263-21962	1.95-2.27	30	20000	2.12	10.0
16.0-49.1	24985-31703	-1.99-2.43	30	30000	2.12	15.0

* k_{ex} was restricted to 100 $s^{-1} \leq k_{ex} \leq$ 10000 s^{-1} to make fitting converge 100% of the time.

Table 2.1: 95% confidence intervals from bootstrap fitting of synthetic $R_{1\rho}$ data to the Laguerre equation.

Grid searching of fast exchange scenarios showed that many local minima appear in the χ^2 surface and generally the entire surface flattens out as $k_{ex}/\Delta\omega$ increases. This increasing indistinguishably between χ^2's as a function of model parameter would predict that bootstrap fitting should result in a large range of solutions. Here, many different model parameters could satisfy a global minimum in

χ^2 depending on the error-corrupted points used to represent the true data distribution. The distribution of fits obtained from bootstrap analysis of 5% noise corrupted data with a p_E of 5% and k_{ex} of 16,000 s^{-1} demonstrates that while the sign of the excited state chemical shift can be determined, more uncertainty exists in the population and chemical shift of the excited state (Figure 2.4, Table 2.1). As exchange becomes even faster with a p_E of 5% and k_{ex} of 30,000 s^{-1}, the only useful information that can be reliably determined is the timescale of the exchange process (Table 2.1). We also note that the ability to obtain robust fitting results in the fast exchange regime is a function of the excited state population. As shown in Table 2.1, higher excited state populations, 30% compared to 5%, result in a larger tolerance for the chemical shift timescales ($k_{ex}/\Delta\omega$) that can be probed by $R_{1\rho}$ relaxation dispersion.

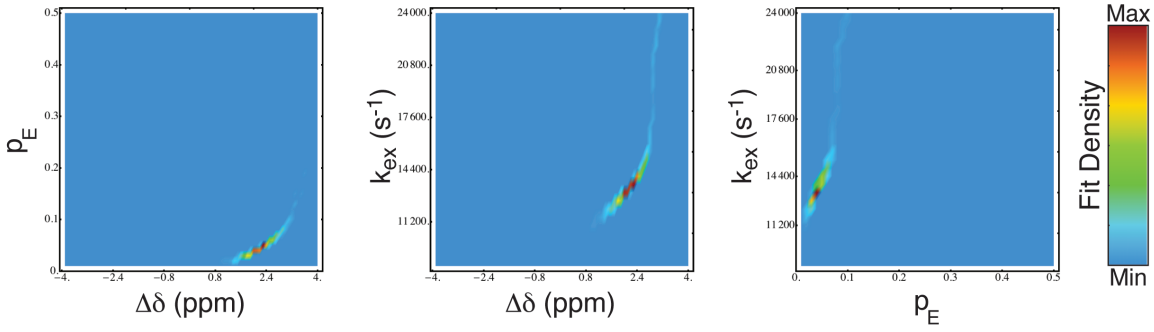

Figure 2.4: Example bootstrap fit for a fast exchange scenario. Plotted are results for the fitting of 1,000 data sets composed of 5% noise corrupted data with p_E = 5%, k_{ex} = 16,000 s^{-1}, $\Delta\omega$ = 2,000 s^{-1}. The color key for fit density is indicated on the right.

Finally, we explored less ideal cases with larger error corruption and increased intrinsic transverse relaxation constants (R_2). As shown in the Laguerre equation (Chapter 1, Equation 12), the contribution to $R_{1\rho}$ relaxation from exchange (R_{ex}) is combined with the intrinsic R_2. Consequently, the influence of conformational exchange on $R_{1\rho}$ decreases relative to R_2 as R_2 becomes larger and larger. First, we carried out simulations on a system mildly in fast exchange ($k_{ex}/\Delta\omega$ =2) with a p_E of 30% and a k_{ex} of 4,000 s^{-1}. R_2's of 20, 30, and 50 s^{-1} were tested in addition to adding 10% noise corruption to a data set with an R_2 of 11 s^{-1}. The

reference bootstrap with an R_2 of 11 s^{-1} and 5% noise corruption showed excellent convergence (Figure 2.5a). Even though increases of R_2 and noise broadened the bootstrap fitting distributions, the fits all converged to a common solution (Figure 2.5b, Table 2.2). Next, we increased the timescale of exchange to $k_{ex}/\Delta\omega = 8$ with a k_{ex} of 16,000 s^{-1}. The reference bootstrap with an R_2 of 11 s^{-1} and 5% noise corruption showed a fairly narrow fitting distribution (Figure 2.5c). However, a systematic increase of R_2 or noise resulted in significant broadening in the bootstrap fitting distribution (Figure 2.5d, Table 2.2) indicating that scenarios far into the fast exchange regime have little tolerance for error, requiring exquisite experimental data to obtain reliable parameters. We also note that these simulations give a best-case scenario since the noise level was held constant for each larger R_2 that was tested. In reality, the signal-to-noise ratio decreases proportional to $(1/R_2)^{1/2}$, therefore many additional experimental scans would be required in order to achieve the same level of noise between different R_2's[18].

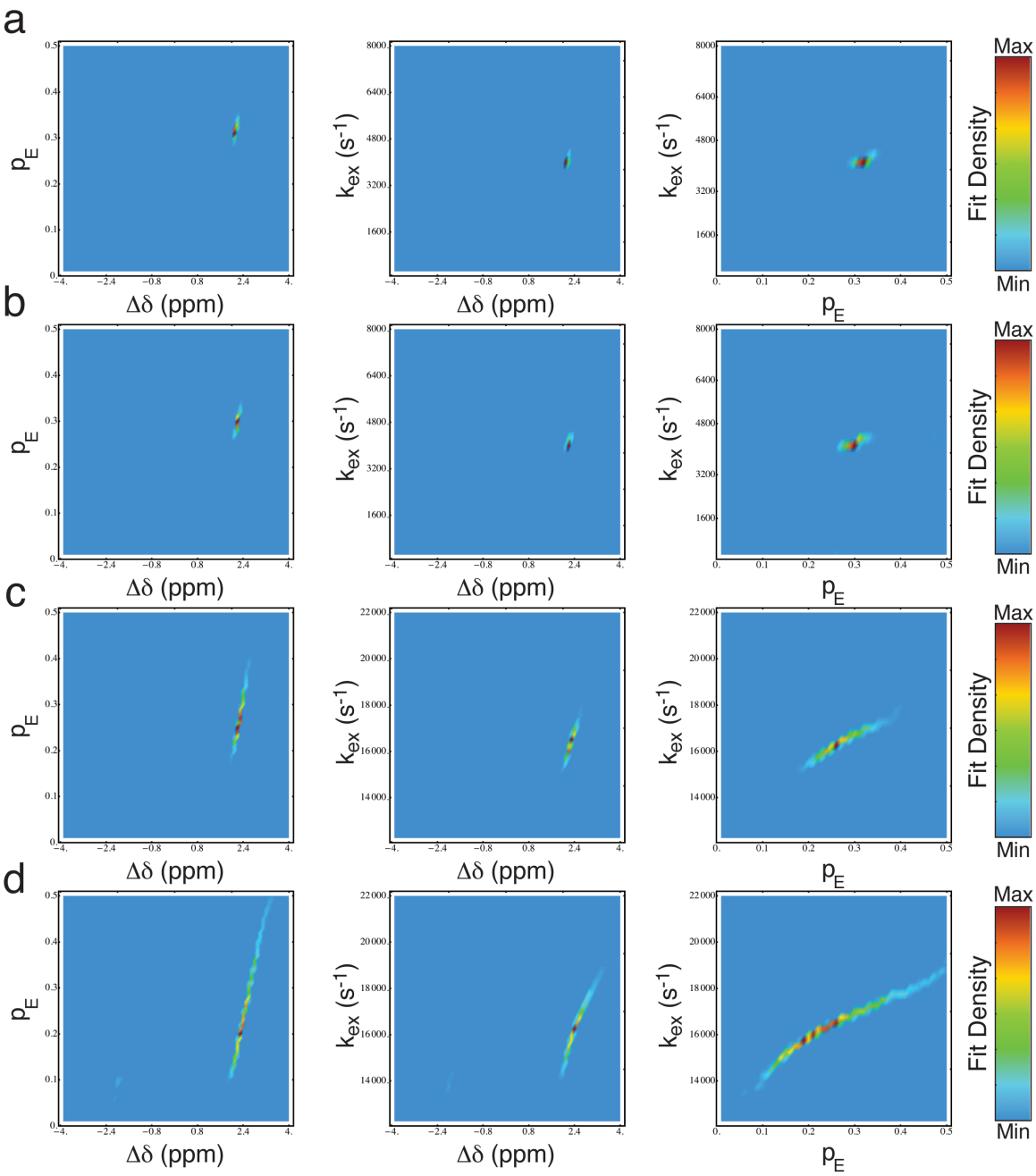

Figure 2.5: Bootstrap fits of 1,000 data sets composed of 5% noise corrupted data with varying R_2. With p_E = 30%, $\Delta\omega$ = 2,000 s^{-1}, and a) k_{ex} = 4,000 s^{-1} and R_2 = 11 s^{-1} b) k_{ex} = 4,000 s^{-1} and R_2 = 50 s^{-1}, c) k_{ex} = 16,000 s^{-1} and R_2 = 11s^{-1}, d) k_{ex} = 16,000 s^{-1} and R_2 = 50 s^{-1}. The color key for fit density is indicated on the right.

p_E (%)	k_{ex} (s^{-1})	Δω (ppm)	Actual p_E (%)	Actual k_{ex} (s^{-1})	Actual Δω (ppm)	R_2 (s^{-1})
29.2-33.6	3848-4025	2.05-2.14	30	4000	2.12	11
30.1-39.1	3848-4135	1.96-2.12	30	4000*	2.12*	11*
27.2-31.7	3909.1-4113	2.09-2.20	30	4000	2.12	20
28.1-34.0	3915-4168	2.04-2.16	30	4000	2.12	30
26.7-32.3	3818-4116	2.08-2.23	30	4000	2.12	50
23.9-31.7	7717-8187	2.09-2.29	30	8000	2.12	11
19.9-32.4	7743-8500	2.09-2.47	30	8000*	2.12*	11*
24.8-36.2	7660-8223	2.02-2.26	30	8000	2.12	20
28.6-41.9	7565-8301	1.96-2.16	30	8000	2.12	30
22.2-36.8	7390-8309	1.99-2.33	30	8000	2.12	50
20.2-35.0	15438-17028	2.04-2.44	30	16000	2.12	11
24.2-49.5	14997-18346	-2.17-2.38	30	16000*	2.12*	11*
24.4-48.1	14657-16524	-1.85-2.24	30	16000	2.12	20
21.3-48.3	13958-16264	1.80-2.29	30	16000	2.12	30
11.4-43.8	14364-18024	1.98-3.07	30	16000	2.12	50

*Noise corrupted 10% instead of 5%

Table 2.2: 95% confidence intervals from bootstrap fitting of synthetic $R_{1\rho}$ data to the Laguerre equation with varied of noise corruption and R_2.

2.3.3 Comparing the limits of $R_{1\rho}$ to CPMG Dispersion

Recently, Vallurupalli *et al.* carried out an analysis of the upper limit of which chemical shift timescales ($k_{ex}/\Delta\omega$) could be fully characterized by CPMG relaxation dispersion. In general, it was found that CPMG dispersion could reliably characterize systems with highly asymmetric populations ($p_E < 5\%$) in the $k_{ex}/\Delta\omega \leq \sim 2$ exchange regime[13]. Here CPMG data was analyzed using a bootstrap method on experimental data collected at multiple magnetic field strengths. In an effort to compare CPMG and $R_{1\rho}$, we examined the ability of $R_{1\rho}$ relaxation dispersion to characterize similar exchange scenarios at a single magnetic field strength.

Vallurupalli *et al.* applied CPMG fitting to two protein systems with residues in fast and slow exchange[13]. The existence of fast and slow exchange within a molecule is a function of both k_{ex} and Δω. For example, a nucleus within a protein that satisfies the fast exchange regime undergoes a conformational transition that results in a small perturbation of its chemical shift (Δω) relative to the timescale of exchange (k_{ex}). The first system studied was a mutated T4 lysozyme protein with an excited state population of 3.6% and a k_{ex} of 6,100 s^{-1} [6]. The average $k_{ex}/\Delta\omega$ for this

system was approximately 2 with a median value of 4.4 for all sites reported in their study. We carried out bootstrap simulations for each $k_{ex}/\Delta\omega$ scenario: 2 and 4.4. Bootstrap analysis of $k_{ex}/\Delta\omega = 2$ resulted in good convergence for all parameters p_E, $\Delta\omega$, and k_{ex} (Table 2.3). Also, bootstrapping for a $k_{ex}/\Delta\omega = 4.4$ scenario resulted in a relatively small fitting distribution (Figure 2.6a, Table 2.3). This example demonstrates that $R_{1\rho}$ dispersion has a slightly larger timescale tolerance into the fast exchange regime than CPMG dispersion. An additional advantage of $R_{1\rho}$ is that model parameters for these scenarios could successfully be quantified with just one field strength compared to two for CPMG.

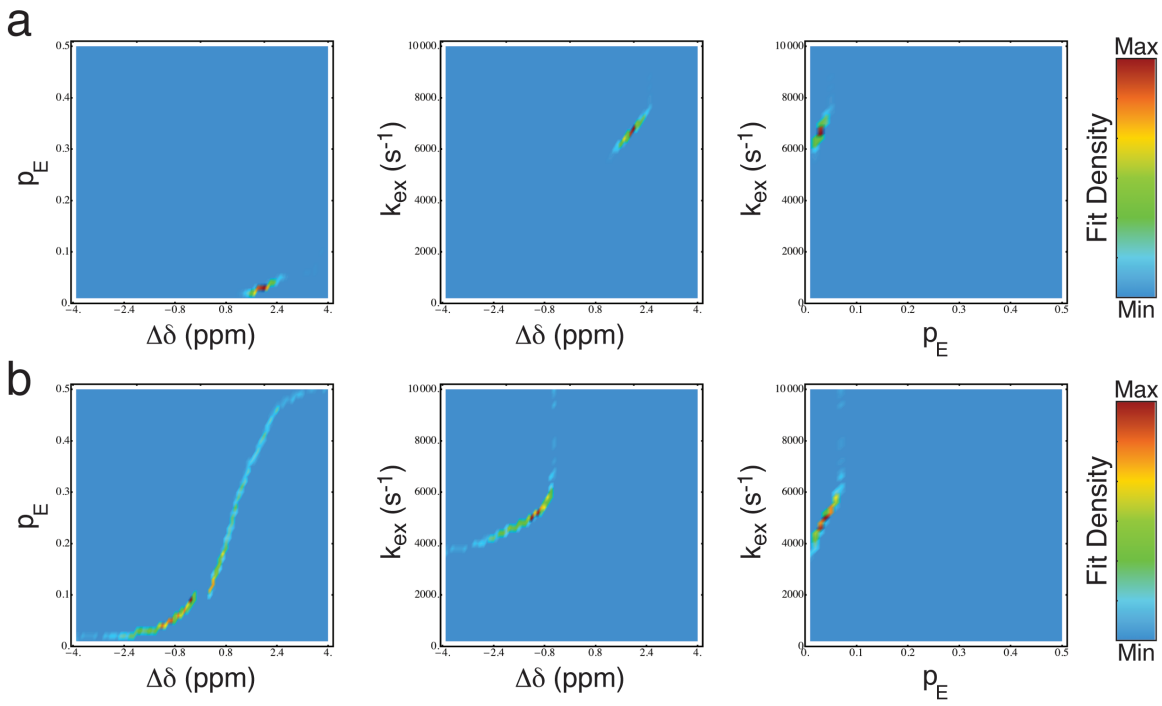

Figure 2.6: Bootstrap fit of 1,000 data sets composed of 5% noise corrupted data. With a) p_E = 3.6%, k_{ex} = 6,100 s^{-1} and $k_{ex}/\Delta\omega = 4.4$ b) p_E = 2.1%, k_{ex} = 5,660 s^{-1} and $k_{ex}/\Delta\omega = 5.6$. The color key for fit density is indicated on the right.

In this study, Vallurupalli *et al.* also investigated another system: the I44A,V67A FF domain, which has a slightly lower excited state population of 2.1% and a k_{ex} of 5,660 s^{-1} [5,19]. We bootstrapped the scenario for the V67A FF domain residue reported with the largest $k_{ex}/\Delta\omega$ (5.6) and found that $R_{1\rho}$ fitting, similar to CPMG, did not converge to a narrow distribution of p_E and $\Delta\omega$ (Figure 2.6b, Table

2.3). Further, we also tried a $k_{ex}/\Delta\omega$ of 4.4, similar to the T4 lysozyme example, and observed a large distribution of solutions for p_E and $\Delta\omega$ (Table 2.3). Lastly, we bootstrapped a $k_{ex}/\Delta\omega$ of 2 and finally observed a narrow distribution of p_E and $\Delta\omega$ (Table 2.3). These examples for the T4 lysozyme and FF domain highlight the influence of the excited state population on the chemical shift timescales that can be robustly probed by $R_{1\rho}$ dispersion. Simulations for our two systems (5% and 30% p_E) and those used by Vallurupalli *et al*. both show that higher excited state populations expand the window of timescales that can be robustly characterized by $R_{1\rho}$. Overall, this analysis shows that $R_{1\rho}$ relaxation dispersion has a slightly higher chemical shift timescale tolerance for the characterization of the populations and chemical shifts of lowly populated excited states than CPMG.

p_E (%)	k_{ex} (s^{-1})	$\Delta\omega$ (ppm)	Actual p_E (%)	Actual k_{ex} (s^{-1})	Actual $\Delta\omega$ (ppm)	$k_{ex}/\Delta\omega$
1.9-2.5	5482-6320	2.76-3.27	2.1	5660	3.00	2.0
1.0-19.4	4912-40563	0.77-3.35	2.1	5660	1.36	4.4
1.4-46.0	3978-38435	-2.78-2.39	2.1	5660	1.07	5.6
3.2-4.1	5963-6467	3.05-3.47	3.6	6100	3.24	2.0
1.5-4.2	5969-7493	1.41-2.43	3.6	6100	1.47	4.4
2.4-35.8	5137-37120	-1.16-2.41	3.6	6100	1.16	5.6

Table 2.3: Comparison of $R_{1\rho}$ fits to scenarios in which CPMG fails. 95% confidence intervals from bootstrap fitting of synthetic $R_{1\rho}$ data to the Laguerre equation.

2.3.4 Fitting Experimental DNA $R_{1\rho}$ Dispersion Data

Finally, we have also applied our bootstrap procedure to experimental dispersion data recorded on a DNA that transiently forms an excited state in the Hoogsteen base-pairing conformation[7]. The Watson-Crick-to-Hoogsteen transition was argued based on a combination of the p_E, k_{ex}, and $\Delta\omega$ determined from dispersion data combined with comparison to chemical shifts of mutated bases in the Hoogsteen conformation. Validation of this hypothesis necessitates that extremely accurate and precise parameters can be obtained from fitting the experimental data. We applied our bootstrap procedure to a guanine that transiently forms the Hoogsteen conformation less than 1% of the time and is in

slow exchange ($k_{ex}/\Delta\omega$ ~0.2). Similar to previous fits of slow-exchange scenarios, we observed that the bootstrap fits did not always converge with approximately 10% having a $\chi^2 > 10$ when scanning a large range of k_{ex} ($0\ s^{-1} < k_{ex} < 50{,}000\ s^{-1}$). However, there was a clear global minimum with the majority of the bootstrap fits converging to a χ^2 of ~1. Guided by the first bootstrap, we restricted the fitting to a maximum k_{ex} of 5,000 s^{-1} and convergence to the global χ^2 minimum was always observed with a χ^2 of ~1. The results from bootstrap fitting in Table 2.4 verifies the robustness of the fitted parameters given their narrow fitting distributions and assures that this data can be used accurately decipher the Watson-Crick to Hoogsteen exchange process. Also, the confidence intervals from our bootstrap analysis are consistent with the reported exchange parameters determined from fitting the full data sets[7].

Nucleus	p_E (%)	k_{ex} (s^{-1})	$\Delta\omega$ (ppm)
G8-C8	0.66-0.71	571-645	-3.14 - -3.07
G8-C1'	0.54-0.72	599-894	-3.90 - -3.70

Table 2.4: 95% confidence intervals for $R_{1\rho}$ bootstrap fits of experimental data collected at 26 °C.

2.4 Conclusions

In this chapter we have combined unbiased grid searching of relaxation dispersion data with bootstrap analysis to gain insight into the limits of which exchange scenarios can be reliably determined by $R_{1\rho}$ relaxation dispersion. Grid searching showed that the χ^2 surface has a distinct global minimum for systems undergoing exchange on slow-to-intermediate timescales. As the exchange timescale approaches the fast regime ($k_{ex}/\Delta\omega >> 1$), the χ^2 surface flattens out and many local minima appear therefore making it difficult to robustly determine the population and chemical shift of the excited state.

Overall, bootstrap simulations revealed that there is not a specific timescale limit ($k_{ex}/\Delta\omega$) for robust parameter determination. Rather, the ability to reliably determine exchange parameters is dependent on many factors including the

population of the excited state, the intrinsic R_2, and the quality of the data. In general, as excited states become more populated, there is an increase in the window of timescales for which excited state chemical shifts and populations can be accurately quantified. Further, as molecular size and R_2 correspondingly increases, the timescale window for $R_{1\rho}$ dispersion narrows where excited state populations and chemical shifts become difficult to determine when in fast exchange. Minimally, reliable parameters could be determined for all scenarios investigated in which $k_{ex}/\Delta\omega \leq \sim 2$ for excited state populations of 2.1, 3.6, 5, and 30%. Whenever the chemical shift timescale of exchange extends further into the fast regime, it would be highly recommended to carry out a bootstrap analysis of experimental data in order to determine if reliable excited state populations and chemical shifts can be extracted from the data.

