

Sequence Specificity of Transient Hoogsteen Base-Pairs in Canonical  
Duplex DNA and Z-DNA Formation

by  
Heidi Suzanne Alvey

A dissertation submitted in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy  
(Chemistry)  
in The University of Michigan  
2014

Doctoral Committee

Professor Hashim M. Al-Hashimi, Chair  
Professor Carol Fierke  
Professor Patrick O'Brien  
Professor Ayyalusamy Ramamoorthy

© Heidi Suzanne Alvey  
2014

---

## **Dedication**

*To my mother and father for teaching me to do what is right over what is easy.*

## Acknowledgements

Thank you to my advisor, Professor Hashim M. Al-Hashimi for his contagious tenacity and perseverance for discovering the truth in nature. Thank you to my past lab mates for setting the bar high and helping me in so many ways over the years, especially Dr. Catherine Eichhorn, Dr. Annette Casiano-Negroni, Dr. Evgenia Nikolova, Dr. Katja Petzold, Dr. Elizabeth Dethoff, Dr. Jeetender Chugh, Dr. Loïc Salmon, Dr. Jameson Bothe, Dr. Joseph Yesselman, Dr. Scott Horowitz and Federico Gottardo. Thank you to my past and current lab mates for your relentless appetite for robust science and brutal honesty when it comes to comments and critiques of my work, especially Dr. Bharathwaj Sathyamoorthy, Dr. Yi Xue, Tony Mustoe, Janghyun (Kevin) Lee, Shan Yang, Huiqing (Jane) Zhou, Isaac Kimsey, Anisha Shakya, Zachary Stein and Yu Chen. The Al-Hashimi Lab is an exceptional environment where great science meets ferocity, encouragement, support and friendship, so thank you to everyone above for the latter as well.

Thank you sincerely to our collaborator, Dr. Ky Lowenhaupt, (formerly) of Professor Alexander Rich's lab at MIT. Her passion for Z-DNA projects has been a saving grace for my thesis work and I am extremely thankful for her willingness and talent to supply us with copious amounts of protein, even when it must be in her free time as of late. Thank you also to Professor Gafni for giving me *carte blanche* to the Gafni Lab CD instrument. Also, thank you to Dr. James Windak for his incredible knowledge

and attentiveness regarding the Chemistry Department CD instrument. The kinetic measurements presented in this thesis would not have been possible without his help.

Graduate school would have been insurmountable without the enormous amount of mentorship and support I've had along the way. Thank you to my undergraduate advisor, Professor David Rueda, for providing an environment where I became obsessed with science, as well as my former Rueda Lab members Professor Elvin Alemán, Dr. Rajan Lamichhane and Dr. May Daher for your continued mentorship, encouragement and friendship. Thank you to my very special friends Professor Arlie Rinaldi, Dr. Mario Blanco, Matt Marek, Matt Kahlscheuer and the whole crew for such fun and supportive times. I am forever indebted to my parents for their unconditional love and support. I would not have made it through college, graduate school, or life without them and I am so thankful to have them in my life in so many ways. Thank you to my brother for his deep friendship and inspiration for the 'spiraling out' that is scientific research. Thank you to my extended family too (yes, all of you) for your love, support and fun times.

## Table of Contents

<b>Dedication .....</b>	<b>ii</b>
<b>Acknowledgements .....</b>	<b>iii</b>
<b>List of Figures.....</b>	<b>x</b>
<b>List of Tables .....</b>	<b>xiv</b>
<b>List of Appendices.....</b>	<b>xv</b>
<b>Abstract.....</b>	<b>xvi</b>
<b>Chapter 1   Introduction.....</b>	<b>1</b>
<b>1.1 Milestones in Discoveries of DNA and their Surrounding Controversy.....</b>	<b>1</b>
1.1.1 The Discovery of Deoxyribonucleic Acid .....	1
1.1.2 History of Base-Pair and Helical Structures of DNA .....	1
<b>1.2 The Surprising Discovery of Z-DNA and its Surrounding Controversy .....</b>	<b>5</b>
1.2.1 A Brief History of Z-DNA Structure .....	5
1.2.2 Biological Relevance of Z-DNA .....	8
1.2.3 B-DNA to Z-DNA Transition Kinetics.....	12
1.2.4 Kinetic Rate Laws.....	14
<b>1.3 Circular Dichroism as a Tool to Measure Biomolecular Conformational</b>	
<b>Change .....</b>	<b>16</b>
<b>1.4 Nuclear Magnetic Resonance Spectroscopy of DNA .....</b>	<b>18</b>

1.4.1 Nuclear Magnetic Resonance Fundamentals .....	18
1.4.2 Chemical Exchange and $R_{1\rho}$ .....	25
<b>1.5 Statistics Used for Model Selection .....</b>	<b>30</b>
<b>1.6 References .....</b>	<b>31</b>

**Chapter 2 | Transient Hoogsteen Base-Pairs Occur Robustly in Canonical Duplex DNA Across Diverse Sequence and Positional Contexts *via* a Late Transition State**

.....	41
<b>2.1 Introduction.....</b>	<b>41</b>
<b>2.2 Methods.....</b>	<b>42</b>
2.2.1 NMR Samples and Resonance Assignments .....	42
2.2.2 NMR $R_{1\rho}$ Relaxation Dispersion.....	44
2.2.3 DNA Melting Monitored <i>via</i> Circular Dichroism .....	46
2.2.4 Phi ( $\Phi$ )-Value Analysis.....	47
<b>2.3 Results and Discussion.....</b>	<b>48</b>
<b>2.4 Conclusions .....</b>	<b>63</b>
<b>2.5 References .....</b>	<b>64</b>

**Chapter 3 | Equilibrium and Transient Kinetics suggest at least Three Species in the B-to-Z-DNA Transition .....**

<b>3.1 Introduction.....</b>	<b>69</b>
<b>3.2 Methods.....</b>	<b>71</b>
3.2.1 DNA Sample Preparation .....	71

3.2.2 Circular Dichroism Thermodynamics.....	71
3.2.3 Circular Dichroism Kinetics .....	72
3.2.4 Calculating Fraction of Z-DNA using $\theta_{254nm}$ .....	73
3.2.5 Fitting of CD Wavelength Spectra to a Linear Combination of Two Species.	74
3.2.6 Determining the B-Z Transition Midpoint using the Boltzmann Equation .....	75
3.2.7 Solving for Residual CD Signal.....	75
3.2.8 Fitting B-to-Z-DNA Transient Kinetics.....	75
3.2.9 Isotopically Labeled DNA Sample Preparation.....	76
3.2.10 NMR Experiments .....	76
<b>3.3 Results and Discussion.....</b>	<b>76</b>
3.3.1 The B-to-Z-DNA Transition Involves more than Two Species.....	76
3.3.2 Residual Signal Demonstrates CG <sub>3</sub> can Adopt at least One Non-B/Z Structure .....	82
3.3.3 Biphasic Nature of the B-Z Transition.....	84
3.3.4 Preliminary Insights into Potential B-to-Z-DNA Transition Kinetic Models .	87
3.3.5 Observation of Additional Species at Intermediate Protein:DNA Ratios.....	97
<b>3.4 Conclusions .....</b>	<b>99</b>
<b>3.5 References .....</b>	<b>101</b>
<b>Chapter 4   Kinetic Analysis of the B-DNA to Z-DNA Transition in Short DNA Sequences; Examining the Consequences of DNA Methylation.....</b>	<b>106</b>
<b>4.1 Introduction.....</b>	<b>106</b>
4.2.1 DNA Sample Preparation .....	107



<b>4.2 Methods.....</b>	<b>107</b>
4.2.2 Circular Dichroism Thermodynamics and Kinetics .....	107
4.2.3 Simulations of the Observed Circular Dichroism Signal.....	107
<b>4.3 Results and Discussion.....</b>	<b>108</b>
4.3.1 Impact of 5-Methylcytosine on Transient Kinetics of the B-to-Z-DNA Transition .....	108
4.3.2 Simulation of a Kinetic Model for the NaCl-Induced B-to-Z-DNA Transition in CG <sub>3</sub> .....	112
<b>4.4 Conclusions.....</b>	<b>119</b>
<b>4.5 References.....</b>	<b>120</b>
<b>Chapter 5   The Impact of C7-Deazaguanine and N1-Methylguanine Modifications on DNA Structure and the Equilibrium Between B-DNA and Z-DNA.....</b>	<b>122</b>
<b>5.1 Introduction.....</b>	<b>122</b>
5.1.1 Importance of C7-Deazaguanine and N1-Methylguanine .....	122
5.1.2 NMR Spectroscopy of Z-DNA .....	123
<b>5.2 Experimental Methods .....</b>	<b>126</b>
5.2.1 DNA Sample Preparation .....	126
5.2.3 Circular Dichroism Equilibrium Measurements.....	126
5.2.4 NMR Experiments .....	127
<b>5.3 Results and Discussion.....</b>	<b>128</b>
5.3.1 Impact of the 7-Deazaguanine Modification on the B-to-Z-DNA Transition	128
5.3.2 Impact of the N1-Methylguanine Modification on DNA Structure.....	133

5.3.3 Preliminary NMR Spectroscopic Investigations of Z-DNA: ( <sup>5</sup> meCG) <sub>3</sub> and a (CG) <sub>3</sub> :Zα Complex.....	139
<b>5.4 Conclusions.....</b>	<b>145</b>
<b>5.5 References.....</b>	<b>146</b>
<b>Chapter 6   Conclusions and Future Perspectives .....</b>	<b>150</b>
<b>6.1 Transient Watson-Crick to Hoogsteen Transitions in Canonical Duplex DNA .....</b>	<b>150</b>
<b>6.2 The B-DNA to Z-DNA Transition and Consequences of Sequence Variations .....</b>	<b>151</b>
<b>6.3 Future Perspective .....</b>	<b>153</b>
<b>6.4 References.....</b>	<b>153</b>
<b>Appendices.....</b>	<b>155</b>

## List of Figures

Figure 1.1   Hoogsteen Base-Pairs in Canonical Duplex DNA. ....	5
Figure 1.2   Atomic Resolution Structure of B-DNA and Z-DNA. ....	7
Figure 1.3   Involvement of DNA Deformation in Cellular Processes .....	11
Figure 1.4   Crystal Structure of Z $\alpha$ Bound to CG <sub>3</sub> .....	11
Figure 1.5   Proposed Schemes for the B-to-Z-DNA Transition. ....	14
Figure 1.6   CD Spectroscopy of the B-to-Z-DNA Transition.....	17
Figure 1.7   DNA Melting as measured by Circular Dichroism .....	18
Figure 1.8   Nuclear Magnetic Resonance Fundamentals.....	21
Figure 1.9   Proton Chemical Shifts of a DNA Duplex .....	21
Figure 1.10   Target DNA Resonances for NMR Spectroscopy .....	22
Figure 1.11   Example of a <sup>1</sup> H- <sup>1</sup> H NOE DNA Assignment .....	24
Figure 1.12   Chemical Exchange .....	26
Figure 1.13   Concepts of R <sub>1<math>\rho</math></sub> Relaxation Dispersion .....	29
Figure 2.1   Probing the occurrence of transient A•T and G•C Hoogsteen Base-Pairs ....	49
Figure 2.2   Characterizing Transient Hoogsteen Base-Pairs using On- and Off-Resonance <sup>13</sup> C and <sup>15</sup> N R <sub>1<math>\rho</math></sub> Relaxation Dispersion .....	50
Figure 2.3   Sequence and Position Dependence of Population and Lifetime of Transient Hoogsteen Base-Pairs .....	52
Figure 2.4   Sequence and Position Dependence of Thermodynamic and Kinetic Parameters Describing the Watson-Crick to Hoogsteen Transition .....	55

Figure 2.5   Circular Dichroism DNA Duplex Melting Experiments .....	56
Figure 2.6   Correlation of Free Energy of Stabilization and Forward Free Energy Barrier for the Watson-Crick-to-Hoogsteen Transition .....	59
Figure 2.7   Correlation between the Free Energy Difference and Forward Free Energy Barrier for Base-Pair Opening as Measured by Imino Proton Exchange in the Highlighted Data taken from Russu and Coworkers in Reference 56. ....	60
Figure 2.8   $\Phi$ -Value Analysis .....	61
Figure 3.1   Monitoring of the B-Z transition in CG <sub>3</sub> DNA.....	77
Figure 3.2   Linear Combination of B- and Z-DNA Spectra for the Salt-Induced Transition .....	78
Figure 3.3   Linear Combination of B- and Z-DNA Spectra for the Z $\alpha$ -induced Transition .....	79
Figure 3.4   Fraction of B- and Z-DNA Species .....	81
Figure 3.5   B-Z Transition Midpoints .....	82
Figure 3.6   Residual Circular Dichroism Signal of the B-to-Z-DNA Transition of CG <sub>3</sub>	83
Figure 3.7   B-Z Transient Kinetics of CG <sub>3</sub> using 3.0 M NaCl.....	85
Figure 3.8   Comparison of Monophasic Fit with and without Weights.....	87
Figure 3.9   NaCl-Induced CG <sub>3</sub> Transient Kinetics .....	88
Figure 3.10   CG <sub>3</sub> Residual Signal for Fits of Transient Kinetics of the NaCl-Induced B- to-Z-DNA Transition .....	90
Figure 3.11   Observed Rate Constants.....	91
Figure 3.12   Transient Kinetics, Controls and Residuals of the NaCl-Induced B-to-Z- DNA Transition in CG <sub>6</sub> .....	92

Figure 3.13   Transient Kinetics, Controls and Residuals of the $Z\alpha$ -Induced B-to-Z-DNA Transition in $CG_3$ .....	93
Figure 3.14   NaCl-Induced Transient Kinetics of $CG_3$ with Two-Fold Concentration Increase .....	94
Figure 3.15   NaCl-Induced Transient Kinetics of $CG_3$ with Two-Fold Concentration Decrease .....	96
Figure 3.16   Observed Rate Constants of $CG_3$ Transition over a Four-Fold Range of DNA Concentrations.....	97
Figure 3.17   Titration of $Z\alpha$ into $^{13}C/^{15}N$ ZJXN at R= 2.....	99
Figure 4.1   Transient Kinetics, Controls and Residuals of the NaCl-Induced B-to-Z-DNA Transition in ( $^{5me}CG$ ) $_3$ Induced by 3.0 M NaCl .....	110
Figure 4.2   Transient Kinetics, Controls and Residuals of the B-to-Z-DNA Transition of ( $^{5me}CG$ ) $_3$ Induced by 4.0 M NaCl.....	111
Figure 4.3   Observed Rate Constant Perturbation Upon C5 Methylation.....	112
Figure 4.4   Simulations of the B-to-Z-DNA Transition in $CG_3$ .....	114
Figure 4.5   Microscopic Rate Constants of the B-to-Z-DNA Transition of $CG_3$ .....	115
Figure 4.6   ( $^{5me}CG$ ) $_3$ NaCl Titration and Residual Signal.....	116
Figure 4.7   Simulation of the B-to-Z-DNA Transition in ( $^{5me}CG$ ) $_3$ .....	117
Figure 5.1   Crystal Structure of the $Z\alpha$ -Stabilized B/Z Junction .....	124
Figure 5.2   Preliminary Transient Kinetics of the B-to-Z-DNA Transition in ( $CG$ ) $_3^{G47dG}$ .....	129
Figure 5.3   Thermodynamic Propensity of N7-Deazaguanine Mutations .....	131
Figure 5.4   pH and Solvent Affects on the B-Z-DNA Equilibrium in ( $C^{7d}G$ ) $_3$ .....	133

Figure 5.5   Structural Impact of N1-Methylguanine on Short DNAs.....	134
Figure 5.6   Changes in $^{13}\text{C}$ and $^1\text{H}$ NMR Chemical Shifts due to the N1-Methylguanine Modification in $\text{CG}_3$ .....	135
Figure 5.7   Changes in $^{13}\text{C}$ and $^1\text{H}$ NMR Chemical Shifts due to the N1-Methylguanine Modification in $\text{A}_6$ .....	136
Figure 5.8   pH Affects on Modified DNAs.....	138
Figure 5.9   NMR Spectroscopy of $(^5\text{meCG})_3$ in Z-Form.....	141
Figure 5.10   Titration of $\text{Z}\alpha$ into $^{13}\text{C}/^{15}\text{N}$ ZJXN .....	144
Figure A1.1   Characterizing transient Hoogsteen base-pairs using on- and off-resonance $^{13}\text{C}$ and $^{15}\text{N}$ $R_{1\rho}$ relaxation dispersion .....	156
Figure A1.2   Melting Temperature Compared to Population and Lifetime of Hoogsteen Transient State.....	161
Figure A2.1   Single-Stranded DNA CD Spectrum Benchmark.....	166
Figure A2.2   Simplest Kinetic Models Giving Rise to Two Observed Phases.....	167
Figure A4.1   Two Most Plausible Mechanistic Schemes for the B-to-Z-DNA Transition .....	169
Figure A4.2   Recognition of $(\text{CG})_3$ , $(\text{CG})_3^{\text{G}^4\text{N}^1\text{meG}}$ and $(\text{C}^{\text{N}^1\text{me}}\text{G})_3$ but not $(\text{TG})_3$ by $\text{Z}\alpha$ ..... .....	170

## List of Tables

Table 1.1   I = 1/2 Nuclei Commonly Studied by NMR Spectroscopy .....	19
Table 2.1   Chemical Exchange Parameters from fits of $R_{1\rho}$ Relaxation Dispersion Data to Eq. 2.1 .....	53
Table 2.2   Experimentally Determined Melting Temperatures .....	55
Table 2.3   $\Psi$ -Wild Type and Sequence Variants for $\Phi$ -Value Analysis .....	62
Table 4.1   Perturbation of Representative Microscopic Rate Constants upon Cytosine C5 Methylation .....	118
Table A1.1   $R_{1\rho}$ RD experimental parameters.....	155
Table A2.1   CG <sub>3</sub> + 2.5 M NaCl B-to-Z-DNA Transition Fitting Values .....	162
Table A2.2   CG <sub>3</sub> + 3.0 M NaCl B-to-Z-DNA Transition Fitting Values .....	163
Table A2.4   CG <sub>3</sub> + 4.0 M NaCl B-to-Z-DNA Transition Fitting Values .....	163
Table A2.5   CG <sub>3</sub> + 4.5 M NaCl B-to-Z-DNA Transition Fitting Values .....	163
Table A2.6   CG <sub>3</sub> + 5.0 M NaCl B-to-Z-DNA Transition Fitting Values .....	164
Table A2.7   5.5 $\mu$ M (0.5 X) CG <sub>3</sub> + 3.0 M NaCl B-to-Z-DNA Transition Fitting Values .....	164
Table A2.8   22 $\mu$ M (2 X) CG <sub>3</sub> + 3.0 M NaCl B-to-Z-DNA Transition Fitting Values .	164
Table A2.9   CG <sub>6</sub> + 4.0 M NaCl B-to-Z-DNA Transition Fitting Values .....	165
Table A2.10   CG <sub>3</sub> + 22 $\mu$ M Z $\alpha$ B-to-Z-DNA Transition Fitting Values .....	165
Table A3.1   Molar Extinction Coefficients Used for Eq. 4.1. ....	168
Table A3.2   Microscopic Rate Comparison for 3.0 M NaCl-Induced ( <sup>5me</sup> CG) <sub>3</sub> .....	168

## List of Appendices

Appendix 1   $R_{1\rho}$ Relaxation Dispersion Experimental Parameters .....	155
Appendix 2   CG <sub>3</sub> B-to-Z-DNA Controls and Transition Fits to Sum of Exponentials..	162
Appendix 3   Kinetic Simulation Parameters for ( <sup>5me</sup> CG) <sub>3</sub> .....	168
Appendix 4   Single-Stranded DNA Circular Dichroism Controls.....	169



## Abstract

Sequence-specific DNA flexibility provides an additional layer of the genetic code. Using Nuclear Magnetic Resonance  $R_{1\rho}$  relaxation dispersion we recently discovered that Watson-Crick (WC) base-pairs in flexible CA/TG and TA/TA steps of DNA transiently morph into Hoogsteen (HG) base-pairs in the context of canonical duplex DNA. CA/TG and TA/TA steps are the most flexible of all 10 dinucleotide steps. We designed DNA duplexes to contain flexible and rigid dinucleotide steps and strikingly obtained evidence for transient Hoogsteen base-pairs in all contexts examined. We observe small, significant variations in the population and lifetime of transient HG base-pairs that are anti-correlated to WC dinucleotide step stabilities. Phi-value analysis suggests a late transition state possibly involving a *syn* purine base. The broad and robust occurrence of transient HG base-pairs suggests their broad presence in the genome. In contrast to the sequence specificity of HG base-pairs, Z-DNA formation is favored in rigid and stable dinucleotide steps, specifically purine-pyrimidine repeats. Interestingly, transient, lowly populated HG base-pairs can still be observed in a Z-DNA forming sequence despite the increased dinucleotide step stability. We monitored the salt- and human adenosine deaminase- (hADAR1-Z $\alpha$ ) induced B-to-Z-DNA transitions in a CpG hexamer using Circular Dichroism (CD) to monitor global DNA conformational changes. Time-resolved CD measurements revealed a biphasic B-to-Z-DNA transition consisting of fast and slow phases for both salt- and Z $\alpha$ -induced transitions. These data are

consistent with a B-to-Z-DNA transition involving at least two sequential steps. Uniform 5-Methylcytosine modification increases the fast phase two-fold while N1-methylguanine and 7-deazaguanine modifications cause unique structural perturbations precluding B-Z transient kinetics that potentially reveal new features of HG dependent DNA flexibility and structure. Overall, the findings presented in this thesis reveal an additional widespread layer to the genetic code and call for a deeper understanding of their behavior when adjacent to one another for further insight into potential roles *in vivo*.

## **Chapter 1**

### **Introduction**

#### **1.1 Milestones in Discoveries of DNA and their Surrounding Controversy**

##### *1.1.1 The Discovery of Deoxyribonucleic Acid*

The most essential molecule of life, deoxyribonucleic acid (DNA), has been surprising us since its accidental discovery by Dr. Friedrich Miescher in 1869 who was investigating the composition of lymphocytes.<sup>1</sup> Near the end of his studies he dismissed the notion that this ‘nuclein’ holds hereditary information<sup>1</sup> and not until many decades later, introduced in 1944 by Avery and confirmed by Hershey and Chase in 1953, would we understand DNA as the carrier of genetic information containing guanine (G), cytosine (C), adenine (A) and thymine (T) nucleotides.<sup>2-5</sup> Two months prior to the final heralded ‘Hershey and Chase’ experiment, Watson and Crick proposed a subjective atomic resolution structure based on their model of DNA made from cardboard and metal<sup>6,7</sup> after Franklin shared her X-ray fiber diffraction data<sup>8,9</sup> on what is now known as the B-form DNA double helix.

##### *1.1.2 History of Base-Pair and Helical Structures of DNA*

Soon after Watson and Crick proposed their double helical model, many efforts were made at uncovering the specific base-pairing partners and geometry among the aforementioned nucleotides. Technological improvements in X-ray instrumentation afforded higher resolution and strategies shifted toward pioneering base-pair patterns in

free purine-pyrimidine (pur-pyr) constructs. In 1959 Hoogsteen presented the first high resolution X-ray structure of crystals containing 9-methyladenine base-paired with 1-methylthymine, where methyl groups were used to cap base moieties where their corresponding sugars were omitted.<sup>10</sup> Interestingly, Hoogsteen did not observe the base-pairing pattern proposed by Watson and Crick.<sup>6,7</sup> Rather, he observed a base-pairing geometry by which the 9-methyladenine employed the opposite side of the base in base-pairing with 1-methylthymine.<sup>10,11</sup> Such a base-pairing motif within the DNA duplex proposed by Watson and Crick would require the glycosyl bond connecting the unique DNA base to its corresponding sugar to rotate 180 ° from the *anti* to *syn* orientation. In these ‘Hoogsteen’ (HG) base-pairs, rather than employing the N1 and exocyclic moiety at carbon 6 as proposed by Watson and Crick, the N7 is used in lieu of N1 while hydrogen bonding with the C6 exocyclic group is still employed (Figure 1.1 A). Despite this HG hydrogen bonding scheme requiring restriction of the DNA double helix by ~2.5 Å as compared to the model Watson and Crick built based on their and Franklin’s data, this discovery brings into question the base-pairing scheme proposed by Watson and Crick. This unique hydrogen bonding pattern was proposed two years earlier by Rich and colleagues in the context of RNA triplexes.<sup>12</sup>

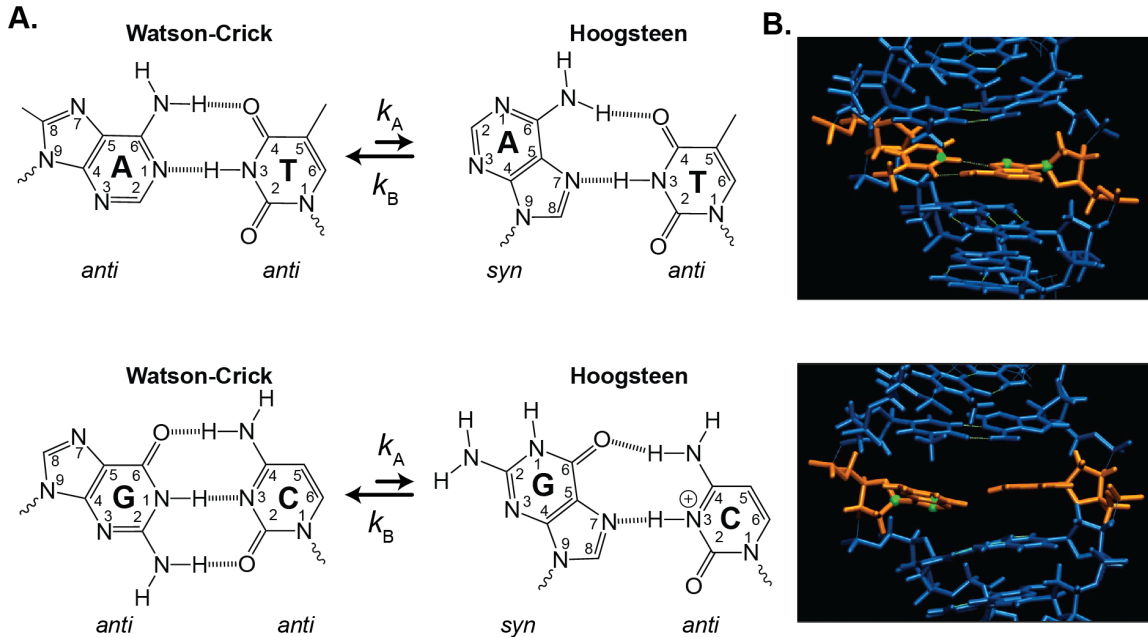
In 1963, one year after Watson and Crick were awarded the Noble Prize in Physiology or Medicine, Rich and coworkers reported a high resolution single crystal structures for G•C base-pairs of 9-ethylguanine with 1-methylcytosine and 1-methyl-5-bromocytosine, respectively.<sup>13</sup> Watson-Crick (WC) base-pairing was observed in the two cases, but skepticism surrounded the proposed base-pairing patterns since A•T base-pairs crystallized opposite to that proposed by Watson and Crick. Adding to the skepticism is

the fact that these modified G•C base-pairs were stabilized by three, not two, hydrogen bonds as proposed by Watson and Crick.<sup>6,7</sup> Pauling and coworkers proposed three hydrogen bonds in a G•C base-pair prior to this experimental evidence presented by Rich and coworkers.<sup>14</sup> Subsequent X-ray diffraction studies of co-crystals involving guanine and cytosine derivatives consistently formed the expected WC base-pairs with three hydrogen bonds.<sup>15</sup> However, to much dismay, attempts to generate co-crystals of adenines and thymine derivatives failed to produce WC base-pairs and, in most cases, favored HG base-pairs.<sup>15-17</sup> In 1968, Guschlbauer and colleagues proposed the formation of protonated G•C HG base-pairs in poly(dG-dC) at pH 3-4 based on optical rotatory dispersion spectra sensing molecule chirality and suggesting guanine adopted a *syn* conformation under such conditions.<sup>18</sup> The first crystallographic observation of a protonated G•C HG base-pair came many years later in 1986 for a DNA duplex in complex with the bisintercalating antibiotic triostin A.<sup>19</sup>

The only experimental indication that both G•C, A•T and A•U base-pairs could form WC base-pairs in duplexes void of crystal packing forces came from subsequent solution NMR studies of tRNA and polynucleotide complexes in the 1970s.<sup>20-26</sup> These studies presented unique chemical shift signatures that were consistent with theoretical predictions for WC rather than HG hydrogen bonding in A•T and A•U base-pairs, further strengthening the original claim by Watson and Crick. Six years later in 1979, significant advancements in chemical synthesis of homogeneous oligonucleotides<sup>27,28</sup> gave rise to a report by Rich and colleagues of the first single crystal X-ray structure of a DNA duplex for the hexameric CpG repeat (CG)<sub>3</sub>.<sup>29</sup> Surprisingly, (CG)<sub>3</sub> crystalized as a duplex with opposite handedness to that consistent with the previous X-ray diffraction patterns

presented in the 1950s and 1960s. This bizarre structure consisted of 7-7.5 base-pairs per turn, which is stereochemically disallowed by the right-handed configuration. It therefore did not give rise to the heralded X-ray diffraction data presented by Watson and Crick<sup>6,7</sup> and Franklin,<sup>8,9</sup> and reignited the skepticism surrounding the DNA duplex structure. It would not be until 1980, nearly 30 years after the original proposal of a right-handed helical DNA structure and particular hydrogen bonding pattern for which a Nobel Prize was awarded, that objective evidence for such is actually achieved by Dickerson and coworkers.<sup>30</sup>

Roughly another 30 years later in 2011, our lab would present evidence for the transient excursion of a WC base-pair within a canonical right-handed helix, like that first published by Dickerson and coworkers,<sup>30</sup> to the HG form first observed by Hoogsteen.<sup>10,11</sup> This report again was met with controversy<sup>31</sup> despite positive controls that used DNA damage of the WC face to ‘trap’ the HG form as a reference and detailed comparison of energetic parameters of the transition as compared to other process involving local base-pair disruption, such as base-pair opening, or ‘breathing’. We observed such instances at two unique base-pairs located at a junction between B-DNA and homopolymeric (A•T)<sub>n</sub> repeats, or ‘A-tracts’. The unique location at the B-DNA/A-tract junction of these specific transient HG base-pairs provides the foundation for a portion of the work presented in this thesis, whereby a wide range of sequence contexts are explored in the context of the transient HG base-pair.



**Figure 1.1 | Hoogsteen Base-Pairs in Canonical Duplex DNA.** (A) A•T (top) and G•C (bottom) Watson-Crick and Hoogsteen base-pairs. (B) Proposed structure of Hoogsteen base-pairs (orange) embedded within Watson-Crick duplex DNA, reported previously (blue).<sup>32</sup> Experimental probes are marked with green spheres.

## 1.2 The Surprising Discovery of Z-DNA and its Surrounding Controversy

### 1.2.1 A Brief History of Z-DNA Structure

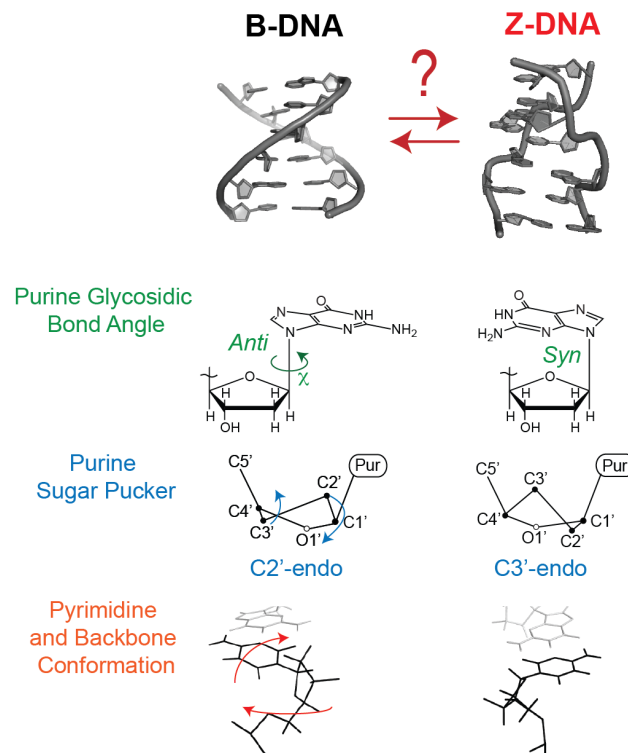
Unlike most biomolecular systems of interest, Z-DNA was discovered at atomic resolution before its precedence was set with biochemical experiments. Z-DNA's accidental discovery was made at an uncanny time when the very nature of the DNA double helix was being questioned by biophysicists (Section 1.1). Add to this the very exotic, unintuitive structure of Z-DNA (Figure 1.2) and it becomes clear why Z-DNA was met with staunch skepticism that continued for decades. Z-DNA would make its first mark on the literature in 1972 when Pohl and Jovin discovered a 'reverse' CD spectrum of poly(dG-dC) upon addition of high concentrations of NaCl, MgCl<sub>2</sub> and other salts.<sup>33</sup>

Their preliminary kinetic studies suggested a high energetic cost (an activation energy of ~ 20 kcal/mol) for the change responsible for the reverse CD spectrum. Seven years later Rich and coworkers would obtain crystals of a CpG dinucleotide with the intercalator acridine orange, in contrast to exhaustive efforts on free nucleosides for the past two decades.<sup>34</sup> They presented evidence that G•C base-pairs were capable of forming the Watson-Crick base-pairs originally proposed, strikingly with three hydrogen bonds instead of the two proposed by Watson and Crick.<sup>6</sup> The ‘Dickerson dodecamer’ that exhibits a full turn of right-handed B-DNA supported the three base-pair scheme of G•C base-pairs.<sup>30</sup>

In their following manuscript, Rich and coworkers explained they were encouraged by the high resolution crystals obtained from the dinucleotide CpG and pursued crystallization of all C- and G-containing tetrameric DNA oligonucleotides. The sequence that diffracted best was that of d(CpGpCpG), but after two years of attempts at solving the structure, Rich and coworkers moved to the crystallization of hexamers. Again, the CpG repeat crystallized best and it was with this diffraction data they solved the proclaimed left-handed structure and published the first understanding of DNA in absence of modifications and intercalators.<sup>29</sup> Rich and coworkers then revisited their tetrameric CpG diffraction data and solved the structure as, again, left-handed.<sup>35</sup> Other groups would immediately begin attempts at understanding how DNA adopts the Z-form.<sup>36,37</sup> After Dickerson and coworkers published the first high resolution crystal structure of a full helical turn of right-handed DNA, it became clear what structural features were distinct for B- and Z-DNA that created their opposite backbone configurations. For B-DNA to convert to Z-DNA, both bases in each base-pair must flip



upside down and maintain (or restore if broken during the transition) WC hydrogen bonds. As in HG base-pairs, the purine base is flipped from the B-like *anti* into the *syn* conformation. The purine sugar pucker also changes from the DNA-like *C2'-endo* to the RNA-like *C3'-endo*. Importantly, in Z-DNA, the flipping of the cytosine base and sugar moiety necessitates translation of the cytosine moiety to restore the WC bonds in the upside down fashion. It is the flipping of the cytosine that causes the dramatic zig-zag in the Z-DNA phosphate backbone (Figure 1.2).