Arguably the most useful parameter that can be obtained from relaxation dispersion experiments is the chemical shift of the excited state, which can ultimately be used to deduce its structural conformation. Determination of excite state structures based on chemical shifts has been done both computationally for proteins and through chemical modifications in nucleic acids[5-7]. Consequently, it is key to consider the certainty at which the excited state chemical shifts can be determined and predicted from experiment and computation, respectively. In the case of proteins, excited state structures are derived using a software package, CS-ROSETTA, which combines protein structure (ROSETTA) and chemical shift prediction (SPARTA+) software[20]. CS-ROSETTA is limited by the ability of SPARTA+ to correctly predict protein chemical shifts where ^{1}H, ^{13}C, and ^{15}N chemical shift predictions are accurate to within ~ 0.5, 1.0, and 2.5 ppm, respectively[21]. Therefore, useful structural constraints might be obtained even for a fast exchange scenario where only a broad distribution (~1 ppm) of possible excited state chemical shifts can be determined experimentally.

Future studies into combining $R_{1\rho}$ dispersion data with additional constraints such as the chemical shifts observed in HSQC and HMQC data may help to narrow the broad chemical shift distributions observed for fast exchange scenarios[13].

Finally, we note the significant sensitivity of nucleic acids' dynamic structure landscape to salt conditions might be utilized to manipulate exchange conditions to be more favorable[22]. Here a slight addition or subtraction of mono or divalent ions might favorably shift the excited state population to allow for reliable characterization of an excited state that exists in the fast chemical shift timescale.

2.5 References

1. A. G. Palmer, NMR characterization of the dynamics of biomacromolecules. *Chem Rev* **104**, 3623 (2004).
2. A. K. Mittermaier, L. E. Kay, Observing biological dynamics at atomic resolution using NMR. *Trends Biochem Sci* **34**, 601 (2009).
3. J. P. Loria, M. Rance, A. G. Palmer, A relaxation-compensated Carr-Purcell-Meiboom-Gill sequence for characterizing chemical exchange by NMR spectroscopy. *J Am Chem Soc* **121**, 2331 (1999).
4. D. M. Korzhnev, L. E. Kay, Probing invisible, low-populated states of protein molecules by relaxation dispersion NMR spectroscopy: An application to protein folding. *Accounts Chem Res* **41**, 442 (2008).
5. D. M. Korzhnev, T. L. Religa, W. Banachewicz, A. R. Fersht, L. E. Kay, A Transient and Low-Populated Protein-Folding Intermediate at Atomic Resolution. *Science* **329**, 1312 (2010).
6. G. Bouvignies *et al.*, Solution structure of a minor and transiently formed state of a T4 lysozyme mutant. *Nature* **477**, 111 (2011).
7. E. N. Nikolova *et al.*, Transient Hoogsteen base pairs in canonical duplex DNA. *Nature* **470**, 498 (2011).
8. A. Ross, M. Czisch, G. C. King, Systematic errors associated with the CPMG pulse sequence and their effect on motional analysis of biomolecules. *J Magn Reson* **124**, 355 (1997).
9. R. Ishima, D. A. Torchia, Accuracy of optimized chemical-exchange parameters derived by fitting CPMG R-2 dispersion profiles when R-2(0a) = R-2(0b). *J Biomol Nmr* **34**, 209 (2006).
10. E. L. Kovrigin, J. G. Kempf, M. J. Grey, J. P. Loria, Faithful estimation of dynamics parameters from CPMG relaxation dispersion measurements. *J Magn Reson* **180**, 93 (2006).
11. W. Myint, R. Ishima, Chemical exchange effects during refocusing pulses in constant-time CPMG relaxation dispersion experiments. *J Biomol Nmr* **45**, 207 (2009).
12. D. F. Hansen, P. Vallurupalli, P. Lundstrom, P. Neudecker, L. E. Kay, Probing chemical shifts of invisible states of proteins with relaxation dispersion NMR spectroscopy: How well can we do? *J Am Chem Soc* **130**, 2667 (2008).
13. P. Vallurupalli, G. Bouvignies, L. E. Kay, Increasing the Exchange Time-Scale That Can Be Probed by CPMG Relaxation Dispersion NMR. *J Phys Chem B* **115**, 14891 (2011).

14. J. R. Bothe *et al.*, Characterizing RNA dynamics at atomic resolution using solution-state NMR spectroscopy. *Nat Methods* **8**, 919 (2011).
15. A. G. Palmer, F. Massi, Characterization of the dynamics of biomacromolecules using rotating-frame spin relaxation NMR spectroscopy. *Chem Rev* **106**, 1700 (2006).
16. H. Motulsky, A. Christopoulos, *Fitting models to biological data using linear and nonlinear regression : a practical guide to curve fitting*. (Oxford University Press, Oxford ; New York, 2004), pp. 351 p.
17. W. H. Press, *Numerical recipes in C : the art of scientific computing*. (Cambridge University Press, Cambridge Cambridgeshire ; New York, 1988), pp. xxii, 735 p.
18. J. Cavanagh, *Protein NMR spectroscopy : principles and practice*. (Academic Press, Amsterdam ; Boston, ed. 2nd, 2007), pp. xxv, 885 p.
19. P. Jemth *et al.*, The structure of the major transition state for folding of an FF domain from experiment and simulation. *J Mol Biol* **350**, 363 (2005).
20. Y. Shen *et al.*, Consistent blind protein structure generation from NMR chemical shift data. *Proc Natl Acad Sci U S A* **105**, 4685 (2008).
21. Y. Shen, A. Bax, SPARTA+: a modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network. *J Biomol Nmr* **48**, 13 (2010).
22. D. E. Draper, D. Grilley, A. M. Soto, Ions and RNA folding. *Annu Rev Biophys Biomol Struct* **34**, 221 (2005).

Chapter 3

Sequence-Specific B-DNA Flexibility Modulates Z-DNA Formation

3.1 Introduction

One of the most dramatic transitions in biology is conversion of right-handed B-DNA into left-handed Z-DNA in sequences that are rich in pyrimidine/purine repeats. While the existence of Z-DNA *in vivo* was originally questioned, and indeed the subject of controversy, overwhelming experimental data accumulated over the past two decades unequivocally support its existence and role in fundamental processes such as transcription and nucleosome positioning[1-3]. In genomes, Z-DNA segments must form dynamically at specific locations with particular lengths and within a sea of B-DNA. The resulting formation of B/Z junctions is accompanied by energetically unfavorable conformational strain due to intersection of the drastically different B-DNA and Z-DNA structures[4]. While the sequence-dependence of Z-DNA formation has been studied extensively and shown to be dependent on both the length of pyrimidine/purine repeats and sequence of neighboring base-pairs[5-7], these studies relied on bulk measurements that do not provide atomic level information about the length of Z-DNA or location of B/Z junctions. Indeed, the one and only X-ray structure of a B/Z junction[4] shows that Z-DNA can extend outside of CG-repeats to include other DNA steps, which based on accepted thermodynamic models[8] of Z-DNA formation is predicted to be highly energetically unfavorable. In addition, current methods used to predict the formation and location of Z-DNA in genomes do not yet account for the sequence specific energetic costs of creating B/Z junctions[9]. Thus, the rules that govern the length of Z-DNA segments and location of B/Z junctions remain poorly understood. Using natural abundance NMR $R_{1\rho}$ carbon relaxation measurements[10] in concert with CD spectroscopy, we show that sequence-specific B-DNA flexibility modulates the thermodynamic propensity to

form Z-DNA. Our results add to a growing view that DNA sequences code for complex functionally important dynamics[11-13] spanning multiple timescales and also suggest that sequence-specific dynamics extends to long sequence repeats.

3.2 Materials and Methods

3.2.1 Sample Preparation

DNA oligonucleotides were purchased from IDT, Inc. (Coralville, IA) and purified by standard desalting. The DNA oligos were suspended in NMR buffer (15 mM phosphate, 25 mM NaCl, 0.1 mM EDTA at pH 7.5) and supplied with 10% D_2O. The concentrations of each single-stranded DNA were determined by UV using extinction coefficients provided by IDT. DNA duplexes were prepared by mixing an equal molar ratio of complimentary strands, heating for 5 minutes at 95°C, followed by gradual cooling to room temperature. The DNA duplexes were further washed 2X with NMR buffer using an Amicon Ultra-4 centrifugal filter (3 kDa cutoff) and concentrated to ~250 μL with a final concentration of 4.1 ± 0.4 mM. DNA samples (5 μM) used in CD studies were incubated with Zα protein in NMR buffer for 1 hour at 25°C. DNA samples (4.5 μM) for high salt circular dichroism (CD) studies were suspended in a buffer containing 10 mM phosphate, 0 or 5.4 M NaCl at pH 7, heated to 95°C for 5 minutes, and allowed to equilibrate overnight.

The Zα domain of ADAR1 was prepared closely following a procedure described previously[14] with the following modifications. Protein expression was induced with 0.2 mM IPTG at 1.0 A_{600} units. The protein was purified using HisSelect resin (Sigma), followed by HiTrapS chromatography, without the intervening thrombin cleavage step.

3.2.2 Spectroscopy

All NMR experiments were performed at 25°C using an Avance Bruker 600 MHz spectrometer equipped with a 5 mm triple-resonance cryogenic probe. NMR spectra were processed using nmrPipe and analyzed using NMRView[15]. Relaxation dispersion experiments were carried out at natural abundance as done in a previous

study[16]. Briefly, two data points (zero and a delay) were collected for each spin lock power. The relaxation constant $R_{1\rho}$ was then determined assuming a mono-exponential decay, Equation 1,

$$R_{1\rho} = \ln(\frac{I_0}{I_t})t^{-1} \quad (1)$$

where I_0 is the peak intensity with a zero delay and I_t is the peak intensity after a delay of time t. The exchangeable and non-exchangeable protons of the DNA duplexes were assigned using conventional NMR methods (^{1}H-^{1}H NOESY) at 25°C in 10% D_2O. CD spectroscopy experiments were carried out using an Aviv 62DS CD spectrometer at 25°C. All CD measurements were performed in duplicate using a 1 cm quartz cell. CD spectra were recorded at 1 nm intervals with signal averaging of 2 seconds at each wavelength.

3.2.3 Data Analysis

All on-resonance relaxation dispersion data was fit using the two-state fast exchange equation,

$$R_{1\rho} = R_2 + \frac{\Phi k_{ex}}{\omega_{eff}^2 + k_{ex}^2} \quad (2)$$

where $\Phi = p_G p_E \Delta\omega^2$, p_E and p_G are the populations of exchanging states, $\Delta\omega$ is their chemical shift difference, ω_1 is the spin lock power, and k_{ex} is the sum of the forward and reverse rate constants[10]. Note that the fast exchange equation is only accurate when $k_{ex} >> \Delta\omega$[10,17] and that use of other equations that allow for slower timescales requires the collection of off-resonance data, which was not practical for our natural abundance based studies. Given the inaccuracy of the fast exchange approximation for slow processes approaching the ms time regime, the obtained fitting constants are likely highly inaccurate. However, the data is sufficient to establish differences in the exchange behavior of various spins in different constructs.

3.3 Results and Discussion

The X-ray structure of a B/Z junction was reported for the sequence 5' GGTTTATGGCGCGCG 3' (Z-JXN) bound to the Z-DNA binding protein domain, Zα, of

the RNA adenosine deaminase (ADAR1) protein[4]. Surprisingly, the left-handed helix extended beyond the boundary of the $(CG)_3$ repeat to include a CC step (Figure 3.1a), which according to a widely accepted Z-DNA thermodynamic model is energetically unfavorable due to the enhanced energetic cost accompanying the *anti*-to-*syn* transition of pyrimidines[8]. This localizes the B/Z junction at the T7-A24 base-pair, which is extruded from the double helix, allowing continuous stacking between B-DNA and Z-DNA (Figure 3.1a,b).

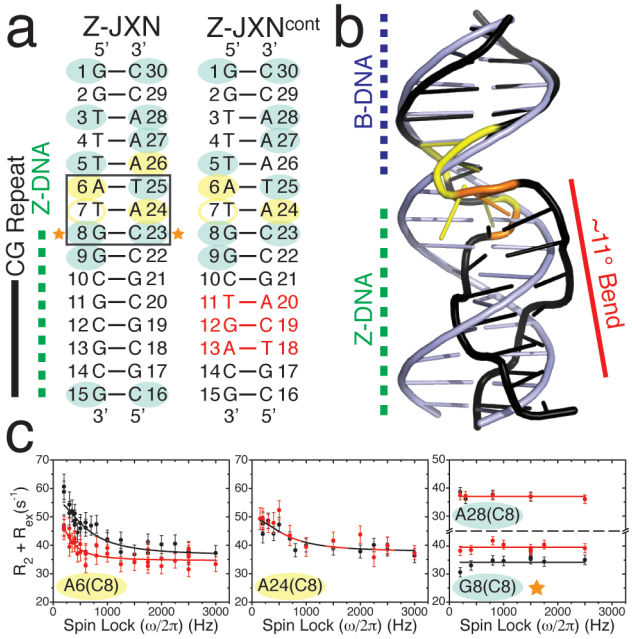

Figure 3.1: NMR observation of sequence-specific B-DNA flexibility in and around B/Z junctions. (a) B/Z junction forming DNA (Z-JXN) and control with disrupted CG-repeat (Z-JXN^{cont}) used in NMR studies. Differences between the two constructs are highlighted in red. Site of B/Z junction formation is highlighted with a box. Residues with detectable carbon chemical exchange in sugar (C1') and base (C8) moieties are highlighted in open and filled yellow circles, respectively. Residues where chemical exchange was not detected are highlighted in blue. Residues with elevated ps-ns motions are highlighted with a star. (b) X-ray structure (PDB: 2ACJ)[4] of the protein bound B/Z junction (black) aligned with a canonical B-form helix (grey). Residues with fast and slow motions when in the B-form state are color coded orange and yellow, respectively. (c) Representative on-resonance relaxation dispersion profiles measured for Z-JXN (black) and Z-JXN^{cont} (red).

To obtain insights into the unusual behavior of Z-JXN upon undergoing the B/Z transition, we used natural abundance NMR $R_{1\rho}$ carbon relaxation dispersion experiments[16,18,19] to characterize its intrinsic dynamic properties in the B-form

state, with no Z-DNA present, over broad (ps-ns and μs-ms) timescales and compared these findings to a control construct (Z-JXN^{cont}) in which the CG repeat was disrupted (Figure 3.1a). The $R_{1\rho}$ NMR experiments revealed chemical exchange involving slow (μs-ms) conformational transitions that likely involve the disruption of stabilizing interactions in B-DNA, such as hydrogen bonding and/or stacking interactions for residues in and around the site of B/Z junction formation (Figure 3.1a,b). Chemical exchange was observed for both the thymine and adenine residues in the T7-A24 base-pair, which becomes extruded upon B/Z junction formation[4]. For T7, the exchange is localized in the sugar (C1') whereas for A24 it is localized in the base (C8). Nowhere else do we observe exchange in both Watson-Crick partners. Thus, the T7-A24 base-pair, which is significantly distorted following the B-to-Z transition, is already dynamic and flexible in the B-state. We also observe chemical exchange in the base moieties of neighboring adenines A6(C8) and A26(C8) but not in their corresponding thymine partner. All of the above residues are part of CA or TA pyrimidine/purine dinucleotide steps, which are known to be the dinucleotide steps with the weakest stacking interactions[20-23]. Thus, the observed exchange likely arises from instability in canonical B-DNA which results in enhanced propensities to access alternative conformations at these sites. The unique exchange observed for T7(C1') is likely encoded by the TAT sequence and may reflect inter-helical bending observed in the context of the extensively studied "TATA" box[24]. Indeed, the TAT site also acts as a hinge for helical bending in the B/Z junction X-ray structure (Figure 3.1b). No exchange was observed for all other residues examined, including A27, A28, and T3 (Figure 3.2). However, this does not rule out the presence of conformational dynamics that falls outside the detection limits of our experiments.

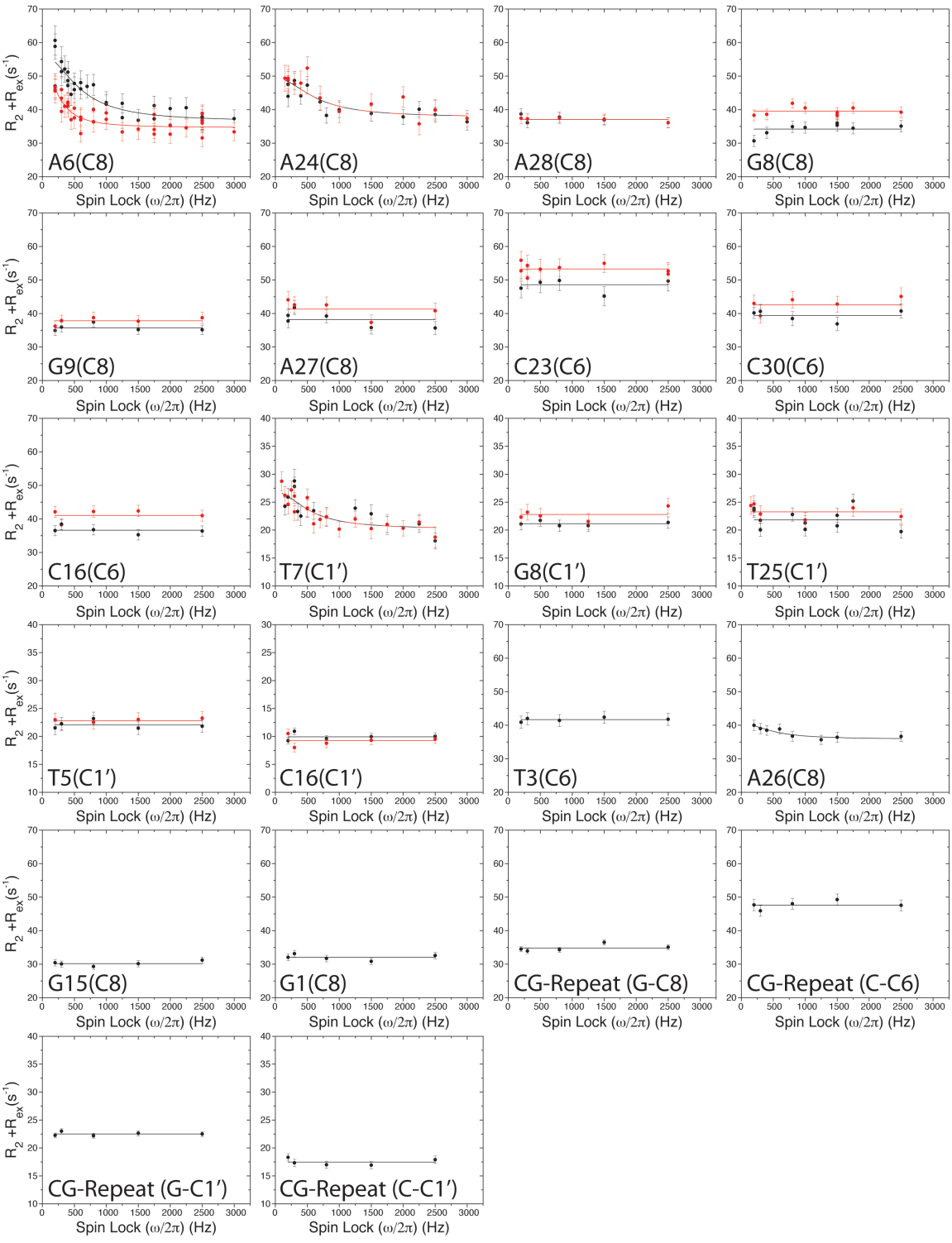

Figure 3.2: Comparison of on-resonance relaxation dispersion data for well resolved resonances in Z-JXN (black) and Z-JXN^{cont} (red). Also included are on-resonance dispersion profiles for the overlapped peaks that correspond to the CG repeat of Z-JXN.

To examine if the CG repeats play a role in the motions observed, we performed natural abundance carbon relaxation dispersion experiments on a control construct, Z-JXN^{cont}, in which the CG repeat was disrupted (Figure 3.1a). Chemical exchange was detected in the same residues of Z-JXN^{cont} which suggests that we observe intrinsic B-DNA flexibility, rather than a unique deformation only accessible to samples able to undergo the B/Z transition. While disruption of the CG repeat did not affect the exchange observed at T7-A24 (Figure 3.1c and 3.2), it significantly diminished the chemical exchange observed at A6(C8) which is more than three base-pairs away from the CG-repeat (Figure 3.1c). It is possible that this long-range effect propagates via the backbone without leading to observable effects on T7-A24, or perhaps because these fall outside the detection limits of the NMR experiment. Disruption of the CG repeat also significantly increased the intrinsic carbon transverse relaxation rate constants ($R_{2,int}$) and therefore decreased the apparent ps-ns motions at G8-C23, which is one of the two base-pairs in the CC step that unexpectedly form Z-DNA despite being outside the CG-repeat (Figure 3.1c and 3.3). This, along with unique chemical shift perturbations induced by CG-repeats (Figure 3.4), strongly suggest that CG-repeats modulate the dynamic properties of its neighboring residues.

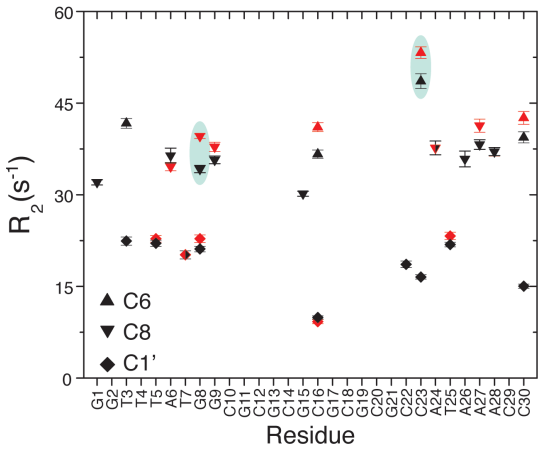

Figure 3.3: Comparison of R_2 measured in Z-JXN (black) and Z-JXN^{cont} (red) constructs. The most significant differences (excluding terminal residues) in R_2, reflecting fast ps-ns motions, are observed for G8(C8) and C23(C6) and are highlighted in blue.

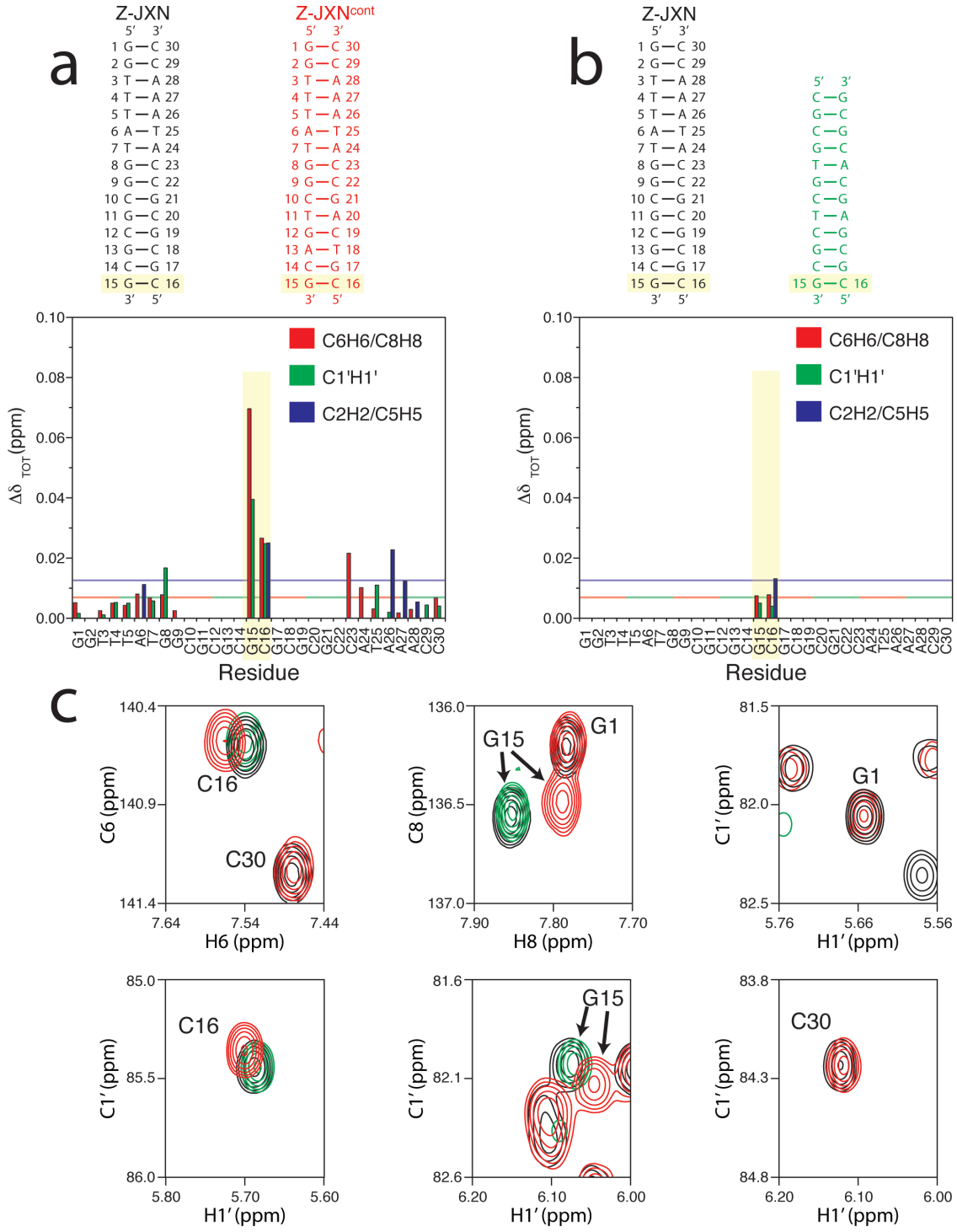

Figure 3.4: CG-induced chemical shift perturbations at terminal residues. (a) Chemical shift differences observed between common residues in Z-JXN and Z-JXN^{cont}. A large difference is observed for C23(C6H6). The largest perturbations, highlighted in yellow, are observed at the 5' end of the $(CG)_3$ repeat. (b) Agreement between resonances belonging to terminal base-pairs in Z-JXN and a DNA construct containing two CG steps (in green). These data provide additional support that the unique perturbations observed at the terminal ends of Z-JXN as compared to Z-

JXN^{cont} are specific to (CG) repeats. Only resonances belonging to the common terminal CG base-pair are compared. The red-green dashed (C6/C8/C1') and blue (C2/C5) lines depict the average chemical shift difference between the Z-JXN and Z-JXN^{cont} constructs. (c) Zoomed in regions of ^{1}H-^{13}C HSQC spectra for terminal peaks in each construct: Z-JXN (black), Z-JXN^{cont} (red), CG Control (green). Chemical shift differences were calculated using Equation 3,

$$\Delta\delta_{TOT} = \sqrt{(\delta_H)^2 + (0.25\delta_C)^2} \quad (3)$$

where δ_H and δ_C is the difference in chemical shift (ppm) in proton and carbon, respectively[25].

Our results suggest that Z-DNA unexpectedly incorporates the C22, C23 step in Z-JXN in part because this makes it possible to localize the B/Z junction at the intrinsically flexible region centered at the T7-A24 site. To test this hypothesis, we measured the propensity to form Z-DNA upon binding to the Z-DNA binding domain, Zα, of ADAR1 in DNA constructs that bear point mutations within and above the junction designed to disrupt the inherent flexibility observed by NMR without disrupting interactions with the Zα protein which exclusively interacts with Z-DNA elements below the B/Z junction[4]. We assessed the propensity to form Z-DNA by recording circular dichroism (CD) spectra following incubation of each DNA construct with Zα protein at a 6:1 DNA:protein ratio and quantifying the extent to which the B-DNA spectrum is inverted as a result of Z-DNA formation[26].

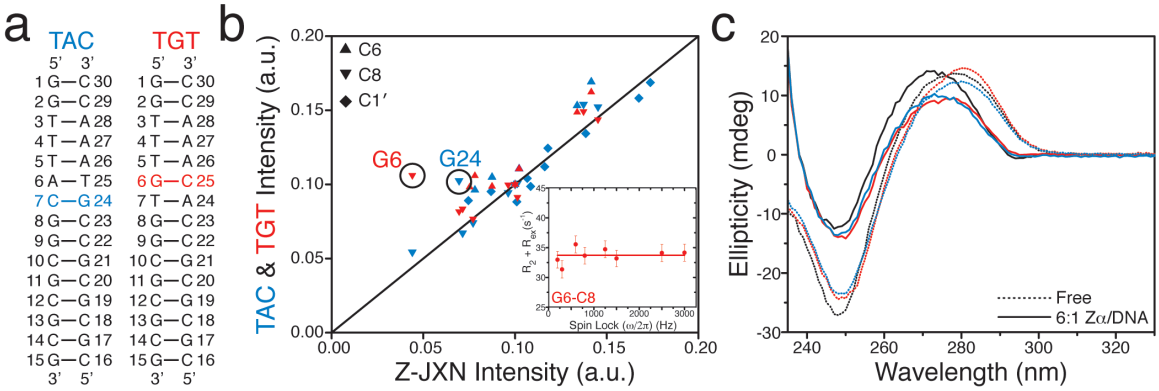

Figure 3.5: Sequence-specific preferences for Z-DNA formation. (a) Mutant constructs of Z-JXN with mutations highlighted in color. (b) Comparison of normalized ^{1}H-^{13}C HSQC peak intensities reveal differences in dynamics between constructs studied by NMR[27]. An A-to-G mutation at A24 (TAC) or A6 (TGT) eliminates exchange broadening and μs-ms motions as confirmed for G6(C8) by relaxation dispersion (inset). Relaxation dispersion measurements could not be carried out on G24(C8) due to spectral overlap. (c) CD spectra of DNA free and

bound to Zα protein color coded according to construct: black, Z-JXN; red, TGT; blue, TAC. CD samples were all the same DNA concentration.

Replacement of the exchange broadened T7-A24 base-pair with a corresponding C7-G24 base-pair that shows no sign of NMR exchange broadening (Figure 3.5b), resulted in a significantly reduced (~15-20%) propensity to form Z-DNA (TAC, Figure 3.5c). Similar results were obtained when replacing the nearby exchange broadened A6-T25 base-pair with a G-C base-pair (TGT, Figure 3.5b,c). Further, in both the TAC and TGT constructs, the exchange broadening at T7(C1') was diminished, indicating that flexibility at this site is an important factor for Z-DNA formation. Both mutants show similar extents of Z-DNA formation consistent with a change in the location of the B/Z junction in which Z-DNA retreats to the expected position at the end of the CG repeat, away from the newly introduced "rigid" residues. The same trends were observed when inducing the B-to-Z transition using conditions of high salt without the Zα protein (Figure 3.6), again confirming that the different propensities observed reflect intrinsic properties of the DNA. Taken together, these data indicate that sequence-specific flexibility in Z-JXN directs Z-DNA to form outside the CG repeat so to position the highly flexible A6-T25 and T7-A24 base-pairs at the junction.

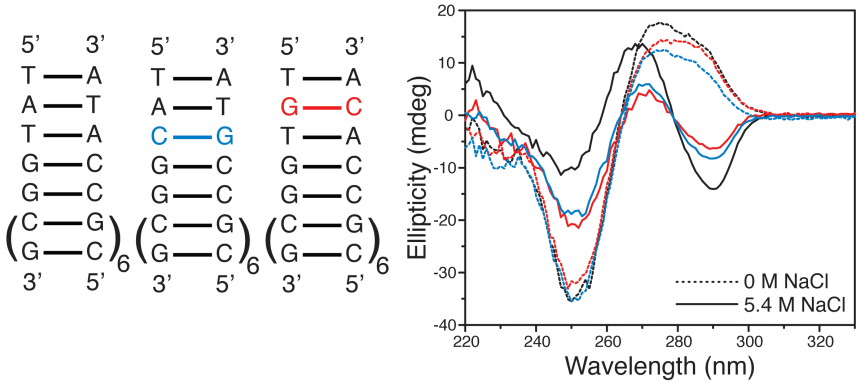

Figure 3.6: Sequence-specific preferences for Z-DNA formation at high salt in B/Z junction containing DNA constructs. Shown are CD spectra, color coded according to construct, at low and high salt concentration. Note that the DNA constructs used in Figure 3.5, which according to CD do not undergo the B-to-Z transition under high salt conditions, had to be modified to allow observation of the B-to-Z transition at high salt. The CG repeats are elongated by six CG base-pairs and the B-form region shortened by four base-pairs.

3.4 Conclusions

In conclusion, our results show that CG-repeats affect the dynamic properties of neighboring residues and that this together with inherent sequence-specific B-DNA flexibility modulates the thermodynamic propensity to form Z-DNA. Sequence-specific B-DNA flexibility may provide a mechanism for controlling the length and location of Z-DNA in genomes.

The work in this chapter was published in the Journal of the American Chemical Society[28].

3.5 References

1. A. Rich, S. Zhang, Timeline: Z-DNA: the long road to biological function. *Nat Rev Genet* **4**, 566 (2003).
2. B. Wong, S. Chen, J. A. Kwon, A. Rich, Characterization of Z-DNA as a nucleosome-boundary element in yeast Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* **104**, 2229 (2007).
3. R. Liu *et al.*, Regulation of CSF1 promoter by the SWI/SNF-like BAF complex. *Cell* **106**, 309 (2001).
4. S. C. Ha, K. Lowenhaupt, A. Rich, Y. G. Kim, K. K. Kim, Crystal structure of a junction between B-DNA and Z-DNA reveals two extruded bases. *Nature* **437**, 1183 (2005).
5. M. A. Fuertes, V. Cepeda, C. Alonso, J. M. Perez, Molecular mechanisms for the B-Z transition in the example of poly[d(G-C)center dot d(G-C)] polymers. A critical review. *Chemical Reviews* **106**, 2045 (2006).
6. R. D. Sheardy *et al.*, Sequence Dependence of the Free-Energy of B-Z Junction Formation in Deoxyoligonucleotides. *Journal of Molecular Biology* **231**, 475 (1993).
7. Z. Reich, P. Friedman, S. Levinzaidman, A. Minsky, Effects of Adenine Tracts on the B-Z Transition - Fine Tuning of DNA Conformational Transition Processes. *Journal of Biological Chemistry* **268**, 8261 (1993).
8. P. S. Ho, M. J. Ellison, G. J. Quigley, A. Rich, A Computer-Aided Thermodynamic Approach for Predicting the Formation of Z-DNA in Naturally-Occurring Sequences. *Embo Journal* **5**, 2737 (1986).
9. H. Li *et al.*, Human genomic Z-DNA segments probed by the Z domain of ADAR1. *Nucleic Acids Research* **37**, 2737 (2009).
10. A. G. Palmer, F. Massi, Characterization of the dynamics of biomacromolecules using rotating-frame spin relaxation NMR spectroscopy. *Chemical Reviews* **106**, 1700 (2006).

11. Z. Shajani, G. Varani, 13C relaxation studies of the DNA target sequence for hhai methyltransferase reveal unique motional properties. *Biochemistry* **47**, 7617 (2008).
12. E. Duchardt, L. Nilsson, J. Schleucher, Cytosine ribose flexibility in DNA: a combined NMR 13C spin relaxation and molecular dynamics simulation study. *Nucleic Acids Res* **36**, 4211 (2008).
13. A. T. Phan, D. J. Patel, Two-repeat human telomeric d(TAGGGTTAGGGT) sequence forms interconverting parallel and antiparallel G-quadruplexes in solution: Distinct topologies, thermodynamic properties, and folding/unfolding kinetics. *J Am Chem Soc* **125**, 15021 (2003).
14. T. Schwartz *et al.*, Proteolytic dissection of Zab, the Z-DNA-binding domain of human ADAR1. *Journal of Biological Chemistry* **274**, 2899 (1999).
15. F. Delaglio *et al.*, Nmrpipe - a Multidimensional Spectral Processing System Based on Unix Pipes. *Journal of Biomolecular Nmr* **6**, 277 (1995).
16. A. L. Hansen, E. N. Nikolova, A. Casiano-Negroni, H. M. Al-Hashimi, Extending the range of microsecond-to-millisecond chemical exchange detected in labeled and unlabeled nucleic acids by selective carbon R(1rho) NMR spectroscopy. *J Am Chem Soc* **131**, 3818 (2009).
17. O. Trott, A. G. Palmer, R-1 rho relaxation outside of the fast-exchange limit. *Journal of Magnetic Resonance* **154**, 157 (2002).
18. F. Massi, E. Johnson, C. Y. Wang, M. Rance, A. G. Palmer, NMR R-1 rho rotating-frame relaxation with weak radio frequency fields. *J Am Chem Soc* **126**, 2247 (2004).
19. D. M. Korzhnev, V. Y. Orekhov, L. E. Kay, Off-resonance R1(p) NMR studies of exchange dynamics in proteins with low spin-lock fields: An application to a fyn SH3 domain. *J Am Chem Soc* **127**, 713 (2005).
20. K. McAteer, M. A. Kennedy, NMR evidence for base dynamics at all TpA steps in DNA. *Journal of Biomolecular Structure & Dynamics* **17**, 1001 (2000).
21. C. Kojima, N. B. Ulyanov, M. Kainosho, T. L. James, Slow motion in the CAA center dot TTG sequence of a DNA decamer duplex studied by NMR. *Biochemistry* **40**, 7239 (2001).
22. J. R. Quintana, K. Grzeskowiak, K. Yanagi, R. E. Dickerson, Structure of a B-DNA Decamer with a Central T-a-Step - C-G-a-T-T-a-a-T-C-G. *Journal of Molecular Biology* **225**, 379 (1992).
23. W. K. Olson, A. A. Gorin, X. J. Lu, L. M. Hock, V. B. Zhurkin, DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 11163 (1998).
24. R. K. Allemann, M. Egli, DNA recognition and bending. *Chemistry & Biology* **4**, 643 (1997).
25. J. Cavanagh, *Protein NMR spectroscopy : principles and practice*. (Academic Press, Amsterdam ; Boston, ed. 2nd, 2007), pp. xxv, 885 p.
26. J. Kypr, I. Kejnovska, D. Renciuk, M. Vorlickova, Circular dichroism and conformational polymorphism of DNA. *Nucleic Acids Research* **37**, 1713 (2009).

27. Q. Zhang, X. Y. Sun, E. D. Watt, H. M. Al-Hashimi, Resolving the motional modes that code for RNA adaptation. *Science* **311**, 653 (2006).
28. J. R. Bothe, K. Lowenhaupt, H. M. Al-Hashimi, Sequence-specific B-DNA flexibility modulates Z-DNA formation. *J Am Chem Soc* **133**, 2016 (2011).

Chapter 4

Incorporation of Non-CG Dinucleotides into Z-DNA: Interplay Between B/Z Junction and Z-DNA Helical Formation

4.1 Introduction

Formation of Z-DNA within the genome, stabilized either by negative supercoiling or proteins, requires the incorporation of a heterogeneous mix of sequences into its left-handed helix and the creation of junctions between Z-DNA and B-DNA[1]. Z-DNA forms sequence-specifically and is known to thermodynamically favor incorporation of sequences rich in pyrimidine/purine dinucleotide steps. For example, it has been estimated that some non-pyrimidine/purine sequences have a several fold higher free energy cost for the B-to-Z transition than the most energetically favorable pyrimidine/purine CG dinucleotide step[2]. Predictions of genomic Z-DNA formation have been carried out using a "thermogenomic"[3] approach where genomic Z-DNA "hot-spots" are identified by calculating the thermodynamic propensity for sequences to form Z-DNA. A sequence's thermodynamic propensity to form Z-DNA is predicted by simultaneously accounting for the free energy cost for its dinucleotide steps to undergo the B-to-Z transition and the formation of B/Z junctions. In general, these computational prediction studies found that Z-DNA favoring sequences are typically localized near transcription start sites, suggesting that Z-DNA may play a role in transcription[4-8]. These *in silico* predictions were recently put to test in an experimental study that probed the occurrence of Z-DNA in human tumor cells[7]. Strikingly, it was found that only 1% of the observed Z-DNA "hot-spots" were localized near transcription start sites.

Thus, despite these previous Z-DNA studies, the sequence-specificity of Z-DNA formation is still not fully understood. In current models, the free energy penalty for junction formation is treated to be equal for all sequences. However,

recent experiments in our lab provided evidence that not all B/Z junctions are created equal and that the B-to-Z transition is highly sensitive to the sequence identity of the B/Z junction[9]. These interesting observations concerning the formation of B/Z junctions warrant further investigation of the influence of sequence-specific effects on the B-to-Z transition.

Very few Z-DNA studies have been reported on model duplex systems amenable to structural characterization for non-$(CG)_x$ repeat DNA's. This is likely due to the major difficulty in inducing the B-to-Z transition when B/Z junctions are present because very specific conditions are necessary for the transition to occur such as the presence of 5.4 M NaCl or proteins that stabilize Z-DNA. Further, depending on the technique used, it is difficult to decipher where B/Z junctions form within mixtures of B and Z-DNA. In this chapter we develop improved methods for determining the location of B/Z junctions within sequences containing mixtures of B and Z-form DNA. Using our technique, we find that multiple energetically unfavorable sequence steps can be incorporated into Z-DNA helices in order to achieve the most energetically favorable B/Z junction. Our results reveal that sequences thought to disfavor Z-DNA formation will readily adopt Z-DNA, suggesting that certain sequence elements in genomes may have a higher propensity to form Z-DNA than previously thought.

4.2 Materials and Methods

4.2.1 Sample Preparation

DNA oligos were purchased from Integrated DNA Technologies (Coraville,IA) and purified by standard desalting with the exception of Z-19 which was purified by HPLC. Single-stranded DNA was suspended in 15 mM phosphate, 25 mM NaCl, 0.1 mM EDTA at pH 7.5. The concentration of each single stranded oligo was determined by UV using extinction coefficients provided by IDT. DNA oligos were then combined in equal molar ratios and heated to 95 °C for 5 minutes and allowed equilibrate at room temperature for several hours to allow duplex formation. The Zα domain of ADAR1 was prepared closely following a procedure described previously with the following modifications[10]. Protein expression was induced with

0.2 mM IPTG at 1.0 A_{600} units. The protein was purified using HisSelect resin (Sigma), followed by HiTrapS chromatography, without the intervening thrombin cleavage step.

4.2.2 Fluorescence Spectroscopy

Steady-state fluorescence spectroscopy experiments were carried out using a Fluormax-2 fluorimeter. Fluorescence spectra were recorded using an excitation wavelength of 320 nm and emission was recorded from 335-500 nm at a temperature of 25 °C. The Zα domain was titrated into DNA samples (2.5 μM DNA) and the complex was allowed to equilibrate such that the fluorescence intensity was constant (typically 30 minutes). Fluorescence spectra were background corrected by subtracting an emission spectrum of the buffer. All sequences studied were found to reach saturation with Zα at a molar ratio less than or equal to 6:1. Relative changes in fluorescence intensities (in fold) were calculated by comparing the ratio of free and fully Zα bound fluorescence intensities at 370 nm.

4.2.3 CD Spectroscopy

Circular dichroism (CD) spectroscopy experiments were performed using an Aviv 62DS CD spectrometer. DNA samples (2.5 μM DNA) were pipetted into a 1 cm quartz cuvette and spectra were recorded in duplicate using a 1 nm interval and averaged for 2 seconds at a temperature of 25 °C. CD spectra were baseline corrected by subtracting reference spectra consisting of either buffer or buffer with Zα protein of the same concentration as bound DNA samples. DNA samples for CD experiments were prepared independently from fluorescence experiments to ensure that the titrated Zα protein concentration was identical to the Zα reference for background subtraction. We note that small sample losses resulted from pipetting over the course of the fluorescence Zα titrations therefore requiring preparation of independent CD and fluorescence samples. Zα bound CD spectra used for fitting were recorded with a 6:1 molar ratio of protein to DNA.