**Figure 1.2 | Atomic Resolution Structure of B-DNA and Z-DNA.** B-DNA PDB ID: 1UQG. Z-DNA PDB ID: 3QBA.

In 1979 Patel and coworkers presented the first evidence for Z-DNA structure in solution using  $^1\text{H}$  and  $^{31}\text{P}$  NMR and confirm this “alternating B-DNA” molecule contained an alternating pattern that repeated itself every other nucleotide, as the name suggests.<sup>38</sup> Contrary to B-DNA, where every non-terminal base forms favorable stacking

interactions with both of its neighbors, bases in Z-form adopt a bifurcated stacking pattern whereby pyr-pur form stacking interactions (i.e. CpG, TpA, etc.) and pur-pyr do not (i.e. GpC, ApT, etc.). The atomic resolution NMR data provided evidence that Z-DNA may be something more than a crystal packing artifact. Despite Z-DNA's initial discovery being in 5 M NaCl or 2 M MgCl<sub>2</sub>, preliminary evidence for a biological role of Z-DNA quickly began accumulating as scientists investigated its behavior under the biologically relevant perturbations of DNA damage<sup>39</sup> and the epigenetic modification (5-methylcytosine).<sup>40</sup> In 1980 Dickerson and colleagues crystallized the so-called Z' form; a left-handed helix with slightly different parameters, including C1'-*exo* purines and ~6° decreased glycosidic angle for the *syn* conformation.<sup>41</sup> The variability between this left-handed helix and that originally published by Rich and coworkers coincides with variability comparing the RNA-like A and A' forms.<sup>42</sup> Since then many left-handed DNAs crystal structures have been solved and Z-DNA is now thought of as a family of left-handed structures,<sup>29,43-47</sup> similar to that of A- and B-forms.<sup>42,48</sup>

### 1.2.2 Biological Relevance of Z-DNA

A decade after the controversial discovery of Z-DNA its structure still seemed biologically unnecessary to many scientists at the time. Z-DNA would experience a curiosity-inspired boom in research in the 1980s with few explicit demonstrations of its involvement *in vivo*. Rather, focus was on probing simply for Z-DNA formation in some cases, and trying to induce the structure in others. Amongst these efforts include discoveries that non-CpG pyr-pur repeats adopt Z-DNA<sup>49-51</sup> and that Z-DNA can accommodate imperfect pyr-pur repeats<sup>52</sup> as well as base-pair mismatches<sup>53-57</sup> and

damage.<sup>58-60</sup> Although the primary investigations of Z-DNA used near-saturating salt concentrations for stabilization,<sup>33,38,39,61</sup> chemists discovered in the 1980s that spermine<sup>46,62-66</sup> and spermidine<sup>67-71</sup> stabilize Z-DNA by binding to the minor groove at physiological (micromolar) concentrations. Moreover, certain types of naturally occurring DNA damage moieties stabilize Z-DNA formation, suggesting DNA damage and Z-DNA formation could be correlated.<sup>72-81</sup> In some rare cases, carcinogens react with B-DNA but not Z-DNA.<sup>59,82</sup>

Wells and coworkers shortly thereafter built a recombinant plasmid containing poly(dG-dC) purified by high performance liquid chromatography (HPLC) and monitored ellipticity from Circular Dichroism (CD), agarose gel migration and phosphorus chemical shifts to demonstrate Z-DNA induction in a relatively short (~150 base-pair) CpG repeat comprising ~1 % of the plasmid was enough to significantly change the overall morphology of the plasmid. Significantly, this study demonstrated B-DNA and Z-DNA can coexist within the same closed circular genomic DNA, suggesting Z-DNA formation and supercoil-driven overall molecular topology are linked.<sup>83</sup> Shortly thereafter Z-DNA was discovered to be stabilized by natural amounts of supercoiling with a high degree of sequence variability, demonstrating a broader range for Z-DNA forming conditions.<sup>39,49,83-88</sup> Rabbits immunized with poly(dG-dC) developed 'antiserum', or antibodies to Z-DNA,<sup>89,90</sup> and importantly autoantibodies of the autoimmune disease Lupus bind preferentially to Z-DNA, providing the first evidence for a relationship between Z-DNA and disease.<sup>91</sup> In the same era, Behe and Felsenfeld presented evidence Z-DNA forms *in vivo*.<sup>40</sup> Groups then reported Z-DNA regions in polytene chromosomes of *Drosophila melanogaster*, the household fruit fly, in

transcriptionally active, or ‘puff’, regions, suggesting Z-DNA forms near transcriptionally active regions in eukaryotes.<sup>92-94</sup> Damaged DNA in the form of alkylated<sup>59,60,95,96</sup> and brominated<sup>97-99</sup> bases was then demonstrated to sufficiently stabilize Z-DNA under native levels of supercoiling in low salt. The aforementioned experiments gave indirect evidence for the biological relevance of Z-DNA.

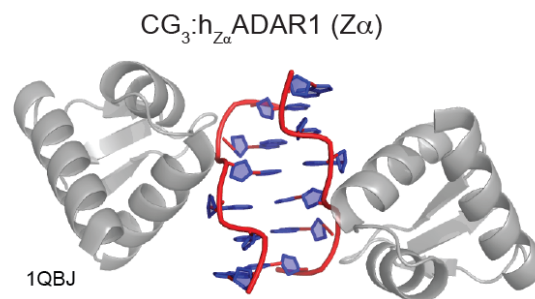
In 1987 Liu and Wang boldly proposed that eukaryotic RNA polymerase (RNAP) does not rotate around the DNA and instead remains rather stationary while it ‘plows through’ the DNA duplex. This model then predicts a large degree of positive supercoiling ahead of RNAP and a large amount of Z-DNA in its wake (Figure 1.3).<sup>100</sup> Shortly thereafter, in 1991, Rich and colleagues would present the first strong evidence for Z-DNA formation *in vivo*.<sup>101-103</sup> Rich and colleagues elegantly encapsulated mammalian cells in agarose beads and mild detergent to lyse the cytoplasmic membrane, permeabilize the nuclear membrane and leave the nucleus intact without altering the rate of replication as compared to that *in vivo*.<sup>101</sup> Using the same assay, Rich and colleagues probed for Z-DNA forming regions in the nucleus using biotinylated Z-DNA antibodies and found that the extent of Z-DNA formation was regulated by torsional strain<sup>101</sup> and correlated with RNA synthesis,<sup>102</sup> confirming the model proposed by Liu and Wang in 1987.<sup>100</sup> In 2004 Levens and coworkers would provide additional evidence for severe DNA distortions in the wake of RNAP.<sup>104,105</sup> Z-DNA has also been associated with regulation of gene expression,<sup>106-112</sup> and more recently Ray and coworkers demonstrated DNA structures driven by supercoiling regulate gene expression of the breast cancer-associated ADAM-12 gene.<sup>109</sup> Rich and Ray confirmed the same gene to have a Z-DNA negative regulatory element.<sup>113</sup> In 2012 Zhao and coworkers demonstrated that

nucleosome remodeling causes Z-DNA formation when Z-DNA forming sequences are present (Figure 1.3).<sup>114</sup>



**Figure 1.3 | Involvement of DNA Deformation in Cellular Processes.** Blue represents DNA. Yellow balls wrapped with DNA under ‘Genome Packaging’ represent nucleosomes.

Perhaps more encouraging than the elaborate experiments to show Z-DNA’s involvement in genome packaging, DNA replication and transcription, is the discovery of a family of eukaryotic Z-DNA binding proteins that are part of the adenosine deaminase ADAR1 in humans, zebra fish, puffer fish and chickens (Figure 1.4).<sup>115-118</sup> These Z-DNA binding proteins were subsequently found to bind many sequences in the Z-form with nanomolar affinity.<sup>119,120</sup> The E3L Z-DNA binding protein produced by vaccinia virus belongs to this family and is essential for mortality in mice.<sup>121</sup>



**Figure 1.4 | Crystal Structure of Zα Bound to CG<sub>3</sub>.** PDB ID: 1QBJ.

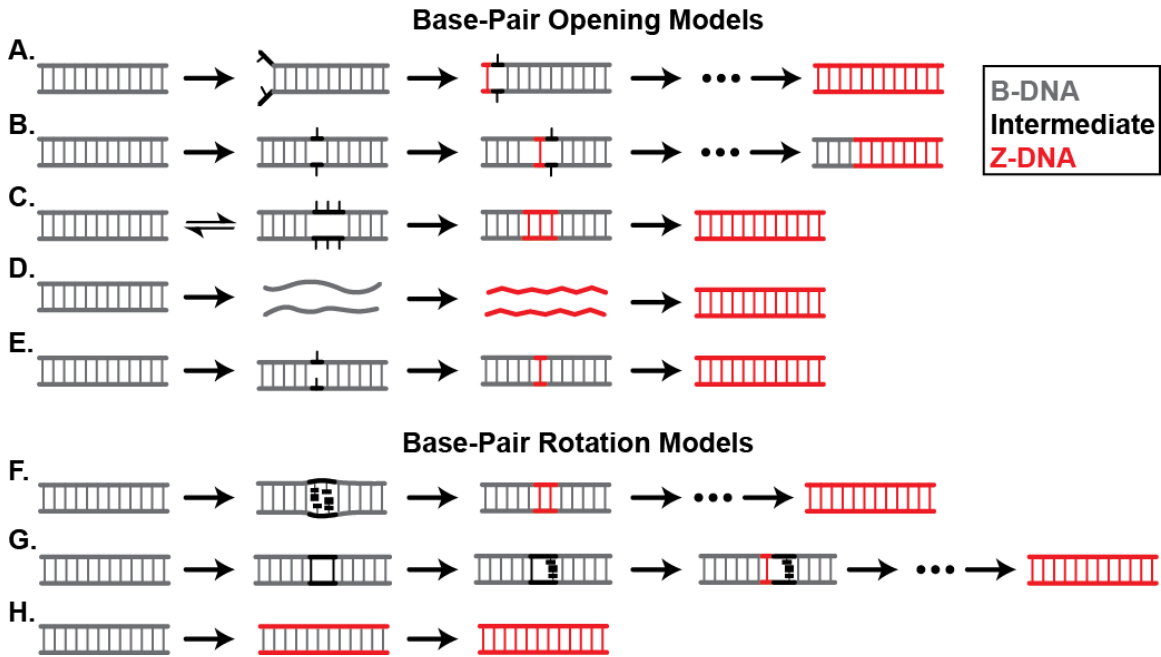
Quite amazingly, despite very elegant and finely tuned experiments *in vivo* and *in situ* that have revealed Z-DNA’s formation during chromatin remodeling,<sup>114</sup> regulation of

gene expression,<sup>106-112</sup> and RNA synthesis by RNAP,<sup>102,122</sup> we still do not fully understand Z-DNA's role in the cell.

### 1.2.3 B-DNA to Z-DNA Transition Kinetics

In the cell, B-DNA must convert to Z-DNA during dynamic, hierarchical, concerted processes and it is therefore believed Z-DNA is in dynamic equilibrium with B-DNA *in vivo*. As such, a fundamental understanding of the B-to-Z-DNA kinetic transition is essential for gaining further insight into Z-DNA's potential role in the aforementioned genetic processes. There are near one dozen proposed molecular models for the B-to-Z-DNA transition (Figure 1.5) but experimental observables hardly merit such discrimination between the models. Models that involve base-pair opening include the all-or-none, Wang, thermal fluctuation base-pair opening, helix-coil and solitary excitations models (Figure 1.5 A-E). Models that are not described with base-pair opening, but rather involve base-pair rotation while retaining WC base-pairs include the theoretical Olson proposal, Harvey and Ansevin-Wang models. We note that none of the aforementioned mechanistic models explicitly address the concept of an observable intermediate. A separate category of B-to-Z-DNA transition models exists whereby stable intermediates are explicitly proposed. These include the zipper, C-DNA-like intermediate, and stretched intermediate models which all define the intermediate as containing disrupted hydrogen bonds, and the Saenger and Heinemann model which proposes an intermediate without hydrogen bonding disruption along the transition. Lastly, two models have been proposed to the end of encompassing all kinetic observations in the literature: the unified biophysical and empirical salt-threshold models.

Proposed by Pohl and Jovin in their original study reporting the ‘reverse’ CD spectrum, the all-or-none model (Figure 1.5 A) involves a high energy nucleation point at the DNA terminus followed by a series of lower energy propagation steps.<sup>33</sup> The Wang model involves a similar energetic description of a nucleation point followed by propagation, with the distinction being that the nucleation point occurs within a duplex and propagates unidirectionally (Figure 1.5 B). The thermal fluctuation base-pair opening model describes a successive series of base-pairs that stochastically open spontaneously, followed by the same base-pairs returning to the confines of the double helix in Z-form to create a nucleation point of several Z-form nucleotides that propagates bidirectionally (Figure 1.5 C). The helix-coil model describes B-DNA as melting to single-stranded coils that adopt a single-stranded Z-form structure and re-anneal as Z-DNA (Figure 1.15 E). The model involving solitary excitations is similar to the Wang model (Figure 1.5 B) except it explicitly requires base flipping to occur extrahelically. For those models describing the B-to-Z-DNA transition restraining nucleotides in the base-paired fashion (Figure 1.15 F-H), the theoretical Olson proposal suggests two base-pairs flip simultaneously such that their stacking faces are pointing away from one another, creating an intermediate point, from which Z-DNA formation proceeds bidirectionally (Figure 1.15 F). The theoretical Harvey model builds off of the Olson model, but describes one base-pair as flipping at a time with the concomitantly created cavity propagating down the helix. The solution of Ansevin-Wang describes the backbone as transforming from B-form to Z-like, while the WC base-pairs remain in a B-like orientation, followed by rearrangement of the bases from B-like to Z-like conformations.

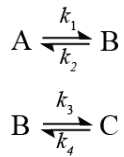


**Figure 1.5 | Proposed Schemes for the B-to-Z-DNA Transition.** Shown are the (A) All-or none, (B) Wang, (C) thermal fluctuation base-pair opening, (D) helix-coil transition, (E) solitary excitations, (F) Olson, (G) Harvey models and (H) the solution of Ansevin-Wang. B-DNA (grey), intermediate (black) and Z-DNA (red) structural features are highlighted, where an intermediate refers to a non-B- and non-Z-like structure.

#### 1.2.4 Kinetic Rate Laws

Consider a sequential two-step reaction whereby species  $A$  is converted to the intermediate  $B$ , and  $B$  is subsequently converted to the product  $C$  in a reversible fashion.

Such a mechanistic model would be described as



where  $k_1$  and  $k_3$  are the microscopic forward rate constants and  $k_2$  and  $k_4$  are the reverse microscopic rate constants. The time dependence of the species' concentrations is given by

$$\frac{dA}{dt} = -k_1A + k_2B \tag{1.1}$$



$$\frac{dB}{dt} = k_1A - (k_2 + k_3)B + k_4C \quad (1.2)$$

$$\frac{dC}{dt} = k_3B - k_4C \quad (1.3)$$

where  $A$ ,  $B$  and  $C$  are the species' concentrations. Conservation of mass is defined by

$$A_0 = A + B + C \quad (1.4)$$

where  $A_0$  is the concentration of species  $A$  when time is equal to zero. Solving Eq. 1.4 for

$C$  and substituting into Eq. 1.2 yields

$$\frac{dB}{dt} = k_1A - (k_2 + k_3)B + k_4(A_0 - A - B) \quad (1.5)$$

The form of Eq.'s 1.1, 1.3 and 1.5 following the Laplace transform and general partition theorem are given by

$$A = A_0 \left[ \frac{k_2k_4}{\alpha\beta} + \frac{k_1(\alpha - k_3 - k_4)}{\alpha(\alpha - \beta)} e^{-\alpha t} + \frac{k_1(\beta - k_3 - k_4)}{\beta(\beta - \alpha)} e^{-\beta t} \right], \quad (1.6)$$

$$B = k_1A_0 \left[ \frac{k_4}{\alpha\beta} + \frac{k_4 - \alpha}{\alpha(\alpha - \beta)} e^{-\alpha t} - \frac{k_4 - \beta}{\beta(\alpha - \beta)} e^{-\beta t} \right], \quad (1.7)$$

and

$$C = k_1k_3A_0 \left[ \frac{1}{\alpha\beta} + \frac{1}{\alpha(\alpha - \beta)} e^{-\alpha t} + \frac{1}{\beta(\alpha - \beta)} e^{-\beta t} \right], \quad (1.8)$$

where

$$\alpha = k_{obs,1} = \frac{k_1 + k_2 + k_3 + k_4 + \sqrt{(k_1 + k_2 + k_3 + k_4)^2 - 4(k_1k_3 + k_2k_4 + k_1k_4)}}{2}, \quad (1.9)$$

$$\beta = k_{obs,2} = \frac{k_1 + k_2 + k_3 + k_4 - \sqrt{(k_1 + k_2 + k_3 + k_4)^2 - 4(k_1k_3 + k_2k_4 + k_1k_4)}}{2}, \quad (1.10)$$

and  $k_{obs,1}$  and  $k_{obs,2}$  are the observed rate constants. Care must be taken when monitoring a kinetic reaction to consider if species' signals overlap such that the acquired time course

contains a superposition of species. The observed signal of the reaction is given by

$$y = A\varepsilon_{\lambda}^A + B\varepsilon_{\lambda}^B + C\varepsilon_{\lambda}^C \quad (1.11)$$

where  $\varepsilon_{\lambda}^A$ ,  $\varepsilon_{\lambda}^B$  and  $\varepsilon_{\lambda}^C$  are the molar absorptivity coefficients for  $A$ ,  $B$  and  $C$  at a given wavelength,  $\lambda$ . Substituting Eq.'s 1.9 and 1.10 into 1.6- 1.8 then the latter into Eq. 1.11 yields a considerably complex equation that is not practical for fitting a kinetic trace, or 'transient'. However, the same observed rate constants describe each of the three species, allowing for a fit to the more general sum of exponentials as given by

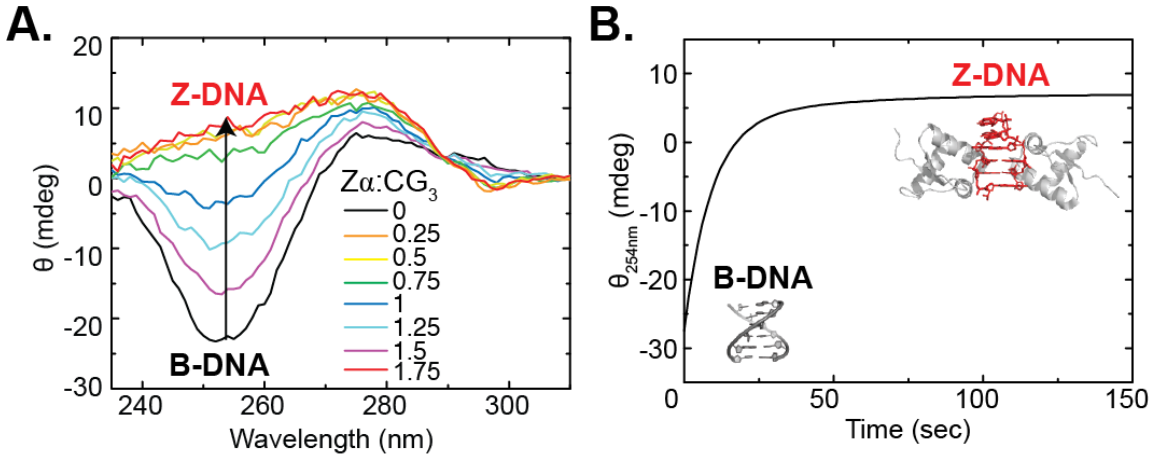
$$y = y_f - \left( A_1 e^{-k_{obs,1}t} + A_2 e^{-k_{obs,2}t} \right) \quad (1.12)$$

where  $y$  is the observed signal as a function of time,  $y_f$  is the final observed signal,  $A_1$  and  $A_2$  represent the signal magnitudes corresponding to each phase and  $k_{obs,1}$  and  $k_{obs,2}$  represent the observed rate constants for each phase of a given transition.

### 1.3 Circular Dichroism as a Tool to Measure Biomolecular Conformational Change

CD is a powerful, robust technique used to assess biomolecular structure due to its unique ability to sense the difference in chirality of molecules.<sup>123-126</sup> B-form, A-form and Z-form DNAs, as well as quadruplexes and hairpins, all give rise to unique CD signals.<sup>127-131</sup> The ability to use low concentrations ( $\sim 5 \mu\text{M}$  for 12-mers) affords rapid data acquisition, especially when many solution conditions are used. Moreover, nucleic acids and proteins absorb circularly polarized light in different spectral regions, allowing for trivial background subtraction of protein-containing solutions. In the case of the protein-induced B-to-Z-DNA transition, a near-inversion of the CD spectrum is observed due to the change from a right- to left-handed structure (Figure 1.6 A). Perhaps not as common, but equally powerful is the potential to measure the CD signal of a given

conformational change over time due to CD's acute sensitivity to DNA structural changes (Figure 1.6 B).

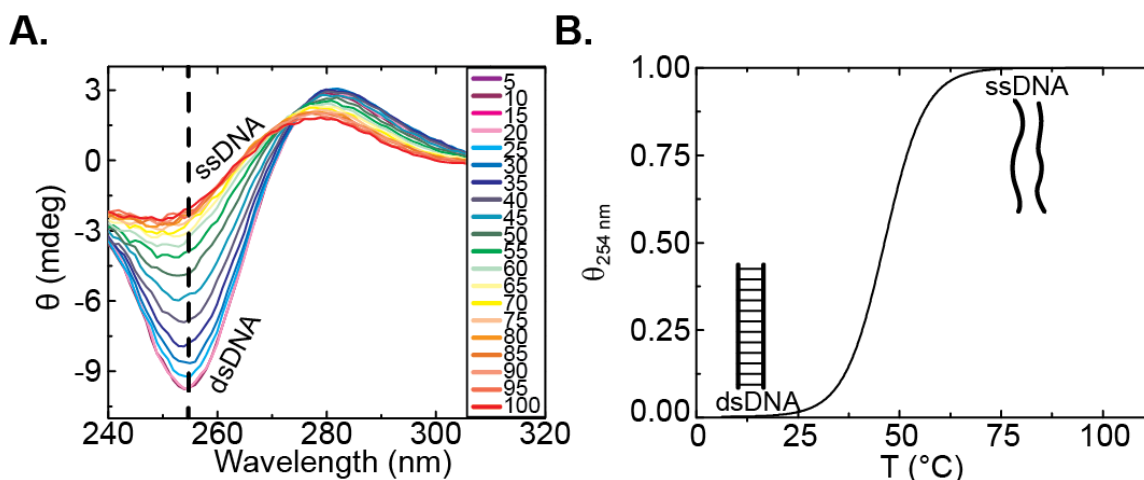


**Figure 1.6 | CD Spectroscopy of the B-to-Z-DNA Transition.** (A) Wavelength scan showing large spectral change near 254 nm upon Z-DNA formation. (B) Example time course curve at 254 nm. B-DNA PDB ID: 1UQC. Z-DNA PDB ID: 1QBJ. See Chapter 3 for details.

CD is also a useful tool for monitoring DNA melting for its sensitivity, precision and stability. As temperature is increased, the signal magnitude due to B-DNA structure decreases (Figure 1.7 A). The fraction of signal change observed at a given wavelength is plotted as a function of temperature (Figure 1.7 B) and can be described as

$$\theta = LL + \frac{UL - LL}{1 + e^{-\frac{T - T_m}{a}}} \quad (1.13)$$

where  $\theta$  is the ellipticity at 254 nm normalized to the signal change magnitude,  $LL$  and  $UL$  are the lower and upper limits of the transition, respectively,  $a$  is the Hill slope,  $T_m$  is the melting temperature defined as the point of inflection of the melting curve and  $T$  is the independent variable temperature.



**Figure 1.7 | DNA Melting as measured by Circular Dichroism.** (A) Wavelength scans and (B) single wavelength measurement as a function of increasing temperature. Legend in ‘A’ represents °C. See Chapter 2 for details.

## 1.4 Nuclear Magnetic Resonance Spectroscopy of DNA

### 1.4.1 Nuclear Magnetic Resonance Fundamentals

NMR compounds are active due to their possession of nuclear spin, which is a complex phenomenon dictated by the nuclear spin quantum number, or nuclear spin angular momentum vector,  $\mathbf{I}$ . For a nucleus in which  $\mathbf{I}$  is non-zero, a non-zero nuclear magnetic moment exists,  $\boldsymbol{\mu}$ . The vector  $\boldsymbol{\mu}$  is projected along the spin angular momentum vector such that they are collinear (Figure 1.8 A) and

$$\vec{\mu} = \gamma \vec{I} \tag{1.14}$$

where  $\gamma$  is the gyromagnetic ratio, an intrinsic constant for a given nucleus (Table 1.1).

**Table 1.1 | I = 1/2 Nuclei Commonly Studied by NMR Spectroscopy.**<sup>132</sup>

Nucleus	$\gamma$ (T•s) <sup>-1</sup>	Natural Abundance (%)
<sup>1</sup> H	2.6752 x10 <sup>8</sup>	99.99
<sup>13</sup> C	6.728 x10 <sup>7</sup>	1.07
<sup>15</sup> N	-2.713 x10 <sup>7</sup>	0.37

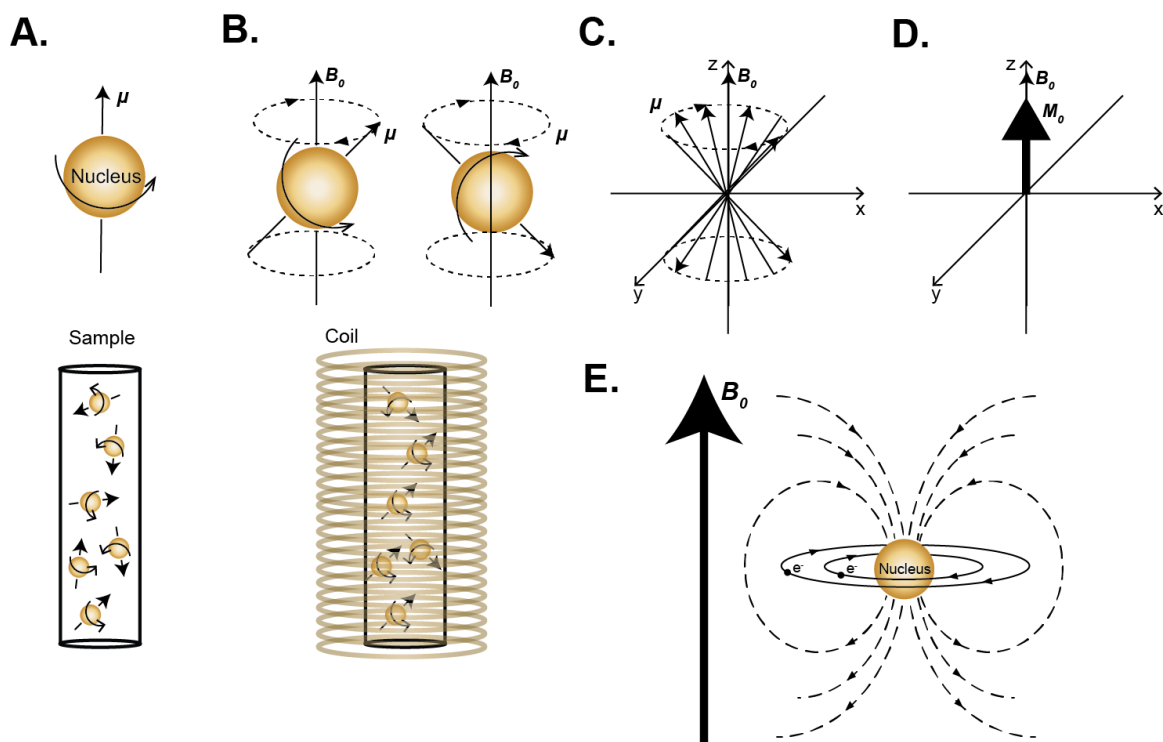
$|I| = 1/2$  is the most fruitful NMR active spin because they have relatively long lifetimes compared to their quadrupolar counterparts ( $|I| > 1/2$ ), and also lack electric quadrupolar moments arising from asymmetric distribution of charges in the nucleus.

Once put in an NMR magnet, the nuclear spins in a given sample will be polarized and a bias in their z component will be instilled either parallel or antiparallel to the external magnetic field of the NMR spectrometer,  $\mathbf{B}_0$  (Figure 1.8 B). The minimum energy of a given spin angular momentum is achieved with maximum projection along the external magnetic field. This orientation is therefore the most likely state for a given vector to adopt. However, the projection of  $\boldsymbol{\mu}$  along  $\mathbf{B}_0$  follows a Boltzmann distribution of orientations of  $\boldsymbol{\mu}$  projected parallel or perpendicular to  $\mathbf{B}_0$ . This phenomenon is known as ‘Zeeman splitting’ and occurs when a sample is placed inside an NMR magnet. Because there is no bias in the x or y dimensions, the sum of x and y components of the nuclear spin vectors is zero. Once a Boltzmann distribution of spin vector orientations is reached, the spin vectors are summed and the slight bias in populations across Boltzmann energy levels in an orientation parallel to the external magnetic field is described as the ‘bulk magnetization’ vector,  $\mathbf{M}_0$  (Figure 1.8 D). Under free precession in the absence of applied radio frequency (RF) pulses,  $\mathbf{M}_0$  is given by the Bloch equations<sup>133,134</sup> as

$$\frac{d}{dt} \begin{pmatrix} M_x(t) \\ M_y(t) \\ M_z(t) \end{pmatrix} = \begin{pmatrix} -R_2 & -\Omega & \omega_1 \\ \Omega & -R_2 & -\omega_1 \\ -\omega_1 & \omega_1 & -R_1 \end{pmatrix} \begin{pmatrix} M_x(t) \\ M_y(t) \\ M_z(t) \end{pmatrix} + R_1 \begin{pmatrix} 0 \\ 0 \\ M_0 \end{pmatrix} \quad (1.15)$$

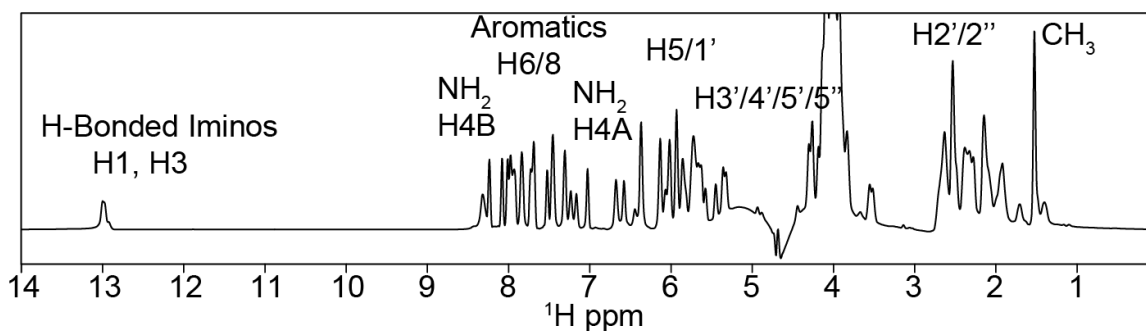
where  $\mathbf{M}_x$ ,  $\mathbf{M}_y$  and  $\mathbf{M}_z$  are the x, y and z components of the bulk magnetization vector, respectively,  $\Omega$  is the offset in Hz relative to the carrier frequency and  $\omega_1$  is the frequency at which the bulk magnetization precesses about the effective magnetic field and  $\mathbf{M}_0$  describes the bulk magnetization vector at thermal equilibrium under the influence of an external magnetic field.  $R_1$  and  $R_2$  are the longitudinal and transverse relaxation rate constants that describe the decay of the bulk magnetization along z and in the xy plane, respectively.

From Table 1.1, one would predict all nuclei of the same identity have the same NMR observable frequency, or chemical shift. Indeed each nucleus type has unique chemical shifts relative to one another. However, when in context of a biomolecule, every individual nucleus of a given type will experience a slightly different effective magnetic field induced by electron currents in molecular orbitals in close proximity (Figure 1.8 E). Though this effect is exceptionally small, with high field magnets atomic resolution is increasingly afforded and such resonances become resolved from one another in the NMR spectrum, even for biomolecules which possess many nuclear spins compared to small molecules. The magnitude of a given electronic current is proportional to the local electron density. For such affects creating higher local magnetic fields, the affected spin is 'shielded' and will be observed 'upfield' at a lower chemical shift relative to the theoretical, non-shielded chemical shift. Similarly, the opposite is true for a spin experiencing a lower local magnetic field; it is 'deshielded' and observed 'downfield' with a higher chemical shift.



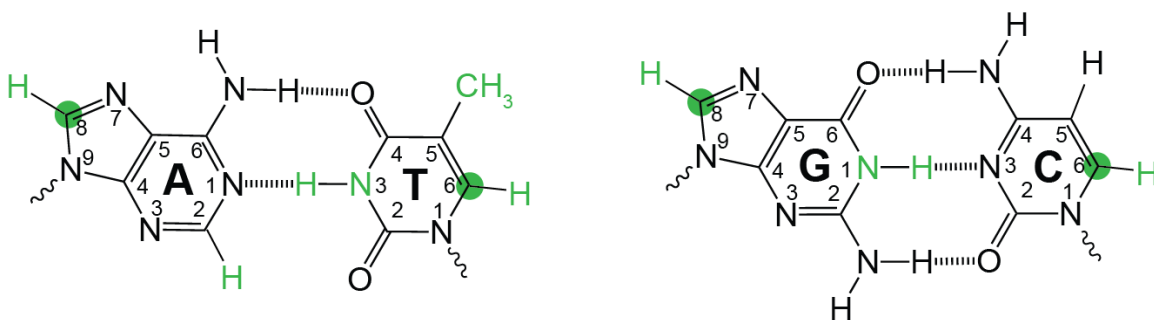
**Figure 1.8 | Nuclear Magnetic Resonance Fundamentals.** (A) A nucleus with magnetic moment,  $\mu$ , shown (top) individually and (bottom) in the context of an NMR sample. (B) Effect of an external magnetic field on  $\mu$  upon placing the sample into an NMR magnet (Coil). (C) Representation of  $\mu$  for many spins within a sample at equilibrium in the laboratory frame. (D) Sum over all nuclear magnetic moments to arrive at  $M_0$ , the bulk magnetization vector at equilibrium in the rotating frame.

Finally, the consequence of each nuclear spin type possessing a unique  $\gamma$ , and each nuclear spin in a biomolecule experiencing a unique local magnetic field is observed in Figure 1.9, a 1D  $^1\text{H}$  NMR spectrum in which a DNA duplex is measured.



**Figure 1.9 | Proton Chemical Shifts of a DNA Duplex.**

The first step in studying a biological macromolecule by NMR is to assign the resonances of interest. For the studies presented herein, we will focus on adenine C8-H8 and C2-H2, thymine N3-H3, C6-H6 and C5 methyl, guanine C8-H8 and N1-H1 and lastly cytosine C6-H6 (Figure 1.10).



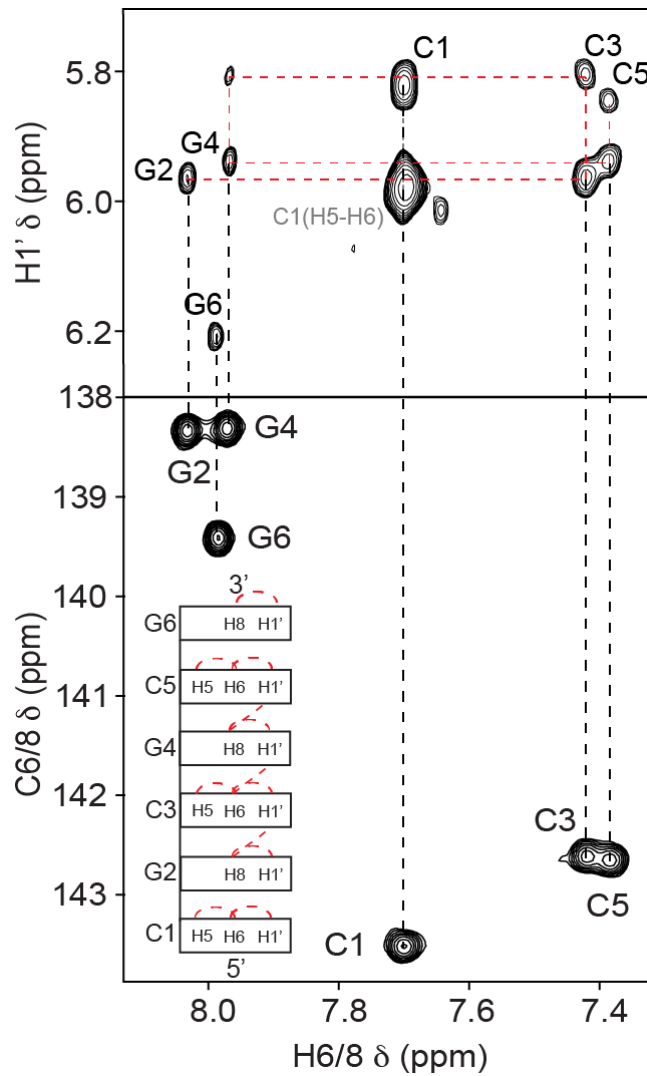
**Figure 1.10 | Target DNA Resonances for NMR Spectroscopy**

Resonance assignments of DNA are well established and have been reviewed extensively by Wijmenga and coworkers.<sup>135,136</sup> After significant advancements in synthesizing isotopically labeled (<sup>13</sup>C and/or <sup>15</sup>N) DNA *in vitro*,<sup>137</sup> elaborate resonance assignment strategies were developed that rely on transfer of magnetization from <sup>1</sup>H to <sup>13</sup>C and/or <sup>15</sup>N (reviewed in Wijmenga and coworkers references above). While this synthesis method greatly simplified obtaining isotopically labeled DNA, its Achilles heel would be the necessity of gel electrophoresis to separate tens of milligrams of template from labeled product such that extremely short DNAs (<8 nucleotides) cannot be effectively resolved in such large quantities. In the studies presented in this thesis we focus much attention on six-nucleotide DNAs, or ‘6-mers’, thus precluding reliable purification by the method presented by Zimmer and Crothers<sup>137</sup> despite exhaustive efforts. We therefore employ the conventional <sup>1</sup>H-<sup>1</sup>H assignment technique for its ability to bypass isotopic enrichment.