4.3 Results and Discussion

4.3.1 2-Aminopurine as a Z-DNA Probe

Bulk spectroscopic techniques such as circular dichroism (CD)[11,12] and Raman[13,14] spectroscopy have previously been employed to characterize B/Z junctions, but interpretation of the data is often difficult due to a lack of site-specific resolution. Recently, Kim *et al.* used 2-aminopurine (2AP) as a site-specific probe of B/Z junctions and observed significant structural perturbations at the site of junction formation, consistent with the B/Z junction crystal structure[15]. 2AP, which reports on base stacking interactions[16], is considered an ideal structural probe as its incorporation induces miniscule structural deviations compared to adenine[17,18]. When 2AP loses base stacking interactions, its steady-state fluorescence intensity increases ~5-50 fold making it an ideal probe of B/Z junctions where bases become extruded and unstacked (Figure 4.1a). The method to probe B/Z junctions reported by Kim *et al.* involves inducing the B-to-Z transition by binding DNA to the Zα domain of the RNA adenosine deaminase (ADAR1) protein. The Zα domain is ideal for studying the formation of Z-DNA as it has been shown to be Z conformation specific rather than sequence-specific by binding non-$(CG)_x$ and non-pyrimidine/purine repeats in the Z conformation[19,20].

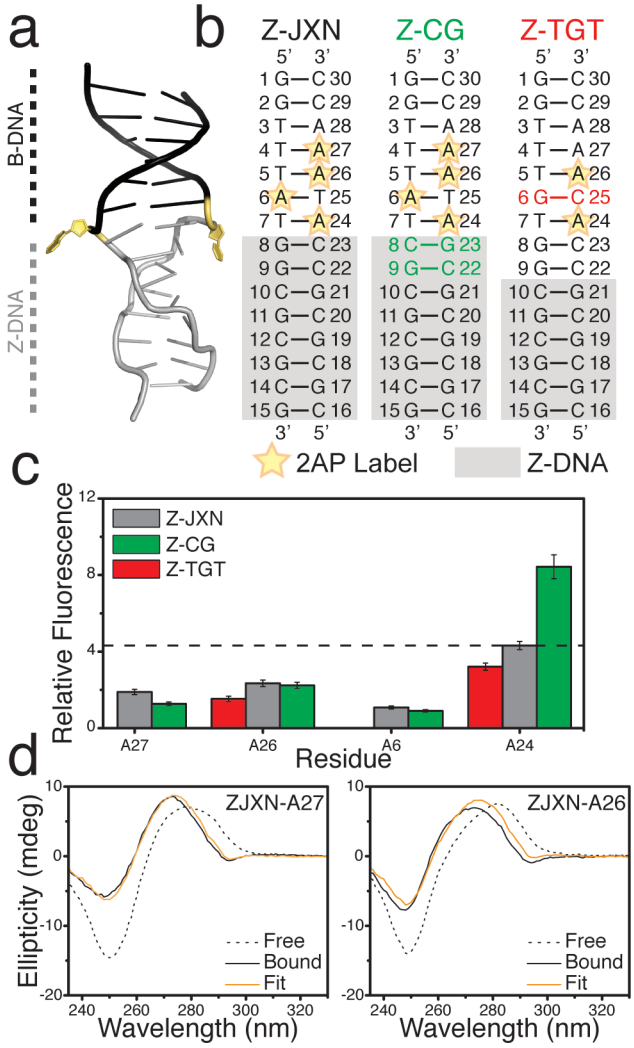

Figure 4.1: Combined CD and fluorescence data to probe B/Z junctions. (a) Crystal structure of a B/Z junction composed of B-DNA (black), Z-DNA (grey) and extruded bases (yellow), PDB: 2ACJ[21]. (b) DNA constructs with 2AP fluorescent probes indicated with a star. Each DNA construct had one 2AP label, so there were 10 independently labeled samples. (c) Relative change in fluorescence (Zα bound intensity/free intensity) at 370 nm upon undergoing the B-to-Z transition for Z-JXN, grey; Z-CG, green; Z-TGT, red. The fluorescence increase for A24 in Z-JXN is highlighted with a dashed line. (d) Example fits of CD data for Z-JXN with a 2AP incorporated at bases A27 (left) and at A26(right).

Using 2AP labeling we sought out to gain new insights into B/Z junction formation but first validated its use on the well-characterized B/Z junction crystal structure sequence Z-JXN (Figure 4.1b). We inserted a 2AP at the site of the extruded base (A24) in Z-JXN and observed the expected significant 4.3X increase in fluorescence intensity upon inducing the B-to-Z transition by binding to the Zα

domain (Figure 4.1c). Incorporation of the 2AP at other sites (Figure 4.1b,c) resulted in smaller (1.1-2.5X) changes upon adoption of Z-DNA, confirming the expectation that the biggest change in fluorescence is localized at the B/Z junction.

While our fluorescence studies allowed us to probe the formation of a B/Z junction, structural interpretation of changes in 2AP steady-state fluorescence have proven to be elusive in many systems[22-24]. A change in fluorescence intensity indicates that a structural change may have occurred, but the identity of the structural change cannot be determined based purely on steady-state fluorescence. There are several conformational changes that could lead to the observed changes in steady-state fluorescence intensity. Thus, while use of multiple 2AP probes allowed us to pinpoint the location of the junction in Z-JXN, the conformation of the junction cannot be deduced. Aside from the looped out conformation reported in the X-ray structure, the junction could in principle also correspond to either a Z-DNA or B-DNA like conformation.

Previously, it was shown that 2AP undergoes a large enhancement in fluorescence upon undergoing a B-to-Z helical transition[25]. We therefore examined the effects on 2AP fluorescence intensity arising due to a conformational change from B-DNA to Z-DNA induced by the Zα domain. Here we incorporated a 2AP into the center of a $(CA)_6$ duplex and observed a large 6.2X increase in fluorescence intensity upon formation of Z-DNA (Figure 4.2), consistent with the large enhancement previously observed[25]. This significant change in fluorescence can be attributed to differences in base stacking interactions between B-DNA and Z-DNA. When the 2AP is in the B conformation, it stacks with both its 3' and 5' neighboring bases; however, while in the Z conformation it only has a strong stacking interaction with its 5' neighboring base (Figure 4.3,4, red purine). This loss of stacking interactions when in Z-form results in a large increase in fluorescence intensity compared to B-DNA. Thus, without the use of multiple 2AP labels we cannot definitively pinpoint the location of a B/Z junction based purely on fluorescence because of the large changes observed for both B/Z junctions and Z-DNA.

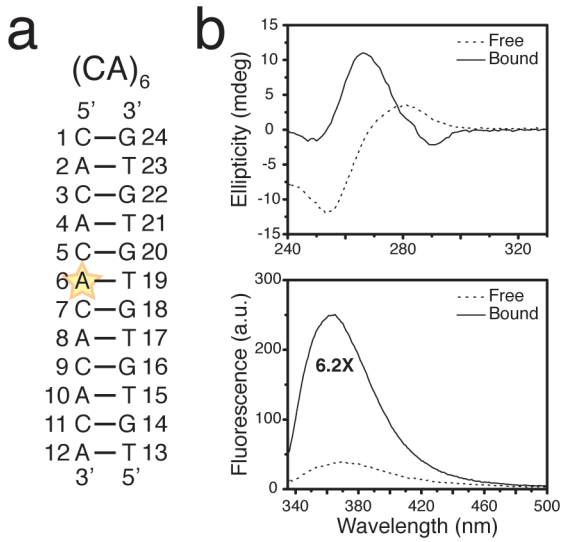

Figure 4.2: CD and fluorescence spectra of a 2AP labeled $(CA)_6$. (a) Sequence of $(CA)_6$ with 2AP (indicated with a star) installed in the middle of the sequence. (b) CD and fluorescence spectra of $(CA)_6$ when free (dashed) and bound (solid) to the Zα domain.

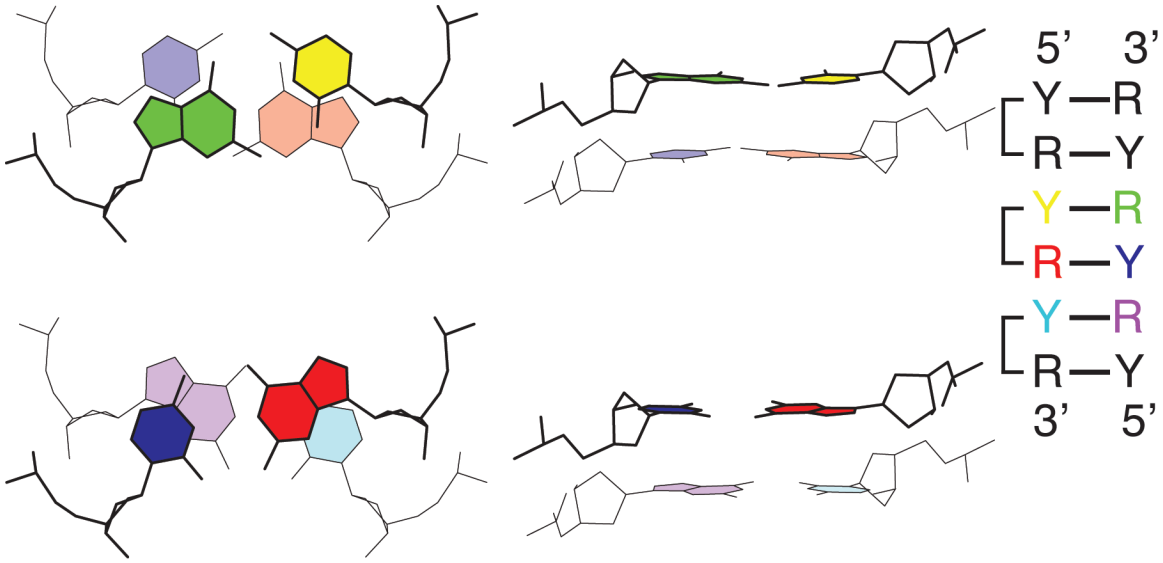

Figure 4.3: B-DNA base stacking patterns for pyrimidine (Y)/purine (R) repeats.

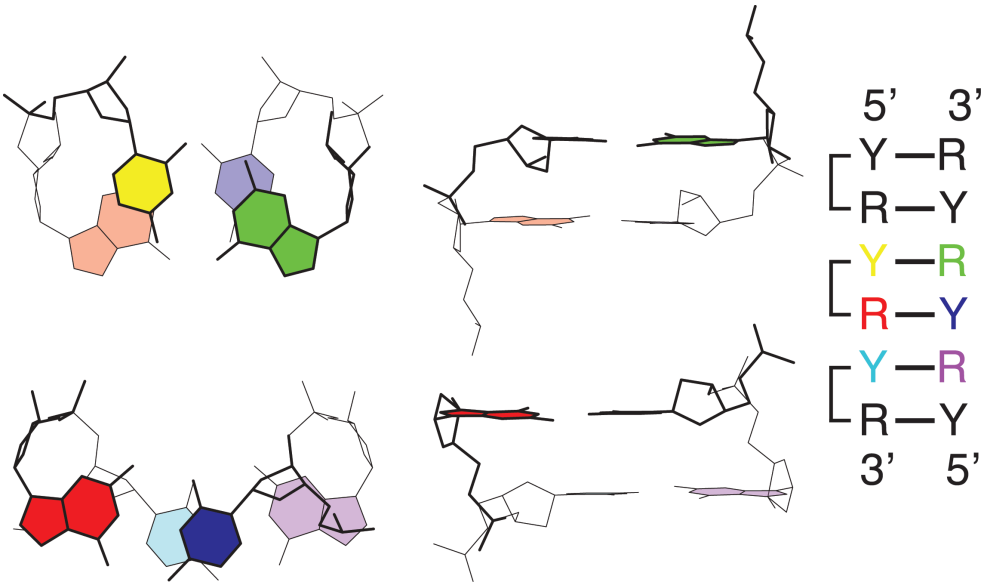

Figure 4.4: Z-DNA base stacking patterns for pyrimidine (Y)/purine (R) repeats.

4.3.2 CD Spectroscopy as a Probe of B/Z Junction Formation

Although the incorporation of 2AP into Z-DNA or B/Z junctions results in similar fluorescence signatures, these structural transitions are expected to be quite different when observed by CD. Thus, we combined fluorescence with CD spectroscopy to better pinpoint the location of B/Z junctions in DNA duplexes with mixtures of B and Z-DNA. CD spectroscopy is one of the most commonly used methods to monitor the B-to-Z transition, where the CD spectrum of Z-DNA is inverted compared to B-DNA[26,27]. Additional complexity arises when studying systems that contain a B/Z junction, because the CD spectrum is composed of a mixture of B-DNA, Z-DNA, and a B/Z junction. Surprisingly, very few studies have employed quantitative methods such as linear combinations[28] or singular value decomposition[29] to analyze the B-to-Z transition by CD.

To better define the extent of Z-DNA formation via CD, we quantitatively analyzed our data in a manner similar to protein studies where the CD spectrum is assumed to be a linear combination of its structural components. For proteins, linear combinations of secondary structure (alpha helix, beta sheet, etc) reference spectra are made to best fit an experimental CD spectrum, and the resulting

structural composition fits can achieve up to ~5-15% accuracy[30]. These principles were used for fitting,

$$\theta_{EXP} = \alpha(\theta_B) + \beta(\theta_Z) \quad (1)$$

where θ_{EXP} is the experimental CD spectrum , α is the percentage of B-DNA, θ_B is the free B-form CD spectrum, β is the percentage of Z-DNA, and θ_Z is a pure Z-DNA spectrum (Figure 4.7, $(CG)_6$). Here, we ignore contributions from the B/Z junction, which cannot be independently measured, but is expected to account for at most 10% of the CD signal for the size of constructs studied. Also, unlike the chemically equivalent protein backbone, the spectral region (220-230 nm) used for nucleic acids reports on the local structure of the chemically heterogeneous nucleobases of which each has slightly different optical properties. Therefore, CD provides a rough estimate of how much Z-DNA is present, allowing for a better verification of junction location when combined with fluorescence data.

We used Z-JXN, which has a known X-ray structure, as a benchmark for CD fitting analysis[21]. First, the DNA contribution to the Zα domain bound CD spectra was isolated by subtracting a spectrum of the free Zα domain. The Zα domain undergoes very small structural changes upon binding Z-DNA; therefore, its free and bound CD spectra are likely to be nearly identical making it a valid background subtraction[31]. After Zα subtraction, CD spectra were fit using Equation 1 for wavelengths spanning 240-310 nm. The wavelength cutoff of 240 nm was chosen for fitting because the Zα domain has a much stronger CD signal than DNA at lower wavelengths therefore making background subtraction difficult. Fits to the individually 2AP labeled Z-JXN samples (Figure 4.1b, Table 4.1) yielded 6.5-8.5 base-pairs in the Z-DNA conformation, which is generally consistent with the 8 Z-DNA base-pairs observed in the crystal structure of Z-JXN[21]. Example fits to experimental CD data are presented in Figure 4.1d. One of the better fits to experimental spectra is shown in Figure 4.1d where the best fit for Zα bound Z-JXN-A27 indicates that 7.3 base-pairs are in the Z-DNA conformation. In contrast, some spectra are difficult to fit as is shown for Z-JXN-A26 where the best fit was obtained with 6.5 base-pairs in Z-DNA (Figure 4.1d). The difficulty in fitting the Z-JXN-A26 Zα bound spectrum

arises because each half of the spectrum results in different estimations of Z-DNA, where < 270 nm indicates less while > 270 indicates more Z-DNA. This inability to perfectly fit the spectra could arise from sequence-dependent effects or possibly the signal arising from the B/Z junction.

Location of 2AP label	Base-pairs in Z conformation
A24	8.52
A6	6.49
A26	7.53
A27	7.29

Table 4.1: CD fitting results for each independently 2AP labeled Z-JXN construct bound to the Zα domain.

To further test our CD fitting procedure we studied an additional construct similar to Z-JXN where the thermodynamically unfavorable 5' CC 3' step in Z-JXN was substituted with a Z-DNA energetically favorable CG step to assure junction localization at A24 (Figure 4.1b, Z-CG). Fluorescence experiments were carried out on Z-CG with 2AP labeling identical to that of Z-JXN and similar trends were observed (Figure 4.1b,c). As expected, the largest increase in fluorescence intensity occurred at A24 (8.4X) while changes in other residues that remain in B-form ranged from 0.9-2.2X. Upon validating by fluorescence that Z-JXN and Z-CG have similar structural perturbations upon undergoing the B-to-Z transition, we compared their changes in CD. Fits to the individually 2AP labeled Z-CG samples ranged from 9.0-11.1 base-pairs in Z-DNA where 8 base-pairs are expected from the fluorescence data (Table 4.2). This likely overestimation of Z-DNA content for some Z-CG samples helps to define the limits of using just CD to determine the location of a B/Z junction. CD fitting can be used to estimate the junction location to within a few base-pairs. As DNA length increases, the fitting accuracy is expected to become slightly worse because the B-to-Z change in CD signal per nucleotide becomes smaller as a function of sequence length. We also note that the magnitude of the unknown junction signal with respect to B and Z CD signals becomes smaller as DNA length increases. Taken together, the large ranges observed from CD fitting and

uncertainty in the structural interpretation of fluorescence data necessitates their combination to obtain the best estimate of the location of B/Z junctions within DNA duplexes that contain mixtures of B and Z-DNA.

Location of 2AP label	Base-pairs in Z conformation
A24	10.49
A6	11.13
A26	9.24
A27	8.99

Table 4.2: CD fitting results for each independently 2AP labeled Z-CG construct bound to the Zα domain.

4.3.3 B/Z Junction Mutations Modulate Z-DNA Formation

With a combined fluorescence and CD method to determine the length of Z-DNA segments in B/Z containing sequences in hand, we sought to explore the formation of B/Z junctions upon binding the Zα domain in multiple sequence contexts. As a basis for our study, we started by investigating sequence-dependent effects in Z-JXN. Based on the thermodynamic preferences for Z-DNA formation, the B/Z junction would be predicted to most likely form at the end of Z-JXN's CG repeat[2]. However, the X-ray structure reveals a different junction location where a thermodynamically unfavorable 5' CC 3' step is incorporated into the Z-DNA helix therefore localizing the junction at a site multiple nucleotides downstream of the CG repeat (Figure 4.1a,b). In current Z-DNA prediction models, the free energy for junction formation is treated to be equal for all sequences. However, in a recent study we showed that instability in the B conformation enhances Z-DNA formation likely by tuning the energetic cost of B/Z junction formation (Chapter 3)[9]. Here we showed that the B/Z junction observed for Z-JXN is localized at a site that is highly flexible when in the B-form state. Mutations that rigidified this intrinsic flexibility (Z-TGT without 2AP, Figure 4.1a) resulted in a reduction in the propensity to form Z-DNA. In our previous study we used CD spectroscopy to probe the extent of Z-DNA formation for both Z-JXN and Z-TGT without 2AP labeling. We observed ~10-15% more Z-DNA in Z-JXN compared to Z-TGT upon inducing the B-to-Z transition with

the Zα domain suggesting that Z-TGT likely did not incorporate the thermodynamically unfavorable CC step to localize the junction at A24.

In this chapter we have further probed B/Z junction formation in Z-TGT by incorporating multiple 2AP labels. A label placed at A26 underwent a small 1.5X change in fluorescence as expected upon undergoing the B-to-Z transition and is consistent with Z-JXN where A26 remains in the B conformation (Figure 4.1b,c). However, in contrast to Z-JXN, a 2AP placed at A24 undergoes a less significant 3.2X increase in fluorescence intensity further confirming that A24 of Z-TGT may not become extruded upon Z-DNA formation. In addition, the best fit of the CD spectrum for Z-TGT with a 2AP label at A24 was achieved with 5.2 base-pairs in the Z conformation (Figure 2.1b). This is consistent with the fluorescence data that shows A24 does not undergo as significant of a change in Z-TGT as in Z-JXN. Both fluorescence and CD data of 2AP labeled Z-TGT constructs support that upon binding the Zα domain, either the junction does not localize at A24 or there is a reduced extent of Z-DNA formation.

4.3.4 Incorporation of CC steps into Z-DNA Helices

Next, we sought to explore the incorporation of more non-pyrimidine/purine steps into Z-DNA helices in the presence of Z-JXN's B/Z junction favoring sequence element (5' CCA 3'). We added an additional CC step to Z-JXN (Z-17) and labeled A28 which is comparable to A24 in Z-JXN and monitored changes in fluorescence upon undergoing the B-to-Z transition with the Zα domain (Figure 4.5a,b). A significant 11.1X increase in fluorescence intensity was observed indicating that the additional CC step was very likely incorporated into the Z-DNA helix (Figure 4.5b). Fitting of the of Z-17 CD data indicated that 9.6 base-pairs were in Z-form, consistent with 10 base-pairs necessary to localize the junction at A28 (Figure 4.5a). Next, we added two CC steps to Z-JXN (Z-19) and installed a 2AP at A30 which is comparable to A24 in Z-JXN. We observed a significant 5.6X increase in fluorescence intensity upon inducing the B-to-Z transition with the Zα domain, indicating that all CC steps had been incorporated into the Z-DNA helix (Figure 4.5a,b). Also, fitting of the Z-19 CD

data indicated that 10.8 base-pairs are in the Z-form, which is consistent with 12 base-pairs necessary to localize the junction at A30 (Figure 4.5a). The variation in the steady-state fluorescence intensity increases observed with different additions of CC steps (Z-JXN, Z-17, Z-19) likely reflects differences in local stability at the 2AP when in the B-form state. We did not explore further additions of CC steps as this results in long continuous stretches of guanine that have a high propensity for quadruplex formation making them difficult to synthesize. By combining changes in 2AP fluorescence with CD fitting we were able to pinpoint the B/Z junction in Z-17 and Z-19 using only one 2AP label. These results suggest that dinucleotide steps outside the energetically favorable pyrimidine/purine sequence space willingly adopt the Z conformation likely in order to achieve the most thermodynamically favorable B/Z junction.