The classical  $^1\text{H}$ - $^1\text{H}$  assignment first polarizes the protons by applying an RF pulse that tips  $\mathbf{M}_0$  from the z axis into the xy, or transverse, plane. The spins are allowed to precess and thus become encoded with their unique frequencies, similar to those shown in Figure 1.9, then brought back to the z axis where a ‘mixing’ phenomenon occurs, followed finally by tipping the frequency- and mixing- encoded spins back to the transverse plane for detection. The mixing phenomenon is known as the nuclear Overhauser effect (NOE), which describes the transfer of a given spin’s magnetic polarization through space to other extremely close ( $<5 \text{ \AA}$ ) spins, or cross relaxation. The result is a 2D NMR spectrum where each axis represents all proton chemical shifts excited by the experiment. A diagonal exists for each ‘auto’ peak and all off diagonal, or ‘cross’ peaks represent cross relaxation events between the two protons represented by each axis of the spectrum. The highly repetitive nature of DNA nucleotides strung together to make one strand of a DNA helix allows for very straightforward assignment of sugar (H1’) and aromatic (G and A H8, A H2, T and C H6) resonances. The aromatic resonances within a nucleoside unit will always experience cross relaxation with, or ‘talk to’ each other as colloquially referred to by NMR spectroscopists. In a unidirectional manner (5’ to 3’ along one DNA strand) the 5’ aromatic resonance will ‘talk’ to the 3’ H1’ resonance, provided one or both resonances are in a canonical duplex formation and are not experiencing significant chemical exchange. This, of course, means no cross relaxation between a terminal aromatic resonance and its neighboring H1’ will be observed. The lack of a cross peak between these two helps to provide a clue to the experimentalist assigning the DNA resonances. Once the terminal resonances are identified, one can ‘walk’ up the helix as indicated by red dotted lines in Figure 1.11.

Upon full assignment of aromatic and sugar proton resonances (Figure 1.11, top spectrum), assignments can then be extrapolated to separate NMR experiments involving heteronuclear correlations between protons and carbons or nitrogens (HSQCs or HMQCs) as shown in the bottom spectrum of Figure 1.11. Once DNA resonances of interest are assigned, measurements of DNA dynamics may be implemented.



**Figure 1.11 | Example of a  $^1\text{H}$ - $^1\text{H}$  NOE DNA Assignment.**

### 1.4.2 Chemical Exchange and $R_{lp}$

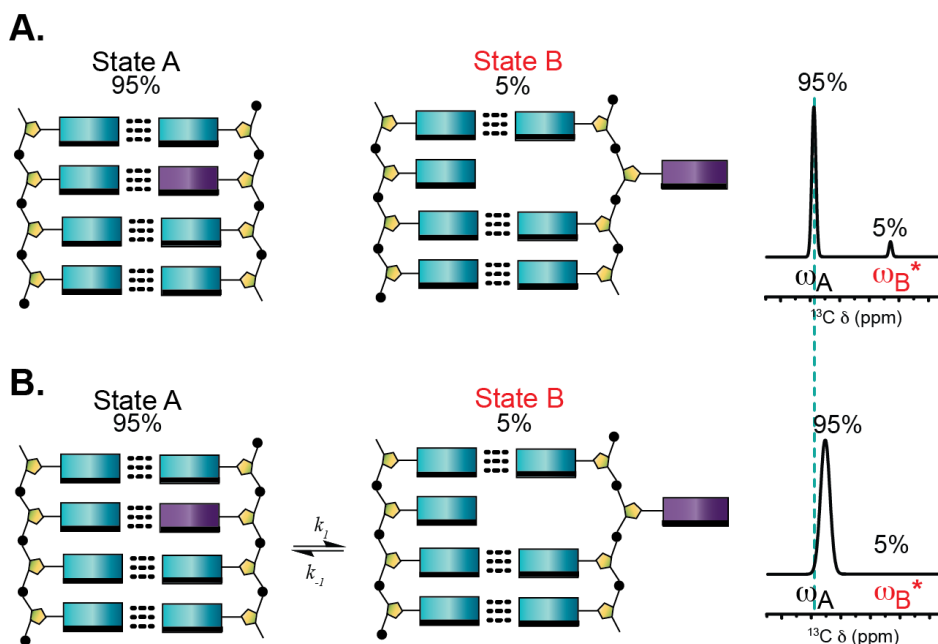
DNA dynamics are exhibited on a wide range of timescales from picoseconds to minutes, reviewed by Bothe and coworkers.<sup>138</sup> Those relative to many biological events, such as ligand binding, conformational changes and protein recognition, occur on the NMR timescale of microseconds to milliseconds. Chemical exchange refers to the process of a given NMR-observable nucleus experiencing two distinct electronic (and subsequently chemical) environments on the NMR timescale. To account for magnetization that is undergoing chemical exchange, the Bloch equations (Eq. 1.15) were subsequently modified. Here, the so-called Bloch-McConnell equations<sup>139,140</sup> to describe the magnetization of a system in two-state exchange (Figure 1.9)

$$\frac{d}{dt} \begin{pmatrix} M_{Ax}(t) \\ M_{Bx}(t) \\ M_{Ay}(t) \\ M_{By}(t) \\ M_{Az}(t) \\ M_{Bz}(t) \end{pmatrix} = \begin{pmatrix} -k_{AB} - R_2 & k_{BA} & -\delta_A & 0 & 0 & 0 \\ k_{AB} & -k_{BA} - R_2 & 0 & -\delta_B & 0 & 0 \\ \delta_A & 0 & -k_{AB} - R_2 & k_{BA} & -\omega_1 & 0 \\ 0 & \delta_B & k_{AB} & -k_{BA} - R_2 & 0 & -\omega_1 \\ 0 & 0 & \omega_1 & 0 & -k_{AB} - R_1 & k_{BA} \\ 0 & 0 & 0 & \omega_1 & k_{AB} & -k_{BA} - R_1 \end{pmatrix} \begin{pmatrix} M_{Ax}(t) \\ M_{Bx}(t) \\ M_{Ay}(t) \\ M_{By}(t) \\ M_{Az}(t) \\ M_{Bz}(t) \end{pmatrix} + R_1 \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ M_{A0} \\ M_{B0} \end{pmatrix} \quad (1.16)$$

where  $\delta_A = \Omega_A - \omega_{rf}$  and  $\delta_B = \Omega_B - \omega_{rf}$  are the resonance offsets of the ground (A) and transient (B) states in the rotating frame.  $\Omega_A$  and  $\Omega_B$  are the respective Larmor frequencies of A and B and  $\omega_{rf}$  is the frequency of the applied RF field.  $k_{AB}$  and  $k_{BA}$  are the rates of exchange between the two states, respectively.  $R_1$  and  $R_2$  are assumed to be identical for the ground and transient states.

When a given nuclear spin exists in electronically and chemically distinct states A and B in a mutually exclusive manner, each state gives rise to an observed resonance with volumes and heights proportional to each state's population (Figure 1.12 A). If exchange were to occur between states A and B, peak broadening occurs due to the chemical

exchange phenomenon. In cases such as DNA chemical exchange, state B is usually <5 %, causing resonance B to broaden out to the point where it evades direct detection. However, the distinct chemical exchange between states A and B has still left its mark on resonance A (Figure 1.12 B). Moreover, when exchange between states A and B occurs on a timescale faster than that of the NMR experiment ( $\mu\text{sec}$ -  $\text{msec}$ ), there is no ability to distinguish between the two and a population-weighted average peak is observed, instead.



**Figure 1.12 | Chemical Exchange.** A resonance (violet) exists in two conformations with (A) no chemical exchange and (B) the presence of chemical exchange. Shown on the right are observable spectra for each scenario. Dashed line represents  $\omega_A$ .

Relaxation dispersion experiments have been developed to measure such chemical exchange, especially Carr-Purcell-Meiboom-Gill (CPMG) and rotating frame relaxation dispersion ( $R_{1\rho}$ ).<sup>141,142</sup>  $R_{1\rho}$  relaxation dispersion measures the relaxation rate of the transverse magnetization decay due to  $R_2$  as well as that due to chemical exchange ( $R_{\text{ex}}$ ). When an RF pulse is applied to tip the bulk magnetization into the transverse plane, creating  $\mathbf{M}_{xy}$ , the spins are said to be in a ‘phase coherence’ whereby there is relatively

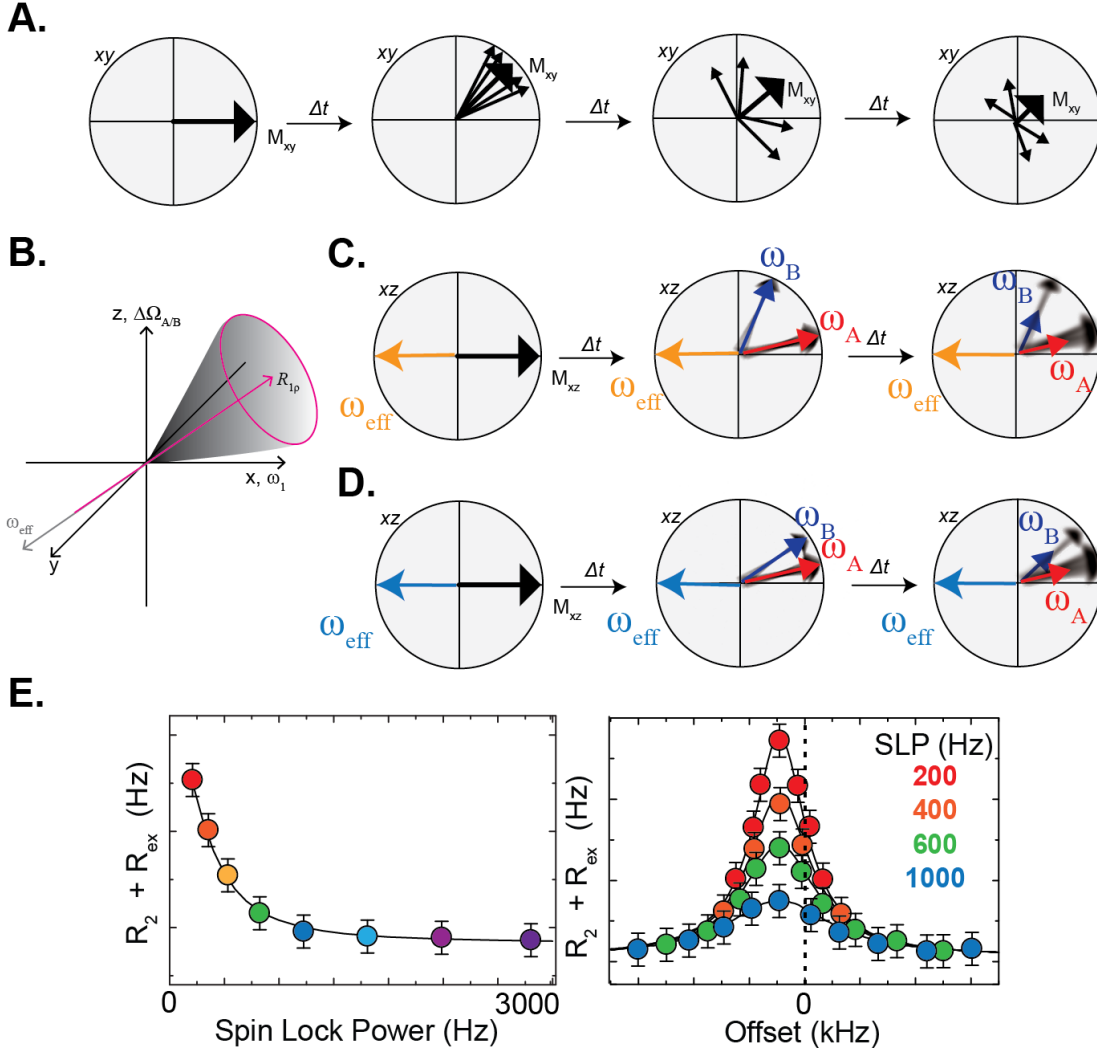
little inhomogeneity in the nuclear spin orientations comprising  $\mathbf{M}_{xy}$ . After a time period,  $\Delta t$ , corresponding to one period of precession, some phase coherence will be lost due to inhomogeneity of the external magnetic field, differences in the local magnetic field that cause unique chemical shifts, imperfections in the RF pulse(s) used to create the xy component of  $\mathbf{M}$  and relaxation due to  $R_2$  and  $R_{ex}$ . As phase coherence is lost among spins,  $\mathbf{M}_{xy}$  decays at a monoexponential rate (Figure 1.13 A).  $R_{1\rho}$  relaxation dispersion measures the projection of magnetization along the effective magnetic field,  $\omega_{eff}$ . To gain information about the exchange rates and populations of states A and B, the  $R_{1\rho}$  relaxation dispersion experiment probes the intensity of the observed resonance (Figure 1.10 B) as a function of the two experimental parameters, spin lock power ( $\omega_1$ ) and resonance offset ( $\Omega$ ).  $\omega_1$  is incrementally increased to suppress the extent of chemical exchange, and consequently the contribution of  $R_{ex}$  to  $\mathbf{M}_{xz}$  relaxation. The observed peak intensity of the target peak will therefore decay faster as a result of more chemical exchange at a lower  $\omega_1$  (Figure 1.13 C) as compared to that at higher  $\omega_1$  (Figure 1.13 D). By fitting the peak intensity to a monoexponential decay we directly obtain  $R_{1\rho}$ .  $\omega_1$  and  $\Omega$  are varied and peak intensity is measured for each combination of the two. The lowest  $\omega_1$  used is limited by the minimum amplitude required to suppress unwanted  $^{13}\text{C}$ - $^{13}\text{C}$  scalar couplings and the highest is limited by the capability of the NMR probe hardware to produce hard pulses with high amplitude for an extended period of time.  $\Omega$  is varied from negative to positive values. When  $\Omega = \Omega_B$ ,  $R_{ex}$  is maximum and the greatest sensitivity to the chemical exchange process occurs under these parameters. The  $R_{1\rho}$ , and consequently  $R_2 + R_{ex}$ , profile collection of many ( $\omega_1$ ,  $\Omega$ ) combinations is given by a combination of distinct values describing the unique targeted chemical exchange process (Figure 1.11 E).

While the Bloch-McConnell equations fully describe a two-state exchange process as written in Eq. 1.2, they are not practically implemented as a fitting technique due to the high computational cost associated with doing so. Considerable theory has been developed to approximate the real solution to Eq. 1.2 using a single algebraic expression under a variety of scenarios taking into account  $p_B$  and the relative magnitudes of  $k_{ex}$  and  $|\Delta\omega_{AB}|$ .<sup>132,141,143-147</sup> The Laguerre approximation is the most accurate of these, provided  $p_A$  and  $p_B$  are highly skewed ( $p_B < \sim 30\%$ ),<sup>146</sup> as it accounts for  $p_A$ ,  $p_B$ ,  $k_{ex}$ ,  $\Delta\omega_{AB}$ ,  $R_1$  and  $R_2$  each individually,<sup>146</sup> which is not the case in such other scenarios.<sup>147</sup>

$$R_{1\rho} = \frac{R_1 \cos^2 \theta + R_2 \sin^2 \theta + \sin^2 \theta p_A p_B \Delta\omega^2 k_{ex}}{\omega_A^2 \omega_B^2 / \omega_{eff}^2 + k_{ex}^2 - \sin^2 \theta p_A p_B \Delta\omega^2 \left( 1 + \frac{2k_{ex}^2 (p_A \omega_A^2 + p_B \omega_B^2)}{\omega_a^2 \omega_b^2 + \omega_a^2 k_{ex}^2} \right)} \quad (1.17)$$

where  $\omega_{eff}^2 = \Omega^2 + \omega_1^2$ ,  $\omega_A^2 = (\Omega_A - \omega_{rf})^2 + \omega_1^2$  and  $\omega_B^2 = (\Omega_B - \omega_{rf})^2 + \omega_1^2$ .  $\theta = \arctan(\omega_1 / \Omega)$  where  $\omega_1$  is the spin lock power strength.  $\Omega = \Omega_{obs} - \omega_{rf}$  where  $\Omega$  is the offset of the spin lock carrier frequency ( $\omega_{rf}$ ) from the observed resonance frequency ( $\Omega_{obs}$ ).  $\Omega_{obs} = p_A \Omega_A + p_B \Omega_B$  where  $p_A$  and  $p_B$  are the ground and transient state populations and  $p_A + p_B = 1$ .  $k_{ex}$  is the chemical exchange rate constant for a two-state exchange process where  $k_{ex} = k_A + k_B$ ,  $k_A = k_{ex} p_B$  and  $k_B = k_{ex} p_A$ . Plots of  $R_{1\rho}$  data are presented as  $R_{2eff}$  ( $R_{2eff} = R_2 + R_{ex}$ ), where

$$R_{2,eff} = \frac{R_{1\rho}}{\sin^2 \theta} - \frac{R_1}{\tan^2 \theta}. \quad (1.4)$$



**Figure 1.13 | Concepts of  $R_{1\rho}$  Relaxation Dispersion.** (A) Representation of the bulk magnetization in the transverse ( $xy$ ) plane,  $\mathbf{M}_{xy}$ , before free precession begins (left) followed by precession of  $\mathbf{M}_{xy}$  (large arrow). Representative individual spins are shown with small arrows. Each subsequent plot is after a period  $\Delta t$ , representing one full revolution of  $\mathbf{M}_{xy}$ . Small arrows depict loss of coherence of individual spins, contributing to the decay of  $\mathbf{M}_{xy}$ . (B) Vector diagram representing the effective field,  $\omega_{\text{eff}}$ , along which  $R_{1\rho}$  is projected. (C-E) The decay of the bulk magnetization component along  $R_{1\rho}$  is measured as a function of spin lock power (SLP),  $\omega_1$ , and offset,  $\Omega$ . (C-D) Vector diagrams depicting chemical exchange between  $\omega_A$  and  $\omega_B$  during an  $R_{1\rho}$  experiment. Blurred black arrows depict individual spins that are stochastically exchanging from  $\omega_A$  to  $\omega_B$  and *vice versa*. Examples are shown at (C) low and (D) high spin lock power. (E) Theoretical plots of  $R_2 + R_{\text{ex}}$  calculated from  $R_{1\rho}$  using Eq. 1.4. (Left) On- and (Right) off-resonance plots. In practice the magnitude of  $\omega_B$  is much less than that of  $\omega_A$  for the systems studied in this thesis.

## 1.5 Statistics Used for Model Selection

A set of data must often be fit to a model to gain valuable information for the interpretation of the event causing the data. The type of model chosen will dictate the potential interpretations, so comparing how well each model fits is of utmost importance. Moreover, a model with more variables (the full model) will always fit a set of data better than the restricted model. To assess if the full model fits significantly better, thus meriting the imposition of an increasingly complex model onto the data, statistical tests have been developed. Here, we use two such tests, the F-test and Akaike Information Criterion (AIC) test.

The F-test<sup>148</sup> compares two nested models fit to the same data. Two models are nested if the restricted model can be obtained from setting one or more variables in the full model equal to zero. The F-Test operates under the null hypothesis that the residual sum of squares of the restricted model is not significantly larger than that of the full model, where

$$F = \frac{\Delta SS / \Delta DF}{SS_F / DF_F} \quad (1.18)$$

$\Delta SS$  is the difference in the residual sum of squares of the full model as compared to the restricted model,  $\Delta DF$  is the difference in the number of degrees of freedom of the full model as compared to the restricted model,  $SS_F$  is the residual sum of square of the full model, and  $DF_F$  is the number of degrees of freedom of the full model. The F value that is calculated from Eq. 1.18 is then compared to standardized F values taking into account the number of degrees of freedom in the restricted and full models with a particular confidence level (here, 0.95). If the calculated F value is greater than the tabulated F



value, the null hypothesis fails and the residual sum of squares of the restricted model are significantly larger than that of the full model. The  $p$  value can then be calculated for a given F-test. A  $p$  value less than the significance level (here, 0.05) means the null hypothesis is rejected. If the null hypothesis is rejected, the full model describes the process that gives rise to the data significantly better than the restricted model.

The AIC test<sup>149</sup> minimizes the amount of information lost when describing the process that generates the data by using a likelihood function, where

$$AIC = N \ln(L) + 2K . \quad (1.19)$$

$N$  is the number of trials,  $K$  is number of free parameters and  $L$  is the likelihood function, where

$$L(\theta|x) = P(X = x; \theta) , \quad (1.20)$$

or the likelihood of a parameter  $\theta$  given the outcome  $x$  of  $X$ . The likelihood ranges from 0 to 1. Because the natural log is taken of the likelihood, a smaller AIC value means the corresponding model is more likely to describe the act that gives rise to the data.

## 1.6 References

- 1 Dahm, R. *Hum Genet* Discovering DNA: Friedrich Miescher and the early years of nucleic acid research **122**, 565-581, (2008).
- 2 Avery, O. T., MacLeod, C. M. & McCarty, M. *J Exp Med* Studies on the chemical nature of the substance inducing transformation of pneumococcal types **79**, 137-158, (1944).
- 3 Hershey, A. D. & Chase, M. *Cold Spring Harb Symp Quant Biol* Genetic recombination and heterozygosis in bacteriophage **16**, 471-479, (1951).
- 4 Hershey, A. D. & Chase, M. *J Gen Physiol* Independent functions of viral protein and nucleic acid in growth of bacteriophage **36**, 39-56, (1952).
- 5 Hershey, A. D., Dixon, J. & Chase, M. *J Gen Physiol* Nucleic acid economy in bacteria infected with bacteriophage T2. I. Purine and pyrimidine composition **36**, 777-789, (1953).

- 6 Watson, J. D. & Crick, F. H. *Nature* Molecular structure of nucleic acids; a structure for deoxyribonucleic acid **171**, 737-738, (1953).
- 7 Watson, J. D. & Crick, F. H. *Nature* Genetical implications of the structure of deoxyribonucleic acid **171**, 964-967, (1953).
- 8 Franklin, R. E. & Gosling, R. G. *Nature* Molecular configuration in sodium thymonucleate **171**, 740-741, (1953).
- 9 Franklin, R. E. & Gosling, R. G. *Nature* Evidence for 2-chain helix in crystalline structure of sodium deoxyribonucleate **172**, 156-157, (1953).
- 10 Hoogsteen, K. *Acta Crystallographica* The structure of crystals containing a hydrogen-bonded complex of 1-methylthymine and 9-methyladenine **12**, 822-823, (1959).
- 11 Hoogsteen, K. *Acta Crystallographica Section A, Foundations of crystallography* The crystal and molecular structure of a hydrogen-bonded complex between 1-methylthymine and 9-methyladenine **16**, 907, (1963).
- 12 Felsenfeld, G. & Rich, A. *Biochim Biophys Acta* Studies on the formation of two- and three-stranded polyribonucleotides **26**, 457-468, (1957).
- 13 Sobell, H. M., Tomita, K. I. & Rich, A. *Proc Natl Acad Sci USA* The crystal structure of an intermolecular complex containing a guanine and a cytosine derivative **49**, 885-892, (1963).
- 14 Corey, R. B. & Pauling, L. *Arch Biochem Biophys* Specific hydrogen-bond formation between pyrimidines and purines in deoxyribonucleic acids **65**, 164-181, (1956).
- 15 Haschemeyer, A. E. & Sobell, H. M. *Acta Crystallogr* The crystal structure of a hydrogen bonded complex of deoxyguanosine and 5-bromodeoxycytidine **19**, 125-130, (1965).
- 16 Haschemeyer, A. E. & Sobell, H. M. *Proc Natl Acad Sci USA* The crystal structure of an intermolecular nucleoside complex: adenosine and 5-bromouridine **50**, 872-877, (1963).
- 17 Mathews, F. S. & Rich, A. *J Mol Biol* The molecular structure of a hydrogen bond complex of N-ethyl adenine and N-methyl uracil **8**, 89-95, (1964).
- 18 Courtois, Y., Fromageot, P. & Guschlbauer, W. *Eur J Biochem* Protonated polynucleotide structures. An optical rotatory dispersion study of the protonation of DNA **6**, 493-501, (1968).
- 19 Quigley, G. J. *et al. Science* Non-Watson-Crick G.C and A.T base pairs in a DNA-antibiotic complex **232**, 1255-1258, (1986).
- 20 Kearns, D. R., Patel, D. J. & Shulman, R. G. *Nature* High resolution nuclear magnetic resonance studies of hydrogen bonded protons of tRNA in water **229**, 338-339, (1971).
- 21 Cross, A. D. & Crothers, D. M. *Biochemistry* A proton magnetic resonance study of single-stranded and double-helical deoxyribooligonucleotides **10**, 4015-4023, (1971).
- 22 Wong, Y. P., Kearns, D. R., Reid, B. R. & Shulman, R. G. *J Mol Biol* Investigation of exchangeable protons and the extent of base pairings in yeast phenylalanine transfer RNA by high resolution nuclear magnetic resonance **72**, 725-740, (1972).

- 23 Shulman, R. G. *et al. Proc Natl Acad Sci USA* Determination of secondary and tertiary structural features of transfer RNA molecules in solution by nuclear magnetic resonance **70**, 2042-2045, (1973).
- 24 Crothers, D. M., Hilbers, C. W. & Shulman, R. G. *Proc Natl Acad Sci USA* Nuclear magnetic resonance study of hydrogen-bonded ring protons in Watson-Crick base pairs **70**, 2899-2901, (1973).
- 25 Patel, D. J. & Tonelli, A. E. *Biopolymers* Assignment of the proton Nmr chemical shifts of the T-N3H and G-N1H proton resonances in isolated AT and GC Watson-Crick base pairs in double-stranded deoxy oligonucleotides in aqueous solution **13**, 1943-1964, (1974).
- 26 Kallenbach, N. R., Daniel, W. E., Jr. & Kaminker, M. A. *Biochemistry* Nuclear magnetic resonance study of hydrogen-bonded ring protons in oligonucleotide helices involving classical and nonclassical base pairs **15**, 1218-1224, (1976).
- 27 Arentzen, R., Van Boeckel, C., van der Marel, G. A. & van Boom, J. H. *Synthesis*, 137-139, (1979).
- 28 de Rooij, J. F., Wille-Hazeleger, G., Burgers, P. M. & van Boom, J. H. *Nucleic Acids Res* Neighbouring group participation in the unblocking of phosphotriesters of nucleic acids **6**, 2237-2259, (1979).
- 29 Wang, A. H.-J. *et al. Nature* Molecular structure of a left-handed double helical DNA fragment at atomic resolution **282**, 680-686, (1979).
- 30 Wing, R. *et al. Nature* Crystal structure analysis of a complete turn of B-DNA **287**, 755-758, (1980).
- 31 Frank-Kamenetskii, M. D. *Artif DNA PNA XNA* DNA breathes Hoogsteen **2**, 1-3, (2011).
- 32 Nikolova, E. N. *et al. Nature* Transient Hoogsteen base pairs in canonical duplex DNA **470**, 498-502, (2011).
- 33 Pohl, F. M. & Jovin, T. M. *J Mol Biol* Salt-induced Co-operative Conformational Change of a Synthetic DNA: Equilibrium and Kinetic Studies with Poly(dG-dC) **67**, 375-396, (1972).
- 34 Wang, A. H.-J., Quigley, G. J. & Rich, A. *Nucleic Acids Research* Atomic resolution analysis of a 2:1 complex of CpG and acridine orange **6**, 3879-3890, (1979).
- 35 Crawford, J. L. *et al. Proc Natl Acad Sci USA* The tetramer d(CpGpCpG) crystallizes as a left-handed double helix **77**, 4016-4020, (1980).
- 36 Gupta, G., Bansal, M. & Sasisekharan, V. *Proc Natl Acad Sci USA* Conformational flexibility of DNA: polymorphism and handedness **77**, 6486-6490, (1980).
- 37 Jayaraman, S. & Yathindra, N. *Biochem Biophys Res Commun* Theoretical analysis of conformational possibilities of Z-DNA **97**, 1407-1419, (1980).
- 38 Patel, D. J., Canel, L. L. & Pohl, F. M. *Proc Nat Acad Sci USA* "Alternating B-DNA" conformation for the oligo (dG-dC) duplex in high-salt solution **76**, 2508-2511, (1979).
- 39 Kuhnlein, U., Tsang, S. S. & Edwards, J. *Nature* Cooperative structural transition of PM2 DNA at high ionic strength and its dependence on DNA damages **287**, 363-364, (1980).

- 40 Behe, M. & Felsenfeld, G. *Proc Natl Acad Sci USA* Effects of methylation on a  
synthetic polynucleotide: the B-Z transition in poly(dG-m5dC).poly(dG-m5dC)  
**78**, 1619-1623, (1981).
- 41 Drew, H., Takano, T., Tanaka, S., Itakura, K. & Dickerson, R. E. *Nature* High-  
salt d(CpGpCpG), a left-handed Z' DNA double helix **286**, 567-573, (1980).
- 42 Arnott, S. & Hukins, D. W. *Biochem Biophys Res Commun* Optimised parameters  
for A-DNA and B-DNA **47**, 1504-1509, (1972).
- 43 Brzezinski, K. *et al. Nucleic Acids Res* High regularity of Z-DNA revealed by  
ultra high-resolution crystal structure at 0.55 Å **39**, 6238-6248, (2011).
- 44 Fenn, T. D. *et al. Structure* Reintroducing electrostatics into macromolecular  
crystallographic refinement: application to neutron crystallography and DNA  
hydration **19**, 523-533, (2011).
- 45 Dauter, Z. & Adamiak, D. A. *Acta Crystallogr D Biol Crystallogr* Anomalous  
signal of phosphorus used for phasing DNA oligomer: importance of data  
redundancy **57**, 990-995, (2001).
- 46 Egli, M., Williams, L. D., Gao, Q. & Rich, A. *Biochemistry* Structure of the pure-  
spermine form of Z-DNA (magnesium free) at 1-Å resolution **30**, 11388-11402,  
(1991).
- 47 Gessner, R. V., Frederick, C. A., Quigley, G. J., Rich, A. & Wang, A. H. *J Biol  
Chem* The molecular structure of the left-handed Z-DNA double helix at 1.0-Å  
atomic resolution. Geometry, conformation, and ionic interactions of  
d(CGCGCG) **264**, 7921-7935, (1989).
- 48 Dickerson, R. E. *et al. Science* The anatomy of A-, B-, and Z-DNA **216**, 475-485,  
(1982).
- 49 Nordheim, A. & Rich, A. *Proc Natl Acad Sci USA* The sequence (dC-dA)<sub>n</sub>(dG-  
dT)<sub>n</sub> forms left-handed Z-DNA in negatively supercoiled plasmids **80**, 1821-  
1825, (1983).
- 50 Suggs, J. W. & Wagner, R. W. *Nucleic Acids Res* Nuclease recognition of an  
alternating structure in a d(AT)<sub>14</sub> plasmid insert **14**, 3703-3716, (1986).
- 51 Ikuta, S. & Wang, Y. S. *Nucleic Acids Res* Conformation and dynamics of Z-  
DNA oligomer duplex of d[(CG)<sub>3</sub>TATA(CG)<sub>3</sub>] in solution **17**, 4131-4144,  
(1989).
- 52 Feigon, J., Wang, A. H., van der Marel, G. A., van Boom, J. H. & Rich, A.  
*Science* Z-DNA forms without an alternating purine-pyrimidine sequence in  
solution **230**, 82-84, (1985).
- 53 Benevides, J. M., Wang, A. H., van der Marel, G. A., van Boom, J. H. & Thomas,  
G. J., Jr. *Biochemistry* Effect of the G.T mismatch on backbone and sugar  
conformations of Z-DNA and B-DNA: analysis by Raman spectroscopy of crystal  
and solution structures of d(CGCGTG) and d(CGCGCG) **28**, 304-310, (1989).
- 54 Brown, T., Kneale, G., Hunter, W. N. & Kennard, O. *Nucleic Acids Res* Structural  
characterisation of the bromouracil.guanine base pair mismatch in a Z-DNA  
fragment **14**, 1801-1809, (1986).
- 55 Orbons, L. P., van der Marel, G. A., van Boom, J. H. & Altona, C. *J Biomol  
Struct Dyn* An NMR study of the polymorphous behavior of the mismatched  
DNA octamer d(m5C-G-m5C-G-T-G-m5C-G) in solution. The B, Z, and hairpin  
forms **4**, 939-963, (1987).

- 56 Otokiti, E. O. & Sheardy, R. D. *Biochemistry* Effect of base pair A/C and G/T mismatches on the thermal stabilities of DNA oligomers that form B-Z junctions **36**, 11419-11427, (1997).
- 57 Yang, X. L. & Wang, A. H. *Biochemistry* Structural analysis of Z-Z DNA junctions with A:A and T:T mismatched base pairs by NMR **36**, 4258-4267, (1997).
- 58 Orbons, L. P., van der Marel, G. A., van Boom, J. H. & Altona, C. *Eur J Biochem* An NMR study of polymorphous behaviour of the mismatched DNA octamer d(m5C-G-m5C-G-A-G-m5C-G) in solution. The B-duplex and hairpin forms **170**, 225-239, (1987).
- 59 Marrot, L., Hebert, E., Saint-Ruf, G. & Leng, M. *Nucleic Acids Res* Comparison of the reactivity of B-DNA and Z-DNA with two isosteric chemical carcinogens: 2-N,N-acetoxyacetylaminofluorene and 3-N,N-acetoxyacetyl-amino-4,6-dimethyldipyrido-[1,2-a:3',2'-d] imidazole **15**, 5629-5642, (1987).
- 60 Vogt, N., Marrot, L., Rousseau, N., Malfoy, B. & Leng, M. *J Mol Biol* Chloroacetaldehyde reacts with Z-DNA **201**, 773-776, (1988).
- 61 Castelijns, M. M. C. F., van Aken, D., Schipper, P., van Lier, J. J. C. & Buck, H. M. *Recueil des Travaux Chimiques des Pays-Bas* Salt-induced conversions of P(V) into P(IV) in solvents of different polarity. A tentative model for the initiation of conformational changes in DNA via the P(IV) residues **99**, 380-383, (1980).
- 62 Shafer, R. H., Brown, S. C., Delbarre, A. & Wade, D. *Nucleic Acids Res* Binding of ethidium and bis(methidium)spermine to Z DNA by intercalation **12**, 4679-4690, (1984).
- 63 Basu, H. S., Shafer, R. H. & Marton, L. J. *Nucleic Acids Res* A stopped-flow H-D exchange kinetic study of spermine-polynucleotide interactions **15**, 5873-5886, (1987).
- 64 Banville, D. L., Feuerstein, B. G. & Shafer, R. H. *J Mol Biol* <sup>1</sup>H and <sup>31</sup>P nuclear magnetic resonance studies of spermine binding to the Z-DNA form of d(m5CGm5CGm5CG)<sub>2</sub>. Evidence for decreased spermine mobility **219**, 585-590, (1991).
- 65 Bancroft, D., Williams, L. D., Rich, A. & Egli, M. *Biochemistry* The low-temperature crystal structure of the pure-spermine form of Z-DNA reveals binding of a spermine molecule in the minor groove **33**, 1073-1086, (1994).
- 66 Howell, M. L., Schroth, G. P. & Ho, P. S. *Biochemistry* Sequence-dependent effects of spermine on the thermodynamics of the B-DNA to Z-DNA transition **35**, 15373-15382, (1996).
- 67 Basu, H. S., Feuerstein, B. G., Zarling, D. A., Shafer, R. H. & Marton, L. J. *J Biomol Struct Dyn* Recognition of Z-RNA and Z-DNA determinants by polyamines in solution: experimental and theoretical studies **6**, 299-309, (1988).
- 68 Thomas, T. J., Meryhew, N. L. & Messner, R. P. *Arthritis Rheum* Enhanced binding of lupus sera to the polyamine-induced left-handed Z-DNA form of polynucleotides **33**, 356-365, (1990).
- 69 Gunnia, U. B., Thomas, T. & Thomas, T. J. *Immunol Invest* The effects of polyamines on the immunogenicity of polynucleotides **20**, 337-350, (1991).

- 70 Thomas, T. J., Gunnia, U. B. & Thomas, T. *J Biol Chem* Polyamine-induced B-DNA to Z-DNA conformational transition of a plasmid DNA with (dG-dC)n insert **266**, 6137-6141, (1991).
- 71 Thomas, T. J. & Thomas, T. *Biochem J* Polyamine-induced Z-DNA conformation in plasmids containing (dA-dC)n.(dG-dT)n inserts and increased binding of lupus autoantibodies to the Z-DNA form of plasmids **298 (Pt 2)**, 485-491, (1994).
- 72 Moller, A., Nordheim, A., Nichols, S. R. & Rich, A. *Proc Natl Acad Sci USA* 7-Methylguanine in poly(dG-dC).poly(dG-dC) facilitates z-DNA formation. **78**, 4777-4781, (1981).
- 73 Santella, R. M., Grunberger, D., Nordheim, A. & Rich, A. *Biochem Biophys Res Commun* N-2-Acetylaminofluorene modification of poly(dG-m5dC).poly(dG-m5dC) induces the Z-DNA conformation **106**, 1226-1232, (1982).
- 74 Bailleul, B., Galiegue-Zouitina, S. & Loucheux-Lefebvre, M. H. *Nucleic Acids Res* Conformations of poly(dG-dC).poly(dG-dC) modified by the O-acetyl derivative of the carcinogen 4-hydroxyaminoquinoline 1-oxide **12**, 7915-7927, (1984).
- 75 Bourtayre, P. *et al. IARC Sci Publ* Z-form induction in DNA by carcinogenic nickel compounds: an optical spectroscopy study, 227-234, (1984).
- 76 Bichara, M. & Fuchs, R. P. *J Mol Biol* DNA binding and mutation spectra of the carcinogen N-2-aminofluorene in Escherichia coli. A correlation between the conformation of the premutagenic lesion and the mutation specificity **183**, 341-351, (1985).
- 77 Grunberger, D., Santella, R. M., Hanau, L. H. & Erlanger, B. F. *Carcinog Compr Surv* Stabilization of Z-DNA conformation by chemical carcinogens **10**, 465-480, (1985).
- 78 Durlach, J., Bara, M., Guiet-Bara, A. & Collery, P. *Anticancer Res* Relationship between magnesium, cancer and carcinogenic or anticancer metals **6**, 1353-1361, (1986).
- 79 Hingerty, B. E. & Broyde, S. *J Biomol Struct Dyn* Energy minimized structures of carcinogen-DNA adducts: 2-acetylaminofluorene and 2-aminofluorene **4**, 365-372, (1986).
- 80 Xu, Y., Ikeda, R. & Sugiyama, H. *J Am Chem Soc* 8-Methylguanosine: a powerful Z-DNA stabilizer **125**, 13519-13524, (2003).
- 81 Heavner, S. & Gannett, P. M. *J Biomol Struct Dyn* Molecular dynamics and free energy calculations of the B and Z forms of C8-arylguanine modified oligonucleotides **23**, 203-220, (2005).
- 82 Rio, P. & Leng, M. *Nucleic Acids Res* Preferential binding of the chemical carcinogen N-hydroxy-2-aminofluorene to B-DNA as compared to Z-DNA **11**, 4947-4956, (1983).
- 83 Klysik, J., Stirdivant, S. M., Larson, J. E., Hart, P. A. & Wells, R. D. *Nature* Left-handed DNA in restriction fragments and a recombinant plasmid **290**, 672-677, (1981).
- 84 Nordheim, A. *et al. Cell* Negatively supercoiled plasmids contain left-handed Z-DNA segments as detected by specific antibody binding **31**, 309-318, (1982).

- 85 Singleton, C. K., Klysik, J., Stirdivant, S. M. & Wells, R. D. *Nature* Left-handed Z-DNA is induced by supercoiling in physiological ionic conditions **299**, 312-316, (1982).
- 86 Azorin, F., Nordheim, A. & Rich, A. *EMBO J* Formation of Z-DNA in negatively supercoiled plasmids is sensitive to small changes in salt concentration within the physiological range **2**, 649-655, (1983).
- 87 Di Capua, E. *et al.* *EMBO J* Torsional stress induces left-handed helical stretches in DNA of natural base sequence: circular dichroism and antibody binding **2**, 1531-1535, (1983).
- 88 Stockton, J. F. *et al.* *EMBO J* Left-handed Z-DNA regions are present in negatively supercoiled bacteriophage PM2 DNA **2**, 2123-2128, (1983).
- 89 Lafer, E. M., Moller, A., Nordheim, A., Stollar, B. D. & Rich, A. *Proc Natl Acad Sci USA* Antibodies specific for left-handed Z-DNA **78**, 3546-3550, (1981).
- 90 Malfoy, B. & Leng, M. *FEBS Lett* Antiserum to Z-DNA **132**, 45-48, (1981).
- 91 Thomas, T. J., Meryhew, N. L. & Messner, R. P. *Arthritis Rheum* DNA sequence and conformation specificity of lupus autoantibodies. Preferential binding to the left-handed Z-DNA form of synthetic polynucleotides **31**, 367-377, (1988).
- 92 Nordheim, A. *et al.* *Nature* Antibodies to left-handed Z-DNA bind to interband regions of Drosophila polytene chromosomes **294**, 417-422, (1981).
- 93 Arndt-Jovin, D. J. *et al.* *Proc Natl Acad Sci USA* Left-handed Z-DNA in bands of acid-fixed polytene chromosomes **80**, 4344-4348, (1983).
- 94 Pardue, M. L., Nordheim, A., Lafer, E. M., Stollar, B. D. & Rich, A. *Cold Spring Harb Symp Quant Biol* Z-DNA and the polytene chromosome **47 Pt 1**, 171-176, (1983).
- 95 Coll, M. *et al.* *Nucleic Acids Res* Effects of 5-fluorouracil/guanine wobble base pairs in Z-DNA: molecular and crystal structure of d(CGCGFG) **17**, 911-923, (1989).
- 96 Hertzberg, R. P., Caranfa, M. J. & Hecht, S. M. *Biochemistry* Degradation of structurally modified DNAs by bleomycin group antibiotics **27**, 3164-3174, (1988).
- 97 Tashiro, R. & Sugiyama, H. *Nucleic Acids Res Suppl* Photochemistry of bromouracil containing Z-DNA, 69-70, (2003).
- 98 Viegas-Pequignot, E., Malfoy, B., Sabatier, L. & Dutrillaux, B. *Hum Genet* Different reactivity of Z-DNA antibodies with human chromosomes modified by actinomycin D and 5-bromodeoxyuridine **75**, 114-119, (1987).
- 99 Vorlickova, M., Chladkova, J. & Kypr, J. *Nucleic Acids Res* Conformational transitions of poly(dA-bromo5dU) and poly(dA-iodo5dU) in solution **20**, 1109-1112, (1992).
- 100 Liu, L. F. & Wang, J. C. *Proc Natl Acad Sci USA* Supercoiling of the DNA template during transcription **84**, 7024-7027, (1987).
- 101 Wittig, B., Dorbic, T. & Rich, A. *J Cell Biol* The level of Z-DNA in metabolically active, permeabilized mammalian cell nuclei is regulated by torsional strain **108**, 755-764, (1989).
- 102 Wittig, B., Dorbic, T. & Rich, A. *Proc Natl Acad Sci USA* Transcription is associated with Z-DNA formation in metabolically active permeabilized mammalian cell nuclei **88**, 2259-2263, (1991).

- 103 Wolfl, S., Wittig, B. & Rich, A. *Biochim Biophys Acta* Identification of transcriptionally induced Z-DNA segments in the human c-myc gene **1264**, 294-302, (1995).
- 104 Kouzine, F., Liu, J., Sanford, S., Chung, H. J. & Levens, D. *Nat Struct Mol Biol* The dynamic response of upstream DNA to transcription-generated torsional stress **11**, 1092-1100, (2004).
- 105 Kouzine, F., Sanford, S., Elisha-Feil, Z. & Levens, D. *Nat Struct Mol Biol* The functional response of upstream DNA to dynamic supercoiling in vivo **15**, 146-154, (2008).
- 106 Bayele, H. K. *et al. Blood* HIF-1 regulates heritable variation and allele expression phenotypes of the macrophage immune response gene SLC11A1 from a Z-DNA forming microsatellite **110**, 3039-3048, (2007).
- 107 Rothenburg, S., Koch-Nolte, F. & Haag, F. *Immunol Rev* DNA methylation and Z-DNA formation as mediators of quantitative differences in the expression of alleles **184**, 286-298, (2001).
- 108 Searle, S. & Blackwell, J. M. *J Med Genet* Evidence for a functional repeat polymorphism in the promoter of the human NRAMP1 gene that correlates with autoimmune versus infectious disease susceptibility **36**, 295-299, (1999).
- 109 Ray, B. K., Dhar, S., Shakya, A. & Ray, A. *Proc Natl Acad Sci USA* Z-DNA-forming silencer in the first exon regulates human ADAM-12 gene expression **108**, 103-108, (2011).
- 110 Taka, S., Gazouli, M., Politis, P. K., Pappa, K. I. & Anagnou, N. P. *Mol Biol Rep* Transcription factor ATF-3 regulates allele variation phenotypes of the human SLC11A1 gene **40**, 2263-2271, (2013).
- 111 Watson, F., Deavall, D. G., Macro, J. A., Kiernan, R. & Dimaline, R. *Biochem J* Transcriptional activation of vesicular monoamine transporter 2 in the pre-B cell line Ea3.123 **337 ( Pt 2)**, 193-199, (1999).
- 112 Lenert, P. *Clin Exp Immunol* Nucleic acid sensing receptors in systemic lupus erythematosus: development of novel DNA- and/or RNA-like analogues for treating lupus **161**, 208-222, (2010).
- 113 Ray, B. K., Dhar, S., Henry, C., Rich, A. & Ray, A. *Cancer Res* Epigenetic regulation by Z-DNA silencer function controls cancer-associated ADAM-12 expression in breast cancer: cross-talk between MeCP2 and NF1 transcription factor family **73**, 736-744, (2013).
- 114 Mulholland, N., Xu, Y., Sugiyama, H. & Zhao, K. *Cell Biosci* SWI/SNF-mediated chromatin remodeling induces Z-DNA formation on a nucleosome **2, 3**, (2012).
- 115 Herbert, A., Lowenhaupt, K., Spitzner, J. & Rich, A. *Proc Natl Acad Sci USA* Chicken double-stranded RNA adenosine deaminase has apparent specificity for Z-DNA **92**, 7550-7554, (1995).
- 116 Herbert, A. *et al. Proc Natl Acad Sci USA* A Z-DNA binding domain present in the human editing enzyme, double-stranded RNA adenosine deaminase **94**, 8421-8426, (1997).
- 117 Liu, Y., Herbert, A., Rich, A. & Samuel, C. E. *Methods* Double-stranded RNA-specific adenosine deaminase: nucleic acid binding properties **15**, 199-205, (1998).



- 118 Kim, Y. G. *et al. J Biol Chem* The Zab domain of the human RNA editing enzyme ADAR1 recognizes Z-DNA when surrounded by B-DNA **275**, 26828-26833, (2000).
- 119 Herbert, A. *et al. Nucleic Acids Res* The Zalpha domain from human ADAR1 binds to the Z-DNA conformer of many different sequences **26**, 3486-3493, (1998).
- 120 Seo, Y. J. *et al. FEBS Lett* Sequence discrimination of the Zalpha domain of human ADAR1 during B-Z transition of DNA duplexes **584**, 4344-4350, (2010).
- 121 Kim, Y. G. *et al. Proc Natl Acad Sci USA* A role for Z-DNA binding in vaccinia virus pathogenesis **100**, 6974-6979, (2003).
- 122 Wittig, B., Wolfl, S., Dorbic, T., Vahrson, W. & Rich, A. *EMBO J* Transcription of human c-myc in permeabilized nuclei is associated with formation of Z-DNA in three discrete regions of the gene **11**, 4653-4663, (1992).
- 123 Hegde, A. H. & Seetharamappa, J. *Mol Biol Rep* Fluorescence and circular dichroism studies on binding and conformational aspects of an anti-leukemic drug with DNA **41**, 67-71, (2014).
- 124 Li, Z. *et al. J Am Chem Soc* Reversible plasmonic circular dichroism of Au nanorod and DNA assemblies **134**, 3322-3325, (2012).
- 125 Liu, Z. J., Si, Y. K. & Chen, X. G. *Yao Xue Xue Bao* Application of circular dichroism to the study of interactions between small molecular compounds and DNA **45**, 1478-1484, (2010).
- 126 Uemura, M., Yoshikawa, Y., Chikuma, M. & Komeda, S. *Metallomics* A circular dichroism study uncovers a two-step interaction of antitumor azolato-bridged dinuclear platinum(II) complexes with calf thymus DNA **4**, 641-644, (2012).
- 127 Kypr, J., Kejnovska, I., Renciuik, D. & Vorlickova, M. *Nucleic Acids Res* Circular dichroism and conformational polymorphism of DNA **37**, 1713-1725, (2009).
- 128 Miyahara, T., Nakatsuji, H. & Sugiyama, H. *J Phys Chem A* Helical structure and circular dichroism spectra of DNA: a theoretical study **117**, 42-55, (2013).
- 129 Vorlickova, M., Kejnovska, I., Bednarova, K., Renciuik, D. & Kypr, J. *Chirality* Circular dichroism spectroscopy of DNA: from duplexes to quadruplexes **24**, 691-698, (2012).
- 130 Randazzo, A., Spada, G. P. & da Silva, M. W. *Top Curr Chem* Circular dichroism of quadruplex structures **330**, 67-86, (2013).
- 131 Chang, Y. M., Chen, C. K. & Hou, M. H. *Int J Mol Sci* Conformational Changes in DNA upon Ligand Binding Monitored by Circular Dichroism **13**, 3394-3413, (2012).
- 132 Cavanagh, J., Fairbrother, W. J., Palmer III, A. G., Rance, M. & Skelton, N. J. *Principles and Practice Protein NMR Spectroscopy*. Second edn, (Elsevier Academic Press, 2007).
- 133 Bloch, F. *Physical Review* Nuclear Induction **70**, 460-474, (1946).
- 134 Bloch, F. *Phys Rev* Generalized Theory of Relaxation **105**, 1206-1222, (1956).
- 135 Wijmenga, S. S. & van Buuren, N. M. *Prog NMR Spectrosc* The use of NMR methods for conformational studies of nucleic acids **32**, 287-387, (1998).
- 136 Furtig, B., Richter, C., Wohnert, J. & Schwalbe, H. *Chembiochem* NMR spectroscopy of RNA **4**, 936-962, (2003).

- 137 Zimmer, D. P. & Crothers, D. M. *Proc Nat Acad Sci USA* NMR of enzymatically synthesized uniformly  $^{13}\text{C}^{15}\text{N}$ -labeled DNA oligonucleotides **92**, 3091-3095, (1995).
- 138 Bothe, J. R. *et al. Nat Methods* Characterizing RNA dynamics at atomic resolution using solution-state NMR spectroscopy **8**, 919-931, (2011).
- 139 McConnell, H. M. *J Chem Phys* **28**, 430-431, (1958).
- 140 McConnell, H. M. *The J Chem Phys* Reaction rates by nuclear magnetic resonance **28**, 430-431, (2004).
- 141 Palmer, A. G. & Massi, F. *Chem Rev* Characterization of the dynamics of biomacromolecules using rotating-frame spin relaxation NMR spectroscopy **106**, 1700-1719, (2006).
- 142 Korzhnev, D. M., Skrynnikov, N. R., Millet, O., Torchia, D. A. & Kay, L. E. *J Am Chem Soc* An NMR experiment for the accurate measurement of heteronuclear spin-lock relaxation rates **124**, 10743-10753, (2002).
- 143 Abergel, D. & Palmer, A. G. *Chemphyschem* Approximate solutions of the Bloch-McConnell equations for two-site chemical exchange **5**, 787-793, (2004).
- 144 Trott, O. & Palmer, A. G., 3rd. *J Magn Reson* Theoretical study of R(1rho) rotating-frame and R2 free-precession relaxation in the presence of n-site chemical exchange **170**, 104-112, (2004).
- 145 Abergel, D. & Palmer, A. G., 3rd. *J Phys Chem B* A Markov model for relaxation and exchange in NMR spectroscopy **109**, 4837-4844, (2005).
- 146 Miloushev, V. Z. & Palmer, A. G., 3rd. *J Magn Reson* R1rho relaxation for two-site chemical exchange: General approximations and some exact solutions **117**, 221-227, (2005).
- 147 Davis, D., Perlman, M. & London, R. *J Magn Res, Series B* Direct Measurements of the Dissociation-Rate Constant for Inhibitor-Enzyme Complexes via the T1 rho and T2 (CPMG) methods **104**, 266-275, (1994).
- 148 Lomax, R. G. & Hahs-Vaugh, D. L. *One-Factor Analysis of Variance: Fixed-Effect Model*, (Taylor & Francis Group, 2007).
- 149 Akaike, H. *Automatic Control, IEEE Transactions on* A new look at the statistical model identification **19**, 716-723, (1974).

## Chapter 2

### **Transient Hoogsteen Base-Pairs Occur Robustly in Canonical Duplex DNA Across Diverse Sequence and Positional Contexts *via* a Late Transition State\***

#### **2.1 Introduction**

We recently showed<sup>1-3</sup> using Nuclear Magnetic Resonance (NMR)  $R_{1\rho}$  relaxation dispersion techniques<sup>4-7</sup> that A•T and G•C Watson-Crick (WC) base pairs in CA/TG and TA/TA steps of duplex DNA can transiently form Hoogsteen (HG) base-pairs (base-pair)<sup>1,8,9</sup> (Figure 1A) with populations in the range of 0.14–0.49 % and lifetimes of 0.3–2.5 ms at pH ~6. This, together with previous studies examining the occurrence<sup>10-12</sup> and energetics<sup>11,13-15</sup> of HG base-pair in DNA duplexes and their involvement in DNA-protein recognition,<sup>16,17</sup> DNA replication,<sup>18,19</sup> and DNA damage repair,<sup>20-22</sup> raise the possibility that HG base-pairs may exist in much greater abundance in the genome where they can carry out functions inaccessible to Watson-Crick base-pairs. Here, we use <sup>13</sup>C and <sup>15</sup>N NMR  $R_{1\rho}$  relaxation dispersion to more broadly examine the occurrence of transient HG base-pair across diverse sequence and positional contexts in canonical duplex DNA.<sup>23</sup>

$R_{1\rho}$  relaxation dispersion NMR experiments offer a unique opportunity to characterize conformations of biomolecules that exist in low abundance and for short

---

\*HSA, FLG, and HMA conceived the idea. HSA prepared ZJXN and CG<sub>3</sub>. FLG and ENN prepared A<sub>5</sub> and CA<sub>3</sub>. ENN prepared A<sub>2</sub>, A<sub>4</sub>, A<sub>6</sub> and E. HAS measured NMR data on ZJXN and CG<sub>3</sub>. FLG measured NMR data on A<sub>2</sub>, A<sub>4</sub>, A<sub>5</sub> and A<sub>6</sub>. ENN measured NMR data on A<sub>2</sub>, A<sub>4</sub>, A<sub>6</sub> and E. HSA carried out the data analysis with help from FLG, ENN, and HMA.

period of times.<sup>1,2,6,24-37</sup> Transitions between ground states (GS) and higher energy transient states of biomolecules with populations ( $p_B$ ) > 0.05 % and lifetimes ( $\tau_B$ ) between ~0.1 and ~10 ms can result in exchange broadening of NMR resonances. In the  $R_{1\rho}$  relaxation dispersion experiment, this chemical exchange contribution ( $R_{ex}$ ) to the observed rotating frame relaxation rate constant ( $R_{1\rho}$ ) is modulated by varying the offset frequency ( $\Omega$ ) and power ( $\omega_1$ ) of an applied radiofrequency (RF) spin lock field. The  $R_{1\rho}$  relaxation dispersion data measured as a function of  $\Omega$  and  $\omega_1$  can be fit using established algebraic expressions describing 2-state chemical exchange to extract the chemical shift ( $\omega_B$ ), population ( $p_B$ ) and lifetime ( $\tau_B$ ) of the transient state.<sup>5,28,35</sup> We have adapted low spin lock field  $R_{1\rho}$  relaxation dispersion experiments developed for protein applications<sup>4,5,38,39</sup> to allow application of  $^{13}\text{C}$ ,<sup>1,6</sup> and more recently  $^{15}\text{N}$ ,<sup>2</sup>  $R_{1\rho}$  relaxation dispersion experiments in studies of nucleic acid conformational exchange with broad exchange rate ( $k_{ex}$ ) timescale sensitivity between ~60 s<sup>-1</sup> and ~100,000 s<sup>-1</sup>.<sup>2,6</sup> Using these experiments, we broadly probed for the occurrence of transient HG base-pair in 22 A•T and 16 G•C base-pairs in 8 DNA duplexes ranging in size from 6 to 15 base-pairs. These constructs encompass a variety of sequence motifs including (A•T)<sub>n</sub> repeats of varying length (n=2, 4, 5 and 6), a (CA)<sub>3</sub> repeat, a duplex sequence that forms HG base-pair upon binding to the drug echinomycin,<sup>40,41</sup> a (CG)<sub>3</sub> repeat capable of forming Z-DNA,<sup>42</sup> and a B/Z junction forming sequence.<sup>26,43</sup>

## **2.2 Methods**

### 2.2.1 NMR Samples and Resonance Assignments

All DNA samples were prepared as previously described<sup>1,2</sup> in NMR buffer containing 15 mM Sodium phosphate pH 5.2 (A<sub>5</sub>), 5.4 (A<sub>2</sub>, A<sub>4</sub>, A<sub>6</sub>, CA<sub>3</sub> and CG<sub>3</sub>), 6.8 (E) or 7.5 (ZJXN), 25 mM NaCl, 25 mM NaCl, 0.1 mM EDTA, 10 % D<sub>2</sub>O. CG<sub>3</sub> and E are unlabeled DNA samples, the T-rich strand of A<sub>5</sub> is <sup>13</sup>C/<sup>15</sup>N labeled while A<sub>6</sub>, A<sub>4</sub>, A<sub>2</sub> CA<sub>3</sub> and ZJXN are fully <sup>13</sup>C/<sup>15</sup>N-labeled.

Unlabeled DNA samples were purchased from Integrated DNA Technologies, Inc. (IDT, Inc.) and prepared as described previously<sup>1</sup> at 2-4 mM concentrations. The semi- <sup>13</sup>C/<sup>15</sup>N-labeled A<sub>5</sub>-DNA duplex was prepared by annealing an unlabeled modified strand to a labeled unmodified strand. The fully <sup>13</sup>C/<sup>15</sup>N-labeled, unmodified DNA duplexes were prepared by annealing two labeled strands together. All labeled single strands were synthesized *in vitro* by the method of Zimmer and coworkers,<sup>44</sup> using a DNA hairpin template with a 5' overhang corresponding to the complement of the target labeled strand and a 3' ribose (IDT, Inc.). In this study we used the same hairpin sequence as Zimmer and coworkers,<sup>44</sup> Klenow fragment DNA polymerase (NEB, Inc.) NEB2 buffer (NEB, Inc.), and uniformly <sup>13</sup>C/<sup>15</sup>N-labeled dNTPs (Isotec, Sigma-Aldrich and Silantes). Base- and heat-catalyzed cleavage separated the hairpin template from the <sup>13</sup>C/<sup>15</sup>N-labeled synthesized product. The single-stranded DNA product was purified by 20 % denaturing polyacrylamide gel electrophoresis, isolated by passive elution from crushed gel pieces and desalted on a C18 reverse-phase column (Sep-Pak, Waters). The oligo was lyophilized and suspended in NMR buffer. The semi-labeled DNA samples were prepared by titrating the unlabeled strand directly into an NMR tube containing the <sup>13</sup>C/<sup>15</sup>N-labeled strand and monitoring the disappearance of single-stranded DNA peaks using quick HSQC experiments. The fully labeled samples were annealed in a similar

fashion. Sample pH was adjusted as described above.

All NMR experiments were performed on a Bruker Avance 600 MHz NMR spectrometer equipped with a 5mm triple-resonance cryogenic probe. Standard 2D  $^1\text{H}$ - $^1\text{H}$  NOESY experiments at 26 °C (A<sub>2</sub>, A<sub>4</sub>, A<sub>5</sub>, A<sub>6</sub>, CA<sub>3</sub>, E) or 25 °C (CG<sub>3</sub> and ZJXN) and pH 5.2 (A<sub>5</sub>), 5.4 (A<sub>2</sub>, A<sub>4</sub>, A<sub>5</sub>, A<sub>6</sub>, CA<sub>3</sub>, CG<sub>3</sub>), 6.8 (E) or 7.5 (ZJXN) were used to assigned resonances as described previously.<sup>1,26</sup>

### 2.2.2 NMR $R_{1\rho}$ Relaxation Dispersion

Carbon and nitrogen rotating frame ( $R_{1\rho}$ ) relaxation dispersion profiles for guanine/adenine C8, guanine/adenine C1', cytosine C6, guanine N1 and thymine N3 were measured using a 1D acquisition scheme which uses selective Hartmann-Hahn polarization transfer<sup>45</sup> to selectively excite one C-H or N-H spin system at a time and extends the detectable exchange regime to tens of milliseconds and analyzed as previously reported.<sup>1,2</sup> See Supplementary Table 4 for spin lock power and offset frequencies. The following delays were used: A<sub>2</sub>: A3 C1', G10 C1', A17 C1', A7/21 C1'; A<sub>4</sub>: A5 C1', G10 C1', A16 C1', A19 C1'; CA<sub>3</sub>: G10 C1'; A<sub>5</sub>: A3 C1', C5 C1' [0, 4, 8, 12, 18, 26, 34, 42, 12, 42]; A<sub>5</sub> G11 C8 [0, 12, 32, 26, 32]; A<sub>4</sub> A17 C8, A<sub>6</sub> A17 C8 [0, 4, 12, 32, 26, 32]; A<sub>2</sub>: C2 C6, T9 C6, G11 C8, A16 C8, A17 C2; CA<sub>3</sub>: A16 C8, C17 C6, C19 C6, A21 C1'; A<sub>4</sub> C15 C6, G11 C8; A<sub>5</sub>: C9 C6, G10 C8; A<sub>6</sub>: G10 C8, A16 C8 [0, 4, 8, 12, 16, 20, 26, 32, 12, 32]; A<sub>2</sub>: T8 N3, G10 N1; A<sub>5</sub>: T4 N3, T5 N3, T6 N3, T7 N3, T8 N3; A<sub>2</sub> G23 N1; A<sub>6</sub> G10 N1 [0, 8, 16, 24, 36, 48, 60, 80, 100, 16, 70, 100]; CG<sub>3</sub>: G4 C8 [0, 60, 60]; ZJXN: A6 C8, A24 C8 [0, 4, 8, 12, 16, 20, 24, 30, 12, 30]; E A5 C8: [0, 48, 48].

Data points meeting C-C Hartmann-Hahn matching conditions were omitted as described previously.<sup>6</sup> Data were processed using NMRPipe<sup>46</sup> and  $R_{1\rho}$  values were determined from monoexponential decay fits of the resonance intensities using a script<sup>47</sup> in Mathematica 9 (Wolfram Research, Inc.). On- and off-resonance  $R_{1\rho}$  data were fit to the Laguerre equation (Eq. 2.1) with the Levenberg-Marquardt algorithm<sup>48</sup> using Origin 8.6 (OriginLab).

The Laguerre equation is given by,

$$R_{1\rho} = \frac{R_1 \cos^2 \theta + R_2 \sin^2 \theta + \sin^2 \theta p_A p_B \Delta \omega^2 k_{ex}}{\omega_A^2 \omega_B^2 / \omega_{eff}^2 + k_{ex}^2 - \sin^2 \theta p_A p_B \Delta \omega^2 \left( 1 + \frac{2k_{ex}^2 (p_A \omega_A^2 + p_B \omega_B^2)}{\omega_a^2 \omega_b^2 + \omega_a^2 k_{ex}^2} \right)} \quad (2.1)$$

where  $\omega_{eff}^2 = \Omega^2 + \omega_1^2$ ,  $\omega_A^2 = (\Omega_A - \omega_{rf})^2 + \omega_1^2$  and  $\omega_B^2 = (\Omega_B - \omega_{rf})^2 + \omega_1^2$ .  $R_1$  and  $R_2$  are the intrinsic longitudinal and transverse relaxation rate constants, respectively, and are assumed to be identical for the ground (A) and transient (B) states.  $\theta = \arctan(\omega_1 / \Omega)$  where  $\omega_1$  is the spin lock power strength.  $\Omega = \Omega_{obs} - \omega_{rf}$  where  $\Omega$  is the offset of the spin lock carrier frequency ( $\omega_{rf}$ ) from the observed resonance frequency ( $\Omega_{obs}$ ).  $\Omega_{obs} = p_A \Omega_A + p_B \Omega_B$  where  $p_A$  and  $p_B$  are the ground and transient state populations and  $p_A + p_B = 1$ .  $k_{ex}$  is the chemical exchange rate constant for a two-state exchange process where  $k_{ex} = k_1 + k_2$ ,  $k_1 = k_{ex} p_B$  and  $k_2 = k_{ex} p_A$ . Plots of  $R_{1\rho}$  data are presented as  $R_{2eff}$  ( $R_{2eff} = R_2 + R_{ex}$ ), where

$$R_{2,eff} = \frac{R_{1\rho}}{\sin^2 \theta} - \frac{R_1}{\tan^2 \theta}. \quad (2.2)$$

In cases where large errors were accompanied by small  $R_{ex}$ , model selection comparing presence and absence of exchange was carried out using the F- and Akaike's

Information Criterion (AIC)<sup>49</sup> tests to discriminate between exchange and no detectable exchange (data not shown). The F-Test compares two nested models fit to the same data under the null hypothesis that the residual sum of squares of the less complex, restricted model is not significantly larger than that of the more complex, full model. The AIC Test assesses the likelihood of a model given the data and seeks to minimize loss of information embedded within the data. For A<sub>2</sub>G<sub>23</sub>N<sub>1</sub> no exchange is the selected model under the conditions used. Because of lower sensitivity to exchange, we did not interpret absence of <sup>15</sup>N dispersion as evidence for absence of transient HG base-pair.

### 2.2.3 DNA Melting Monitored via Circular Dichroism

DNA duplexes (IDT, Inc.) were prepared as described previously<sup>1</sup> in 90 % NMR buffer (15 mM Phosphate buffer, 0.1 mM EDTA, 25 mM NaCl) pH 5.4, 10 % D<sub>2</sub>O. Duplexes were prepared by diluting complementary single-stranded stocks to 50 μM in the same tube with a final volume of 200 μL. Samples were denatured at 95 °C for 5 min followed by annealing of at least 10 min on the bench top. Samples were transferred to a 1 mm cuvette (Starna Cells), mineral oil was added to the top of the solution and the cuvette was capped. Melting experiments were performed on a Jasco Spectropolarimeter equipped with a recirculating water bath and Peltier temperature control unit. Temperature ramps were performed from 5 °C to 80 °C with a bandwidth of 5 nm (1 nm for ZJXN), ramp rate of 1 °C/min, equilibration time of 20 sec and sensitivity of 100 mdeg. Wavelength scans used the same sensitivity and bandwidth as melting runs. Spectral measurements were performed from 220 nm to 330 nm with a scan rate of 100



nm/min. Temperature ramp profiles were fit to the Boltzmann model,<sup>50</sup> which has previously been used to determine nucleic acid melting temperatures,<sup>51</sup>

$$\theta = LL + \frac{UL - LL}{1 + e^{-\frac{T_m - T}{a}}} \quad (2.3)$$

where  $\theta$  is the ellipticity at 254 nm normalized to the signal change magnitude,  $LL$  and  $UL$  are the lower and upper limits of the transition, respectively,  $a$  is the Hill slope,  $T_m$  is the melting temperature defined as the point of inflection of the melting curve and  $T$  is the independent variable temperature. The thermodynamic model<sup>52,53</sup> has also been used to characterize melting temperatures of nucleic acids, but we focus on the Boltzmann model in order to simply rank the inflection/midpoint of the DNA sequences used herein as was done previously.<sup>50</sup>

#### 2.2.4 Phi ( $\Phi$ )-Value Analysis

$\Phi$ -value analysis was carried out by computing:

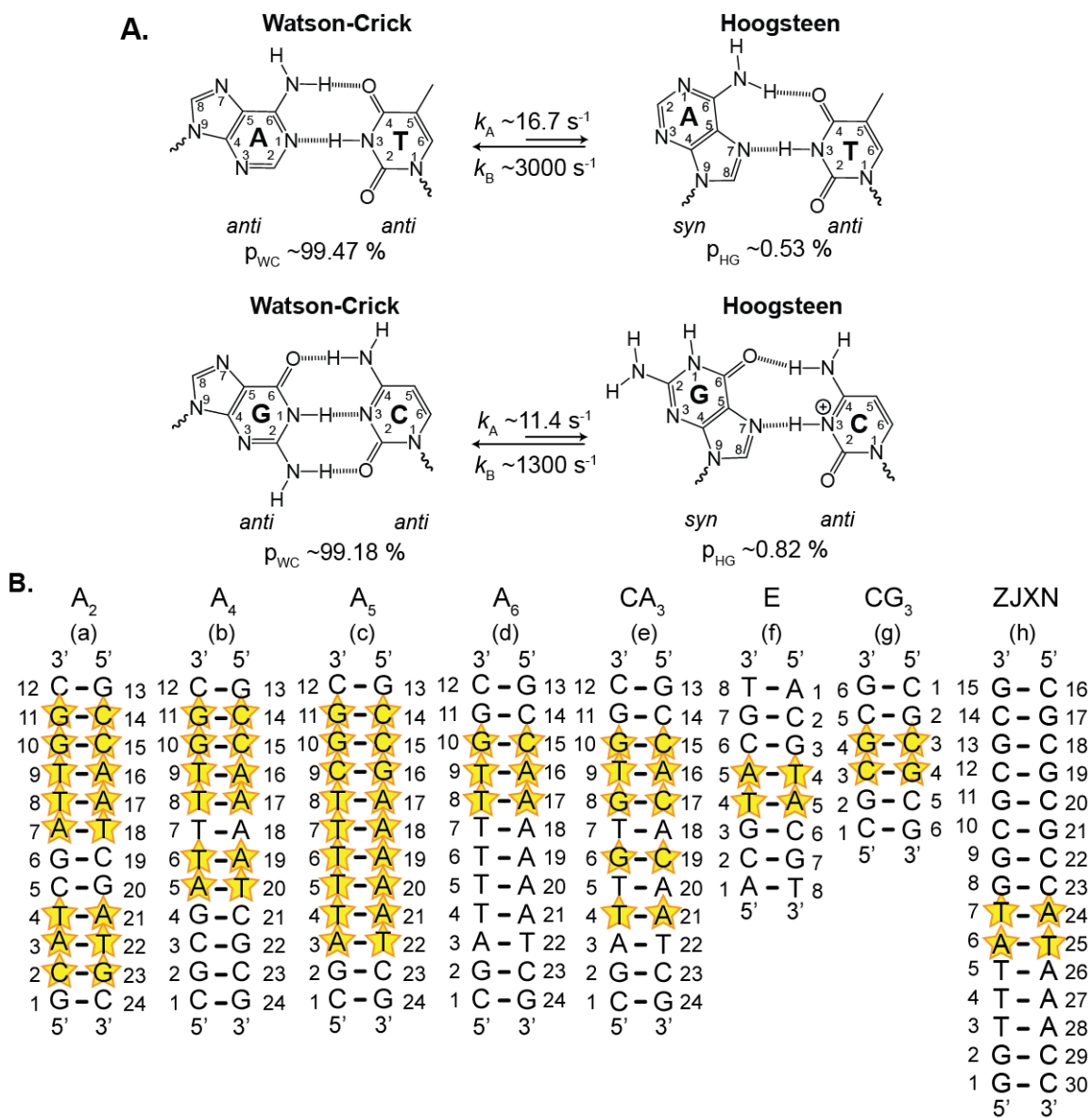
$$\Phi = \Delta\Delta G_{TS-WC} / \Delta\Delta G_{WC-HG}, \quad (2.4)$$

where  $\Delta\Delta G_{TS-WC} = \Delta G_{WC-HG,mut}^\ddagger - \Delta G_{WC-HG,\Psi-WT}^\ddagger$  and  $\Delta\Delta G_{WC-HG} = \Delta G_{WC-HG,mut} - \Delta G_{WC-HG,\Psi-WT}$  are the change in free energy difference between the transition and WC, and WC and HG states, respectively, upon introduction of one or multiple ‘mutations’, or sequence variants, as compared to an arbitrarily defined ‘wild type’ ( $\Psi$ -WT), or reference base-pair. Sequence variants describe changing the neighboring sequence context and position relative to terminal ends. Reference base-pairs are defined as A<sub>4</sub> T4 N3 and A<sub>2</sub> G10 N1 for A•T and G•C base-pair, respectively. To control for systematic errors in  $\Phi$  arising due to  $\Psi$ -WT referencing, we also performed  $\Phi$ -value analysis assigning a <sup>13</sup>C resonance

as  $\Psi$ -WT for A•T and G•C base-pairs in a mutually exclusive manner (data not shown).  $\Delta G_{\text{WC-HG,mut}}^{\ddagger}$  and  $\Delta G_{\text{WC-HG},\Psi\text{-WT}}^{\ddagger}$  are the free energy barriers of the WC-HG transition for mutant and  $\Psi$ -WT base-pair, respectively.  $\Phi \approx 0$  when  $\Delta\Delta G_{\text{TS-WC}} \approx \Delta\Delta G_{\text{WC-HG}}$  and  $\Phi \approx 1$  when  $\Delta\Delta G_{\text{TS-WC}} \approx \Delta\Delta G_{\text{WC-HG}}$ .  $\Phi \approx 0$  or 1 suggests an ‘early’ (WC-like) or late (HG-like) transition state in terms of free energy, from which structural inferences of the transition state can be made.