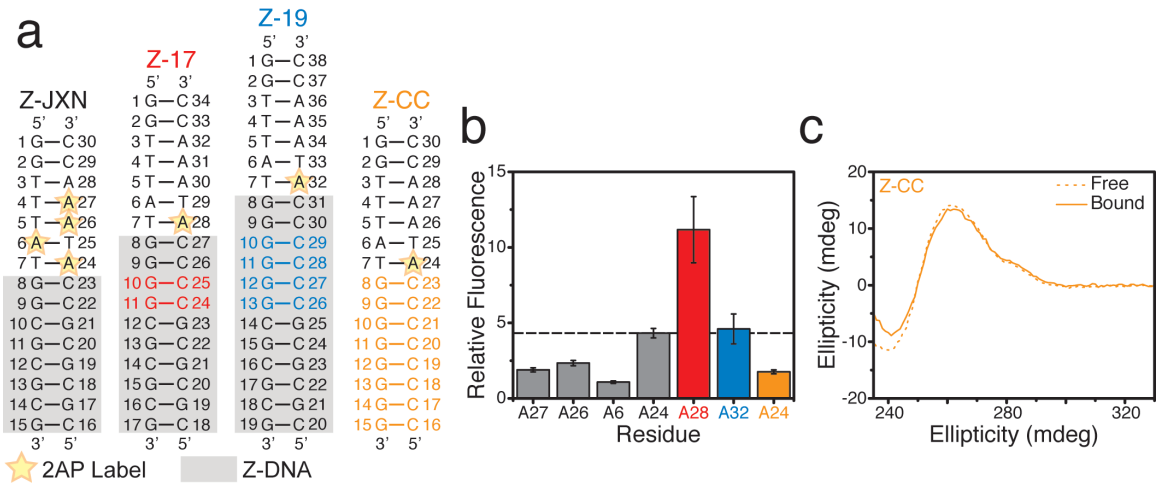

Figure 4.5: Probing the incorporation of CC steps into Z-DNA. (a) DNA constructs where the location of a 2AP fluorescent probe is indicated with a star. (b) Relative change in fluorescence (bound intensity/free intensity) upon inducing the B-to-Z transition with the Zα domain for Z-JXN, grey; Z-17, red; Z-19, blue; Z-CC, orange. The fluorescence increase for A24 in Z-JXN is indicated with a dashed line. (c) CD spectra of Z-CC when free (dashed) and bound (solid) to the Zα domain.

The surprising incorporation of unfavorable sequences into Z-DNA with simultaneous B/Z junction formation led us to probe additional sequences in order to better understand this effect. One possible explanation for the incorporation of CC steps into Z-JXN, Z-17, and Z-19 is that CC steps would readily form Z-DNA upon binding the Zα domain. We tested if continuous CC steps adopt the Z conformation

without being adjacent to a CG repeat (Figure 4.5a, Z-CC). Interestingly, titration of the Zα domain with the fluorescently labeled Z-CC resulted in a binding curve with saturation at a ratio of 4:1 Zα domain to DNA indicating that the protein was somehow binding the DNA (data not shown). However, the 2AP labeled A24 underwent a 1.8X increase in fluorescence intensity, which is smaller than the changes in fluorescence observed in the B-form region of Z-JXN (Figure 4.5b: Z-JXN-A28,A27,A6). CD spectra of Z-CC showed minimal changes between free and bound spectra indicating that the structure of Z-CC does not change significantly upon binding the Zα domain (Figure 4.5c). Therefore, the Zα domain likely undergoes non-specific interactions with Z-CC. We note that the CD spectrum of Z-CC has spectral features that indicate quadruplex-like structures may be present to some extent (Figure 4.6). However, regardless of some quadruplex formation, the unchanged CD spectrum in Figure 4.5c indicates that the Zα domain does not induce a helical Z transition in Z-CC. The failure of the Zα domain to induce the B-to-Z transition in Z-CC shows that binding the Zα domain does not force CC steps into the Z conformation. Therefore, the unfavorable CC steps in Z-JXN, Z-17, and Z-19 are only incorporated into Z-form helices when coupled with thermodynamically favorable $(CG)_x$ stretches.

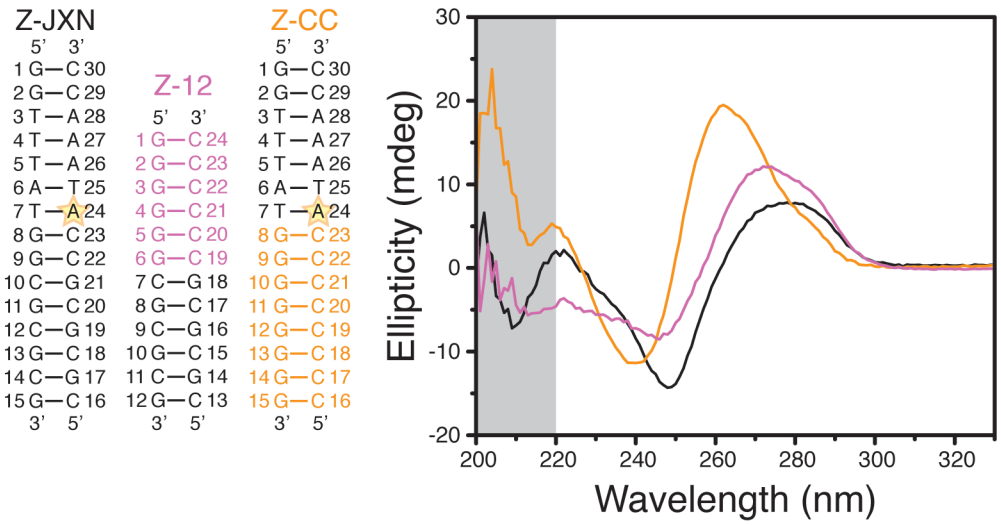

Figure 4.6: CD conformational analysis of control constructs in their Zα free state. Plotted are CD spectra of Z-JXN (black), Z-12 (pink), Z-CC (orange) in their free form with no Zα present. The CD spectrum of Z-JXN is shown as a B-form reference. Highlighted in grey is the spectral region that can be used to identify A-form or quadruplex structures. Both A-form and parallel quadruplex structures have similar CD spectra with the exception of the grey region, where parallel quadruplex and A-form structures have positive and negative ellipticities, respectively[27,32]. The CD spectrum of Z-CC free appears to be a mix of quadruplex and B-DNA based on the positive signal in the grey region. Z-12 appears to be a mix of A and B-DNA given the negative CD spanning 200-220 nm.

We then sought to determine if the favorable junction in Z-19 drove extension of the Z-DNA helix to incorporate the $(CC)_3$ repeat. Here we decoupled the Z-DNA helix in Z-19 from its favorable junction, localized at a CA step, by independently inducing the B-to-Z transition in a sequence composed of the Z-DNA helix portion of Z-19 (Figure 4.7a, Z-12). In order to determine the extent of Z-DNA formation in Z-12 we carried out a titration with the Zα domain and monitored it with CD spectroscopy (Figure 4.7b). We note that the Zα free CD spectrum of Z-12 has some A-form spectral characteristics, probably because the $(CC)_3$ region likely adopts an A-like structure (Figure 4.6)[33]. The best fit of the fully bound Z-12 CD spectrum was obtained with 10.4 base-pairs in Z-DNA suggesting that the whole Z-12 sequence did not transition into Z-DNA and that likely only some of the CC repeat was incorporated into the Z-form helix. However, we do not put much weight on the

fit of Z-12 because the quality of the fit is low, possibly because of the A like structure that forms in the $(CC)_3$ region.

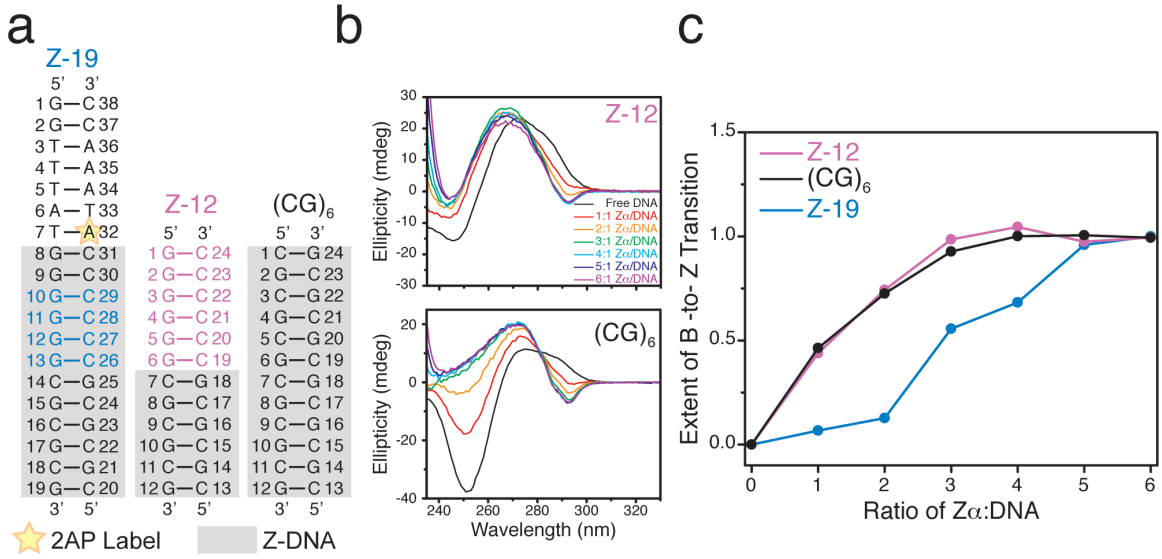

Figure 4.7: Probing the structure and stoichiometry of Zα domain/DNA complexes. (a) DNA constructs used in titrations where the location of a 2AP fluorescent probe in Z-19 is indicated with a star. (b) Titrations of Z-12 (top) and $(CG)_6$ (bottom) with the Zα domain monitored by CD spectroscopy. (c) Plot of the extent of the B-to-Z transition upon binding the Zα domain. The extent of the transition was determined by monitoring the change in 2AP fluorescence at 370 nm for Z-19 and change in CD at 293 nm for Z-12 and $(CG)_6$.

Although there is significant uncertainty in the specific location of the junction given the absence of fluorescence measurements, the binding titration of the Zα domain to Z-12 provides additional evidence that the B/Z junction likely forms near the end of the CG repeat. A titration of the Zα domain with Z-12 revealed that the transition completed at a ~4:1 ratio of Zα domain to Z-12 (Figure 4.7b,c). As a control, we carried out a titration with the most thermodynamically favorable 12 base-pair Z-DNA helix, $(CG)_6$, and observed a similar binding profile over the course of the B-to-Z transition where the transition completes at a ~4:1 ratio of Zα domain to DNA (Figure 4.7b,c). Comparison of both binding curves reveals nearly identical binding profiles. Previously it was shown that binding of the Zα domain was generally consistent with the sequence-specific B-to-Z free energy cost for dinucleotide steps[34]. Thus, if Z-12 adopted a pure Z-DNA helix lacking a B/Z

junction, the transition would complete at a greater ratio of Zα domain to DNA than $(CG)_6$ because Z-12 has an approximately two fold higher cost in free energy for the B-to-Z transition than $(CG)_6$.

A $(CG)_3$ sequence, which is found within Z-12, has been shown to bind two Zα domains and reach titration saturation at 2:1 Zα domain to DNA[35]. The higher binding stoichiometry observed for Z-12 compared to a $(CG)_3$ likely reflects the additional free energy cost of junction formation required when the Zα domain binds the $(CG)_3$ element within Z-12. Both fitting of CD data and the titration curve for Zα domain binding to Z-12 provide strong evidence that Z-12 forms a B/Z junction and does not incorporate a $(CC)_3$ repeat into a Z-DNA helix. It is likely that Z-12 forms a full Z-DNA helix in the context of Z-19 because the energetic cost of favorable junction formation, localized at A30, combined with incorporation of unfavorable CC steps into Z-DNA is less than the cost of junction formation adjacent to the CG repeat when binding the Zα domain.

In all sequences studied (Z-JXN, Z-17, and Z-19), formation of a favorable B/Z junction required incorporation unfavorable CC steps into a Z-DNA helix. The unfavorable CC steps in Z-JXN, Z-17, and Z-19 are only incorporated into Z-form helices when coupled with thermodynamically favorable $(CG)_x$ stretches. Once the Z-form helix is adopted by a stretch of pyrimidine/purines, more thermodynamically unfavorable steps propagate the Z-form helix in order to form the most favorable B/Z junction when binding the Zα domain.

4.3.5 Propagation of Z-DNA Helices with Additional Non-Pyrimidine/Purine Steps

Next we examined if dinucleotide steps other than CC that generally disfavor Z-DNA formation could be incorporated into Z-DNA helices upon binding the Zα domain. Samples were designed to have the same 5′ CCA 3′ junction as Z-17 while mutating the dinucleotide step downstream of the $(CG)_3$ repeat (Figure 4.8). Specifically, we inserted a GC step (Figure 4.8a, Z-17GC) which is predicted to be approximately two-fold in free energy more unfavorable for undergoing the B-to-Z transition than a CC step[2]. A small (2.9X) change in fluorescence was observed upon

undergoing the B-to-Z transition for Z-17GC and CD fitting indicated only 4.8 base pairs were in the Z conformation (Figure 4.8b,c). Here the B/Z junction is localized at the end of the $(CG)_3$ repeat where the thermodynamic advantage of forming the favorable CCA junction is outweighed by the penalty resulting from incorporation of a GC step into the Z-form helix.

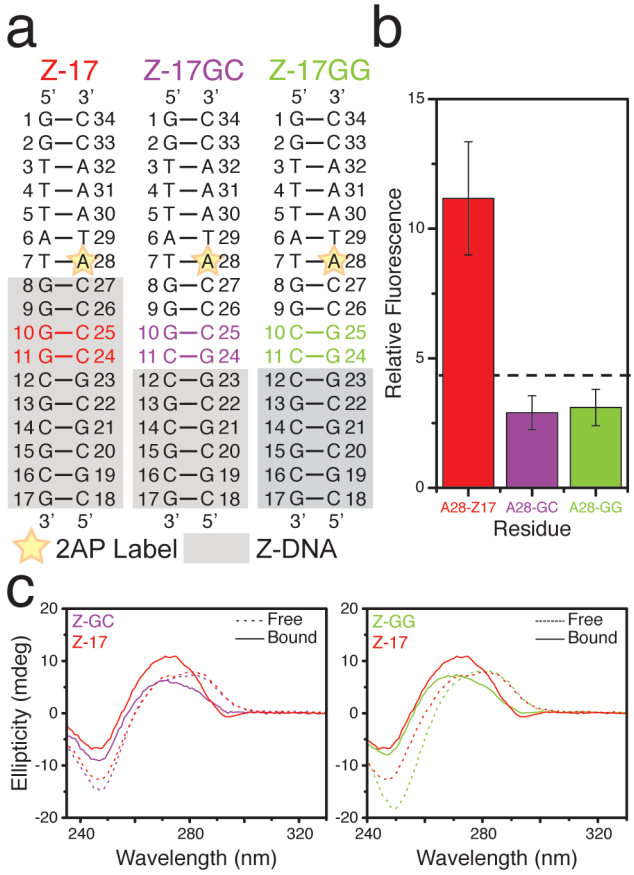

Figure 4.8: Z-DNA preferences for non-pyrimidine/purine repeats upon binding the Zα domain. (a) Fluorescently labeled constructs with 2AP fluorescent probes indicated with a star. (b) Relative change in fluorescence (bound intensity/free intensity) at 370 nm upon inducing the B-to-Z transition with the Zα domain for Z-17, red; Z-17GC, purple; Z-17GG, green. As a reference, the fluorescence increase for A24 in Z-JXN is indicated with a dashed line. (c) Comparison of free (dashed) and Zα bound (solid) CD spectra of Z-17 (red) to Z-17GC (purple, left) and Z-17GG (green, right).

We made a similar mutation in another construct, Z-17GG where the CC step of Z-17 was replaced with a GG. Based on the currently available B-to-Z transition thermodynamic data, the free energy cost of GG and CC B-to-Z helical transitions are

considered equal[2]. Thus, the junction would be expected to localize in the same location for both Z-17 and Z-17GG. Interestingly, similar to Z-17GC we found that the Z-form helix did not incorporate the non-pyrimidine/purine GG step (Figure 4.8a). Both fluorescence (3.1X increase, Figure 4.8b) and CD (7.4 base-pairs in Z-form, Figure 4.8c) experiments of Z-17GG indicate that the junction does not localize at A28, rather it is likely at the end of the CG repeat (Figure 4.8a). Our result suggests that either CC and GG steps do not have the same energetic cost for the B-to-Z transition or that the Zα domain may interact differently with each sequence. Both of these examples further highlight the energetic tug of war between junction formation and extension of Z-DNA helices where the lowest energy conformation wins. For Z-17GC and Z-17GG the Z-DNA helix only extended to the end of the $(CG)_3$ repeat, while in Z-17 multiple unfavorable CC steps were incorporated into Z-DNA in order to localize the junction beyond the CG repeat.

4.4 Conclusions

In conclusion, we have developed a method to determine the location of B/Z junctions within DNA duplexes that contain a mixture of B and Z-DNA by combining fluorescence and CD spectroscopy. Application of this combined approach revealed that sequences thought to disfavor the Z conformation are readily incorporated into Z-DNA in order to achieve the most favorable B/Z junction. Therefore, both the sequence dependence of B-to-Z helical conversions and B/Z junction formation play major roles in dictating the B-to-Z transition. These new surprising preferences for Z-DNA formation may expand the sequence space predicted to be available to Z-DNA in genomes. Further, the formation of Z-DNA *in vivo* necessitates the formation of two B/Z junctions. Consequently, this second junction might cause an even larger enhancement of the effects observed in our studies.

4.5 References

1. A. Rich, S. Zhang, Timeline: Z-DNA: the long road to biological function. *Nat Rev Genet* **4**, 566 (2003).

2. P. S. Ho, M. J. Ellison, G. J. Quigley, A. Rich, A Computer-Aided Thermodynamic Approach for Predicting the Formation of Z-DNA in Naturally-Occurring Sequences. *Embo Journal* **5**, 2737 (1986).
3. P. S. Ho, Thermogenomics: thermodynamic-based approaches to genomic analyses of DNA structure. *Methods* **47**, 159 (2009).
4. G. P. Schroth, P. J. Chou, P. S. Ho, Mapping Z-DNA in the Human Genome - Computer-Aided Mapping Reveals a Nonrandom Distribution of Potential Z-DNA-Forming Sequences in Human Genes. *Journal of Biological Chemistry* **267**, 11846 (1992).
5. P. C. Champ, S. Maurice, J. M. Vargason, T. Camp, P. S. Ho, Distributions of Z-DNA and nuclear factor I in human chromosome 22: a model for coupled transcriptional regulation. *Nucleic Acids Research* **32**, 6501 (2004).
6. P. Khuu, M. Sandor, J. DeYoung, P. S. Ho, Phylogenomic analysis of the emergence of GC-rich transcription elements. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 16528 (2007).
7. H. Li *et al.*, Human genomic Z-DNA segments probed by the Z domain of ADAR1. *Nucleic Acids Research* **37**, 2737 (2009).
8. D. Zhabinskaya, C. J. Benham, Theoretical Analysis of the Stress Induced B-Z Transition in Superhelical DNA. *Plos Computational Biology* **7**, (2011).
9. J. R. Bothe, K. Lowenhaupt, H. M. Al-Hashimi, Sequence-Specific B-DNA Flexibility Modulates Z-DNA Formation. *Journal of the American Chemical Society* **133**, 2016 (2011).
10. T. Schwartz *et al.*, Proteolytic dissection of Zab, the Z-DNA-binding domain of human ADAR1. *The Journal of biological chemistry* **274**, 2899 (1999).
11. R. D. Sheardy, S. A. Winkle, Temperature-Dependent Cd and Nmr-Studies on a Synthetic Oligonucleotide Containing a B-Z Junction at High Salt. *Biochemistry* **28**, 720 (1989).
12. M. J. Doktycz, A. S. Benight, R. D. Sheardy, Energetics of B-Z Junction Formation in a 16 Base-Pair Duplex DNA. *Journal of Molecular Biology* **212**, 3 (1990).
13. Z. Dai, G. A. Thomas, E. Evertsz, W. L. Peticolas, The Length of a Junction between the B-Conformations and Z-Conformations in DNA Is 3 Base-Pairs or Less. *Biochemistry* **28**, 6991 (1989).
14. W. L. Peticolas, Z. Dai, G. A. Thomas, The Use of Raman-Spectroscopy to Characterize Double B/Z Conformational Junctions in DNA. *Journal of Molecular Structure* **242**, 135 (1991).
15. D. Kim *et al.*, Base extrusion is found at helical junctions between right- and left-handed forms of DNA and RNA. *Nucleic Acids Research* **37**, 4353 (2009).
16. D. C. Ward, E. Reich, L. Stryer, Fluorescence studies of nucleotides and polynucleotides. I. Formycin, 2-aminopurine riboside, 2,6-diaminopurine riboside, and their derivatives. *The Journal of biological chemistry* **244**, 1228 (1969).
17. T. M. Nordlund *et al.*, Structure and dynamics of a fluorescent DNA oligomer containing the EcoRI recognition sequence: fluorescence, molecular dynamics, and NMR studies. *Biochemistry* **28**, 9095 (1989).

18. A. Dallmann *et al.*, 2-Aminopurine incorporation perturbs the dynamics and structure of DNA. *Angewandte Chemie* **49**, 5989 (2010).
19. A. Herbert *et al.*, The Zalpha domain from human ADAR1 binds to the Z-DNA conformer of many different sequences. *Nucleic Acids Research* **26**, 3486 (1998).
20. S. C. Ha *et al.*, The structures of non-CG-repeat Z-DNAs co-crystallized with the Z-DNA-binding domain, hZ alpha(ADAR1). *Nucleic Acids Research* **37**, 629 (2009).
21. S. C. Ha, K. Lowenhaupt, A. Rich, Y. G. Kim, K. K. Kim, Crystal structure of a junction between B-DNA and Z-DNA reveals two extruded bases. *Nature* **437**, 1183 (2005).
22. Y. V. Reddy, D. N. Rao, Binding of EcoP15I DNA methyltransferase to DNA reveals a large structural distortion within the recognition sequence. *Journal of Molecular Biology* **298**, 597 (2000).
23. H. Gowher, A. Jeltsch, Molecular enzymology of the EcoRV DNA-(Adenine-N (6))-methyltransferase: kinetics of DNA binding and bending, kinetic mechanism and linear diffusion of the enzyme on DNA. *Journal of Molecular Biology* **303**, 93 (2000).
24. C. Beck, A. Jeltsch, Probing the DNA interface of the EcoRV DNA-(adenine-N6)-methyltransferase by site-directed mutagenesis, fluorescence spectroscopy, and UV cross-linking. *Biochemistry* **41**, 14103 (2002).
25. R. Tashiro, H. Sugiyama, A nanothermometer based on the different pi stackings of B- and Z-DNA. *Angewandte Chemie* **42**, 6018 (2003).
26. F. M. Pohl, T. M. Jovin, Salt-induced co-operative conformational change of a synthetic DNA: equilibrium and kinetic studies with poly (dG-dC). *Journal of Molecular Biology* **67**, 375 (1972).
27. J. Kypr, I. Kejnovska, D. Renciuk, M. Vorlickova, Circular dichroism and conformational polymorphism of DNA. *Nucleic Acids Research* **37**, 1713 (2009).
28. V. Ivanov, K. Grzeskowiak, G. Zocchi, Evidence for an intermediate state in the B-to-Z transition of DNA. *Journal of Physical Chemistry B* **107**, 12847 (2003).
29. R. D. Sheardy *et al.*, Sequence dependence of the free energy of B-Z junction formation in deoxyoligonucleotides. *Journal of Molecular Biology* **231**, 475 (1993).
30. N. J. Greenfield, Using circular dichroism spectra to estimate protein secondary structure. *Nature protocols* **1**, 2876 (2006).
31. M. Schade *et al.*, The solution structure of the Z alpha domain of the human RNA editing enzyme ADAR1 reveals a prepositioned binding surface for Z-DNA. *Proc Natl Acad Sci U S A* **96**, 12465 (1999).
32. J. Kypr, M. Fialova, J. Chladkova, M. Tumova, M. Vorlickova, Conserved guanine-guanine stacking in tetraplex and duplex DNA. *Eur Biophys J* **30**, 555 (2001).
33. M. McCall, T. Brown, O. Kennard, The crystal structure of d(G-G-G-G-C-C-C-C). A model for poly(dG).poly(dC). *J Mol Biol* **183**, 385 (1985).

34. Y. G. Kim *et al.*, The zab domain of the human RNA editing enzyme ADAR1 recognizes Z-DNA when surrounded by B-DNA. *The Journal of biological chemistry* **275**, 26828 (2000).
35. T. Schwartz, M. A. Rould, K. Lowenhaupt, A. Herbert, A. Rich, Crystal structure of the Zalpha domain of the human editing enzyme ADAR1 bound to left-handed Z-DNA. *Science* **284**, 1841 (1999).

Chapter 5

Preliminary Studies of Supercoiled DNA: Development and Optimization of a DNA Minicircle System

5.1 Introduction

The current DNA structure-function paradigm is largely based on atomic-level studies of DNA structure under the experimentally accessible relaxed duplex state; it has largely ignored the forces to which DNA is constantly subjected to in the native cellular environment. In particular, supercoiling is key for packaging DNA into small volumes in prokaryotic and eukaryotic cells[1,2]. Transcription, replication, and other genetic processes also give rise to transient supercoiling since DNA must be unwound for polymerase action. The torsional stress resulting from DNA supercoiling can be absorbed by non-B-form structures such as Z-DNA, H-DNA, triplex DNA, and cruciforms[2]. The stress may also be relieved through more local deformations of the double helix such as kinking and extrusion of bases[1-3].

Supercoiling induced stress has been shown to play a variety of functional roles, most of which are believed to involve the modulation of DNA structure and dynamics. For example, it is widely accepted that supercoiling is a major transcription control element in bacteria[4]. Varying levels of supercoiling enhance structural deformations at transcription start sites, thus enabling a gene expression control system that is highly sensitive to supercoiling. Further, it was recently demonstrated for the first time in eukaryotic cells that transcription-induced supercoiling modulates gene expression through formation of non-canonical DNA structures[5,6]. Using the human MYC gene as a reporter, it was shown that dynamic supercoiling can propagate genetic signals over thousands of base-pairs.

A major challenge in studying supercoiled DNA at the atomic scale, or even at nucleotide resolution, is that supercoiled DNA plasmids contain thousands of

nucleotides and are therefore too large for characterization by techniques such as X-ray crystallography and NMR. Recently, supercoiled DNA minicircles consisting of less than 100 base-pairs have been prepared and used as a model system in biochemical studies to investigate the influence of superhelical forces DNA structure and stability[7-9]. These studies suggest the existence of pronounced supercoiling dependent deformations in small minicircles. Further, molecular dynamics simulations of minicircles show large perturbations involving inter-helical kinking and severe base-pair rupturing[10,11]. In this chapter, we establish procedures for the synthesis and purification of minicircles composed of a single topoisomer. Further, we carried out a variety of experiments to test if the sequence-specific changes in local dynamics and conformational transitions observed for linear duplexes in Chapters 3 and 4 manifest into observable changes in this supercoiled minicircle system.

5.2 Materials and Methods

5.2.1 Synthesis and Purification of DNA Minicricles

DNA oligonucleotides were purchased from IDT, Inc. (Coralville, IA), 5' phosphorylated, and PAGE purified. Oligonucleotides were designed to be consistent for ligation using the recently reported DNA Circularization by Long Cohesive Ends (DCLCE) method (Figure 5.1)[7]. Briefly, two double stranded sequences are designed so that they are half complimentary such that they can form nicked minicircles as depicted in Figure 5.1a,b. The nicked minicircles are then ligated with a DNA ligase protein (Figure 5.1c,d). The initial report of the DCLCE method carried out ligation reactions using 4 strands of two double stranded sequences, however we observed a higher yield of the desired minicircle product when using just two of the four strands. Therefore, in our synthesis of minicircles we only used 2 single strands as depicted in Figure 5.1a.

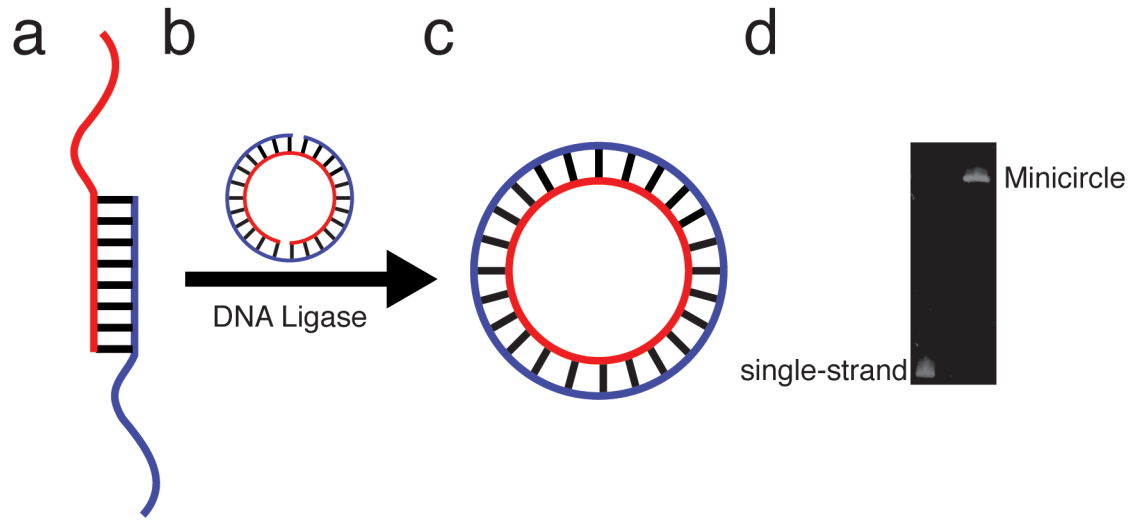

Figure 5.1: Schematic representation of DNA minicircle synthesis. a) Two single strands anneal together with floppy ends that are complimentary to each other. b) The complimentary floppy ends anneal to each other to form a nicked minicircle that is subsequently ligated with a ligase protein. c) DNA minicircle after ligation of nicks. d) Denaturing polyacrylamide gel of DNA before (single strand) and after (minicircle) the ligation reaction.

Ligation of DNA to produce supercoiled minicircles was achieved using either Taq or 9° North DNA ligase purchased from New England Biolabs (Beverly, MA). 9° North DNA ligase was preferred because it was more cost effective and produced negative topoisomers more readily compared to Taq ligase. Ligation reactions were carried out in PCR tubes using ~0.3 μM of each single strand oligonucleotide and ~1.2 Units/μL of either Taq or 9° North DNA ligase in the corresponding ligase reaction buffer (NEB) with volumes varying between 25μL – 200μL. We found that these conditions (protein/DNA ratio) resulted in good yield of the desired minicircle product. Generally, as the DNA concentration decreases, a larger ratio of minicircles to undesired products is obtained. However, a tradeoff of doing ligation reactions in very dilute conditions is that it takes many reactions to obtain large quantities of DNA minicircles. Ligation reactions were carried out using a Thermo Scientific Px2 thermal cycler with the following temperature cycling scheme consistent with the previously reported Ligation Assisted Minicircle Accumulation (LAMA) method: 30 seconds at 95 °C, followed by 4 °C for 1 minute and then ligation at 65 °C for 25 minutes[7]. Ligation reactions were typically carried out for 7 full temperature cycles

for an optimal combination of yield and time. We observed lower yields with fewer cycles and found that products no longer accumulate at a certain point likely due to loss in ligase activity.

After optimization of LAMA for a given minicircle sequence, a large scale preparation of minicircles was carried out followed by minicircle purification. When carrying out large scale reactions (1-2 mL of total ligation mixture), the LAMA procedure was run overnight with 14 cycles. The completed ligation reactions were then combined and concentrated to a small volume (~250-500 uL) for gel purification using an Amicon Ultra-4 centrifugal filter unit with a 3 kDa cutoff. DNA minicircles were purified from concentrated ligation reaction mixtures using 10% (w/v) denaturing polyacrylamide gel electrophoresis (PAGE) with 8M urea and TBE buffer. The minicircles were cut from the gel with the assistance of UV shadowing and extracted from the gel using a crush and soak method. Here the minicircle gel slice was chopped into small pieces and allowed to soak overnight in 50 mL of 0.5 M sodium acetate pH 8 buffer with gentle shaking. The acetate buffer containing minicircles was separated from the gel pieces and another 50 mL of acetate buffer was added to the pieces for further extraction of DNA. The sodium acetate solution containing DNA minicircles was then reduced to a volume of approximately 1 mL using an Amicon Ultra-15 centrifugal filter unit with a 3 kDa cutoff. The DNA minicircles were then further cleaned by incubating with Exonuclease I and III proteins (NEB) at 37 °C for 30 minutes to remove any single or double stranded non-circular DNA that was cut out with the DNA minicircle. Finally, the DNA minicircles were purified from the Exonuclease proteins and buffer exchanged into TE buffer using a Qiagen MinElute PCR purification kit. The yield calculated from input single stranded DNA to final purified DNA minicircles was approximately 10-15%.

5.2.2 Analysis of Topoisomer Distributions

Topoisomer distributions were determined by analyzing gel electrophoresis bands of minicircle ligation reactions. Minicircle ligation reactions were run on a 10% (w/v) denaturing polyacrylamide gel, stained with SYBR Gold (Invitrogen), and

imaged using a Kodak EDAS 290 imaging system. Topoisomers were quantified by determining the luminosity of their corresponding electrophoresis bands in black and white gel pictures using Photoshop CS4. Here the minicircle luminosity was determined by calculating the luminosity of each band followed by subtraction of the average surrounding luminosity.

5.2.3 BAL-31 Nuclease Minicircle Digestion Reactions

DNA minicircle digestion by BAL-31 nuclease (NEB) was monitored by measuring the luminosity of minicircle gel bands as a function of time upon addition of BAL-31. The reaction conditions for positively supercoiled minicircles was 50 nM DNA minicircle, 0.06 U/μL BAL-31 nuclease, 20 mM Tris-HCl, 300 mM NaCl, 6 mM $CaCl_2$, 6 mM $MgCl_2$, 0.5 mM EDTA at pH 8. The minicircles were incubated at 37 °C for 5 minutes prior to the addition of BAL-31 and the reactions were carried out at 37 °C. The rate of digestion was significantly faster for negatively supercoiled minicircles, therefore the reactions were run with the same conditions except at a lower BAL-31 concentration (0.005 U/μL) and temperature (25 °C).

5.3 Results and Discussion

5.3.1 Preparation of DNA Minicircles with Pure Topoisomers

The first step in most biophysical studies is to generate large quantities of uniform samples amenable to structural or dynamical characterization. Further, the ideal system would be small enough such that the effect of a sequence-specific property is observable within a sea of DNA. One major hurdle for studying small DNA minicircles is the challenge in synthesizing closed-circular DNAs that are less than 100 base-pairs. Recently, Du *et al.* reported a method, LAMA, to synthesize small minicircles in high yield (Figure 5.1) so we used this method to synthesize DNA mincircles[7]. First, we sought out to produce negatively supercoiled minicircles motivated by the presence of negative supercoiling in eukaryotic DNA as a result of genomic packaging and the stabilization of Z-DNA by negative supercoiling.

Supercoiling in DNA can be quantified by the linking difference "ΔLk"

$$\Delta Lk = Lk - Lk^{\circ} \quad (1)$$

where Lk is the linking number and Lk° is a reference linking number[1]. The linking number, Lk, describes the number of times double helical single strands are linked (like a chain) upon the creation of closed-circular DNA by formation of 5'-3' phosphodiester bonds to close each strand. Each helical turn results in one link between the single strands and Lk must be an integer because exact helical turns are necessary to properly phase the 5'-3' ends of DNA strands for ligation. Lk° is the reference linking number for the relaxed linear duplex form of the closed-circular DNA. It is described by the number of helical turns present in a relaxed duplex DNA

$$^{\circ}Lk = \frac{N}{h} \quad (2)$$

where N is the number of bases and h is the number of base-pairs per turn for a B-DNA helix (~10.5)[1]. Therefore, the linking difference, ΔLk (Equation 1), quantifies the strain put on the double helix to properly phase the 5'-3' ends for ligation where the helix must under or over wind in order to form closed-circular DNA.

The linking number can be broken down into two components:

$$Lk = Tw + Wr \quad (3)$$

where Tw (twist) describes the twisting of single strands about the helical axis and Wr (writhe) accounts for coiling of the helical axes in three dimensional space. It has been predicted and observed that small minicircles (< 100 bp) are planar and therefore have no writhe (Wr = 0)[12-14]. Thus, the linking difference for small minicircles is primarily determined by helical twist

$$\Delta Lk_{\text{minicircle}} = Tw - (\frac{N}{h}) \quad (4)$$

where Tw is the integer number of helical twists necessary for phasing the 5'-3' ends correctly for ligation and (N/h) has been substituted for Lk°. A negative linking difference means that the DNA has unwound and is negatively supercoiled and correspondingly a positively supercoiled minicircle is overwound and has a positive linking difference. Closed-circular DNA's with the same sequence, but different ΔLk are called topoisomers.

An advantage of using small minicircles is that their size can be used to tune the properties of supercoiling where the thermodynamically favorable topoisomer is predicted to have the smallest |ΔLk|. This is because supercoiling driven DNA destabilization is proportional to $(\Delta Lk)^2$; therefore smaller |ΔLk|'s result in less superhelical stress[1]. We were interested in producing negatively supercoiled minicircles, so we chose a length of 87 base-pairs which would be predicted to minimize ΔLk upon ligation favoring the negative topoisomer with ΔLk = -0.3 compared to ΔLk = 0.7 for the positive topoisomer.

The first 87 base-pair minicircle we synthesized, Mini87, had a sequence identical to a sequence previously investigated by Du *et al.* (Table 5.1)[7]. Upon carrying out ligation reactions we observed a large distribution of products (Figure 5.2) consistent with those reported by Du *et al.*[7]. Starting from the bottom of the gel in Figure 5.2a, we typically observed single-stranded DNA that failed to be ligated regardless of the number of cycles. Next, in Figure 5.1b, we observed the formation of single-stranded minicircles. These were confirmed to be single stranded minicircles by their resistance to cleavage by Exonuclease I and III and using a nicking reaction procedure similar to Du *et al.*[7]. The band observed in Figure 5.2c corresponds to the desired minicircle product, but upon careful inspection there are two bands, which indicates the formation of multiple topoisomers. This was a surprising observation as we predicted to only form one topoisomer (the negative topoisomer) based on the size of the minicircle. Finally, we observed very high molecular weight products (Figure 5.2d) that were stuck in the loading lane. We believe that these are also closed-circular DNA since they are not susceptible to cleavage by Exonuclease I and III. It is likely that during the cycling process many strands become ligated into large single strands that are multiples of the parent strand, and they eventually ligate with each other to form large closed-circular systems. Last, the streaking observed in all three lanes (Figure 5.1) disappeared upon introduction of Exonuclease I and III, indicating that the streaking corresponded to single or double stranded linear DNA that had been ligated during the LAMA procedure.

Minicircle	Sequence
Mini87	5´ CCT TGA TGT CGA TAG CAT TGC ACG GGT CTT GTT CGA TCG AAG ATA TCC TCA CCT GCA TCC TGA ACC CAT TGA CTC CCA AGC GAT AGG 3´
MiniZJXN	5´ CCT TGA TGT CGA TAG CAT TGC ACG GGT CTT GTT CGA TCG AAG ATA TCC TCA CCT GCA TGG TTT ATG GCG CGC GTC CCA AGC GAT AGG 3´
MiniZJXN^{cont}	5´ CCT TGA TGT CGA TAG CAT TGC ACG GGT CTT GTT CGA TCG AAG ATA TCC TCA CCT GCA TGG TTT ATG GCT GAC GTC CCA AGC GAT AGG 3´
MiniZJXN^{cont2}	5´ CCT TGA TGT CGA TAG CAG TGC ACG GGT CTT GTC CGA TCG AAG ATA TCC TCA CCT GCA TGG TTT ATG GCT GAC GTC CCA AGC GAT AGG 3´
MiniZA1	5´ CCT TGA TGT CGA TAG CAT TGC ACG GGT CTT GTT CGA TCG AAG ATA TCC TCA CCT GCA CAA AAA CGC GCG CGC GCG CCA AGC GAT AGG 3´
MiniZA2	5´ CCT TGA TGT CGA TAG CAT TGC ACG GGT CTT GTT CGA TCG AAG ATA TCC TCA CCT GCA CAA AAA GCG CGC GCG CGC CCA AGC GAT AGG 3´
Mini84	5´ CCT TGA TGT CGA TAG CAT TGC ACG GGT CTT GTT CGA TAA GAT ATC CTC ACC TGC ATC CTG AAC CCA TTG ACT CCC AGC GAT AGG 3´
Mini84^{TGT}	5´ CCT TGA TGT CGA TAG CAT TGC ACG GGT CTT GTT CGA CAA GAT ATC CTC ACC TGC ATC CTG AAC CCA TTG ACT CCC AGC GAT AGG 3´

Table 5.1: DNA Minicircle sequences. Mutations made on Mini87 and Mini84 are highlighted in yellow for each sequence.

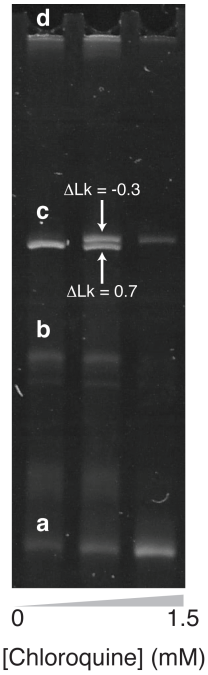

Figure 5.2: Denaturing PAGE of typical ligation reaction products with increasing concentrations of chloroquine (0.0 - 1.5 mM) from left to right. a) single-stranded DNA, b) single-stranded minicircles, c) DNA minicircles with positive and negative topoisomers d) high molecular weight products.

Upon observation of multiple topoisomers, we set out to identify the topoisomer associated with each band. Topoisomers in supercoiled DNA can be

identified by carrying out ligation reactions in the presence of an intercalating agent[1]. Intercalators such as ethidium bromide and chloroquine unwind DNA upon binding, therefore resulting in the formation of negative topoisomers upon ligation. We used chloroquine as an intercalator so as to not interfere with gel imaging, where ethidium bromide strongly fluoresces upon binding DNA. As shown in Figure 5.2, increasing the concentration of chloroquine increasingly drove the ligation product distribution into a single minicircle topoisomer corresponding to the slower migrating band in the gel. Therefore, the top band in Figure 5.2c corresponds to the negative topoisomer (ΔLk = -0.3). Generally for all minicircles studied, we observed the top band to be the negative topoisomer and found that the presence of ~1 mM chloroquine was sufficient to result in formation of only the negative topoisomer using our reaction conditions. Once the ligation conditions for producing pure topoisomers was determined, we carried out ligation reactions on a large scale as described in materials and methods. Figure 5.3 shows an example of an 87 base-paired negative topoisomer that was purified from all ligation products after carrying out the described purification protocol.