### 2.3 Results and Discussion

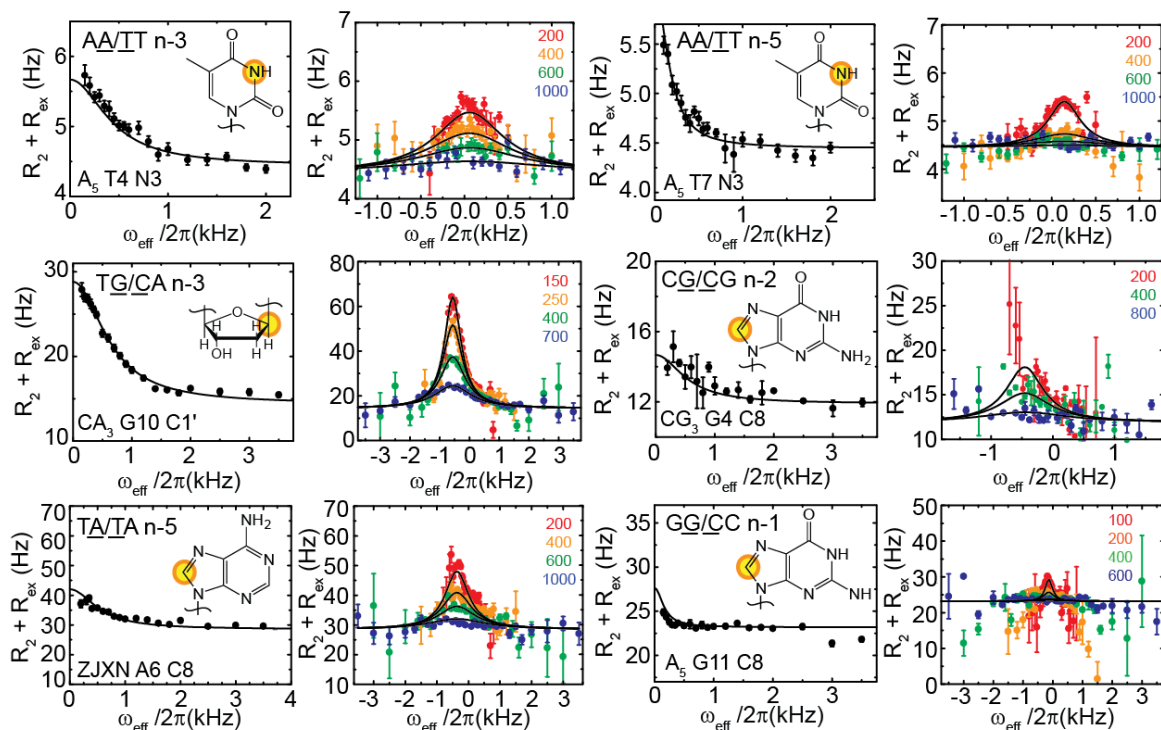
$R_{1\rho}$  relaxation dispersion experiments were performed at pH 5.2–7.5 and 5.2–5.4 for A•T and G•C base-pairs, respectively. We use low pH conditions for transient (and protonated) G•C base-pairs in order to increase the WC-to-HG chemical exchange signature detected by NMR  $R_{1\rho}$  relaxation dispersion.<sup>3</sup> We recently reported a detailed analysis of the pH dependence of chemical exchange corresponding to transient HG base-pair formation and how measurements at such lower pH conditions can be qualitatively interpolated to assess exchange at higher pH.<sup>3</sup>



**Figure 2.1 | Probing the occurrence of transient A•T and G•C Hoogsteen Base-Pairs.** (A) The equilibrium between Watson-Crick and Hoogsteen base-pairs. Shown are the average forward ( $k_A$ ) and reverse ( $k_B$ ) rate constants and populations obtained in this study. (B) DNA duplexes sequences used in this study. Base-pairs targeted for relaxation dispersion measurements are highlighted with a star.

We used  $R_{1\rho}$  relaxation dispersion experiments on well-resolved sugar (C1') or base (C6/8, N1/3) resonances to probe for the occurrence of transient HG base-pairs in the eight duplexes studied (Figure 2.1 B). A total of 41  $R_{1\rho}$  relaxation dispersion profiles were measured targeting 21 and 13 unique A•T and G•C base-pairs, respectively,

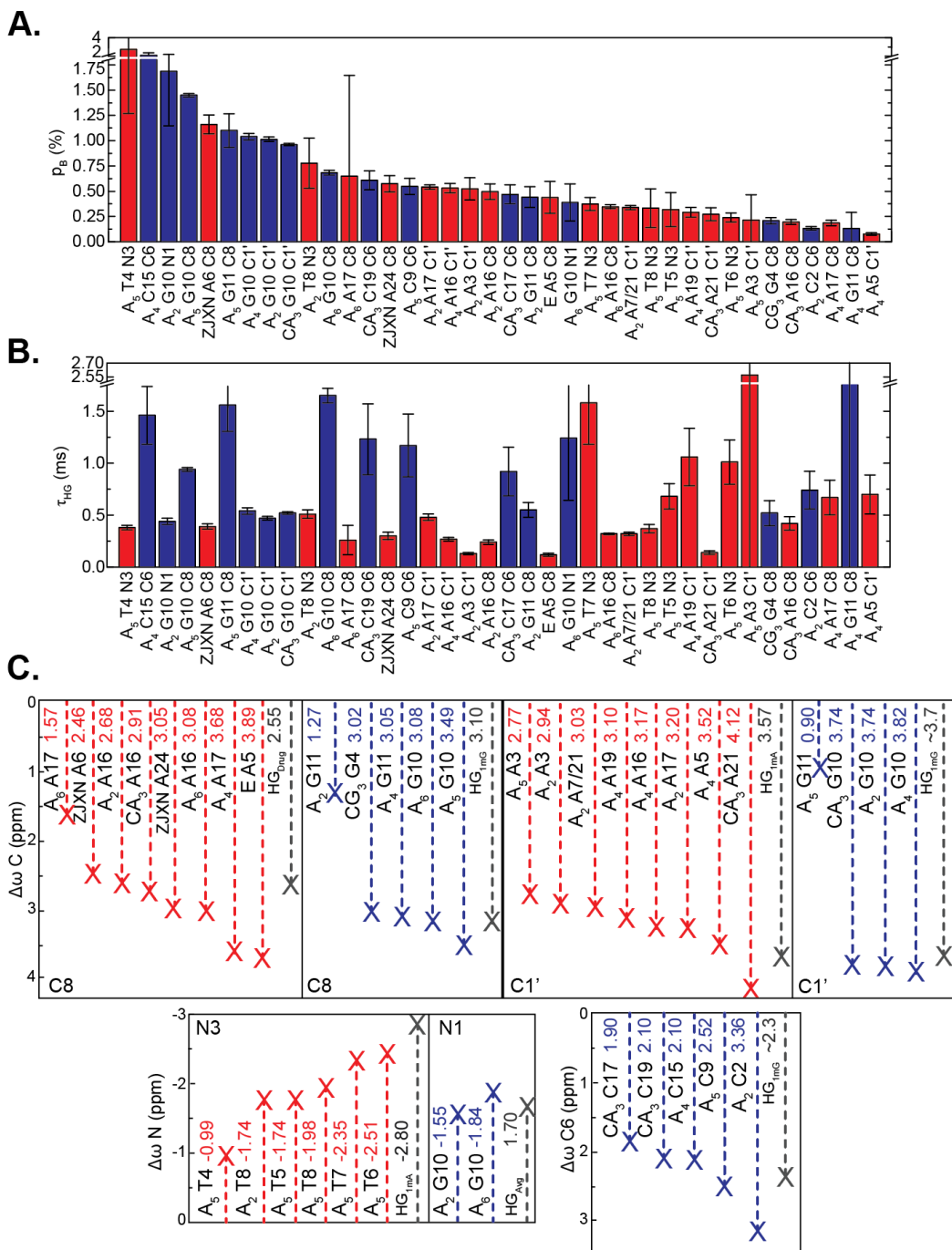
embedded within 6/10 distinct dinucleotide steps– 11x AA/TT, 4x GA/TC, 3x GG/CC, 3x CG/CG, 2x TA/TA and 11x TG/CA or 7x CA/TG, where the underlined nucleotide is the purine of the base-pair for which  $R_{1\rho}$  relaxation dispersion was measured. These base-pairs encompass 16/64 unique trinucleotide sequences and vary with respect to position relative to the terminal end ranging between n-1 and n-6, where n is the numbers of base-pair from the closest terminal end.



**Figure 2.2 | Characterizing Transient Hoogsteen Base-Pairs using On- and Off-Resonance  $^{13}\text{C}$  and  $^{15}\text{N}$   $R_{1\rho}$  Relaxation Dispersion.** Shown are representative examples of relaxation dispersion profiles for different sequence and positional contexts fit to Eq. 2.1 with spin lock powers (Hz) indicated in plot insets. Error bars represent experimental uncertainty (one standard deviation) estimated from monoexponential fitting of duplicate  $R_{1\rho}$  data. Buffer conditions are 15 mM Sodium phosphate, 25 mM NaCl, 0.1 mM EDTA, 10 %  $\text{D}_2\text{O}$  pH 5.2 ( $\text{A}_6$  C15 C6,  $\text{A}_5$ ,  $\text{A}_4$ : C15 C6, A16 C1', G10 C1'), 5.4 ( $\text{A}_2$ ,  $\text{A}_4$ ,  $\text{A}_6$ ,  $\text{CA}_3$ ,  $\text{CG}_3$ ), 6.8 (E) or 7.5 (ZJXN), 26 °C ( $\text{A}_2$ ,  $\text{A}_4$ ,  $\text{A}_5$ ,  $\text{A}_6$ ,  $\text{CA}_3$ , E) or 25 °C ( $\text{A}_2$  A3 C1',  $\text{CG}_3$ , ZJXN). 26 °C ( $\text{A}_2$ ,  $\text{A}_4$ ,  $\text{A}_5$ ,  $\text{CA}_3$ , E) or 25 °C (ZJXN).

For the 34 unique base-pairs examined, we measured significant  $R_{1\rho}$  relaxation dispersion indicating chemical exchange consistent with transient HG base-pair formation (representative plots in Figure 2.2, all plots in Figure A1.1). A two-state analysis ( $A \xrightleftharpoons[k_B]{k_A} B$ ) of the  $R_{1\rho}$  relaxation dispersion data using the Laguerre<sup>54</sup> equation yielded the downfield shifted purine C8 ( $\Delta\omega \approx 2.86$  ppm), purine C1' ( $\Delta\omega \approx 3.17$  ppm), cytosine C6 ( $\Delta\omega \approx 2.40$  ppm), and upfield shifted imino nitrogen N1/3 ( $\Delta\omega \approx -1.84$  ppm) chemical shifts that are expected for HG base-pair (Figure 2.3 C).<sup>1,2</sup> Consistent with transient HG base-pair formation, we do not observe chemical exchange at adenine C2 and thymine C6, which do not undergo large chemical shift changes upon HG formation<sup>1</sup> (Figure A1.1). These results establish the existence of transient HG base-pairs broadly outside CA/TG and TA/TA steps.

We observe significant variations in the populations (0.08-2.73 % for A•T and 0.13-2.11 % G•C) and lifetimes (0.12-2.57 ms for A•T and 0.40–2.08 ms for G•C base-pair) of the transient HG base-pair (Figure 2.3 and Table 2.1).<sup>1,2</sup> Variations are also observed in the differences between WC and HG chemical shifts (C8  $\Delta\omega = 1.27$ -3.89 ppm, C6  $\Delta\omega = 1.90$ -3.36 ppm, C1'  $\Delta\omega = 0.90$ -4.12 ppm, N1/3  $\Delta\omega = -0.99$ - -2.51 ppm) that could reflect sequence-specific contributions to chemical shifts or variations in WC and/or HG geometry (Figure 2.3 C).



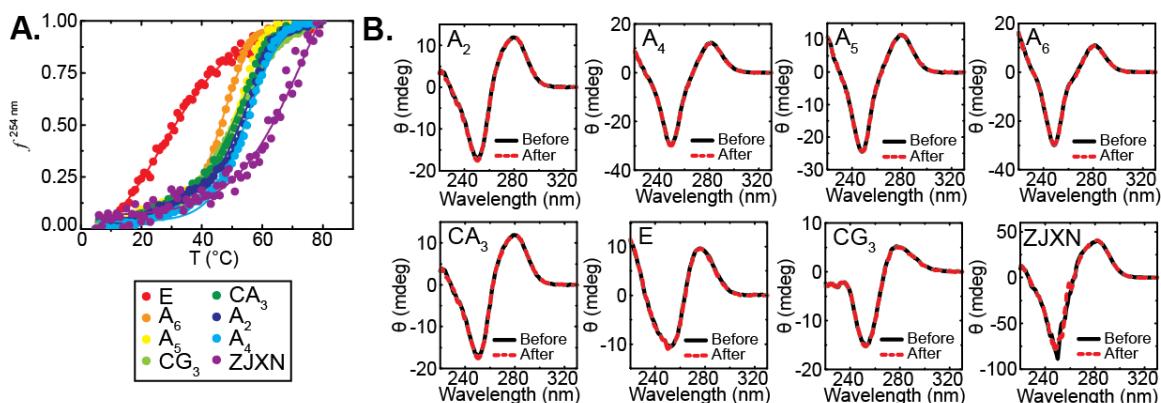
**Figure 2.3 | Sequence and Position Dependence of Population and Lifetime of Transient Hoogsteen Base-Pairs.** Shown are (A) populations ( $p_B$ ) and (B) lifetimes ( $\tau_B$ ) of transient HG A•T (red) and G•C (blue) base-pairs. (C) Transient Hoogsteen state chemical shift differences ( $\Delta\omega_{AB}$ ) relative to the Watson-Crick state obtained from  $R_{1\rho}$  relaxation dispersion for A•T (red) and G•C (blue) base-pairs, respectively. Trapped Hoogsteen base-pair references<sup>1,2</sup> are shown in grey. Error bars represent experimental uncertainty (one SD) using propagation of errors from monoexponential fitting of duplicate  $R_{1\rho}$  relaxation dispersion measurements and analysis of signal-to-noise.

**Table 2.1 | Chemical Exchange Parameters from fits of  $R_{1\rho}$  Relaxation Dispersion Data to Eq. 2.1.** Error bars represent experimental uncertainty (one SD) estimation based on monoexponential fitting of duplicate  $R_{1\rho}$  relaxation dispersion measurements and analysis of signal-to-noise. Error is not determined by fit for values equal to zero.

Probe	$p_B$ (%)	$k_{ex}$ ( $s^{-1}$ )	$\Delta\omega_{AB}$ (ppm)	$R_1$ ( $s^{-1}$ )	$R_2$ ( $s^{-1}$ )	$\tau_B$ (ms)
A <sub>2</sub> C2 C6	0.13±0.02	1357±335	3.36±0.23	2.90±0.03	26.74±0.09	0.74±0.18
A <sub>2</sub> A3 C1'	0.52±0.11	7946±699	2.94±0.29	1.86±0.03	15.37±0.33	0.13±0.01
A <sub>2</sub> T8 N3	0.78±0.25	1962±156	-1.74±0.29	1.44±0.02	4.51±0.03	0.51±0.04
A <sub>2</sub> G10 C1'	1.01±0.02	2157±86	3.74±0.06	1.72±0.06	14.85±0.20	0.47±0.02
A <sub>2</sub> G10 N1	1.88±0.63	2542±152	-1.55±0.26	1.78±0.03	4.92±0.05	0.40±0.02
A <sub>2</sub> G11 C8	0.44±0.10	1828±238	1.27±0.16	2.31±0.02	21.40±0.09	0.55±0.07
A <sub>2</sub> A16 C8	0.50±0.08	4264±394	2.68±0.20	2.22±0.05	22.64±0.29	0.24±0.02
A <sub>2</sub> A17 C1'	0.54±0.02	2074±132	3.20±0.10	1.81±0.05	14.74±0.15	0.48±0.03
A <sub>2</sub> A7/A21 C1'	0.34±0.02	3157±161	3.03±0.11	1.87±0.02	15.14±0.11	0.32±0.02
A <sub>4</sub> A5 C1'	0.08±0.01	1439±387	3.52±0.30	1.82±0.03	15.83±0.08	0.70±0.19
A <sub>4</sub> G10 C1'	1.04±0.03	1865±108	3.82±0.08	2.17±0.09	14.41±0.30	0.54±0.03
A <sub>4</sub> G11 C8	0.13±0.16	482±695	3.05±0.33	2.18±0.03	20.33±0.12	2.08±3.00
A <sub>4</sub> C15 C6	2.11±0.26	698±132	2.10±0.11	3.99±0.16	28.03±0.52	1.46±0.28
A <sub>4</sub> A16 C1'	0.53±0.05	3905±272	3.17±0.16	2.05±0.06	14.94±0.29	0.26±0.02
A <sub>4</sub> A17 C8	0.19±0.03	1493±369	3.68±0.29	2.39±0.06	21.03±0.17	0.67±0.17
A <sub>4</sub> A19 C1'	0.29±0.05	945±247	3.10±0.18	1.87±0.08	14.55±0.17	1.06±0.28
A <sub>5</sub> A3 C1'	0.21±0.25	390±512	2.77±0.20	1.24±0.05	16.31±0.06	2.57±3.37
A <sub>5</sub> T4 N3	2.73±1.46	2724±144	-0.99±0.26	1.28±0.01	4.44±0.02	0.38±0.02
A <sub>5</sub> T5 N3	0.32±0.17	1474±264	-1.74±0.48	1.26±0.01	4.40±0.02	0.68±0.12
A <sub>5</sub> T6 N3	0.24±0.05	990±208	-2.51±0.24	1.35±0.01	4.47±0.02	1.01±0.21
A <sub>5</sub> T7 N3	0.37±0.06	637±161	-2.35±0.14	1.27±0.01	4.45±0.02	1.58±0.40
A <sub>5</sub> T8 N3	0.33±0.19	2712±297	-1.98±0.58	1.28±0.01	4.37±0.03	0.37±0.04
A <sub>5</sub> C9 C6	0.55±0.08	863±223	2.52±0.18	2.87±0.08	29.74±0.30	1.17±0.30
A <sub>5</sub> G10 C8	1.45±0.02	1077±21	3.49±0.02	2.14±0.03	22.10±0.11	0.94±0.02
A <sub>5</sub> G11 C8	1.10±0.17	647±106	0.90±0.07	2.29±0.02	23.18±0.08	1.56±0.26
A <sub>6</sub> G10 C8	0.68±0.02	611±26	3.08±0.02	2.44±0.01	23.04±0.04	1.65±0.07
A <sub>6</sub> G10 N1	0.39±0.18	811±391	-1.84±0.48	1.60±0.03	4.74±0.04	1.24±0.60
A <sub>6</sub> A16 C8	0.34±0.02	3172±157	3.08±0.09	2.33±0.02	23.80±0.11	0.32±0.02
A <sub>6</sub> A17 C8	0.65±1.00	3940±2173	1.57±1.23	2.38±0.19	20.97±0.76	0.26±0.14
CA <sub>3</sub> G10 C1'	0.96±0.01	1947±43	3.74±0.03	1.93±0.03	14.37±0.12	0.52±0.01
CA <sub>3</sub> A16 C8	0.20±0.02	2389±362	2.91±0.23	2.46±0.04	25.02±0.18	0.42±0.06
CA <sub>3</sub> C17 C6	0.47±0.09	1091±277	1.90±0.24	3.21±0.07	28.62±0.28	0.92±0.23
CA <sub>3</sub> C19 C6	0.61±0.09	819±227	2.10±0.18	3.34±0.10	28.69±0.28	1.23±0.34
CA <sub>3</sub> A21 C1'	0.27±0.06	7190±981	4.12±0.48	2.82±0.03	20.44±0.43	0.14±0.02
E A5 C8	0.44±0.16	8177±1012	3.89±0.59	4.54±0.21	16.96±0.76	0.12±0.02
CG <sub>3</sub> G4 C8	0.21±0.03	1938±451	3.02±0.34	4.22±0.09	11.90±0.21	0.52±0.12
ZJXN A6 C8	1.16±0.09	2582±167	2.46±0.13	1.46±0.09	28.55±0.30	0.39±0.03
ZJXN A24 C8	0.57±0.08	3377±411	3.05±0.26	2.02±0.10	29.49±0.43	0.30±0.04
A <sub>2</sub> T9 C6	0	0	0	3.18±0.08	26.04±0.11	0
A <sub>2</sub> A17 C2	0	0	0	2.59±0.05	25.49±0.07	0

We previously reported<sup>2</sup> that owing to smaller  $\Delta\omega$  and  $R_{ex}$  contributions,  $^{15}\text{N}$  dispersion data alone may not allow for a reliable determination of the population of HG transient states. Moreover, moderate differences in pH may explain some of the observed variations. However, we observe significant variations in the populations (0.08-0.65 % for A•T and 0.13-2.11 % G•C) and lifetimes (0.13-1.06 ms for A•T and 0.47-2.08 ms for G•C base-pair) even when focusing on  $^{13}\text{C}$   $R_{1\rho}$  relaxation dispersion data only measured at a very narrow pH range (pH= 5.2-5.4). The WC-to-HG variations are also not systematic across different duplexes and no correlation between exchange parameters and melting temperatures as measured by Circular Dichroism is observed (Figure 2.4 and Table 2.2, also see Figure A1.2 for  $T_m$  according to  $p_{HG}$  and  $\tau_{HG}$ ).

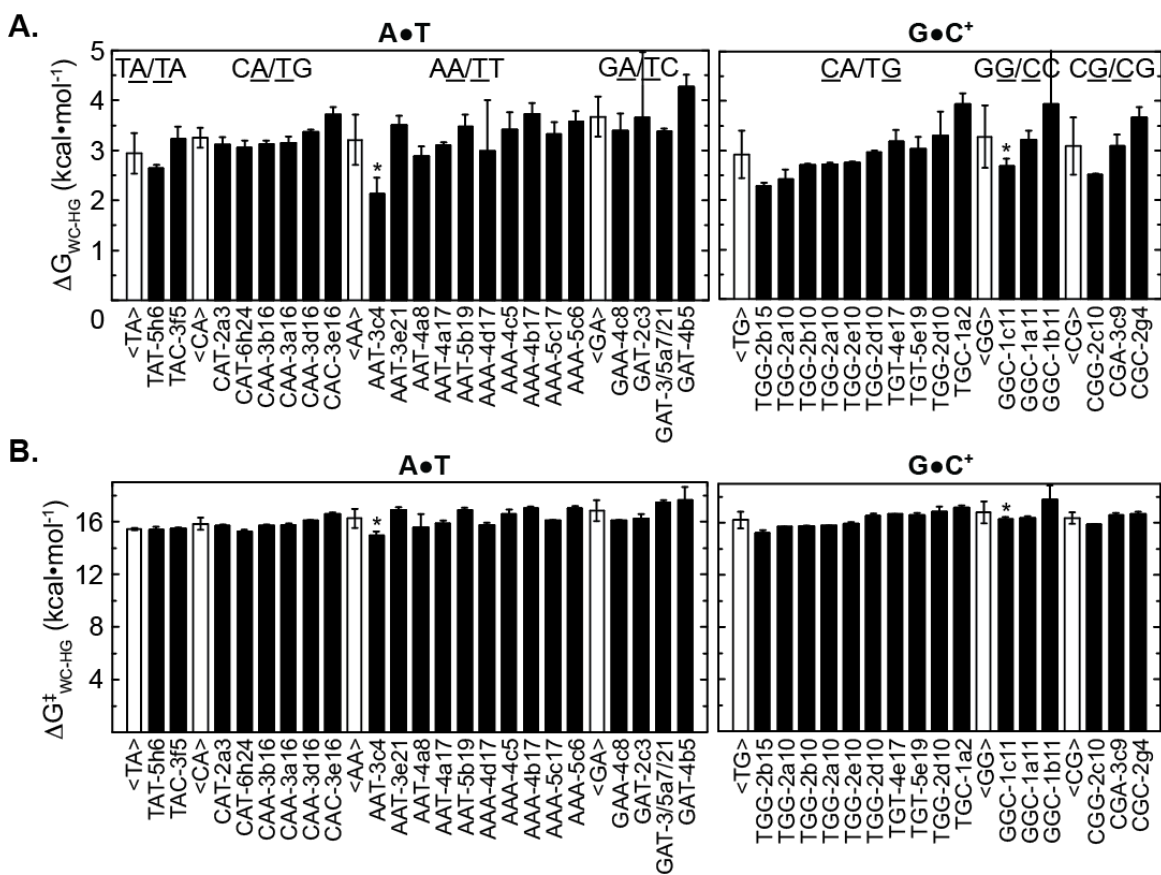




**Figure 2.4 | Circular Dichroism DNA Duplex Melting Experiments.** (A) Fraction of the melting transition calculated from ellipticity at 254 nm as a function of increasing temperature. Data are fit to Eq. 3. (B) Circular Dichroism spectra measured at 25 °C are shown in solid black and dotted red lines, before and after melting experiments respectively. Buffer conditions are 15 mM Sodium phosphate, 25 mM NaCl, 0.1 mM EDTA, 10 % D<sub>2</sub>O pH 5.2 (A<sub>6</sub> C15 C6, A<sub>5</sub>, A<sub>4</sub>: C15 C6, A16 C1', G10 C1'), 5.4 (A<sub>2</sub>, A<sub>4</sub>, A<sub>6</sub>, CA<sub>3</sub>, CG<sub>3</sub>), 6.8 (E) or 7.5 (ZJXN), 26 °C (A<sub>2</sub>, A<sub>4</sub>, A<sub>5</sub>, A<sub>6</sub>, CA<sub>3</sub>, E) or 25 °C (A<sub>2</sub> A3 C1', CG<sub>3</sub>, ZJXN).

**Table 2.2 | Experimentally Determined Melting Temperatures.** The buffer conditions are 15 mM phosphate buffer pH 5.4, 25 mM NaCl, 0.1 mM EDTA, 10 % D<sub>2</sub>O. Error bars reported are from a  $\chi^2$  minimization performed using Eq. 3 in Origin 8.6 (Origin Lab).

Sequence	T <sub>m</sub> (°C)
A <sub>2</sub>	53.9±0.4
A <sub>4</sub>	55.0±0.3
A <sub>5</sub>	51.1±0.3
A <sub>6</sub>	46.5±0.2
CA <sub>3</sub>	51.4±0.4
E	26.8±0.8
CG <sub>3</sub>	51.3±0.2
ZJXN	73.6±3.6



**Figure 2.5 | Sequence and Position Dependence of Thermodynamic and Kinetic Parameters Describing the Watson-Crick to Hoogsteen Transition.** Black bars represent (A) free energy difference ( $\Delta G_{WC-HG}$ ) and (B) forward barrier ( $\Delta G_{WC-HG}^{\ddagger}$ ) of the Watson-Crick to Hoogsteen transition for base-pairs probed. Labels follow the nomenclature: sequence context (i.e. TAT), base-pairs from nearest terminal end (i.e. -5), sequence (i.e. h), residue number (i.e. 6) such as TAT-5h6. Error bars represent experimental uncertainty (one standard deviation) estimated from propagation of errors from monoexponential fitting of duplicate  $R_{1\rho}$  relaxation dispersion measurements and analysis of signal-to-noise. Average and standard deviation for dinucleotide steps are shown using white bars.

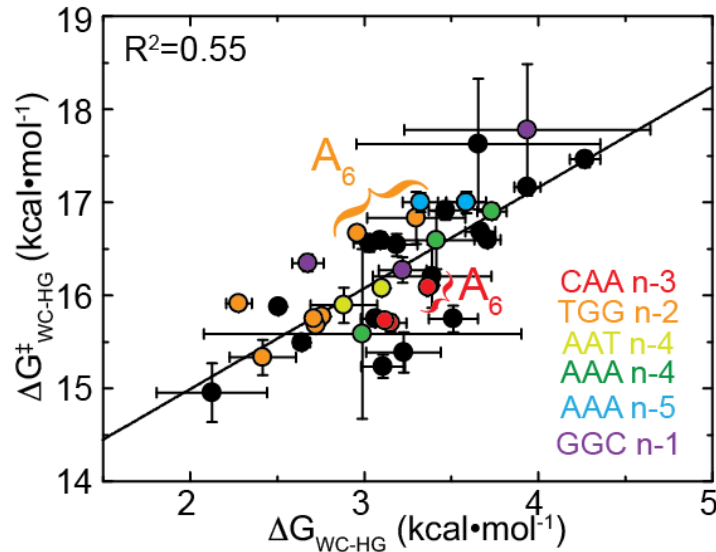
Strikingly, we find that the sequence dependent contribution to WC-to-HG equilibrium can be ascribed in part to energetic interactions with neighboring base-pairs at the dinucleotide level. In Figure 2.5, we plot the free energy difference ( $\Delta G_{WC-HG} = G_{HG} - G_{WC}$ ) and forward free energy barrier ( $\Delta G_{WC-HG}^{\ddagger} = G_{TS} - G_{WC}$ ) for the WC-to-HG transition classified by each target base-pair's individual dinucleotide step. Presenting our

data in terms of free energies allows for comparison with those previously reported for WC stability<sup>55</sup> and other dynamic transitions involving DNA, such as base opening.<sup>56</sup> Although we do observe variations in  $\Delta G_{\text{WC-HG}}$  and  $\Delta G_{\text{WC-HG}}^{\ddagger}$  for the same dinucleotide steps related to both the position of the step and broader sequence context, on average, the measured stability of transient HG A•T base-pair (TA/TA>CA/TG>AA/TT>GA/TC, where the underlined base-pair represents the transient HG base-pair probed) is nearly inverted relative to WC dinucleotide stability (GA/TC≥CA/TG>AA/TT>TA/TA).<sup>55</sup>

This suggests that steps that are less energetically favorable in the WC state, such as CA/TG and TA/TA steps, are more likely to form HG base-pairs as compared to more stable WC dinucleotide steps such as GA/TC, CG/CG and GG/CC. The preference for forming HG base-pair at TA/TA steps observed here is consistent with a large body of data showing that HG base-pair are favored in A-T rich sequences, in particular TA/TA steps (reviewed by Nikolova and coworkers). Likewise, the order of stability for transient G•C HG base-pair (TG/CA≥CG/CG≥GG/CC) is nearly inverted relative to WC dinucleotide stability<sup>55</sup> (CG/CG>GG/CC>TG/CA) though it should be noted that the stability of transient G•C HG base-pair also depends on the  $pK_a$  of cytosine N3 and can vary due to small variations in pH.<sup>1-3</sup> These results suggest that the observed variations in kinetic and thermodynamic parameters for the WC-to-HG transition may arise in part due to sequence-specific variations in WC stability.

In addition, as might be expected, base-pairs within the same trinucleotide sequence context and identical position relative to duplex ends (e.g. CAA/TTA n-3, AAT/ATT n-4, AAA/TTT n-4, AAA/TTT n-5, GGC/GCC n-1, TGG/CCA n-2; Figure 2.3 A and B and Table 2.1) have similar populations and lifetimes as compared to the

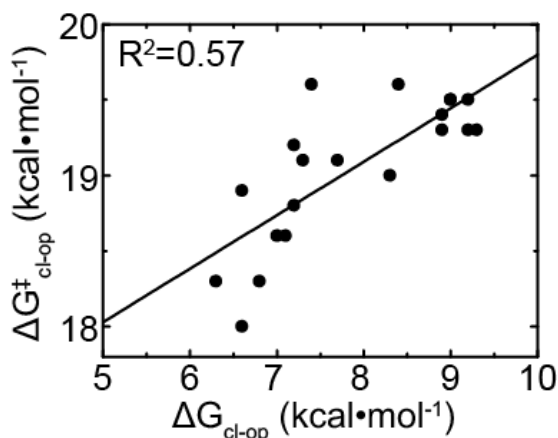
entire distribution (for a free energy description see Figure 2.6). In addition, there are variations that seem to depend position context within a given sequence. Thus, the population and lifetime of transient HG base-pairs can vary significantly for the same trinucleotide sequence context across different positions along a duplex. For example, at the n-3 position, AAT/ATT CA<sub>3</sub> A21 C1' has exchange parameters ( $p_B = 0.27 \pm 0.06$  %,  $\tau_B = 0.14 \pm 0.02$  ms,  $\Delta\omega = 4.12 \pm 0.48$  ppm) that differ from those observed for AAT/ATT in the n-4 position in A<sub>2</sub> A17 C1' ( $p_B = 0.54 \pm 0.02$  %,  $\tau_B = 0.48 \pm 0.03$  ms,  $\Delta\omega = 3.20 \pm 0.10$  ppm). Similarly, AAA/TTT at n-4 and n-5 positions in A<sub>4</sub> A17 C8 and A<sub>4</sub> A19 C1' have significantly different exchange parameters of  $p_B = 0.19 \pm 0.03$  %,  $\tau_B = 0.67 \pm 0.17$  ms and  $\Delta\omega = 3.68 \pm 0.29$  ppm as compared to  $p_B = 0.29 \pm 0.05$  %,  $\tau_B = 1.06 \pm 0.28$  ms and  $\Delta\omega = 3.10 \pm 0.18$  ppm, respectively. In general, the relative stability of HG base-pairs increases as the base-pair is embedded deeper within the double helix away from terminal ends. Our prior studies showed that formation of HG base-pair is accompanied by unfavorable entropy;<sup>1</sup> it is possible that reduced flexibility with the core of duplexes as compared to the terminal ends favors the more rigid HG as compared to WC base-pair.



**Figure 2.6 | Correlation of Free Energy of Stabilization and Forward Free Energy Barrier for the Watson-Crick-to-Hoogsteen Transition.** Shown are corresponding energetic differences ( $\Delta G_{\text{WC-HG}}$  and  $\Delta G^{\ddagger}_{\text{WC-HG}}$ ) for resonances with the same sequence and positional context and the correlation coefficient ( $R^2$ ). Error bars represent experimental uncertainty (one standard deviation) estimated from propagation of errors from monoexponential fitting of duplicate  $R_{1\rho}$  relaxation dispersion measurements and analysis of signal-to-noise.