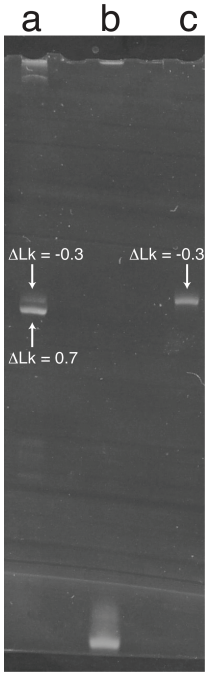

Figure 5.3: Denaturing PAGE of a negatively supercoiled 87 base-paired minicircle purified from ligation reactions. Lanes are: a) Ligation reaction of an

87 base-paired minicircle with the same products as described in Figure 5.2, b) Single stranded DNA c) Negatively supercoiled minicircle after purification protocol.

Surprised by the ligation products of Mini87, we synthesized another minicircle, Mini84, from the study of Du *et al.* (Table 5.1). Mini84 is composed of 84 bases, therefore the lowest energy topoisomer would correspond to a ΔLk of ~0. Interestingly, ligation of Mini84 also resulted in the formation of two topoisomers, with the product distribution skewed toward the top band (Figure 5.4) consistent with the ligation products observed by Du *et al.*[7]. Each topoisomer of Mini84 was determined by chloroquine titration and the top band became more populated upon addition of chloroquine, therefore indicating that the negative topoisomer (ΔLk = -1.0) was the major ligation product (Figure 5.4).

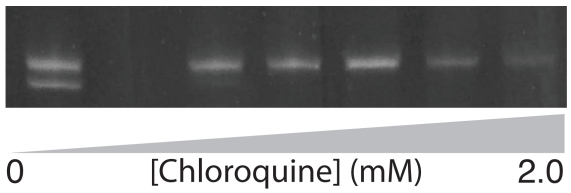

Figure 5.4: Chloroquine titration of the 84 base-pair minicircle Mini84. Lower and upper bands correspond to ΔLk = 0 and ΔLk = -1 topoisomers, respectively.

5.3.2 Probing the Sequence-Specific Formation of Topoisomers

The observation of these interesting topoisomer distributions upon ligation of Mini84 and Mini87 led us to explore if ligation topoisomer distributions are a function of sequence. We generated samples with a series of mutations (Figure 5.5a, Table 5.1), similar to sequences used in Chapters 3 and 4, and measured their topoisomer distribution upon ligation with 9° North ligase. The first mutation we made was inserting the Z-JXN sequence into Mini87 (MiniZJXN) to see if enhancing Mini87's propensity to form Z-DNA influenced the topoisomer distribution. Negative supercoiling enhances Z-DNA formation, therefore we predicted that if the Z-JXN element transiently formed a Z-like conformation, it would shift the topoisomer distribution to favor a ΔLk of -0.3[15,16]. Analysis of the MiniZJXN ligation reactions revealed a topoisomer distribution similar to Mini87, where the negative

topoisomer was populated ~30% of the time (Figure 5.5b). We also inserted ZJXN^{cont}, which does not form Z-DNA even in the presence of the Zα domain of ADAR1, into Mini87 as a Z-DNA control. Again, we observed a similar topoisomer distribution where the negative topoisomer was populated ~25% of the time (Figure 5.5b). One minor difference between MiniZJXN and MiniZJXN^{cont} is that their GC content differs slightly at 53% and 51%, respectively. For an additional control, we inserted additional GC base-pairs into MiniZJXN^{cont} to produce MiniZJXN^{cont2} which has a GC content of 53% (Table 5.1). Ligation of MiniZJXN^{cont2} resulted in a similar topoisomer distribution with the negative topoisomer forming ~20% of the time (Figure 5.5b). In summary, while we did detect slight differences in topoisomer distributions between multiple 87 base-pair minicircle sequences (ranging from 2-9%), these mutations did not cause dramatic sequence dependent changes in topoisomer distributions.

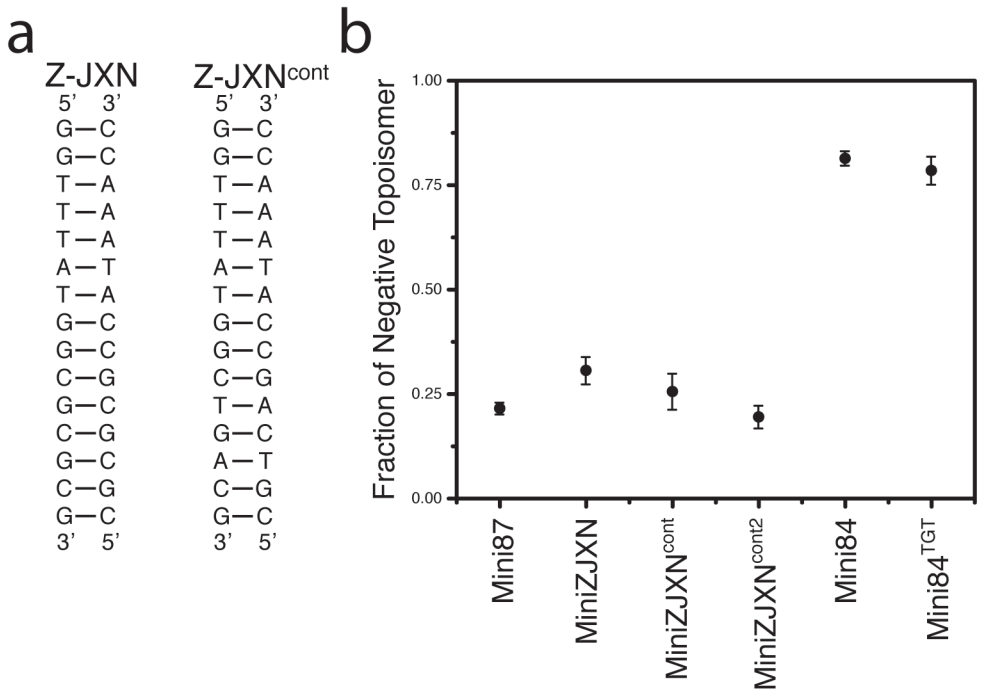

Figure 5.5: Sequence dependence of topoisomer distributions from ligation reactions. a) Sequences of Z-JXN and Z-JXN^{cont} that were inserted into Mini87 to result in MiniJXN and MiniZJXN^{cont}, respectively. b) Fraction of negative topoisomer formation that resulted from ligation reactions.

Next we tested if a sequence mutation in Mini84, which also had surprising ligation products, would perturb its topoisomer distribution. Here, we made a TAT to TGT mutation (Mini84^{TGT}, Table 5.1), which we have shown by NMR to result in a significant rigidification of the DNA duplex[17]. Upon ligation of Mini84^{TGT}, we observed a topoisomer distribution similar to Mini84 where ~80% of minicircles formed the negative topoisomer (Figure 5.5b). Therefore, this TGT mutation, which had a dramatic effect on duplex DNA conformational flexibility, did not significantly influence the distribution of topoisomers resulting from ligation. Overall, the 4 mutations made in the 84 and 87 base-paired minicircles did not significantly influence the distribution of topoisomers resulting from 9° North ligation reactions. For all 6 sequences studied, we only observed minor sequence-specific differences all roughly within error of each other.

5.3.3 Kinetics of Minicircle Digestion by BAL-31 Nuclease

Recent studies employing small DNA minicircles have used BAL-31 nuclease as a probe of deformations in supercoiled DNA[7-9]. Here the stress from supercoiling, either positive or negative, causes disruptions in the double helix and the BAL-31 protein subsequently cleaves the DNA backbone at the distorted site (Figure 5.6a)[18-20]. Generally, disruptions of the double helix are more pronounced in negatively supercoiled DNA since these disruptions can serve to reduce the superhelical stress on the double helix[3]. It has also been shown that the rate of digestion by BAL-31 is proportional to the number of deformed sites in supercoiled DNA[21]. Thus, the BAL-31 nuclease kinetic assay has the potential to provide insight into sequence-specific structural plasticity by probing rates of minicircle digestion (Figure 5.6a).

Similar to our studies of ligation topoisomer distributions, we sought out to determine if minor sequence-specific changes in supercoiled minicircles resulted in unique BAL-31 digestion patterns. Specifically, we inserted Z-DNA favoring sequences into minicircles to probe if minicircles with higher propensities to adopt the Z conformation have altered rates of digestion by BAL-31 nuclease. It has previously been shown that B/Z junctions are sensitive to cleavage by BAL-31 nuclease[20]. Our hypothesis was that supercoiled induced Z-DNA formation might

localize helical unwinding to a single site in negatively supercoiled minicircles; therefore, perturbing the distribution of deformed sites sensitive to BAL-31 digestion.

We began by measuring the digestion rates for a series of positively supercoiled minicircles using sequences with varying propensities to form Z-DNA. In these kinetic experiments, BAL-31 induced digestion of minicircles was monitored by measuring the intensity of gel electrophoresis minicircle bands as a function of digestion time (Figure 5.6a). The kinetic data was then fit to a single mono-exponential decay to determine the digestion rate constant (k_{digest}). We also note that kinetic data for positively supercoiled minicircles was very reproducible between multiple runs (Figure 5.6c).

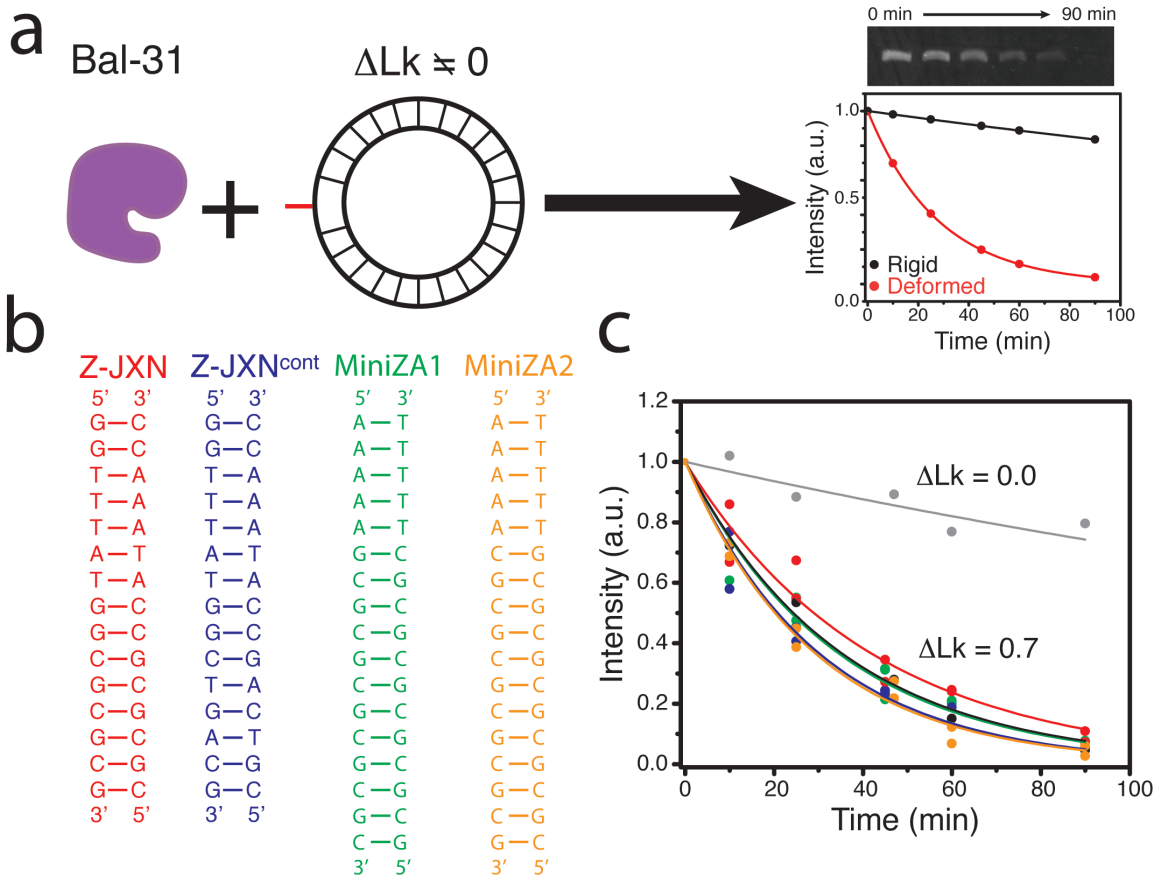

Figure 5.6: BAL-31 kinetic assay and results for positively supercoiled minicircles. a) Schematic description of BAL-31 digestion reactions. b) Sequences that were inserted into Mini87. c) Kinetic profiles for Mini84(grey), Mini87(black), MiniZJXN(red), MiniZJXN^{cont}(blue), MiniZA1 (green), MiniZA2 (orange). The best fit mono-exponential decay is plotted for each corresponding data set.

Kinetic experiments were carried out on five 87 base-pair positively supercoiled minicircles with varying propensities to form Z-DNA, three of which: Mini87, MiniZJXN, MiniZJXN^{cont} were also used for the topoisomer experiments. The two additional minicircles that were studied involved the insertion of $(GC)_6$ or $(CG)_6$ elements adjacent to an A-tract, MiniZA1and MiniZA2, respectively (Table 5.1). The highlighted elements of MiniZA1 and MiniZA2 in Figure 5.6b have higher propensities for Z-DNA formation than MiniJXN because of their longer CG/GC repeats: six compared to three for Z-JXN[22]. Overall, we observed similar rates of digestion for all five positively supercoiled minicircles with the biggest deviation being ~20% faster or slower compared to Mini87 (Table 5.2). Outside of the propensity to form Z-DNA, another source for kinetic differences could be the overall minicircle stability based on GC content (Table 5.2). In summary, while we did detect minor differences in the rates of digestion between 5 positively supercoiled minicircles, we did not observe dramatic sequence dependent effects.

Minicircle	k_{digest} (min^{-1})	ΔLk	% GC
Mini84	0.003 ± 0.001	0.0	50.6
Mini87	0.029 ± 0.001	0.7	50.6
MiniZJXN	0.024 ± 0.002	0.7	52.9
MiniZJXN^{cont}	0.033 ± 0.002	0.7	50.6
MiniZA1	0.031 ± 0.001	0.7	55.2
MiniZA2	0.034 ± 0.001	0.7	55.2

Table 5.2: Rate constants measured for minicircle digestion by BAL-31 nuclease.

As a control, we also carried out kinetic experiments on a minicircle with minimal superhelical stress, Mini84 with a ΔLk of ~0 (Figure 5.6c, grey). Here we observed a significantly decreased rate of digestion compared to positively supercoiled minicircles as would be expected since the destabilization of the double helix from supercoiling is nearly zero for a ΔLk of 0 (Table 5.2).

Finally, we tried to carry out BAL-31 kinetic experiments on negatively supercoiled minicircles. As predicted, the rate of digestion was significantly faster

for negatively supercoiled minicircles compared to their positively supercoiled counterparts. In order to slow down the reactions to a rate measureable by our method, we reduced the temperature to from 37 °C to 25 °C and the concentration of BAL-31 nuclease from 0.06 U/μL to 0.005 U/μL. Multiple kinetic reactions were carried out on MiniZJXN^{Cont} and we found the kinetic traces to be both quite strange and irreproducible. As shown in Figure 5.7, repeated experiments with the same minicircle resulted in a large spread in the data and for some cases a mono exponential decay was not observed. This is in contrast to the kinetic data for positively supercoiled minicircles where we were able to readily reproduce each kinetic trace. We did not run any further kinetic experiments on negatively supercoiled minicircles because of the difficulty in generating reliable data.

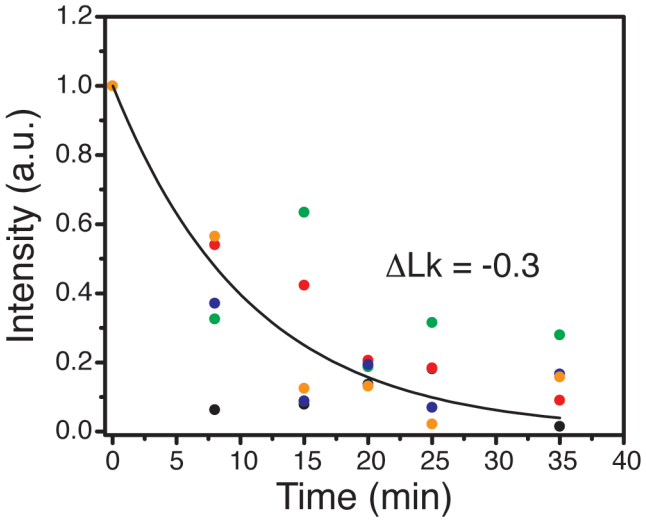

Figure 5.7: BAL-31 kinetic data for a negatively supercoiled MiniJXN^{cont} minicircle (ΔLk = -0.3). Each color (black, red, green, blue, orange) corresponds to a different repetition of the same experiment.

5.4 Conclusions

We have established a procedure for producing topologically pure DNA minicircles in large quantities necessary for biophysical characterization. Our results from minicircle synthesis show that simple assumptions about product formation based on forming the most stable topoisomer upon ligation cannot be

made. For both minicircle sizes studied, 84 and 87 base-pairs, we observed a mixture of topoisomers that were skewed to the opposite of what would be predicted based on simple ΔLk calculations. Therefore, it is necessary to carefully identify each topoisomer that forms from the ligation process. First, it is crucial to run minicircle gels for an ample amount of time to allow for the topoisomers to fully separate. Additionally, one topoisomer might form in a very small amount; therefore it is necessary to use sensitive methods for gel analysis. For example, when 87 base-paired minicircle gels were stained with ethidium bromide, only the highly populated positive topoisomer was observed. Subsequent staining with SYBR gold revealed the presence of a small amount of the negative topoisomer. Once the desired topoisomer is identified, minicircle synthesis can be tweaked for its specific selection. Formation of purely the negative topoisomer can be achieved by driving the distribution of ligation products to the unwound form using an intercalating agent. If the positive topoisomer is desired, BAL-31 nuclease digestion can be used to remove any negative topoisomers that are present after gel purification since they digest at a rate at least an order of magnitude faster than positive topoisomers.

Similar to Chapters 3 and 4 we have explored the influence of sequence-dependent mutations, except on supercoiled DNA. First, we studied the sequence-dependence of topoisomer formation from ligation reactions. Insertion of sequences with differing propensities to form Z-DNA did not significantly alter the product distributions of ligation reactions. Additionally, we made a TAT to TGT mutation shown to significantly rigidify the double helix of linear DNA in Chapter 3, and again did not detect a significant change the ligation topoisomer distribution.

We also implemented a BAL-31 nuclease kinetic assay to probe deformations in DNA minicircles composed of different sequences. Generally, the kinetic assay worked as predicted where the rates of BAL-31 nuclease digestion varied based on the minicircle's ΔLk with $\Delta Lk = 0.0$ being the slowest, $\Delta Lk > 0$ slightly faster, and $\Delta Lk < 0$ being very fast. We were able to successfully determine the rates of digestion for positively supercoiled minicircles with differing GC contents and propensities to form Z-DNA but did not observe any dramatic differences in their digestion rates. Our minicircle systems were designed for Z-DNA forming sequences; therefore, we

were mostly interested in digestion rates of the negatively supercoiled minicircles. However, the kinetics experiments for negatively supercoiled minicircles resulted in inconsistent and unreliable data.

Both the topoisomer and kinetics experiments were unable to detect dramatic sequence-specific differences in DNA minicircles. This does not necessarily suggest that supercoiling does not significantly amplify the sequence-specific structural dynamic properties observed in relaxed duplex DNA. Rather, the sequence-specific effects we attempted to probe with these methods might be too small to be revealed within a sea of nearly 90 nucleotides. The inability to determine sequence effects at the base level further motivates the necessity to produce quantities of minicircles amenable to characterization by higher resolution methods such as NMR spectroscopy.

5.5 References

1. A. D. Bates, A. Maxwell, *DNA topology*. (Oxford University Press, Oxford ; New York, ed. 2nd, 2005), pp. xviii, 198 p.
2. F. Kouzine, D. Levens, Supercoil-driven DNA structures regulate genetic transactions. *Front Biosci* **12**, 4409 (2007).
3. A. V. Vologodski*i, *Topology and physics of circular DNA*. (CRC Press, Boca Raton, 1992), pp. 179 p.
4. C. J. Dorman, C. P. Corcoran, Bacterial DNA topology and infectious disease. *Nucleic Acids Res* **37**, 672 (2009).
5. F. Kouzine, S. Sanford, Z. Elisha-Feil, D. Levens, The functional response of upstream DNA to dynamic supercoiling in vivo. *Nat Struct Mol Biol* **15**, 146 (2008).
6. C. Lavelle, DNA torsional stress propagates through chromatin fiber and participates in transcriptional regulation. *Nat Struct Mol Biol* **15**, 123 (2008).
7. Q. Du, A. Kotlyar, A. Vologodskii, Kinking the double helix by bending deformation. *Nucleic Acids Res* **36**, 1120 (2008).
8. D. Demurtas *et al.*, Bending modes of DNA directly addressed by cryo-electron microscopy of DNA minicircles. *Nucleic Acids Res* **37**, 2882 (2009).
9. T. A. Lionberger *et al.*, Cooperative kinking at distant sites in mechanically stressed DNA. *Nucleic Acids Res* **39**, 9820 (2011).
10. F. Lankas, R. Lavery, J. H. Maddocks, Kinking occurs during molecular dynamics simulations of small DNA minicircles. *Structure* **14**, 1527 (2006).
11. J. S. Mitchell, C. A. Laughton, S. A. Harris, Atomistic simulations reveal bubbles, kinks and wrinkles in supercoiled DNA. *Nucleic Acids Res* **39**, 3928 (2011).