Interestingly, we observe a correlation between  $\Delta G_{\text{WC-HG}}$  and  $\Delta G^{\ddagger}_{\text{WC-HG}}$  (Figure 2.6) and a relatively uniform backward free energy barrier ( $\Delta G^{\ddagger}_{\text{HG-WC}}$ ) of  $\sim 13$  kcal/mol (data not shown). Again this suggests that the sequence-specific variations in the free energy of the TS and HG state are either small or highly correlated and distinct from the sequence/position dependencies of the WC state. If the WC stabilities dominate the observed sequence-specific variations, one might expect a similar correlation to emerge for the WC to base-opening transition, which involves a distinct final state (and possibly TS) that might be expected to have distinct sequence dependencies. Indeed, even though the experiments are performed on different DNA sequences under higher pH conditions with base present as a catalyst of base-pair opening, as typically used in the imino proton

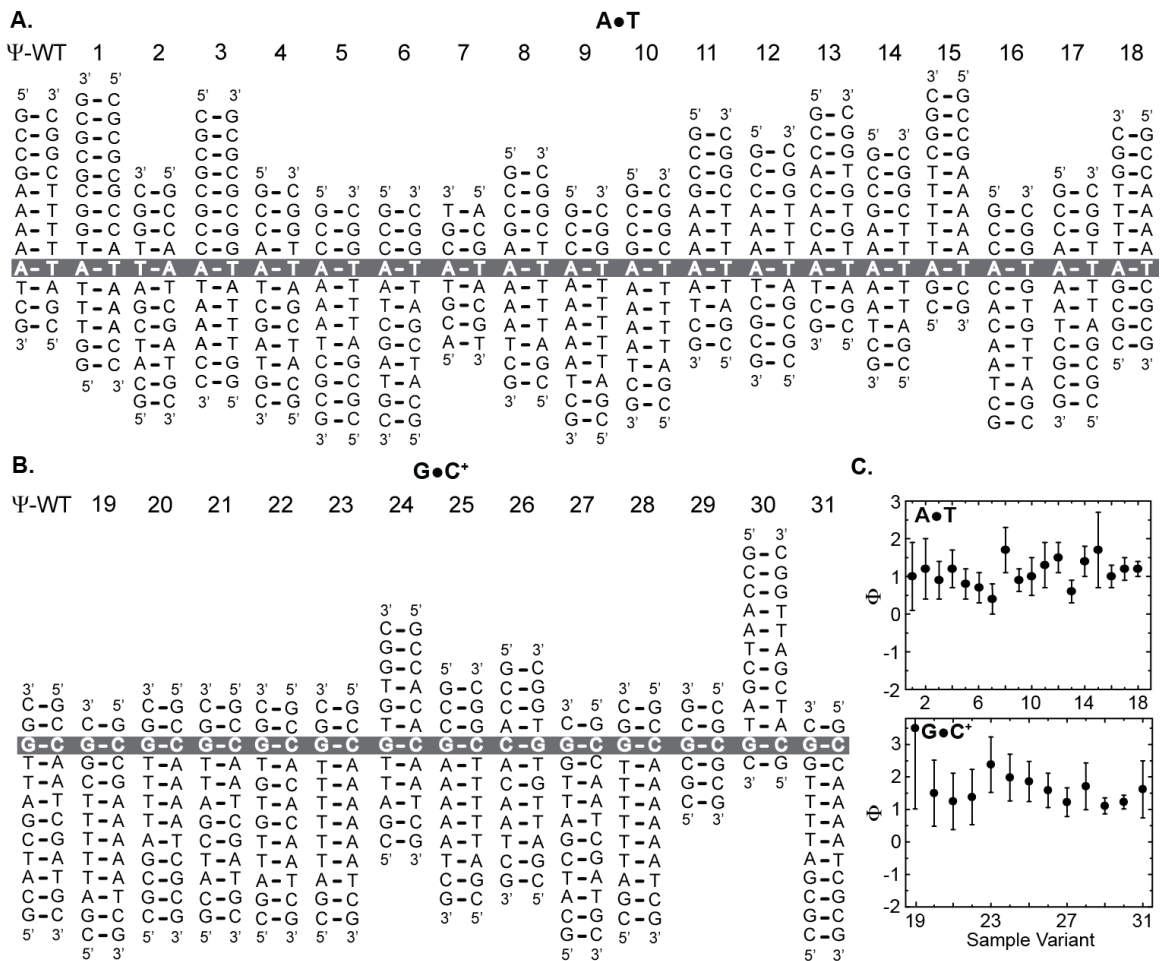
exchange measurements, a similar correlation exists between forward free energy barrier and free energy difference measured for DNA base-opening determined by imino proton exchange by Russu and coworkers (Figure 2.7).<sup>56</sup> Taken together, these data suggest that the sequence-specific variations in energetics of the WC-to-HG transition and base-opening are dominated by sequence specific variations in WC stabilities. The smaller, possibly correlated sequence-specific variations in the free energy stability for the TS and HG state may reflect weaker, less ideal stacking interactions with neighboring base-pairs; however this requires further investigation.



**Figure 2.7 | Correlation between the Free Energy Difference and Forward Free Energy Barrier for Base-Pair Opening as Measured by Imino Proton Exchange in the Highlighted Data taken from Russu and Coworkers in Reference 56.** Shown is the best-fit line and correlation coefficient ( $R^2$ ). For errors and experimental conditions see original study.<sup>56</sup>

$\Phi$ -value analysis<sup>57,58</sup> is employed to assess the energetic similarity of the TS to either the beginning or final state (here, WC and HG, respectively) and subsequently infer qualitative structural features. Indeed, as expected,  $\Phi$ -value analysis of the measured  $\Delta G_{TS-WC}$  and  $\Delta G_{WC-HG}$  values for A•T and G•C base-pairs ( $\Phi = \Delta \Delta G_{TS-WC} / \Delta \Delta G_{WC-HG} \sim 1.1$ .) suggest a ‘late’ HG-like transition state (Figure 2.8 and Table 2.3). The TS may involve a *syn*-like purine given that this structural aspect dominates the underlying carbon

chemical shifts measured here. Indeed, Previous Conjugate Peak Refinement simulations suggest a TS consistent with the *syn* purine orientation.<sup>1</sup> Interestingly, a *syn* conformation is observed for a transition state mimic (Immucillin-H)<sup>59</sup> of base phosphorylation by mammalian purine nucleoside phosphorylases (PNPs), which are essential for proper immune system function,<sup>60</sup> demonstrating the feasibility of a HG-like TS.



**Figure 2.8 |  $\Phi$ -Value Analysis.** Shown is reference ( $\Psi$ -WT) base-pair ( $A_5 T_4 N_3$  for  $A\bullet T$  and  $A_2 G_{10} N_1$  for  $G\bullet C$ ) and ‘mutants’ (1-21 and 22-36) for (A)  $A\bullet T$  and (B)  $G\bullet C$  base-pairs. (C)  $\Phi$ -values for  $A\bullet T$  (top) and  $G\bullet C$  (bottom) base-pairs, respectively. Error bars represent experimental uncertainty (one SD).

**Table 2.3 |  $\Psi$ -Wild Type and Sequence Variants for  $\Phi$ -Value Analysis.**  $\Psi$ -Wild Type and Variants for which  $\Phi$ -Value Analysis was calculated. n.d. indicates  $\Phi$ -value not determined for  $\Psi$ -Wild Type.  $\Phi$ -values were calculated using Eq. 2.4. Error bars represent experimental uncertainty (one SD) using propagation of errors from monoexponential fitting of duplicate  $R_{1\rho}$  Relaxation Dispersion measurements and analysis of signal-to-noise.

<b>Variant</b>	<b>RD Probe</b>	<b><math>\Phi</math></b>
$\Psi$ -WT	A <sub>5</sub> T4 N3	n.d.
1	ZJXN A6 C8	1.0±0.9
2	A <sub>2</sub> T8 N3	1.2±0.8
3	ZJXN A24 C8	0.9±0.5
4	A <sub>2</sub> A17 C1'	1.2±0.5
5	A <sub>4</sub> A16 C1'	0.8±0.4
6	A <sub>2</sub> A16 C8	0.7±0.4
7	E A5 C8	0.4±0.4
8	A <sub>5</sub> T7 N3	1.7±0.6
9	A <sub>6</sub> A16 C8	0.9±0.3
10	A <sub>5</sub> T8 N3	1.0±0.5
11	A <sub>5</sub> T5 N3	1.3±0.6
12	A <sub>4</sub> A19 C1'	1.5±0.5
13	CA <sub>3</sub> A21 C1'	0.6±0.3
14	A <sub>5</sub> T6 N3	1.4±0.4
15	A <sub>5</sub> A3 C1'	1.7±1.7
16	CA <sub>3</sub> A16 C8	1.0±0.3
17	A <sub>4</sub> A17 C8	1.2±0.4
18	A <sub>4</sub> A5 C1'	1.2±0.3
$\Psi$ -WT	A <sub>2</sub> G10 N1	n.d.
19	A <sub>5</sub> G11 C8	3.5±3.0
20	A <sub>4</sub> G10 C1'	1.5±1.1
21	A <sub>2</sub> G10 C1'	1.2±0.9
22	CA <sub>3</sub> G10 C1'	1.4±0.9
23	A <sub>6</sub> G10 C8	2.4±0.9
24	CA <sub>3</sub> C19 C6	2.0±1.0
25	A <sub>5</sub> C9 C6	1.9±0.9
26	CA <sub>3</sub> C17 C6	1.6±0.7
27	A <sub>2</sub> G11 C8	1.2±0.5
28	A <sub>6</sub> G10 N1	1.7±1.1
29	CG <sub>3</sub> G4 C8	1.1±0.3
30	A <sub>2</sub> C2 C6	1.2±0.3
31	A <sub>4</sub> G11 C8	1.6±1.6



Our results show that transient HG base-pairs occur ubiquitously across a wide range of sequence and positional contexts. We observe small but significant variations in the lifetime ( $\sim 20$ -fold) and populations ( $\sim 30$ -fold) of HG base-pairs that depend on both sequence and position context. The sequence dependence can be partially described at the dinucleotide level and the stability of HG relative to WC is inversely correlated to the stability of the WC dinucleotide step. This provides a simple and robust mechanism for achieving sequence-specific DNA recognition of HG base-pairs *via* indirect readout-type mechanisms. Finally, our results suggest a late TS for the WC-to-HG transition that is energetically similar to the base opening TS and possibly involving a *syn* purine base. The broad and sequence-specific occurrence of transient HG base-pairs in canonical duplex DNA suggests a potentially broader role in biology.

## 2.4 Conclusions

We recently showed using  $R_{1\rho}$  NMR relaxation dispersion that Watson-Crick base-pairs in CA/TG and TA/TA steps of canonical duplex DNA transiently form short-lived ( $\sim 0.3$ - $2.5$  ms) and sparsely populated ( $p_B \sim 0.14$ - $0.49$  %) Hoogsteen base-pairs. Here, we show that transient Hoogsteen base-pairs occur robustly across a wide range of DNA sequence and positional contexts, sampling a wider range of lifetimes ( $\sim 0.12$ - $2.57$  ms) and populations ( $\sim 0.08$ - $2.73$  %) than previously reported. Off-resonance NMR relaxation dispersion experiments performed on eight distinct duplexes provided evidence for transient Hoogsteen base-pairs in 22 A•T (pH= 5.2-7.5) and 16 G•C (pH= 5.2-5.4) base-pairs in all six dinucleotide steps (GA/TC, AA/TT, TA/TA, GG/CC,

CG/CG and TG/CA or CA/TG) and sixteen trinucleotide sequences (TAT/ATA, CAA/TTG, CAT/ATG, CAC/GTG, TAC/GTA, AAA/TTT, AAT/ATT, GAT/ATC, GAA/TTC, CGG/CCG, TGG/CCA, GGC/GCC, CGC/GCG, TGT/ACA, CGA/TCG and TGC/GCA) examined. We observe correlated variations in the energetic stabilities and forward barriers ( $R^2=0.61$ ) for transient Hoogsteen base-pairs that are inversely correlated to Watson-Crick base-pair stability at dinucleotide steps, with Hoogsteen base-pairs being energetically preferred at less stable TA/TA and CA/TG dinucleotide steps. Phi-value analysis suggests a late transition state for the Watson-Crick to Hoogsteen transition for both A•T ( $\phi \sim 1.1$ ) and G•C ( $\phi \sim 1.7$ ) base-pairs that possibly involve the purine in a *syn*-like conformation. The widespread sequence-dependent occurrence of transient Hoogsteen base-pairs in DNA suggests a potentially broader role in DNA-based functions.

## 2.5 References

- 1 Nikolova, E. N. *et al.* *Nature* Transient Hoogsteen base pairs in canonical duplex DNA **470**, 498-502, (2011).
- 2 Nikolova, E. N., Gottardo, F. L. & Al-Hashimi, H. M. *J Am Chem Soc* Probing Transient Hoogsteen Hydrogen Bonds in Canonical Duplex DNA Using NMR Relaxation Dispersion and Single-Atom Substitution **134**, 3667-3670, (2012).
- 3 Nikolova, E. N., Goh, G. B., Brooks, C. L. & Al-Hashimi, H. M. *J Am Chem Soc* Characterizing the Protonation State of Cytosine in Transient G•C Hoogsteen Base Pairs in Duplex DNA **135**, 6766-6769, (2013).
- 4 Korzhnev, D. M., Orekhov, V. Y. & Kay, L. E. *J Am Chem Soc* Off-resonance R(1rho) NMR studies of exchange dynamics in proteins with low spin-lock fields: an application to a Fyn SH3 domain **127**, 713-721, (2005).
- 5 Palmer, A. G. & Massi, F. *Chem Rev* Characterization of the dynamics of biomacromolecules using rotating-frame spin relaxation NMR spectroscopy **106**, 1700-1719, (2006).
- 6 Hansen, A. L., Nikolova, E. N., Casiano-Negroni, A. & Al-Hashimi, H. M. *J Am Chem Soc* Extending the range of microsecond-to-millisecond chemical exchange

- detected in labeled and unlabeled nucleic acids by selective carbon R(1rho) NMR spectroscopy **131**, 3818-3819, (2009).
- 7 Orekhov, V. Y., Korzhnev, D. M. & Kay, L. E. *J Am Chem Soc* Double- and zero-quantum NMR relaxation dispersion experiments sampling millisecond time scale dynamics in proteins **126**, 1886-1891, (2004).
- 8 Nikolova, E. N. *et al. Biopolymers* A historical account of Hoogsteen base-pairs in duplex DNA **99**, 955-968, (2013).
- 9 Hoogsteen, K. *Acta Crystallographica Section A, Foundations of crystallography* The crystal and molecular structure of a hydrogen-bonded complex between 1-methylthymine and 9-methyladenine **16**, 907, (1963).
- 10 Abrescia, N. G., Gonzalez, C., Gouyette, C. & Subirana, J. A. *Biochemistry* X-ray and NMR studies of the DNA oligomer d(ATATAT): Hoogsteen base pairing in duplex DNA **43**, 4092-4100, (2004).
- 11 Cubero, E., Abrescia, N. G., Subirana, J. A., Luque, F. J. & Orozco, M. *J Am Chem Soc* Theoretical study of a new DNA structure: the antiparallel Hoogsteen duplex **125**, 14603-14612, (2003).
- 12 Abrescia, N. G., Thompson, A., Huynh-Dinh, T. & Subirana, J. A. *Proc Natl Acad Sci USA* Crystal structure of an antiparallel DNA fragment with Hoogsteen base pairing **99**, 2806-2811, (2002).
- 13 Cubero, E., Luque, F. J. & Orozco, M. *Biophys J* Theoretical study of the Hoogsteen-Watson-Crick junctions in DNA **90**, 1000-1008, (2006).
- 14 Cubero, E. *et al. J Am Chem Soc* Hoogsteen-based parallel-stranded duplexes of DNA. Effect of 8-amino-purine derivatives **124**, 3133-3142, (2002).
- 15 Cubero, E., Luque, F. J. & Orozco, M. *J Am Chem Soc* Theoretical studies of d(A:T)-based parallel-stranded DNA duplexes **123**, 12018-12025, (2001).
- 16 Patikoglou, G. A. *et al. Genes Dev* TATA element recognition by the TATA box-binding protein has been conserved throughout evolution **13**, 3217-3230, (1999).
- 17 Kitayner, M. *et al. Nat Struct Mol Biol* Diversity in DNA recognition by p53 revealed by crystal structures with Hoogsteen base pairs **17**, 423-429, (2010).
- 18 Nair, D. T., Johnson, R. E., Prakash, S., Prakash, L. & Aggarwal, A. K. *Nature* Replication by human DNA polymerase- $\epsilon$  occurs by Hoogsteen base-pairing **430**, 377-380, (2004).
- 19 Johnson, R. E., Prakash, L. & Prakash, S. *Proc Natl Acad Sci USA* Biochemical evidence for the requirement of Hoogsteen base pairing for replication by human DNA polymerase  $\epsilon$  **102**, 10466-10471, (2005).
- 20 Lu, L., Yi, C., Jian, X., Zheng, G. & He, C. *Nucleic Acids Res* Structure determination of DNA methylation lesions N1-meA and N3-meC in duplex DNA using a cross-linked protein-DNA system **38**, 4415-4425, (2010).
- 21 Zaliznyak, T., Lukin, M., Johnson, F. & de los Santos, C. *Biochemistry* Solution structure of duplex DNA containing the mutagenic lesion 1,N(2)-etheno-2'-deoxyguanine **47**, 4606-4613, (2008).
- 22 Batra, V. K., Shock, D. D., Beard, W. A., McKenna, C. E. & Wilson, S. H. *Proc Natl Acad Sci USA* Binary complex crystal structure of DNA polymerase  $\beta$  reveals multiple conformations of the templating 8-oxoguanine lesion **109**, 113-118, (2012).

- 23 Alvey, H. S., Gottardo, F. L., Nikolova, E. N. & Al-Hashimi, H. M. *Nat Struct Mol Biol* Transient Hoogsteen Base-Pairs Occur Robustly in Duplex DNA **In Submission**, (2014).
- 24 Dethoff, E. A. *et al. Biophys J* Characterizing complex dynamics in the transactivation response element apical loop and motional correlations with the bulge by NMR, molecular dynamics, and mutagenesis **95**, 3906-3915, (2008).
- 25 Hansen, A. L. & Al-Hashimi, H. M. *J Am Chem Soc* Dynamics of Large Elongated RNA by NMR Carbon Relaxation **129**, 16072-16082, (2007).
- 26 Bothe, J. R., Lowenhaupt, K. & Al-Hashimi, H. M. *J Am Chem Soc* Sequence-specific B-DNA flexibility modulates Z-DNA formation **133**, 2016-2018, (2011).
- 27 Zhao, B., Hansen, A. L. & Zhang, Q. *J Am Chem Soc* Characterizing Slow Chemical Exchange in Nucleic Acids by Carbon CEST and Low Spin-Lock Field R1ρ NMR Spectroscopy **136**, 20-23, (2014).
- 28 Bothe, J. R. *et al. Nat Methods* Characterizing RNA dynamics at atomic resolution using solution-state NMR spectroscopy **8**, 919-931, (2011).
- 29 Mulder, F. A., Skrynnikov, N. R., Hon, B., Dahlquist, F. W. & Kay, L. E. *J Am Chem Soc* Measurement of slow (micros-ms) time scale dynamics in protein side chains by (15)N relaxation dispersion NMR spectroscopy: application to Asn and Gln residues in a cavity mutant of T4 lysozyme **123**, 967-975, (2001).
- 30 Mulder, F. A., Mittermaier, A., Hon, B., Dahlquist, F. W. & Kay, L. E. *Nat Struct Biol* Studying excited states of proteins by NMR spectroscopy **8**, 932-935, (2001).
- 31 Korzhnev, D. M. *et al. Nature* Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR **430**, 586-590, (2004).
- 32 Li, Y. *et al. Proc Nat Acad Sci USA* Mechanism of E-cadherin dimerization probed by NMR relaxation dispersion **110**, 16462-16467, (2013).
- 33 O'Connell, N. E. *et al. J Biomol NMR* Partially folded equilibrium intermediate of the villin headpiece HP67 defined by 13C relaxation dispersion **45**, 85-98, (2009).
- 34 Grey, M. J. *et al. J Mol Biol* Characterizing a partially folded intermediate of the villin headpiece domain under non-denaturing conditions: contribution of His41 to the pH-dependent stability of the N-terminal subdomain **355**, 1078-1094, (2006).
- 35 Rinnenthal, J. *et al. Acc Chem Res* Mapping the Landscape of RNA Dynamics with NMR Spectroscopy **44**, 1292-1301, (2011).
- 36 Blad, H., Reiter, N. J., Abildgaard, F., Markley, J. L. & Butcher, S. E. *J Mol Biol* Dynamics and Metal Ion Binding in the U6 RNA Intramolecular Stem-Loop as Analyzed by NMR **353**, 540-555, (2005).
- 37 Hoogstraten, C. G., Wank, J. R. & Pardi, A. *Biochemistry* Active site dynamics in the lead-dependent ribozyme **39**, 9951-9958, (2000).
- 38 Wang, C., Grey, M. J. & Palmer, A. G., 3rd. *J Biomol NMR* CPMG sequences with enhanced sensitivity to chemical exchange **21**, 361-366, (2001).
- 39 Grey, M. J., Wang, C. & Palmer, A. G., 3rd. *J Am Chem Soc* Disulfide bond isomerization in basic pancreatic trypsin inhibitor: multisite chemical exchange quantified by CPMG relaxation dispersion and chemical shift modeling **125**, 14324-14335, (2003).

- 40 Ughetto, G. *et al. Nucleic Acids Res* A comparison of the structure of  
echinomycin and triostin A complexed to a DNA fragment **13**, 2305-2323,  
(1985).
- 41 Gilbert, D. E., van der Marel, G. A., van Boom, J. H. & Feigon, J. *Proc Nat Acad  
Sci USA* Unstable Hoogsteen base pairs adjacent to echinomycin binding sites  
within a DNA duplex **86**, 3006-3010, (1989).
- 42 Schwartz, T., Rould, M. A., Lowenhaupt, K., Herbert, A. & Rich, A. *Science*  
Crystal Structure of the Za Domain of the Human Editing Enzyme ADAR1  
Bound to Left-Handed Z-DNA **284**, 1841-1845, (1999).
- 43 Ha, S. C., Lowenhaupt, K., Rich, A., Kim, Y. G. & Kim, K. K. *Nature* Crystal  
structure of a junction between B-DNA and Z-DNA reveals two extruded bases  
**437**, 1183-1186, (2005).
- 44 Zimmer, D. P. & Crothers, D. M. *Proc Nat Acad Sci USA* NMR of enzymatically  
synthesized uniformly <sup>13</sup>C/<sup>15</sup>N-labeled DNA oligonucleotides **92**, 3091-3095,  
(1995).
- 45 Chiarparin, E., Pelupessy, P. & Bodenhausen, G. *Mol Phys* Selective cross-  
polarization in solution state NMR **95**, 759-767, (1998).
- 46 Delaglio, F. *et al. J Biomol NMR* NMRPipe: a multidimensional spectral  
processing system based on UNIX pipes **6**, 277-293, (1995).
- 47 Spyropoulos, L. *J Biomol NMR* A suite of *Mathematica* notebooks for the  
analysis of protein main chain <sup>15</sup>N NMR relaxation data **36**, 215-224, (2006).
- 48 Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. *Numerical  
Recipes. The Art of Scientific Computing*. (Cambridge University Press, 1986).
- 49 Akaike, H. *Automatic Control, IEEE Transactions on* A new look at the statistical  
model identification **19**, 716-723, (1974).
- 50 Schulz, M. N., Landstrom, J. & Hubbard, R. E. *Anal Biochem* MTSA--a Matlab  
program to fit thermal shift data **433**, 43-47, (2013).
- 51 Doktycz, M. J., Morris, M. D., Dormady, S. J., Beattie, K. L. & Jacobson, K. B. *J  
Biol Chem* Optical melting of 128 octamer DNA duplexes: effects of base pair  
location and nearest neighbors on thermal stability **270**, 8439-8445, (1995).
- 52 Eichhorn, C. D. *et al. Nucleic Acids Res* Unraveling the structural complexity in a  
single-stranded RNA tail: implications for efficient ligand binding in the  
prequeuosine riboswitch **40**, 1345-1355, (2012).
- 53 Sashital, D. G., Allmann, A. M., Van Doren, S. R. & Butcher, S. E. *Biochemistry*  
Structural basis for a lethal mutation in U6 RNA **42**, 1470-1477, (2003).
- 54 Miloushev, V. Z. & Palmer, A. G., 3rd. *J Magn Reson* Rho relaxation for two-  
site chemical exchange: General approximations and some exact solutions **117**,  
221-227, (2005).
- 55 SantaLucia, J., Jr., Allawi, H. T. & Seneviratne, P. A. *Biochemistry* Improved  
nearest-neighbor parameters for predicting DNA duplex stability **35**, 3555-3562,  
(1996).
- 56 Coman, D. & Russu, I. M. *Biophys J* A Nuclear Magnetic Resonance  
Investigation of the Energetics of Basepair Opening Pathways in DNA **89**, 3285-  
3292, (2005).
- 57 Fersht, A. R. & Sato, S. *Proc Natl Acad Sci USA* Phi-Value analysis and the  
nature of protein-folding transition states **101**, 7976-7981, (2004).

- 58 Fersht, A. R., Matouschek, A. & Serrano, L. *J Mol Biol* The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding **224**, 771-782, (1992).
- 59 Fedorov, A. *et al. Biochemistry* Transition State Structure of Purine Nucleoside Phosphorylase and Principles of Atomic Motion in Enzymatic Catalysis **40**, 853-860, (2000).
- 60 Somech, R. *et al. Immunol Res* Purine nucleoside phosphorylase deficiency presenting as severe combined immune deficiency **56**, 150-154, (2013).

## Chapter 3

### Equilibrium and Transient Kinetics suggest at least Three Species in the B-to-Z-DNA Transition \*

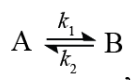
#### 3.1 Introduction

Since the discovery of the exotic left-handed form of DNA nearly half a century ago<sup>1</sup> Z-DNA has been a controversial focal point of DNA structure and dynamics.<sup>2-5</sup> It is now known that Z-DNA formation in pyr-pur repeats is correlated with supercoiling,<sup>6-14</sup> regulation of gene expression,<sup>15-21</sup> DNA damage,<sup>22</sup> the epigenetic modification 5-methylcytosine,<sup>23,24</sup> and the incurable Lupus erythematosus<sup>21,25-28</sup> and Crohn's disease.<sup>29,30</sup> B- and Z-DNA exist in dynamic equilibrium *in vivo*,<sup>31,32</sup> which requires knowledge of the B-to-Z-DNA transition kinetics to understand the role of the transition in biological events. Moreover, understanding the B-Z transition mechanism will allow for a robust understanding of DNA's most peculiar structural change and could motivate the discovery of drugs targeting transiently forming Z-DNA. Z-DNA forming pyrimidine-purine repeat (pyr-pur) sequences undergo the B-Z transition in genomic DNA *in vivo*<sup>23,33</sup> and exist on lengths of ~4 to 16 base-pairs;<sup>34</sup> potentially even longer, as the upper limit of Z-DNA forming sequences in genomic DNA is poorly defined due to ambiguity in structures adopted by long palindromic sequences.<sup>35,36</sup>

---

\*HSA and HMA conceived the idea with help from PJO and CF; HSA prepared samples and carried out the experiments and data analysis with help from HMA.

Kinetic studies seeking to understand the B-to-Z-DNA transition have been reported in the literature, including those for longer Z-DNA forming sequences that robustly describe the B-Z kinetics with at least two states,<sup>37</sup>



and others that exhibit greater complexities.<sup>38,39</sup> Prior to technological advancements in the chemical synthesis of DNA, B-to-Z-DNA transition kinetics employed the use of much longer poly(dG-dC) sequences for ease of isolating from genomic DNA. However, these are often inhomogeneous in length and the pyr-pur repeat nature may not always be uniform, potentially convoluting and masking important features of the transition. While the ability of Z-DNA to accommodate imperfect repeats *in vivo* remains an important question,<sup>40-43</sup> care must be taken when choosing a model system for the purpose of demonstrating robust observables for the B-to-Z-DNA transition. For those reports of complex B-to-Z-DNA transition pathways, including those induced by the Z-DNA binding domain of the human adenosine deaminase enzyme ADAR1 ( $Z\alpha$ ; Section 1.2.1), no observed rate constants are reported which precludes many kinetic models.<sup>44,45</sup> Many B-to-Z-DNA transition mechanistic schemes have been proposed, reviewed by Fuertes and coworkers<sup>46</sup> and discussed in Section 1.2.3, yet none have been explicitly tested. As the majority of these proposed mechanisms postulate intermediate structures that could be discernable by Circular Dichroism (CD), we turn to this powerful and sensitive tool to measure chirality of chemical bonds and subsequently DNA backbone conformation (Figure 1.6). Here, we use CD spectroscopy to focus on characterizing the thermodynamic and kinetic features of the B-Z transition in hexameric pyr-pur repeats, namely 5'-CGCGCG-3' ( $CG_3$ ), anticipating kinetic complexities with the goal of gaining



insights into the B-to-Z-DNA transition mechanism. Moreover, despite extensive proposals of molecular models, a rigorous kinetic model has yet to be presented for the B-to-Z-DNA transition of lengths less than 1,000 base-pairs, leaving much room in the field for variability in analysis. Here we probe the B-to-Z-DNA transition of short (6 nucleotide) CpG repeats to obtain robust experimental observables and hope to provide the conceptual framework for future investigations of the B-Z transition and potentially other biomolecular conformational changes.

## **3.2 Methods**

### 3.2.1 DNA Sample Preparation

Oligonucleotides were purchased from Integrated DNA Technologies with standard desalting. Oligonucleotides were resuspended in milliQ H<sub>2</sub>O and dialyzed against 2-4 L of milliQ H<sub>2</sub>O using a mini dialysis kit with a 1 kDa molecular weight cutoff (GE Healthcare Lifesciences). Dialyzed samples were lyophilized and resuspended in the appropriate phosphate buffer. Oligos for salt-induced transition were resuspended in 15 mM Phosphate Buffer, 0.1 mM EDTA, pH 7.5 (indicated in plots as 'X M' where X is the concentration of NaCl added to induce the B-Z transition). Oligos for Z $\alpha$ -induced transition were resuspended in 15 mM Phosphate Buffer, 0.1 mM EDTA, 25 mM NaCl, pH 7.5. A<sub>260</sub> values were measured and oligo concentrations were calculated using the manufacturer-supplied extinction coefficients. Stock oligos were diluted to 1 mM and stored at -20 °C.

### 3.2.2 Circular Dichroism Thermodynamics

Samples for the thermodynamic NaCl-induced transition were heated to 95 °C for 5 min and allowed to cool to room temperature for at least 10 min before use in sample preparation. Samples for the thermodynamic Z $\alpha$ -induced transition were allowed to reach thermal equilibrium by sitting at room temperature for at least 60 min. CD thermodynamics were measured on a Jasco CD spectropolarimeter equipped with a Peltier temperature control unit hooked up to a recirculating water bath. Wavelength scans were taken from 220-330 nm for salt-induced B-Z transitions and controls, and 240-330 nm for Z $\alpha$ -induced transitions and controls. Scan rate was 100 nm/min, sensitivity was 100 mdeg, data were acquired every 1 nm with an integral time of 1 sec. Samples were allowed 2 minutes to equilibrate to cell temperature and data were acquired at 5 °C. Samples were either 2 mL of 11  $\mu$ M DNA in a 10 mm path length quartz cuvette (Starna Cells) or 200  $\mu$ L of 110  $\mu$ M in a 1 mm path length quartz cell (Starna Cells). Subtraction of background signal was performed using identical solutions for each given measurement excluding DNA in preparation of the blank. A baseline adjustment was performed to make  $\theta_{330\text{nm}} = 0$  where DNA, protein and buffer do not exhibit CD signal.

### 3.2.3 Circular Dichroism Kinetics

CD kinetics were measured on an Aviv 202 CD spectropolarimeter equipped with a Peltier temperature control unit and a recirculating cooling water bath. Temperature was set at 5 °C with a 0.5 °C dead band. Sensitivity was 100 mdeg. Data were acquired every 1 ms for 150 s (CG<sub>3</sub> + 4.0 M, 4.5 M and 5.0 M NaCl), 360 s (CG<sub>3</sub> + 2.5 M, 3.0 M, 3.5 M NaCl-induced) or 1,800 s (CG<sub>3</sub> + 22  $\mu$ M Z $\alpha$ ). Data were acquired every 150 ms for

5,500 s for CG<sub>6</sub> + 4.0 M NaCl. Signal integration times were set equal to data collection times.

1 mM DNA stocks in appropriate buffer containing 0 M NaCl (for salt-induced) or 25 mM NaCl (for Z $\alpha$ -induced) were heated to 95 °C for 5 min and allowed to cool to room temperature for at least 10 minutes then stored on ice until use. 22  $\mu$ L of 1 mM B-DNA was placed in the bottom of a 10 mm path length quartz cuvette. The cuvette was placed in the temperature regulated cell and allowed to equilibrate to the set temperature for 2 min. Salt- and Z $\alpha$ -induced kinetic measurements were performed by rapidly expelling 2 mL of the appropriate inducing solution into the cuvette containing concentrated B-DNA to yield a final DNA concentration of  $\sim$ 11  $\mu$ M. The CD signal was monitored as a function of time at 254 nm. Dead times were  $\sim$ 10 s.

Subtraction of background signal was performed using identical solutions for each given measurement excluding DNA in preparation of the blank. A baseline adjustment was then performed to make  $\theta_{330\text{nm}} = 0$  mdeg according to thermodynamic controls.  $\theta_{254\text{nm}}$  of samples at thermodynamic equilibrium (prepared as described in Section 3.2.2) were acquired to benchmark the exact signal expected at the beginning and end of each transition.

#### 3.2.4 Calculating Fraction of Z-DNA using $\theta_{254\text{nm}}$

The magnitude of signal change at 254 nm is given by<sup>1,37</sup>

$$f_{Z-DNA} = \frac{\theta_i - \theta_{B-DNA}}{\theta_{Z-DNA} - \theta_{B-DNA}} \quad (3.1)$$

where, for a given wavelength,  $f_{Z-DNA}$  represents the fraction of Z-DNA formed,  $\theta_i$  is the CD signal at a particular titration point,  $\theta_{B-DNA}$  and  $\theta_{Z-DNA}$  are the signals of B- and Z-

DNA, respectively. B- and Z-DNA spectra must be chosen with care as outlined in Section 3.2.5.

### 3.2.5 Fitting of CD Wavelength Spectra to a Linear Combination of Two Species.

CD Spectra were acquired as outlined in Section 3.2.2. B-DNA and Z-DNA spectra were assigned as those without and with maximum inducing solution concentrations, respectively. For the NaCl-induced transition B-DNA and Z-DNA references were CG<sub>3</sub> + 0 M NaCl and CG<sub>3</sub> + 5.0 M NaCl both in 15 mM Phosphate, 0.1 mM EDTA, pH 7.5. For the Z $\alpha$ -induced transition B-DNA and Z-DNA references were CG<sub>3</sub> + R=0 Z $\alpha$  and CG<sub>3</sub> + R=2 Z $\alpha$ , both in 15 mM Phosphate, 0.1 mM EDTA, 25 mM NaCl, pH 7.5. Note saturation of signal change was observed for maximum inducing solution scenarios (Figure 3.1). NaCl and Z $\alpha$  titrations of CG<sub>3</sub> were performed separately at 5 °C and the CD signal is represented as

$$\theta = \beta B + \zeta Z \quad (3.2)$$

where  $\theta$  is the observed signal,  $\beta$  is the fraction of B-DNA,  $B$  represents the B-DNA spectrum,  $\zeta$  is the fraction of Z-DNA and  $Z$  represents the Z-DNA spectrum under saturating inducing solution.  $\beta$  and  $\zeta$  were determined using the fitting function *NonLinearModelFit* of Mathematica, the restraint  $\beta + \zeta = 1$  was imposed and  $\beta$  and  $\zeta$  were successively varied by 0.001 in Eq. 3.2 to find the best fit to  $\theta$  when given spectra representing  $B$  and  $Z$ . Error bars were determined from  $\chi^2$  based on fit of Eq. 3.2 to CD signal.

### 3.2.6 Determining the B-Z Transition Midpoint using the Boltzmann Equation

We use the Boltzmann equation for its simplicity in describing conformational changes of biomolecules<sup>47</sup> where

$$f = f_f + \frac{(f_i - f_f)}{1 + e^{(x-x_{0.5})/n'}} \quad (3.3)$$

and  $f$  is the fraction of species at a given concentration of inducing solution,  $x$ ,  $f_i$  and  $f_f$  are the initial and final fraction of the target species,  $x_{0.5}$  is the midpoint of the transition and  $n'$  is a reflection of the cooperativity of the transition.

### 3.2.7 Solving for Residual CD Signal

To solve for the residual CD signal,  $\theta_R$ , we use:

$$\theta_R = \theta_i - (\beta\theta_B + \zeta\theta_Z) \quad (3.4)$$

where  $\theta_i$  is the CD signal under a given set of conditions,  $\theta_B$  and  $\theta_Z$  are the CD spectra of the B and Z species, respectively.

### 3.2.8 Fitting B-to-Z-DNA Transient Kinetics

Time traces of the B-to-Z-DNA transition were fit to the generalized monophasic and biphasic exponential growth functions as given by

$$y = y_f - Ae^{-k_{obs}t} \quad (3.5)$$

and

$$y = y_f - \left( A_1 e^{-k_{obs,1}t} + A_2 e^{-k_{obs,2}t} \right), \quad (3.6)$$

respectively, where  $y$  is the CD signal observed as a function of time,  $y_f$  is the final observed signal,  $A$  (Eq. 3.5),  $A_1$  and  $A_2$  (Eq. 3.6) represent the magnitudes in millidegrees

corresponding to each phase and  $k_{obs}$  (Eq. 3.5),  $k_{obs,1}$  and  $k_{obs,2}$  (Eq. 3.6) represent the observed rate constants for a given transition. Fits were performed without weights in Origin 8.6 (OriginLab) and with  $x^{-2}$  weights in Prism (Graph Pad Software).

### 3.2.9 Isotopically Labeled DNA Sample Preparation

$^{13}\text{C}/^{15}\text{N}$  ZJXN DNA was prepared as described previously<sup>48,49</sup> and in Section 2.2.1.  $Z\alpha$  (9 kDa) was prepared at natural abundance as described previously.<sup>50</sup> His tag cleavage was omitted as his tag does not affect  $Z\alpha$  binding to DNA (data not shown).

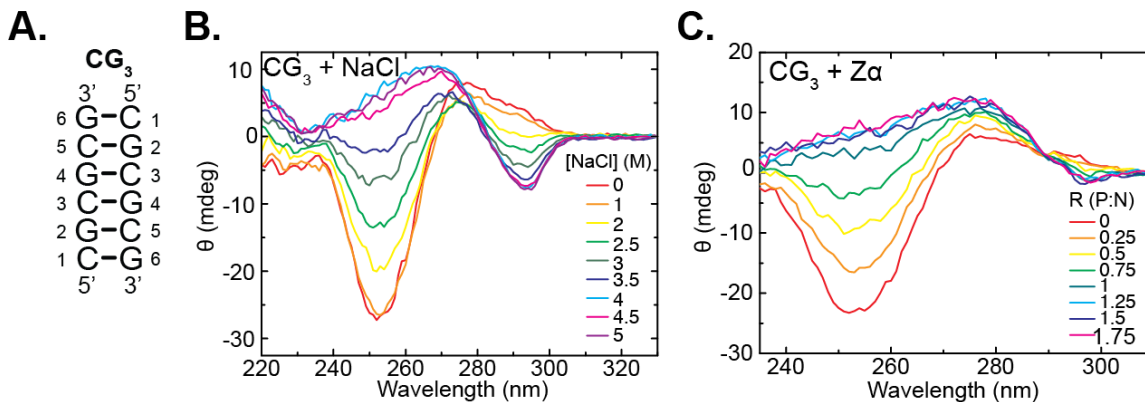
### 3.2.10 NMR Experiments

Buffer was 15 mM Phosphate buffer pH 7.5, 0.1 mM EDTA, 25 mM NaCl for  $^{13}\text{C}/^{15}\text{N}$  ZJXN in absence and presence (R=2) of natural abundance  $Z\alpha$ . Temperature was 25 °C. NMR experiments were ran on ~0.4 mM uniformly  $^{13}\text{C}/^{15}\text{N}$  labeled ZJXN samples. 2D aromatic SOFAST-HMQC experiments were ran on  $Z\alpha$ : $^{13}\text{C}/^{15}\text{N}$  ZJXN R=2 (~27 kDa). Samples were ran on an Agilent 600 MHz NMR spectrometer equipped with a 5 mm triple resonance cryogenic probe.

## **3.3 Results and Discussion**

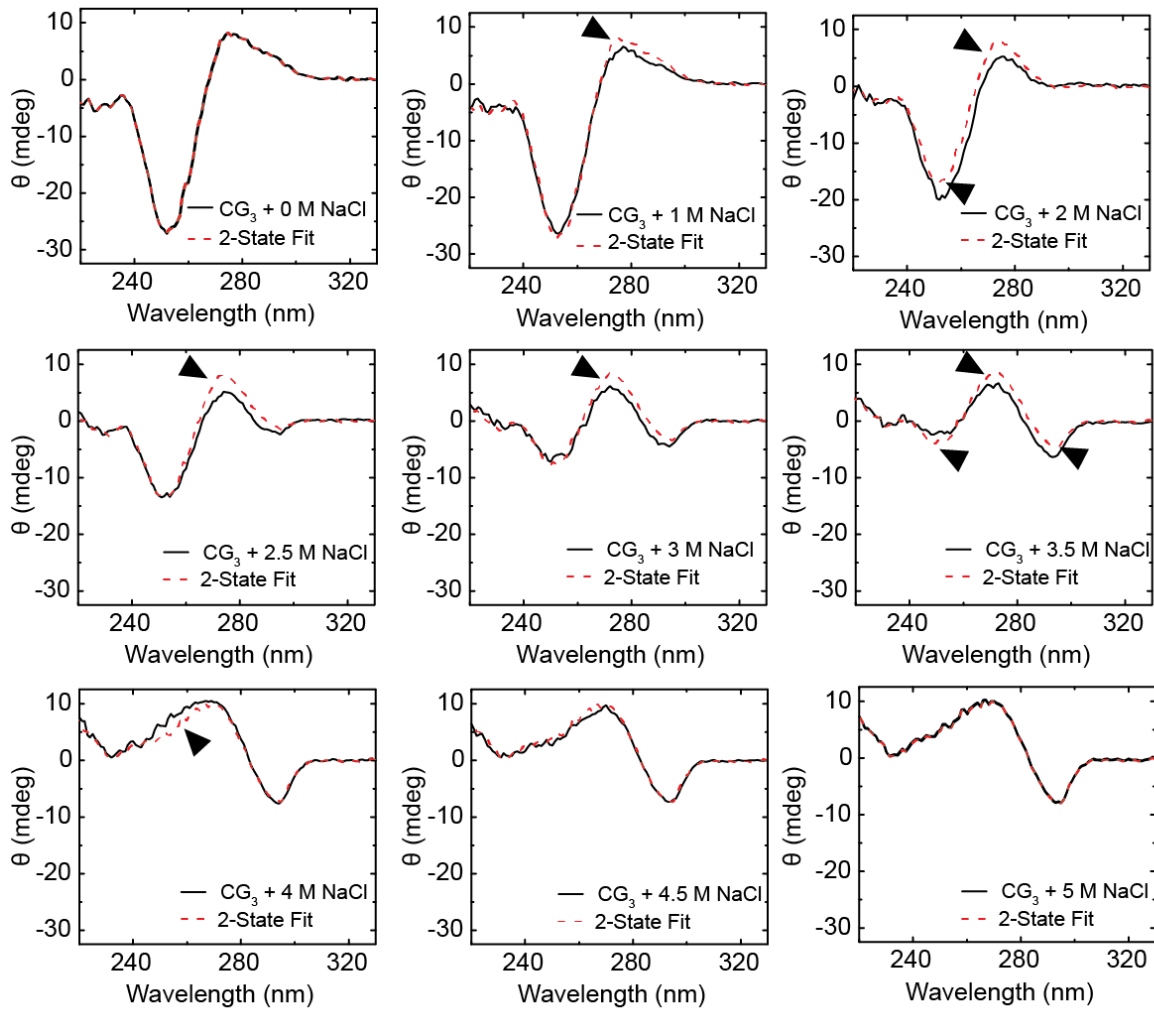
### 3.3.1 The B-to-Z-DNA Transition Involves more than Two Species

We first performed independent titrations of NaCl and  $Z\alpha$  into the pyr-pur repeat  $\text{CG}_3$  under experimental conditions we plan to use for kinetics (Figure 3.1) and observe saturation of the transition at  $\text{NaCl}^{37}$  and  $Z\alpha^{50-52}$  concentrations as demonstrated previously.



**Figure 3.1 | Monitoring of the B-Z transition in CG<sub>3</sub> DNA.** Inducing the B-Z transition in (A) CG<sub>3</sub> with (B) NaCl and (C) Z $\alpha$  where insets annotate inducing solution amount. “R P:N” is the protein:DNA ratio. 15 mM Phosphate buffer pH 7.5, 0.1 mM EDTA, (B) 0 M or (C) 25 mM NaCl, 5 °C.

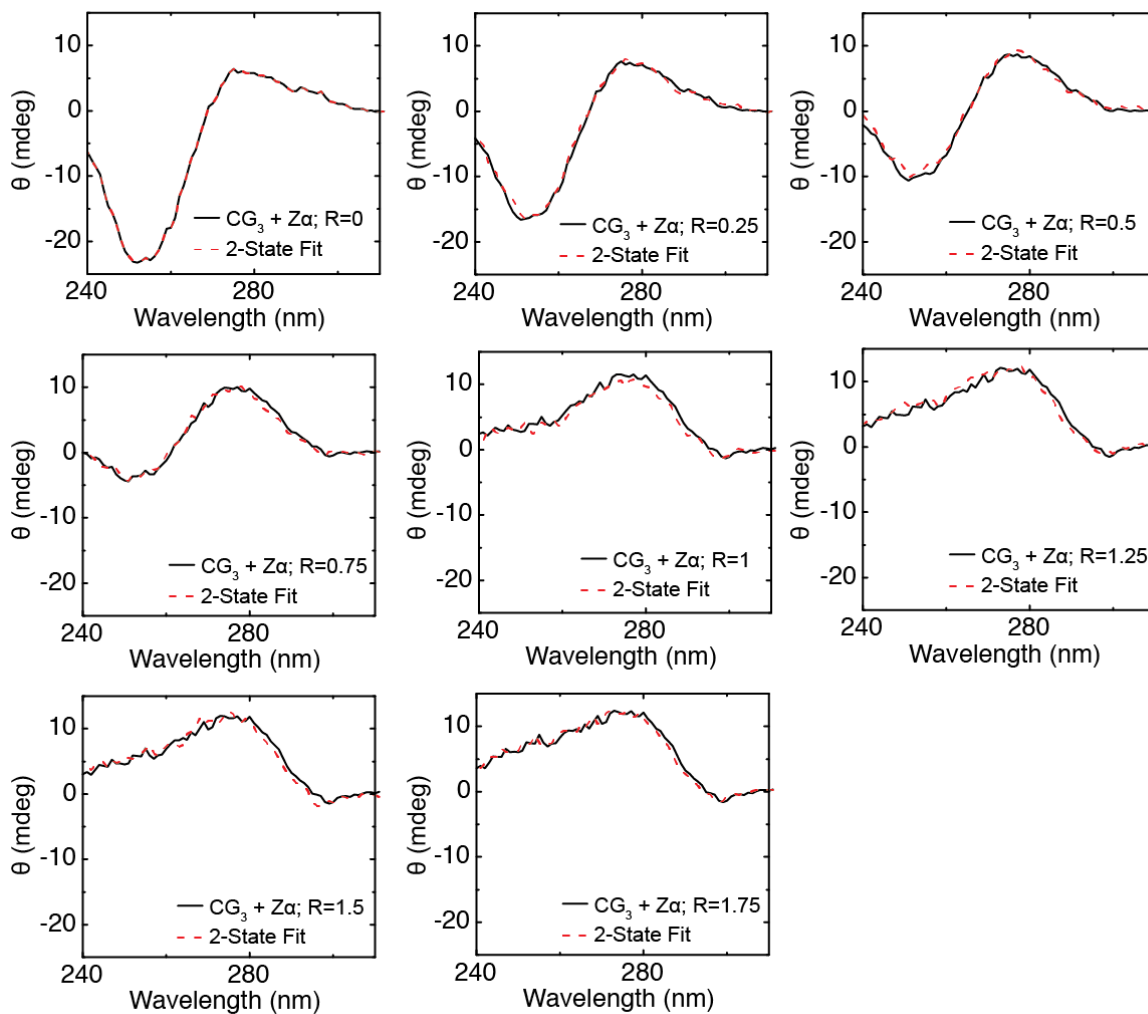
Our lab recently developed a technique to quantitatively determine fractions of B- and Z-DNA accounting for a given CD signal that relies on the concept that the CD signal is a linear combination of the species present in solution.<sup>53</sup> The observed CD signal for a thermodynamic measurement of DNA in equilibrium between B- and Z-forms is given by  $\theta = \beta B + \zeta Z$  (Eq. 3.2) where  $\theta$  is the observed signal,  $\beta$  is the fraction of B-DNA,  $B$  represents the B-DNA spectrum,  $\zeta$  is the fraction of Z-DNA,  $Z$  represents the Z-DNA spectrum under saturating inducing solution and  $\beta + \zeta = 1$ .  $\beta$  and  $\zeta$  were determined using the fitting function *NonLinearModelFit* of Mathematica to iteratively vary  $\beta$  and  $\zeta$ . We fit these titrations to a linear combination of B- and Z-DNA spectra (Section 3.2.5) and next back-calculated a CD signal to represent the fit to Eq. 3.2. Comparing the fits to the observed signal reveals poor recapitulation of NaCl-induced spectra, especially at intermediate salt concentrations, suggesting a 2-state model cannot fully describe the NaCl-induced B-to-Z-DNA transition (Figure 3.2). We note these discrepancies occur primarily in the 255- 274 nm and secondarily near 290 nm regions.



**Figure 3.2 | Linear Combination of B- and Z-DNA Spectra for the Salt-Induced Transition.** NaCl-induced B-Z transition from Figure 3.1 (black) was fit using *NonLinearModelFit* in Mathematica 9.0 as described in Section 3.2.5 and observed signal was calculated using Eq. 3.2 (dashed red). Black arrow heads highlight poor recapitulation of observed signal. 15 mM Phosphate buffer, 0.1 mM EDTA, pH 7.5, 5°C. NaCl concentration indicated in plot inset.



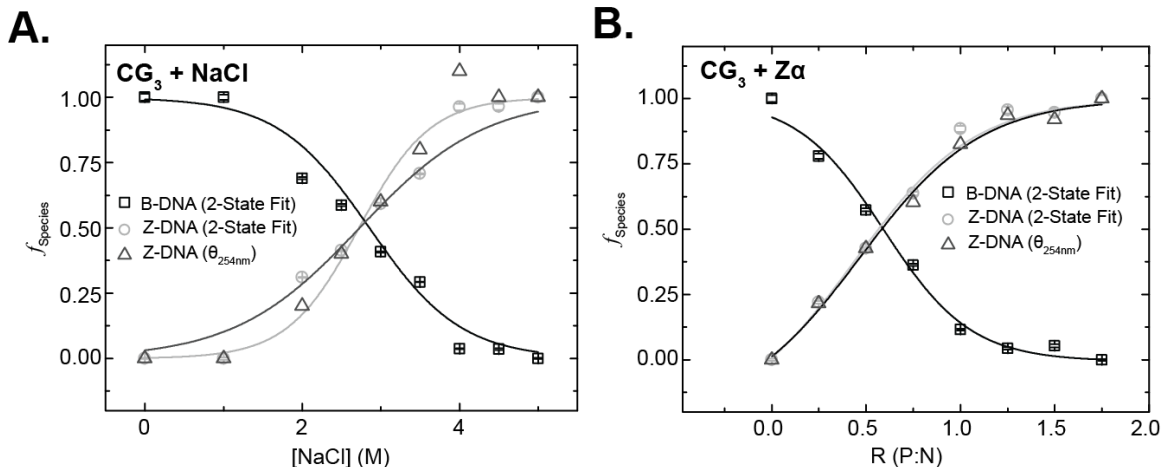
We next examined the fractional species contribution of the  $Z\alpha$ -induced B-Z transition (Figure 3.3) using the method outlined in Section 3.2.5 and used for the NaCl-induced transition. The spectral recapitulation is quite precise and does not merit the introduction of a model containing at least three species but also does not preclude a complex kinetic mechanism beyond a two-state model.



**Figure 3.3 | Linear Combination of B- and Z-DNA Spectra for the  $Z\alpha$ -Induced Transition.**  $Z\alpha$ -induced B-Z transition from Figure 3.1 (black) was fit using *NonLinearModelFit* in Mathematica 9.0 as described in Section 3.2.5 and observed signal was calculated from best fit using Eq. 3.2 (dashed red). 15 mM Phosphate buffer, 0.1 mM EDTA, 25 mM NaCl, pH 7.5, 5°C.  $Z\alpha$ : $CG_3$  ratio denoted as ‘R’.

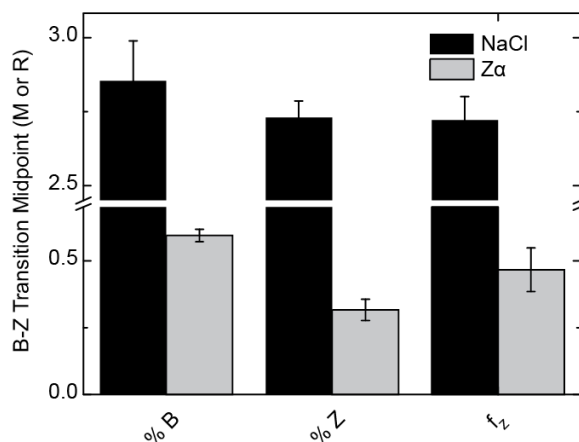
Reasons for no detectable deviation from a 2-state fit for titrations monitored by CD include cases where a third species or intermediate is at very low population (<~5 %), or is unstructured such that its molar absorptivity constant is nearly zero.

From these titrations we represent the extent of transition as the fraction of species obtained from the two-state fits (Figure 3.2 and 3.3) and benchmark these against the conventional method of calculating the fraction of Z-DNA formation as given by  $f_{Z-DNA} = (\theta_i - \theta_{B-DNA}) / (\theta_{Z-DNA} - \theta_{B-DNA})$  where, for a given wavelength,  $f_{Z-DNA}$  represents the fraction of Z-DNA formed,  $\theta_i$  is the CD signal at a particular titration point,  $\theta_{B-DNA}$  and  $\theta_{Z-DNA}$  are the signals of B- and Z-DNA, respectively. We chose 254 nm for its common use as a metric for the B-to-Z-DNA transition.<sup>1,37</sup>  $f_{Z-DNA}$  profiles, as well as those from 2-state fits, qualitatively recapitulate the corresponding transition midpoints previously reported for NaCl-<sup>1</sup> and Z $\alpha$ -<sup>51</sup> induced transitions with midpoints near 2.5 M NaCl for the NaCl-induced transition and a minimal lag before Z-DNA formation in the Z $\alpha$ -induced transition followed by saturation at R=2 (Figure 3.4 triangles).



**Figure 3.4 | Fraction of B- and Z-DNA Species.** (A) NaCl and (B) Zα-induced transitions. B- (black squares) and Z- (light grey circles) were calculated from *NonLinearModelFit* in Mathematica 9.0 as described in Section 3.2.5 and used previously.<sup>53</sup> Error bars were determined from  $\chi^2$  based on fit of Eq. 3.2 to CD signal.  $\theta_{254\text{nm}}$  was used to calculate fraction Z-DNA with Eq. 3.1 (dark grey triangles). Data are fit to Eq. 3.3.

Since we impose the restraint that  $\beta + \zeta = 1$  in Eq. 3.2,  $\beta = \zeta = 0.5$  is defined as the midpoint of the transition. In a 2-state transition, B and Z are expected to exhibit the same midpoint. We therefore determined the midpoints of the NaCl- and Zα-induced B-Z transitions using the Boltzmann equation (Eq. 3.3) to describe the species fractional profiles as the point of inflection of the line of best fit (Figure 3.4).<sup>47,54</sup> We note we also fit the species fractional profiles to the Hill equation,<sup>55</sup> and poorer fits resulted (data not shown;  $R^2 = 0.96-0.98$  as compared to  $0.98-0.99$  for Boltzmann). The percent Z-DNA obtained from the 2-state fitting as compared to the calculation of  $f_{Z-DNA}$  using Eq. 3.1 differ noticeably for the NaCl-induced transition, particularly at points where the DNA is most sensitive to NaCl concentration changes (Figure 3.4 A) which reflects the small but noticeable discrepancies in 2-state fits (Figure 3.2 and 3.3).

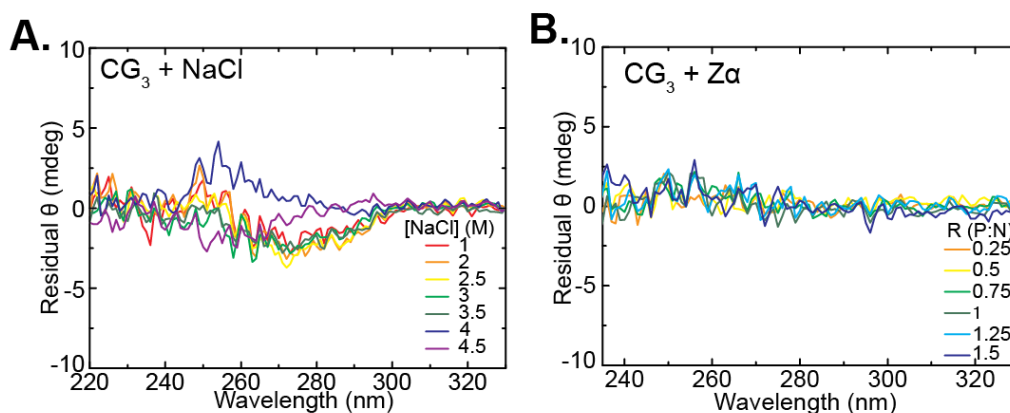


**Figure 3.5 | B-Z Transition Midpoints.** Shown are the B-Z transition midpoints for the two fitting techniques (2-State Fit and  $\theta_{254\text{nm}}$ ). Solid colored and white bars represent NaCl- and Z $\alpha$ -induced transitions, respectively. Error bars were determined from  $\chi^2$  based on fit of Eq. 3.3 to the fraction of species as described in Sections 3.2.4 and 3.2.5.

### 3.3.2 Residual Signal Demonstrates CG<sub>3</sub> can Adopt at least One Non-B/Z Structure

Upon inspection of the titrations in Figure 3.1 it becomes apparent an isosbestic point is absent in the NaCl titration. An isosbestic point is a hallmark of a two-state transition and its absence subsequently suggests the presence of at least a third species at minimally one titration point. To understand why an isosbestic point does not exist in the NaCl titration of CG<sub>3</sub> we calculated the residual signal at each titration point between the B- and Z-DNA spectra as given by  $\theta_r = \theta_i - (\beta\theta_B + \zeta\theta_Z)$  (Eq. 3.4) where  $\theta_i$  is the CD signal under a given set of conditions,  $\theta_B$  and  $\theta_Z$  are the CD spectra of the B and Z species, respectively. Eq. 3.4 reveals measurable residuals above the noise for the NaCl-induced transition, but not so for that of Z $\alpha$ . These results demonstrate CG<sub>3</sub> can adopt more than the B- and Z-DNA conformations in amounts detectable by CD in the case of the NaCl-induced transition. Additionally, alternative structures may exist in the Z $\alpha$ -induced transition at significantly low populations or lifetimes such that they do not

contribute to the CD spectrum. It is also possible any structures on-pathway in the transition resemble B- or Z-DNA to a great degree (for examples see Figure 1.4), owing to why the residuals are so low for NaCl- and either absent or in the noise for the  $Z\alpha$ -induced transition. Overall the analyses of the B-Z transition under equilibrium conditions present evidence for at least three species that are present under equilibrium conditions of the B-to-Z-DNA transition.



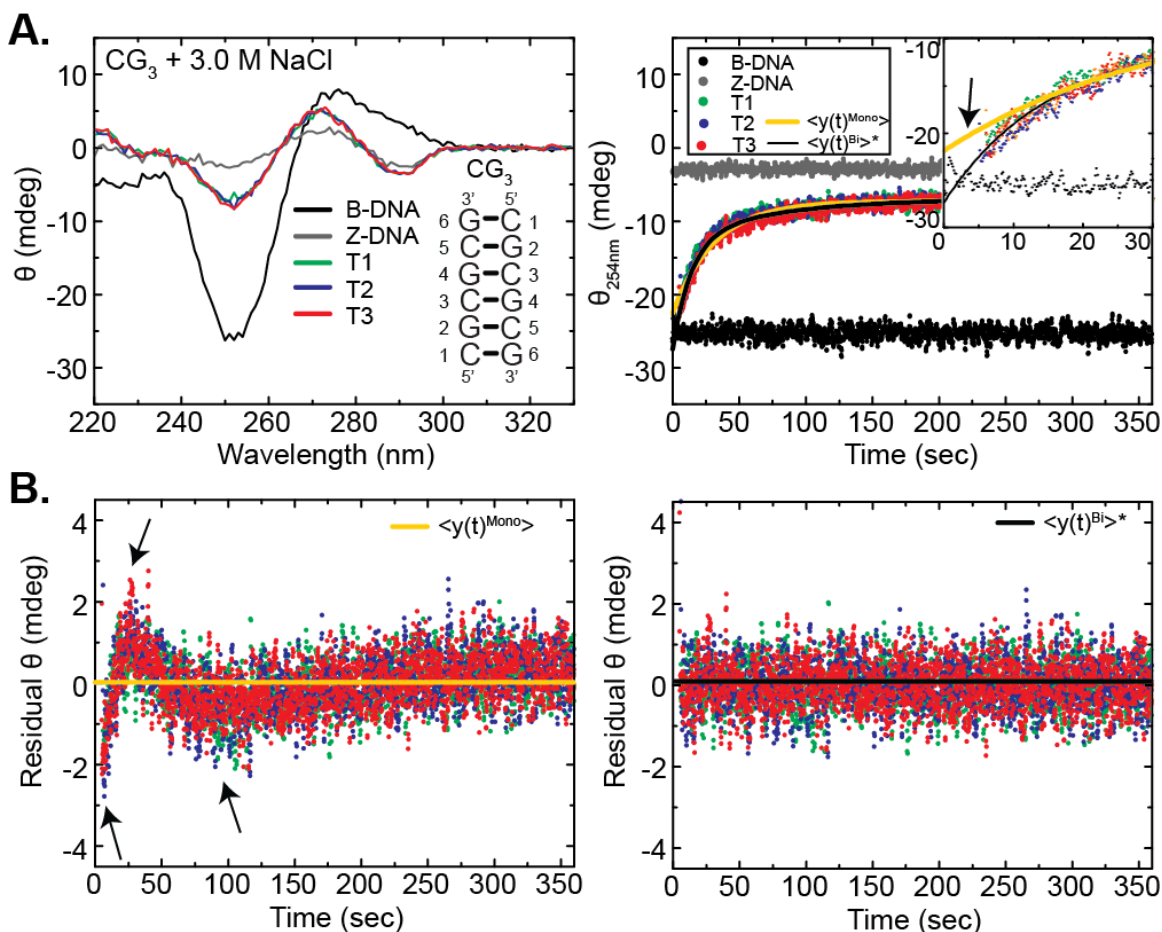
**Figure 3.6 | Residual Circular Dichroism Signal of the B-to-Z-DNA Transition of CG<sub>3</sub>.** (A) NaCl- and (B)  $Z\alpha$ -induced residuals calculated from Eq. 3.4 using B- and Z-DNA references as outlined in Section 3.2.5. 15 mM Phosphate buffer, 0.1 mM EDTA, pH 7.5, 5 °C and (B) 25 mM NaCl.

Although the residual signal of the NaCl-induced transition has a low magnitude as compared to B- and Z-DNA, a unique profile is exhibited distinct from that of B-, A- and Z-forms, quadruplexes and hairpins.<sup>56-60</sup> The literature lacks robust controls for CD signal of single-stranded DNA and there is indirect evidence that single-stranded DNA can adopt a helical structure based on force pulling techniques.<sup>61</sup> This prompted us to investigate the nature of the single-stranded DNA structure and compare our findings with the residual signal of the NaCl-induced transition (Figure 3.6 A). Consistent with the force pulling experiments,<sup>61</sup> we observe a nominal signal of single-stranded DNA, even

under denaturing conditions of 8 M urea and 95 °C (Figure A2.1). The CD signature of single-stranded DNA is B-like in shape with a maximum  $|\theta| \sim 5$  mdeg and is inconsistent with the signature observed for the residuals of the NaCl-induced transition, suggesting the residual CD signature is caused by a unique DNA structural element not previously reported in the literature. We note the residual signal observed for the NaCl-induced transition is not necessarily representative of an entire structure since the B-like and Z-like contributions have been removed from the spectrum, but would rather be a contribution to an overall CD signature of a given CD signal.

### 3.3.3 Biphasic Nature of the B-Z Transition

Thus far we have presented evidence that the B-Z transition contains at least three species using measurements with samples at equilibrium. Next we sought to investigate the kinetic transition from B- to Z-DNA using CD at 254 nm, where the largest signal change is observed. Interestingly, we observe a biphasic transition (Figure 3.7) that is verified using the statistical F- and AIC tests (Section 1.5). The F value takes into account the residual signal after fitting a data set to two nested models relative to the degrees of freedom within the models and tests the null hypothesis that the more complex model does not fit the data significantly better than that of the restricted (Eq. 1.18). The AIC test uses a likelihood function to predict the likelihood of the data for a given model (Eq. 1.19)



**Figure 3.7 | B-Z Transient Kinetics of CG<sub>3</sub> using 3.0 M NaCl.** (A) Wavelength scan controls and corresponding measurements with transients at 254 nm. Inset highlights bias in monoexponential fit that does not accurately capture first 25 seconds of transition and qualitative improvement of fit for biexponential fit. (B) Residuals of monophasic and biphasic fits (Eq.'s 3.5 and 3.6, respectively). “T” indicates respective trials. Arrows highlight poorly distributed residual signal about the fit for monoexponential fit. \* Indicates significantly better fit as determined by AIC and F-Tests with a 95 % confidence level.

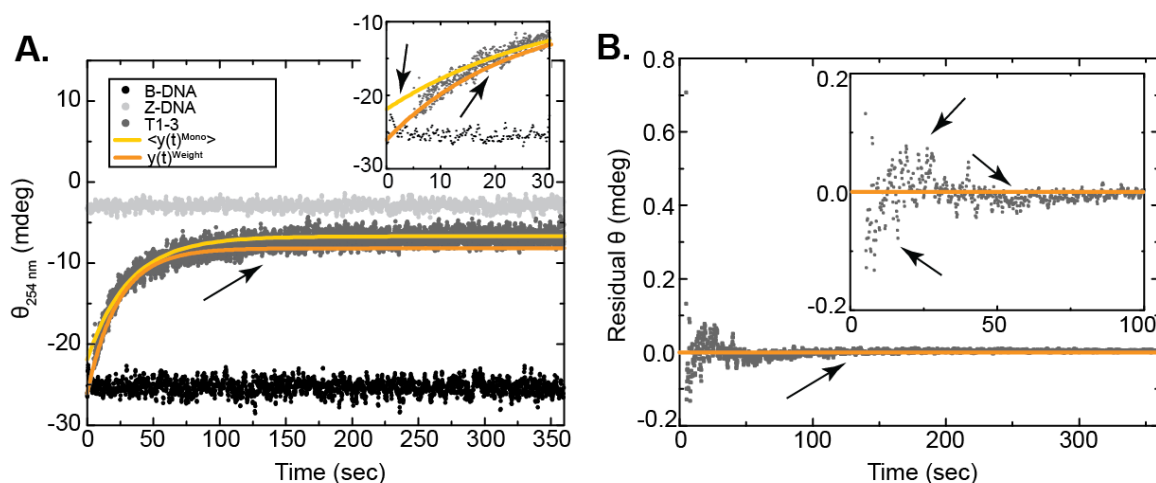
In Figure 3.7 we present a typical kinetics experiment for a given set of experimental conditions. Transient kinetics are performed and recorded repetitively (Figure 3.7 A, right) and thermodynamic controls of beginning B-DNA and end point samples. The same samples are measured with a wavelength scan to control for variations in overall structure (Figure 3.7 A, left). We note controls were performed where one component was systematically omitted from kinetic trials including DNA or high salt during the

rapid mixing just before measurement. In all cases no systematic change consistent with the B-Z transition was observed (data not shown). Data were first fit to a monophasic exponential function as given by  $y = y_f - Ae^{-k_{obs}t}$  (Eq. 3.5) where  $y$  is the CD signal observed as a function of time,  $y_f$  is the final observed signal,  $A$  is the magnitude in millidegrees corresponding to the signal change and  $k_{obs}$  is the observed rate constant for a given transition. Poor fits were observed in the first  $\sim 150$  s, especially during the first 30 s (Figure 3.7 A, right), which inspired us to fit our data to the sum of two exponential terms. The biphasic exponential growth function is given by  $y = y_f - (A_1e^{-k_{obs,1}t} + A_2e^{-k_{obs,2}t})$  (Eq. 3.6) where  $A_1$  and  $A_2$  (Eq. 3.5) and are the amplitudes of signal change for their corresponding phases and  $k_{obs,1}$  and  $k_{obs,2}$  are the observed rate constants for each respective phase (Figure 3.7 A, right). The discrepancies between the monophasic equation and actual data are best observed visually in the residual plots shown in Figure 3.7 B. Residual plots from a fit to a model that describes all the events giving rise to the data should have an equal number of points above and below the line with most points concentrated near residual signal= 0 and should not vary as a function of the dependent variable. Residuals of fits to Eq. 3.5 show clear, systematic biases that are removed when fitting to Eq. 3.6

Despite the more visually appealing Eq. 3.6 fit as compared to that of Eq. 3.5, model selection must be guided by statistics for cases where differences are less than dramatic when moving from a less to more complex model. F- and AIC- tests (Section 1.6 for theory and formulae) selected the more complex, biphasic model to describe the B-to-Z-DNA transient kinetics with a confidence level of 95 % (indicated by a \* in Figure 3.7). Weighted fits can be employed when the standard deviation of a



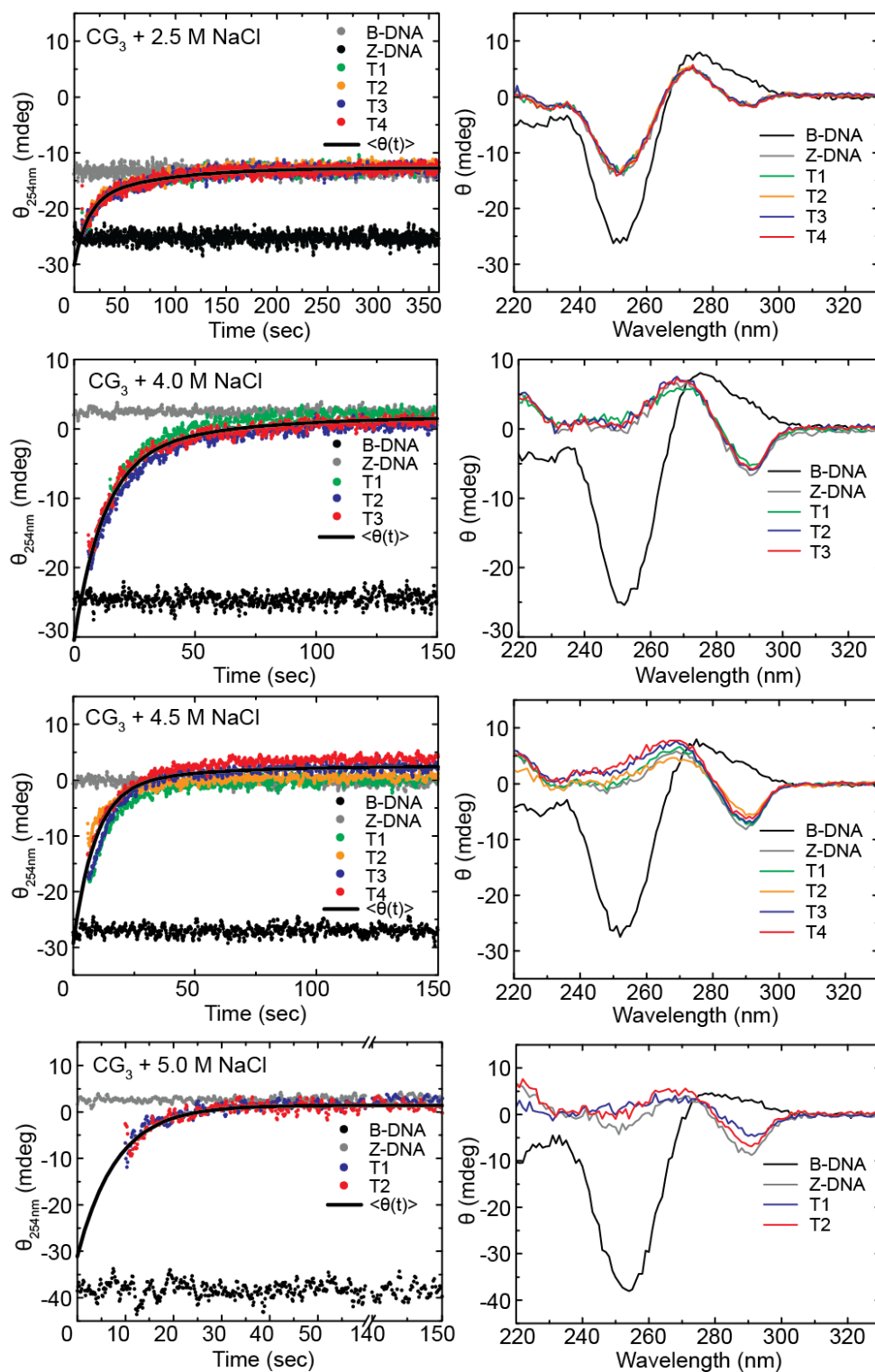
measurement varies as a function of the dependent or independent variable.<sup>62</sup> We used Prism 6 (Graph Pad Software) to fit representative transients with  $x^{-2}$  weighting to account for the fact that the residuals of fits to Eq. 3.5 without weighting (Figure 3.7 B, left) are due to the fast phase present in the transients. Weighting in the form of  $x^{-2}$  will decrease the bias toward fitting the latter portion of the curve and allow for the beginning of the curve to dominate the fit to a greater extent. As expected,  $x^{-2}$  weighted fits of transients to Eq. 3.5 better described a portion of the first phase. However, they fail to capture the subsequent variations in the transient curve (Figure 3.8).