12. M. Le Bret, Twist and writhing in short circular DNAs according to first-order elasticity. *Biopolymers* **23**, 1835 (1984).
13. D. S. Horowitz, J. C. Wang, Torsional rigidity of DNA and length dependence of the free energy of DNA supercoiling. *J Mol Biol* **173**, 75 (1984).
14. D. Shore, R. L. Baldwin, Energetics of DNA twisting. II. Topoisomer analysis. *J Mol Biol* **170**, 983 (1983).
15. C. K. Singleton, J. Klysik, S. M. Stirdivant, R. D. Wells, Left-handed Z-DNA is induced by supercoiling in physiological ionic conditions. *Nature* **299**, 312 (1982).
16. L. J. Peck, A. Nordheim, A. Rich, J. C. Wang, Flipping of cloned d(pCpG)n.d(pCpG)n DNA sequences from right- to left-handed helical structure by salt, Co(III), or negative supercoiling. *Proc Natl Acad Sci U S A* **79**, 4560 (1982).
17. J. R. Bothe, K. Lowenhaupt, H. M. Al-Hashimi, Sequence-specific B-DNA flexibility modulates Z-DNA formation. *J Am Chem Soc* **133**, 2016 (2011).
18. H. B. Gray, Jr., D. A. Ostrander, J. L. Hodnett, R. J. Legerski, D. L. Robberson, Extracellular nucleases of Pseudomonas BAL 31. I. Characterization of single strand-specific deoxyriboendonuclease and double-strand deoxyriboexonuclease activities. *Nucleic Acids Res* **2**, 1459 (1975).
19. P. P. Lau, H. B. Gray, Jr., Extracellular nucleases of Alteromonas espejiana BAL 31.IV. The single strand-specific deoxyriboendonuclease activity as a probe for regions of altered secondary structure in negatively and positively supercoiled closed circular DNA. *Nucleic Acids Res* **6**, 331 (1979).
20. M. W. Kilpatrick, C. F. Wei, H. B. Gray, Jr., R. D. Wells, BAL 31 nuclease as a probe in concentrated salt for the B-Z DNA junction. *Nucleic Acids Res* **11**, 3811 (1983).
21. X. Zheng, A. Vologodskii, Theoretical analysis of disruptions in DNA minicircles. *Biophys J* **96**, 1341 (2009).
22. Z. Reich, P. Friedman, S. Levin-Zaidman, A. Minsky, Effects of adenine tracts on the B-Z transition. Fine tuning of DNA conformational transition processes. *J Biol Chem* **268**, 8261 (1993).

Chapter 6
Conclusions and Future Directions

6.1 Conclusions and Future Directions

Great strides have been made in characterizing the structure and energetics of DNA from the proposed structure of B-DNA based on X-ray diffraction of DNA fibers[1], to the first single crystal structure of DNA[2], to the crystal structures of huge nucleosome complexes[3], to the development of NMR experiments that can probe the structure of lowly populated conformational states[4,5]. However, with each deeper insight made into the minutia of the dynamic structure landscape of biomolecules comes the difficulty in determining their specific biological function. This is particularly challenging because in many cases very controlled conditions are necessary for their characterization.

NMR relaxation dispersion methods allow for an unprecedented insight into conformations that form as little as one percent of the time. The development of advanced relaxation dispersion methods has allowed for extremely accurate structural, thermodynamic, and kinetic information to be obtained about fleeting excited conformational states. In Chapter 2, we explored the limits of which systems can be quantitatively characterized using $R_{1\rho}$ relaxation dispersion. In our analysis we found that systems in slow to moderately fast exchange ($k_{ex}/\Delta\omega < \sim 2$) could always be fully characterized to obtain accurate and precise excited state chemical shifts, populations, and kinetics. In cases of faster exchange ($k_{ex}/\Delta\omega > 2$), it becomes more difficult to precisely determine the exchange parameters, but useful information can still be obtained. One possible avenue for the improvement of parameters obtained in fast exchange scenarios could be to combine $R_{1\rho}$ dispersion with additional NMR measurements. For example, Vallurupalli *et al.* recently demonstrated that differences in peak positions from HSQC and HMQC experiments

(ω_{SQ}-ω_{MC}) could be combined with CPMG data to obtain accurate exchange parameters in fast exchange systems that are beyond the limit of CPMG by itself[6].

Beyond obtaining extremely accurate and precise exchange parameters from dispersion data is the challenge in determining the biological or chemical significance, if any, of these excited states. Relaxation dispersion experiments carried out on enzymes have shown an interesting correlation between the rates of intrinsic conformational dynamics and catalytic rates. It has been proposed that these motions measured by NMR are important for the catalytic process, but this is currently a very controversial topic in the literature[7-10]. Arguably it might be even more difficult to fully comprehend the significance of the intrinsic dynamical properties of nucleic acids. This can be highlighted by considering the influence that ions have on the dynamic structure landscape of nucleic acids. The structure and dynamical properties of both DNA and RNA are extremely sensitive to ionic conditions and the presence of specific mono-to-multivalent ions[11]. Absence of key stabilizing or destabilizing ions such as Mg^{2+} can completely alter the conformational dynamics of the target nucleic acid[12,13]. For example, excited states may be observed in the presence monovalent ions but not when Mg^{2+} is present. Does this shift in conformational equilibrium mean that the excited state conformation populated by monovalent ions is not biologically relevant because it is not sufficiently populated in more a biological context where Mg^{2+} is present? Simple questions like this remain to be answered, but at minimum there is great need for a systematic study to pin down how environmental conditions alter the dynamic structure landscape of nucleic acids.

In Chapter 3 we used relaxation dispersion to identify sites with a high propensity to form non-canonical conformations, when in B-DNA, for the only DNA that has been crystallized with a B/Z junction[14]. Mutations that altered the local stability and dynamics at the flexible B/Z junction forming sites were shown to modulate the B-to-Z transition. In Chapter 4 we further probed this B/Z junction using a combined CD and fluorescence approach and showed that favorable junction sequences can drive the B-to-Z transition allowing for the incorporation of multiple

thermodynamically unfavorable steps into Z-DNA helices. These studies highlight the complex interplay between the formation of B/Z junctions and incorporation of dinucleotides into Z-DNA helices, where the most thermodynamically favorable combination wins.

In Chapter 4 we observed varying sequence-dependent extents of Z-DNA formation and interpreted the data assuming a two-state model. However, we cannot rule out that an ensemble of conformations composed of significantly different extents of Z-DNA exists in solution. Application of advanced multidimensional NMR methods could help to clear up the inherent uncertainties that exist when using bulk CD and fluorescence methods to characterize the formation of B/Z junctions and the B-to-Z transition. NMR has in fact recently been employed to probe the B-to-Z transition[15-18]. These studies utilized 1D NMR experiments to study the transition, but are limited to observing only imino protons that are involved in base-pairing. Future studies should apply more advanced multidimensional methods that allow for the observation of multiple base (C8/C6) and sugar probes (C1′) to better characterize the transition.

Further, NMR studies could provide an atomic resolution insight into the structural and dynamic properties of Z-DNA and B/Z junctions. Most previous high-resolution NMR studies of Z-DNA, in complex with the Zα domain, have focused purely on the Zα protein[19,20]. The DNA perspective could be characterized using isotopically labeled DNA. However, this will be difficult due to the challenge in biochemically synthesizing short duplexes with CG repeats. NMR experiments on a DNA/Zα domain will also be challenging due to the size of the complex (~50 kDa for Z-JXN), possibly requiring the use of methods that take advantage of the favorable relaxation properties of methyl thymines. A full understanding of the formation of these complicated mixtures of B and Z-DNA necessitates the utilization of a wide variety of biophysical methods for their full energetic and structural characterization.

Finally, studies must be carried out on DNA in the most biologically relevant context in order to fully grasp the biological relevance of any observed DNA

structural or dynamical property. In the cell, DNA is bombarded with a complicated network of interactions including a complex mix of ions, proteins, and supercoiling forces. In Chapter 5 we have initiated studies on supercoiled DNA using small minicircles to probe the effects of supercoiling on sequence-dependent DNA properties. Using a combination of biochemical experiments we were unable to detect dramatic sequence-specific differences in DNA minicircles. The inability to detect sequence-specific characteristics using our methods calls for higher resolution studies such as NMR to be carried out on this system.

While we were able to produce significant quantities of DNA minicircles, which would be necessary for NMR, the overall yields were very low 10-15%. At this point it is still unfeasible to produce the quantities of minicircles necessary for study by NMR using conventional methods. One possible avenue for NMR studies with limited material would be to carry out highly sensitive methyl TROSY experiments that have successfully been applied to other systems with high molecular weights (100's of kDa)[21]. For example, a recent report by Kay and coworkers used protein methyl NMR experiments to study full nucleosome complexes[16]. Another important consideration for NMR of DNA minicircles is that uniform isotopic labeling will result in significant spectral overlap. It is possible to generate selectively labeled DNA minicircles, but this adds additional steps to an already lowly yielding process. If these significant biochemical and spectroscopic hurdles can be overcome, we will have a more complete atomic level picture of the already extremely complicated DNA double helix.

6.2 References

1. J. D. Watson, F. H. Crick, Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**, 737 (1953).
2. A. H. Wang *et al.*, Molecular structure of a left-handed double helical DNA fragment at atomic resolution. *Nature* **282**, 680 (1979).
3. K. Luger, A. W. Mader, R. K. Richmond, D. F. Sargent, T. J. Richmond, Crystal structure of the nucleosome core particle at 2.8 angstrom resolution. *Nature* **389**, 251 (1997).
4. D. F. Hansen, P. Vallurupalli, L. E. Kay, Using relaxation dispersion NMR spectroscopy to determine structures of excited, invisible protein states. *J Biomol Nmr* **41**, 113 (2008).

5. J. R. Bothe *et al.*, Characterizing RNA dynamics at atomic resolution using solution-state NMR spectroscopy. *Nat Methods* **8**, 919 (2011).
6. P. Vallurupalli, G. Bouvignies, L. E. Kay, Increasing the Exchange Time-Scale That Can Be Probed by CPMG Relaxation Dispersion NMR. *J Phys Chem B* **115**, 14891 (2011).
7. S. C. Kamerlin, A. Warshel, At the dawn of the 21st century: Is dynamics the missing link for understanding enzyme catalysis? *Proteins* **78**, 1339 (2010).
8. J. Villali, D. Kern, Choreographing an enzyme's dance. *Curr Opin Chem Biol* **14**, 636 (2010).
9. J. D. McGeagh, K. E. Ranaghan, A. J. Mulholland, Protein dynamics and enzyme catalysis: insights from simulations. *Biochim Biophys Acta* **1814**, 1077 (2011).
10. M. Karplus, Role of conformation transitions in adenylate kinase. *P Natl Acad Sci USA* **107**, E71; author reply E72 (2010).
11. D. E. Draper, D. Grilley, A. M. Soto, Ions and RNA folding. *Annu Rev Biophys Biomol Struct* **34**, 221 (2005).
12. A. Casiano-Negroni, X. Sun, H. M. Al-Hashimi, Probing Na(+)-induced changes in the HIV-1 TAR conformational dynamics using NMR residual dipolar couplings: new insights into the role of counterions and electrostatic interactions in adaptive recognition. *Biochemistry* **46**, 6525 (2007).
13. N. V. Eldho, K. T. Dayie, Internal bulge and tetraloop of the catalytic domain 5 of a group II intron ribozyme are flexible: implications for catalysis. *J Mol Biol* **365**, 930 (2007).
14. S. C. Ha, K. Lowenhaupt, A. Rich, Y. G. Kim, K. K. Kim, Crystal structure of a junction between B-DNA and Z-DNA reveals two extruded bases. *Nature* **437**, 1183 (2005).
15. Y. M. Kang *et al.*, NMR spectroscopic elucidation of the B-Z transition of a DNA double helix induced by the Z alpha domain of human ADAR1. *J Am Chem Soc* **131**, 11485 (2009).
16. H. Kato *et al.*, Architecture of the high mobility group nucleosomal protein 2-nucleosome complex as revealed by methyl-based NMR. *P Natl Acad Sci USA* **108**, 12283 (2011).
17. Y. M. Lee *et al.*, NMR Study on the B-Z Junction Formation of DNA Duplexes Induced by Z-DNA Binding Domain of Human ADAR1. *J Am Chem Soc* **134**, 5276 (2012).
18. Y. J. Seo *et al.*, Sequence discrimination of the Zalpha domain of human ADAR1 during B-Z transition of DNA duplexes. *FEBS Lett* **584**, 4344 (2010).
19. M. Schade *et al.*, The solution structure of the Zalpha domain of the human RNA editing enzyme ADAR1 reveals a prepositioned binding surface for Z-DNA. *P Natl Acad Sci USA* **96**, 12465 (1999).
20. J. D. Kahmann *et al.*, The solution structure of the N-terminal domain of E3L shows a tyrosine conformation that may explain its reduced affinity to Z-DNA in vitro. *P Natl Acad Sci USA* **101**, 2712 (2004).
21. D. M. Korzhnev, K. Kloiber, V. Kanelis, V. Tugarinov, L. E. Kay, Probing slow dynamics in high molecular weight proteins by methyl-TROSY NMR

spectroscopy: application to a 723-residue enzyme. *J Am Chem Soc* **126**, 3964 (2004).

Appendix 1

Noise Corruption

(*

This was originally written for Mathematica 8.0.

All parameters that can be modified are in bold.

Setup the inputfile to have 3 columns: 1. offset, 2. r1rho, 3. spinlock power 4. field strength 5. error in R1rho.

*)

```
inputdata = Import["~/Desktop/profile.txt", "Table"];  (*Data to be corrupted*)
size = Length[inputdata];
noisecorrupt = {};
For[i = 1, i < size + 1, i++,
error = 0.05; (*Size of error corruption*)
corupt =RandomReal[NormalDistribution[inputdata[[i, 2]], error*inputdata[[i, 2]]]];
AppendTo[noisecorrupt,{inputdata[[i, 1]], corupt, inputdata[[i, 3]], inputdata[[i, 4]],
inputdata[[i, 5]]}]; ];


Export["~/Desktop/profile-corrupt.txt", noisecorrupt, "Table"]; (*Export noise
corrupted data*)
```

Appendix 2

Grid Search

(*

This was originally written for Mathemaitca 8.0 and takes advantage of its parallel computing packages.

All parameters that can be modified are in bold.

Also, note that Δω is defined here as $\Delta\omega = \omega_G - \omega_E$!

Setup the input file to have 3 columns: 1. offset, 2. r1rho, 3. spinlock power 4. field strength 5. error in R1rho.

*)

```
Expdata = Import["~/Desktop/profile-lag-corrupt.txt", "Table"];  (*Import
      data*)
R2 = 11;
R1 = 1.5;
deltainitial = -3;          (*Minimum chemical shift to search*)
deltafinal = 3;             (*Maximum chemical shift to search*)
deltainc = 0.1;             (*Chemical shift increment*)
painit = 0.5;               (*Minimum G state population to search*)
popinc = 0.005;             (*Population increment*)
kinitial = 8000;            (*Minimum kex to search*)
kfinal = 24000;             (*Maximum kex to search*)
kinc = 1000;                (*Increment for kex*)


numomega = (deltafinal - deltainitial)/deltainc + 1;
numpop = (1 - painit)/popinc;
numk = (kfinal - kinitial)/kinc + 1;
"YOUR GRID SIZE IS:"
gridsize = numomega*numpop*numk
gridparams = Table[{x, y, z}, {x, deltainitial, deltafinal, deltainc},
   {y, popinc, 1 - painit, popinc}, {z, kinitial, kfinal, kinc}];
gridparams = Flatten[gridparams, 2];
offset = Transpose[Expdata][[1]];
expr1rho = Transpose[Expdata][[2]];
spinlock = Transpose[Expdata][[3]];
field = Transpose[Expdata][[4]];
error = Transpose[Expdata][[5]];
params = Transpose[{offset, spinlock, field}];
```

```
expdata = Transpose[{expr1rho, (1/error)^2}];
 Model = Compile[{offset, w1, B0, Δw, pb, kex}, Δwrad = Δw*B0*2*Pi;
   pa = 1 - pb; Ωa = pb*Δw*B0; Ωb = Ωa - Δw*B0;
   weff = (w1*2*Pi)^2 + (offset*2*Pi)^2;
   waeff = (w1*2*Pi)^2 + ((Ωa + offset)*2*Pi)^2;
   wbeff = (w1*2*Pi)^2 + ((Ωb + offset)*2*Pi)^2; If[offset == 0, theta = Pi/2,
    theta = ArcTan[w1/offset]]; Null; RexLag = pa*pb*Δwrad^2*
     (kex/(waeff*(wbeff/weff) + kex^2 - Sin[theta]^2*pa*pb*Δwrad^2*
       (1 + 2*kex^2*((pa*waeff + pb*wbeff)/(waeff*wbeff + weff*kex^2)))));
   r1rho = R1*Cos[theta]^2 + (R2 + RexLag)*Sin[theta]^2, CompilationTarget ->
"C"];

datalength = Length[expdata];
dof = datalength - 4; input = {};
r1rholist = {};
chivals = {};
chivals = ParallelTable[r1rholist = {}; input = {};
   Do[AppendTo[input, Join[params[[j]], gridparams[[i]]]]; , {j, datalength}];
   r1rholist = (Model @@ #1 & ) /@ input;
   chi = Total[(r1rholist -
Transpose[expdata][[1]])^2*Transpose[expdata][[2]]]/dof,
  {i, gridsize}];
grid = Transpose[Insert[Transpose[gridparams], chivals, 4]];
```

Export["~/Desktop/chigrid.txt", grid, "Table"] (*Export grid to text file*)

Appendix 3

Bootstrap Fitting

(*

This was originally written for Mathemaitca 8.0 and takes advantage of its parallel computing packages.

All parameters that can be modified are in bold.

Also, note that Δω is defined here as $\Delta\omega = \omega_G - \omega_E$!

Setup the input file to have 3 columns: 1. offset, 2. r1rho, 3. spinlock power 4. field strength 5. error in R1rho.

*)

```
Expdata = Import["~/profilekex8000pb30-lag-R2-20-corrupt.txt", "Table"];
(*Import data*)

Off[CompiledFunction::cfsa];
ParallelEvaluate[Off[CompiledFunction::cfsa]];
datalength = Length[Expdata];
dof = datalength - 6; r1rholist = {};
chivals = {};
offset = Transpose[Expdata][[1]];
expr1rho = Transpose[Expdata][[2]];
spinlock = Transpose[Expdata][[3]];
field = Transpose[Expdata][[4]];
error = Transpose[Expdata][[5]];
params = Transpose[{offset, spinlock, field}];
expdata = Transpose[{expr1rho, (1/error)^2}];
Model = Compile[{offset, w1, B0, Δw, pb, kex, R1, R2},
   (Δwrad = Δw*B0*2*Pi; pa = 1 - pb; Ωa = pb*Δw*B0; Ωb = Ωa - Δw*B0;
    weff = (w1*2*Pi)^2 + (offset*2*Pi)^2;
    waeff = (w1*2*Pi)^2 + ((Ωa + offset)*2*Pi)^2;
    wbeff = (w1*2*Pi)^2 + ((Ωb + offset)*2*Pi)^2; If[offset == 0, theta = Pi/2,
     theta = ArcTan[w1/offset]]; ); RexLag = pa*pb*Δwrad^2*
     (kex/(waeff*(wbeff/weff) + kex^2 - Sin[theta]^2*pa*pb*Δwrad^2*
       (1 + 2*kex^2*((pa*waeff + pb*wbeff)/(waeff*wbeff + weff*kex^2)))));
   r1rho = R1*Cos[theta]^2 + (R2 + RexLag)*Sin[theta]^2, CompilationTarget ->
"C"];
ChiSqr[Δw_, pb_, kex_, R1_, R2_, params_, expdata_] :=
 (input = {}; gridparams = {Δw, pb, kex, R1, R2};
  Do[AppendTo[input, Join[params[[j]], gridparams]]; , {j, datalength}];
```

```
  r1rholist = (Model @@ #1 & ) /@ input;
  Total[(r1rholist - Transpose[expdata][[1]])^2*Transpose[expdata][[2]]]/dof)

R2min = 10;          (*Minimum allowed R2*)
R2max = 30;          (*Maximum allowed R2*)
R1min = 0.5;         (*Minimum allowed R1*)
R1max = 4;           (*Maximum allowed R1*)
kexmin = 100;        (*Minimum allowed kex*)
kexmax = 50000;      (*Maximum allowed kex*)
pbmin = 0.001;       (*Minimum allowed population*)
pbmax = 0.999;       (*Maximum allowed population*)
Δwmin = -7;          (*Minimum allowed chemical shift difference*)
Δwmax = 7;           (*Maximum allowed chemical shift difference*)

Resample[list_] := list[[Table[RandomInteger[{1, Length[list]}], {Length[list]}]]];
bootcount = 1000; bootlist = Table[Resample[Expdata], {bootcount}];
dof = datalength - 6; fitlist = {}; chilist = {}; sortthis = {};
fitlist = ParallelTable[Clear[Δw]; Clear[pb]; Clear[kex]; Clear[R1]; Clear[R2];
   Expdata = bootlist[[h]]; offset = Transpose[Expdata][[1]];
   expr1rho = Transpose[Expdata][[2]]; spinlock = Transpose[Expdata][[3]];
   field = Transpose[Expdata][[4]]; error = Transpose[Expdata][[5]];
   params = Transpose[{offset, spinlock, field}];
   expdata = Transpose[{expr1rho, (1/error)^2}];
   Fitthisstuff = Minimize[{ChiSqr[Δw, pb, kex, R1, R2, params, expdata],
     Δwmin <= Δw <= Δwmax && pbmin <= pb <= pbmax &&
      kexmin <= kex <= kexmax && R1min <= R1 <= R1max && R2min <= R2 <=
R2max},
    {Δw, pb, kex, R1, R2}, WorkingPrecision -> 15];
   Join[{Δw, pb, kex, R1, R2} /. Fitthisstuff[[2]], {Fitthisstuff[[1]]}],
  {h, bootcount}];

Export["~/bootfitkex8000pb30-R2-20.txt", fitlist, "Table"]; (*Export bootstrap
fits to a text file*)
```