**Figure 3.8 | Comparison of Monophasic Fit with and without Weights.** Shown are (A) best monoexponential fits to Eq. 3.5 without (yellow) and with (orange)  $x^{-2}$  weights. (B) Residual signal from orange line in (A) is shown for a representative trial. Arrows highlight (A) exceptionally poor regions of fit or (B) strong biases in residuals. “T” indicates respective trials. Transients and controls are for CG<sub>3</sub> B-to-Z-DNA transition induced by 3.0 M NaCl in 15 mM Phosphate buffer pH 7.5, 0.1 mM EDTA, 5°C.

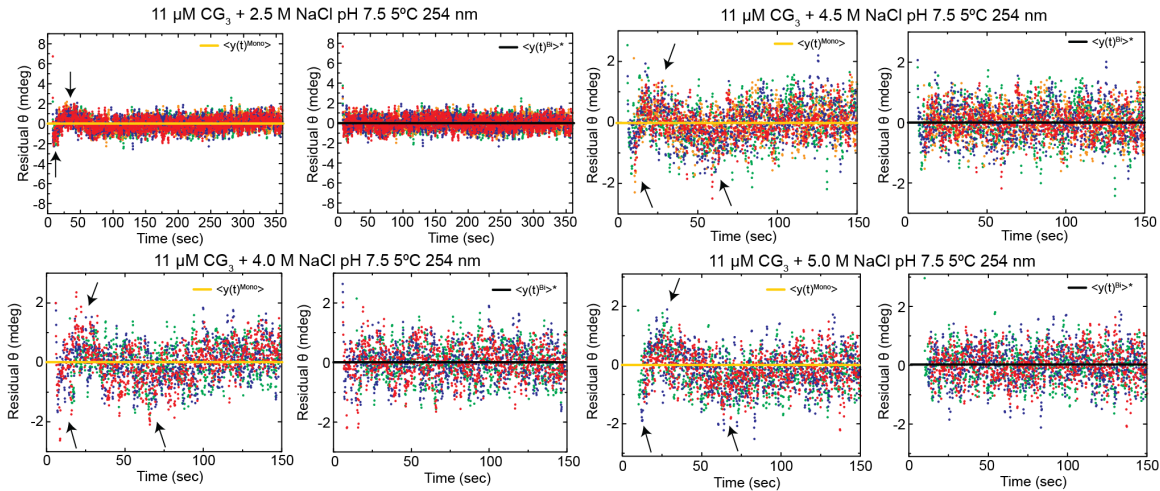
### 3.3.4 Preliminary Insights into Potential B-to-Z-DNA Transition Kinetic Models

Of course, numerous models can give rise to biphasic behavior, including (at least) two reactions in series or parallel (Figure A2.2). To gain insight into potential models of the B-to-Z-DNA transition, transient kinetics were performed using a wide range of NaCl concentrations to induce the B-to-Z-DNA transition (Figure 3.9).



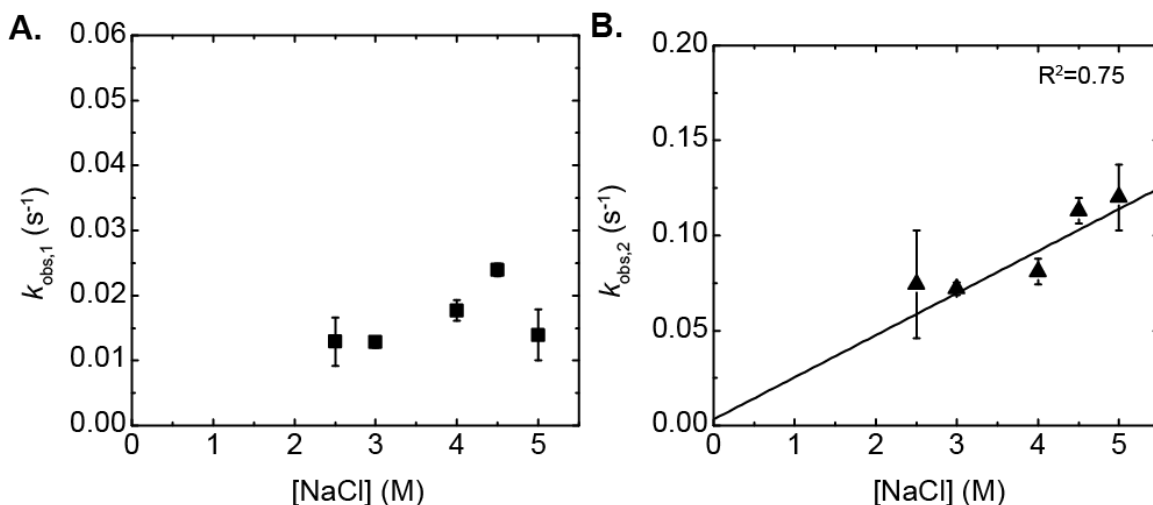
**Figure 3.9 | NaCl-Induced CG<sub>3</sub> Transient Kinetics.** Shown are (left) transient kinetics measured at 254 nm and (right) wavelength scan controls for corresponding samples. Parameters obtained from best fit of individual trials to the biphasic Eq. 3.6 were averaged and plotted (black line). “T” indicates respective trials.

In all cases from 2.5 M- 5.0 M NaCl inducing solution the kinetic traces exhibited clear biphasic behavior (Figure 3.10). The robust and uniform nature of the biases in residuals of fits to Eq. 3.5 support the notion that the B-to-Z-DNA transition is, in fact, giving rise to the additional phase. Moreover, the change in time points of the maximum residual biases correlate with the observed rate constants, i.e. the faster the transition, the earlier the biases in residuals of monoexponential fits appear (arrows in Figure 3.10). These observations suggest that the biases in residuals are not due to an anomaly such as poor temperature adjustment before rapid mixing of DNA and inducing solution. If such a scenario gave rise to the biphasic nature, the residual signature would remain constant so long as temperature does. Here, all our kinetics are performed at 5 °C and residual biases do not remain constant. These results present evidence for a mechanism beyond a single kinetic step and suggest at least two steps (hereon out referred to simply as two steps) occur during the B-to-Z-DNA transition, but alone do not yield any information on the order of the two steps.



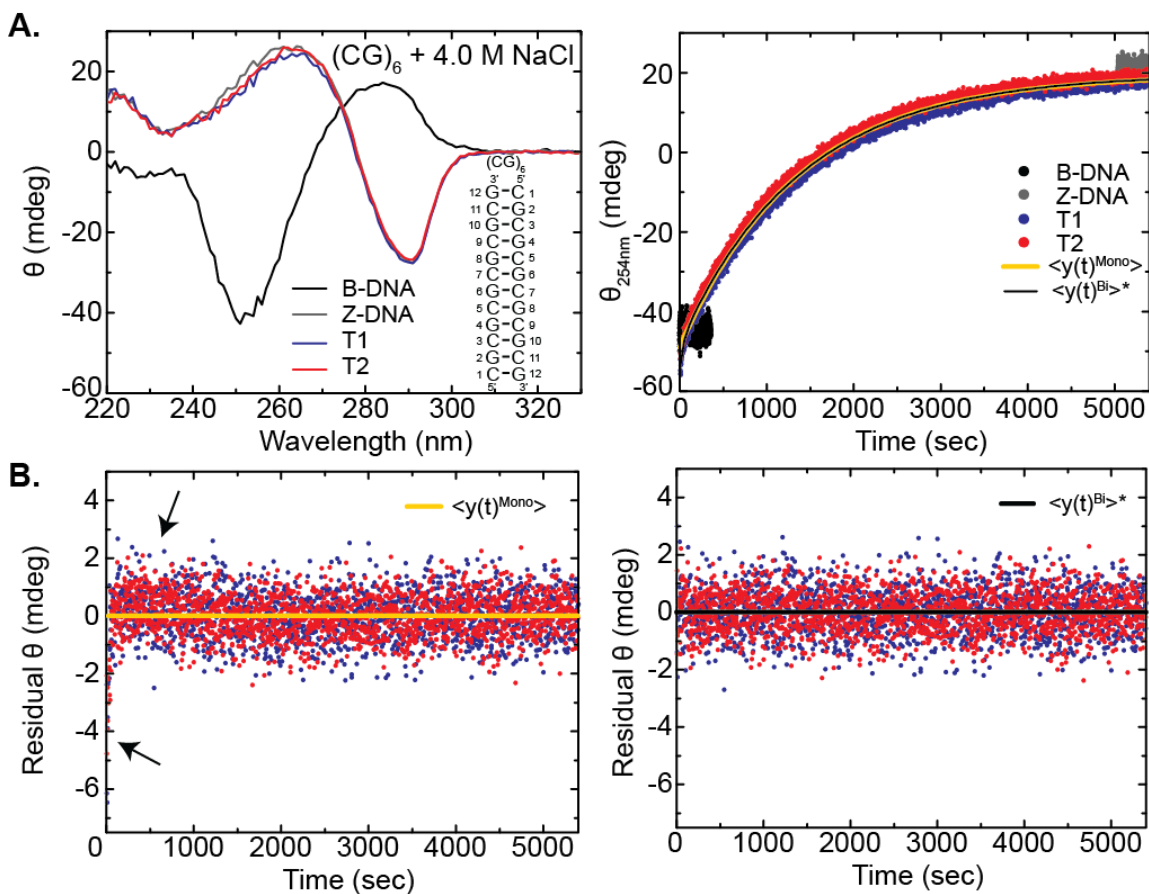
**Figure 3.10 | CG<sub>3</sub> Residual Signal of Fits of Transient Kinetics of the NaCl-Induced B-to-Z-DNA Transition.** Residuals for monophasic (left) and biphasic (right) best fit lines to Eq.'s 3.5 and 3.6, respectively. Arrows denote biases from the best fit to the monophasic Eq. 3.5. Data from respective trials are shown in different colors.

As a quality control we plot the observed rate constants versus concentration of inducing solution (here, NaCl) to potentially reveal key mechanistic features of a transition (Figure 3.11). The faster phase ( $k_{obs,2}$ ) has a linear dependence on NaCl and the slower phase could resemble that of a hyperbolic function, consistent with a unimolecular two-step reversible reaction.<sup>63</sup> We note that this assessment is a first approximation due to the high variability of  $k_{obs,1}$  relative to potential hyperbolic functions that describes the slower observed rate constant behavior as a function of inducing solution for a unimolecular reversible two-step reaction.



**Figure 3.11 | Observed Rate Constants.** Average (A)  $k_{\text{obs},1}$  and (B)  $k_{\text{obs},2}$  obtained from fits of transients to Eq. 3.6. Error bars represent standard deviation if independent trials. 15 mM Phosphate buffer, 0.1 mM EDTA, pH 7.5, 5 °C. For amplitudes see Appendix 2.

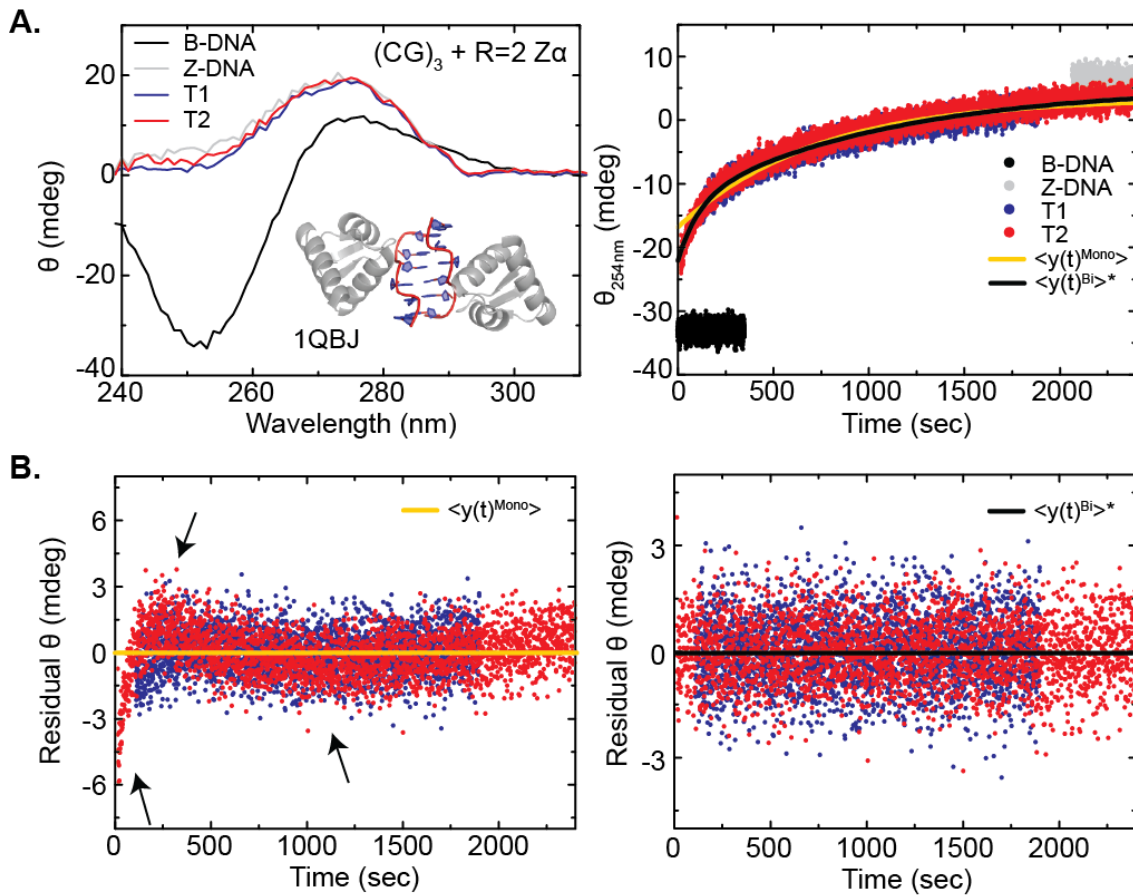
To test for the possibility that the biphasic nature of the B-to-Z-DNA transition is an abnormality due to the very short nature of CG<sub>3</sub>, we monitored the NaCl-induced transition of CG<sub>6</sub> (Figure 3.12 A). Again, a reproducible and systematic bias is observed in the residuals of the monophasic fit to Eq. 3.5 (Figure 3.12 B) and AIC and F-tests (Section 1.5) both select for the biphasic fit to Eq. 3.6 (data not shown). We note that we cannot discern whether the biphasic nature of CG<sub>6</sub> + 4.0 M NaCl as compared to that of CG<sub>3</sub> is due to a similar or distinct process. Additional experiments that perturb the kinetics of the B-to-Z-DNA transition in both systems would help resolve this point. The CG<sub>6</sub> + 4.0 M NaCl experiment does, however, demonstrate that hexameric and dodecameric CpG repeats both possess complex mechanistic features of the B-to-Z-DNA transition.



**Figure 3.12 | Transient Kinetics, Controls and Residuals of the NaCl-Induced B-to-Z-DNA Transition in  $CG_6$ .** (A) Wavelength scan controls and corresponding measurements with transients at 254 nm. (B) Residuals of monophasic and biphasic fits (Eq.'s 3.5 and 3.6, respectively). Arrows highlight poorly distributed residual signal about the fit for monoexponential fit. “T” indicates respective trials. \* Indicates significantly better fit as determined by AIC and F-Tests with a 95 % confidence level.

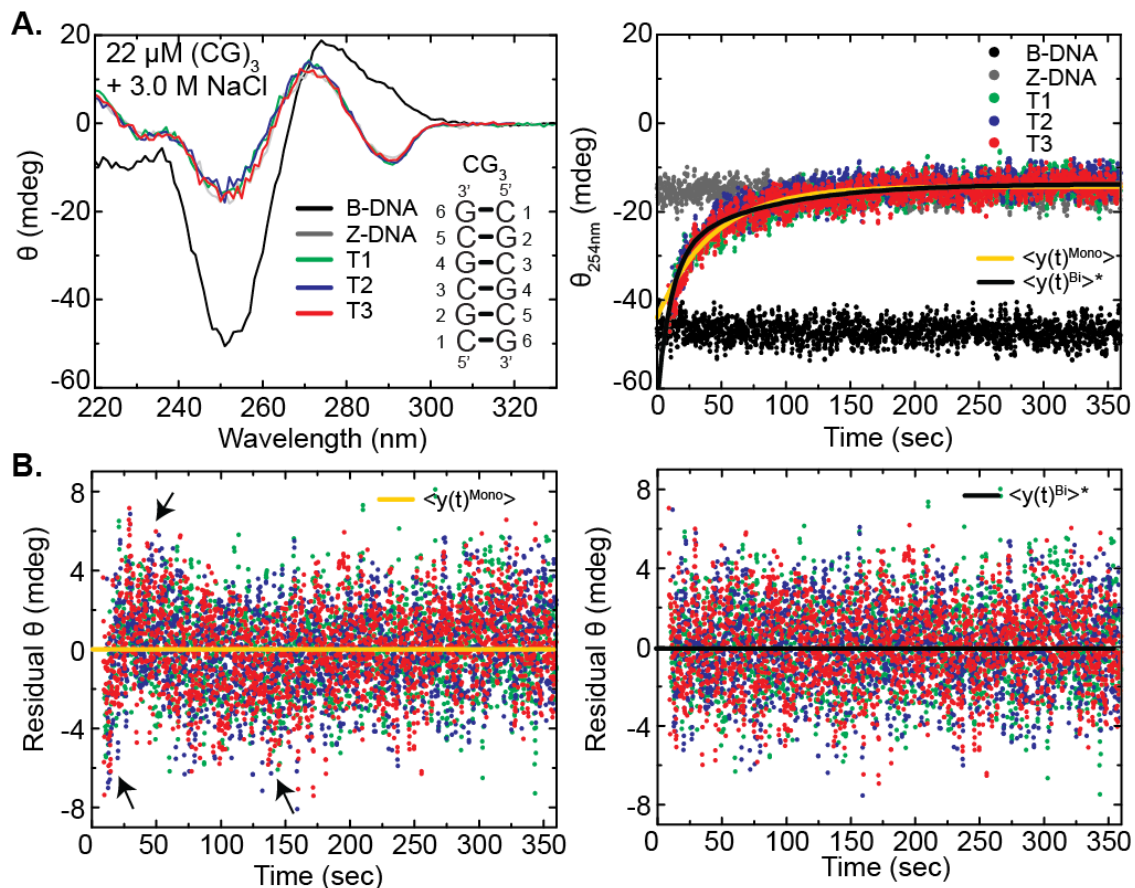
To test for the prospect that the biphasic nature of the B-to-Z-DNA transition is possibly due to the physiologically unlikely salt concentrations used to induce the transition, we monitored the  $Z\alpha$ -induced transition of  $CG_3$  (Figure 3.13 A). Again, clear biphasic behavior was observed, here with larger magnitudes observed in the residuals of monophasic fits than the salt-induced transitions (Figure 3.13 B). The overall transition occurs on a similar timescale to that observed previously for C-G rich hexamers induced by  $Z\alpha$ , albeit at 10 °C in different buffer and protein:DNA ratios.<sup>45</sup> We note the

mechanistic model may have changed for either  $CG_6 + 4.0 \text{ M NaCl}$  or  $CG_3 + 22 \mu\text{M Z}\alpha$  scenarios and that further characterization of these two additional systems is necessary to robustly understand their differences from that of  $CG_3 + \text{NaCl}$ . It would be informative to measure the  $Z\alpha$ -induced kinetics of  $CG_6$  to gain further insight into the nature of the relative amplitudes of the two phases as a function of sequence length and inducing solution.



**Figure 3.13 | Transient Kinetics, Controls and Residuals of the  $Z\alpha$ -Induced B-to-Z-DNA Transition in  $CG_3$ .** (A) DNA and 9 kDa protein used to measure kinetics. PDB ID: 1QBJ. (B) Wavelength scan controls and corresponding measurements with transients at 254 nm. (C) Residuals of monophasic and biphasic fits (Eq.s 3.5 and 3.6, respectively). Arrows highlight poorly distributed residual signal about the fit for monoexponential fit. “T” indicates respective trials. \* Indicates significantly better fit as determined by AIC and F-Tests with a 95 % confidence level.

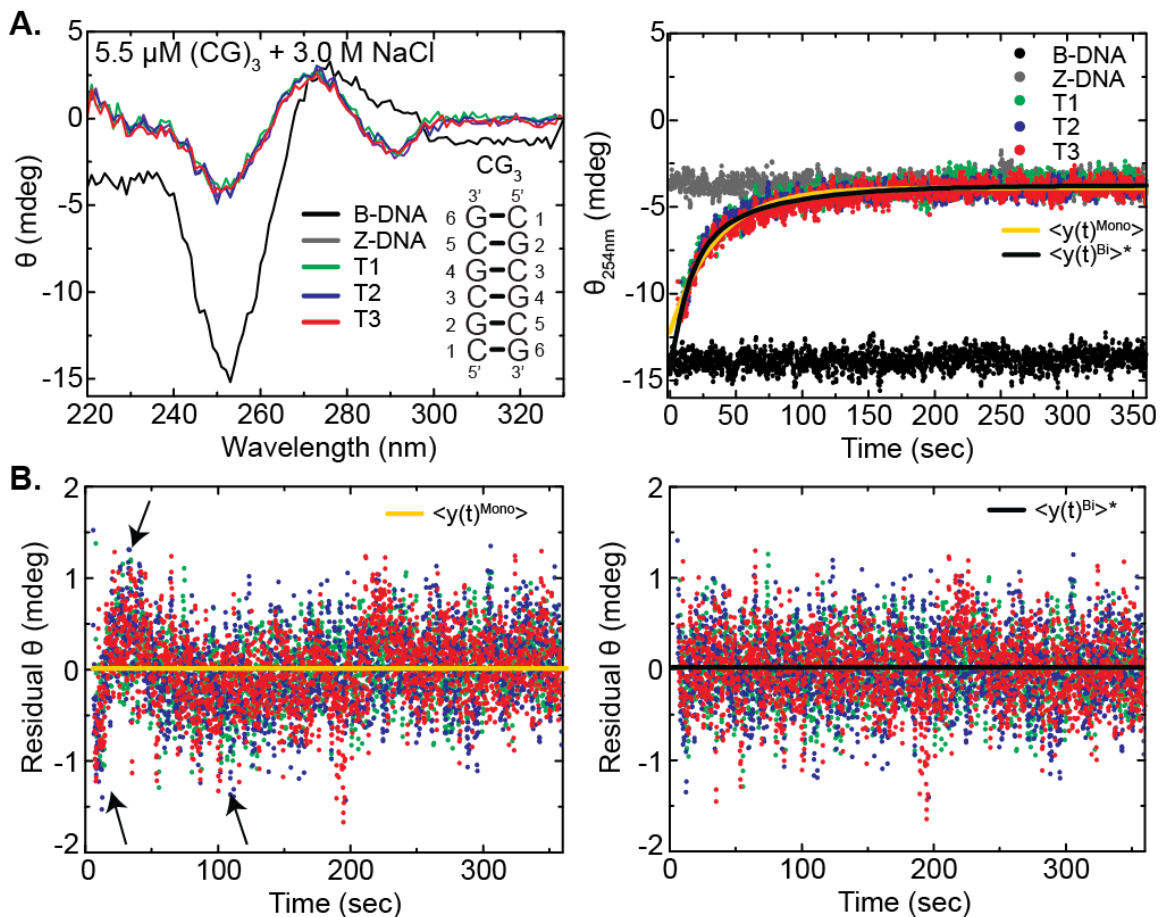
Finally, to confirm the B-to-Z-DNA transition is unimolecular for CG<sub>3</sub> we varied the DNA concentration in the NaCl-induced kinetics (here, 3.0 M NaCl). We increased the concentration two-fold and note concentration increase was limited by the upper limit on the dynode response of the CD instrument. The dynode response is proportional to the signal magnitude and loses the ability to detect CD signal accurately when saturated.



**Figure 3.14 | NaCl-Induced Transient Kinetics of CG<sub>3</sub> with Two-Fold Concentration Increase.** (A) Wavelength scan controls and corresponding measurements with transients at 254 nm. (B) Residuals of monophasic and biphasic fits (Eq.'s 3.5 and 3.6, respectively). Arrows highlight poorly distributed residual signal about the fit for monoexponential fit. "T" indicates respective trials. \* Indicates significantly better fit as determined by AIC and F-Tests with a 95 % confidence level.



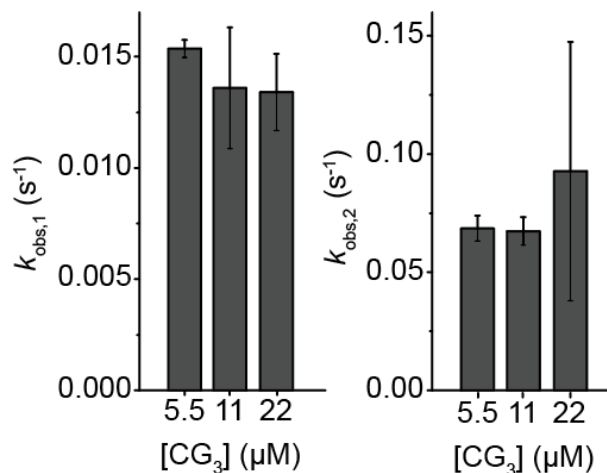
We successively diluted CG<sub>3</sub> down from the initial concentration of 110 μM and used wavelength scans to determine if the dynode was saturated. We found 22 μM was the highest DNA concentration for which the dynode response was not saturated for the conditions and spectropolarimeter used. 22 μM CG<sub>3</sub> + 3.0 M NaCl transient kinetic traces still exhibited strong biphasic characteristics consistent with those observed insofar (Figure 3.14). Similarly, to test a lower DNA concentration we began with 1 μM DNA and increased the DNA concentration successively, using wavelength scans to assess the signal-to-noise, and found 5.5 μM to be the lowest acceptable for kinetic studies. 5.5 μM CG<sub>3</sub> + 3.0 M NaCl kinetics also exhibited clear biphasic behavior with residuals of monoexponential fits similar to those reported at the principal 11 μM concentration used herein (Figure 3.15).



**Figure 3.15 | NaCl-Induced Transient Kinetics of  $\text{CG}_3$  with Two-Fold Concentration Decrease.** (A) Wavelength scan controls and corresponding measurements with transients at 254 nm. (B) Residuals of monophasic and biphasic fits (Eq.'s 3.5 and 3.6, respectively). Arrows highlight poorly distributed residual signal about the fit for monoexponential fit. "T" indicates respective trials. \* Indicates significantly better fit as determined by AIC and F-Tests with a 95 % confidence level.

These concentration variations allowed us to control for a 4-fold range in concentration. Within this range, no difference in either observed rate constant was observed (Figure 3.16). We note that the observed rate constants are complex convolutions of the microscopic rate constants for which the exact solution is given by Eq.'s 1.9 and 1.10. This finding rules out any model containing an intermediate or transition state with single-stranded DNA as others have proposed for the B-to-Z-DNA transition.<sup>64</sup> It is not

surprising single-stranded DNA is not involved in the B-to-Z-DNA transition of a canonical, pure CpG repeat since the barrier for duplex denaturation is high for dinucleotide steps with strong stacking energies.<sup>65,66</sup>

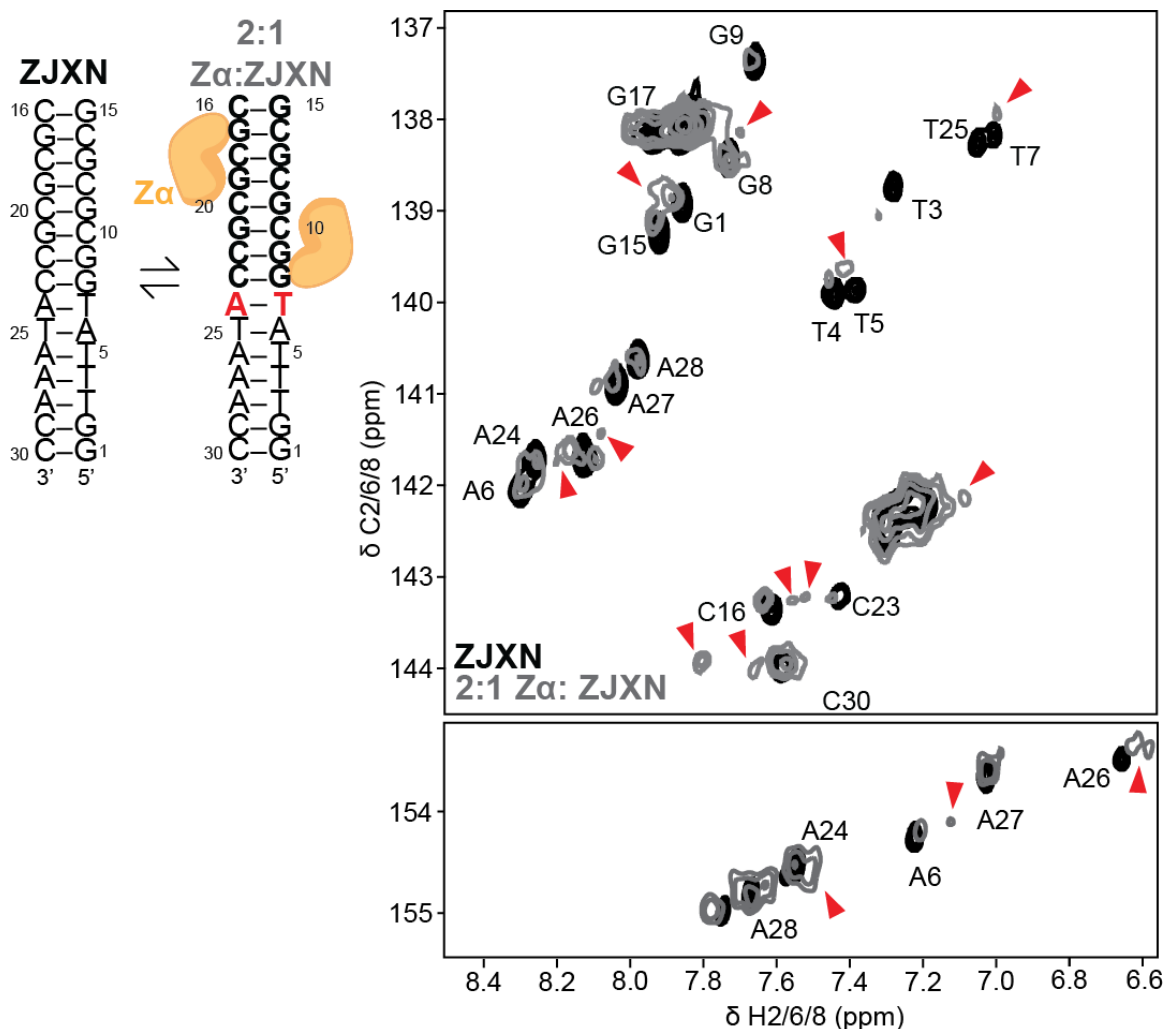


**Figure 3.16 | Observed Rate Constants of CG<sub>3</sub> Transition over a Four-Fold Range of DNA Concentrations.**

### 3.3.5 Observation of Additional Species at Intermediate Protein:DNA Ratios

As mentioned previously, the observation of two phases during the measurement of transient kinetics suggests at least three species, but more knowledge about the system is needed to understand the mechanistic relationship between the three species. In a completely orthogonal study we turned to NMR for its ability to readily give atomic resolution structural information on biomolecular complexes. We prepared the uniformly isotopically enriched <sup>13</sup>C/<sup>15</sup>N ZJXN sequence and titrated in natural abundance Zα to a protein:DNA ratio of 2:1. A crystal structure of this complex was solved by Rich and coworkers and shows that the Zα domain binds ZJXN at a 4:1 ratio.<sup>67</sup> Prior titration experiments demonstrate saturation of signal change near a ratio of 6:1.<sup>53</sup> We chose to create a Zα:<sup>13</sup>C/<sup>15</sup>N ZJXN complex near the midpoint of the transition to probe for

formation of an alternative structure under equilibrium conditions. Indeed, in our 2D heteronuclear correlation experiments we observe extra peaks in the guanine and cytosine regions, which correspond to the portion of the sequence expected to undergo the B-to-Z-DNA transition. We note  $Z\alpha$  only contacts the phosphate backbone, precluding such a widespread observation of new peaks upon protein binding.<sup>67</sup> To our knowledge, the aromatic carbon chemical shifts of Z-DNA have not been reported in the literature, but the *anti-to-syn* transition of Z-DNA guanines suggests guanine resonances will shift downfield upon Z-DNA formation based on the  $\sim 2.5$  ppm downfield shift observed for the Watson-Crick (WC) to Hoogsteen (HG) transition as shown in Chapter 2. In fact, no such changes in chemical shift are observed for the 2:1 complex in which the midpoint ( $\sim 3:1$ ) has not been reached yet. It is possible Z-DNA has not been formed in an appreciable amount such that chemical shifts are observed, or that the guanine chemical shifts change in a manner that cannot be predicted based on our knowledge of the WC-to-HG chemical exchange signatures. A portion of the work presented in Chapter 5 aims to benchmark Z-DNA aromatic carbon chemical shifts and further investigate the signatures of the structural changes.



**Figure 3.17 | Titration of  $Z\alpha$  into  $^{13}C/^{15}N$  ZJXN at  $R=2$ .** Aromatic SOFAST-HMQC spectra of ZJXN (black) and a  $Z\alpha$ :ZJXN 2:1 complex (grey) complex. 15 mM Phosphate buffer pH 7.5, 0.1 mM EDTA, 25 mM NaCl, 25 °C. B-DNA chemical shifts are labeled according to previously published assignments.<sup>68</sup> Blue arrowheads denote shift of overlapped guanine resonances. Red arrowheads denote B-DNA chemical shifts of junctional base-pair.

### 3.4 Conclusions

We applied our 2-state fitting approach<sup>53</sup> to the NaCl- and  $Z\alpha$ -induced B-to-Z-DNA transitions and obtained  $f_{Z-DNA}$  profiles similar to those published previously.<sup>1,51</sup> The DNA signal in the NaCl-induced transition cannot be accounted for by the B-DNA and Z-DNA species only, suggesting a two-state model is insufficient to describe the

NaCl-induced B-to-Z-DNA transition in CG<sub>3</sub>. CD spectra of a short CpG repeat at equilibrium show residual signal beyond that accounted for by B-DNA and Z-DNA, also providing evidence for at least a third species under equilibrium conditions of the transitions reported herein. Perhaps not surprisingly, the B-to-Z-DNA transition in this short pyr-pur repeat is biphasic. Many mechanistic models have been presented for the B-to-Z-DNA transition over the past few decades (reviewed by Fuertes and coworkers<sup>46</sup>), most with little experimental support. We have demonstrated, for the first time, high temporal resolution transient kinetic data on a short CpG repeat. Transient kinetics are biphasic across DNA and NaCl concentrations, DNA length and inducing solution identity (i.e. NaCl or Z $\alpha$ ). We note it is possible NaCl- and Z $\alpha$ -induced B-Z transitions undergo different mechanisms.

Future studies could investigate strategies to improve the resolving power in the observed rate constants by changing experimental conditions such that the two observed phases are further different from each other, for example the change of pH or addition of viscosogens to the already-low temperature. Nevertheless, a lack of change in observed rate constants disproves mechanisms involving single-stranded DNA as a potential intermediate of the B-to-Z-DNA transition. The structures involved in the B-DNA-Z-DNA equilibrium in the cell may be even more structurally diverse than previously thought. Lastly, and perhaps most significantly, evidence for an additional species along the B-to-Z-DNA transition was directly observed using NMR spectroscopy. These findings contribute to the field a robust framework for which a new kinetic transition may be interrogated. The NMR studies of Z-DNA call for benchmarks of Z-DNA carbon

chemical shifts, especially aromatic chemical shifts, which fueled a portion of the work in Chapter 5.

### 3.5 References

- 1 Pohl, F. M. & Jovin, T. M. *J Mol Biol* Salt-induced Co-operative Conformational Change of a Synthetic DNA: Equilibrium and Kinetic Studies with Poly(dG-dC) **67**, 375-396, (1972).
- 2 Hill, R. J. *J Cell Sci* Z-DNA: a prodrome for the 1990s **99**, 675-680, (1991).
- 3 Hill, R. J. *BioEssays* On the status of the Z-DNA question for animal chromosomes **1**, 244-249, (1984).
- 4 Krishna, P., Fritzler, M. J. & Van de Sande, J. H. *Clin Exp Immunol* Interactions of anti-DNA antibodies with Z-DNA **92**, 51-57, (1993).
- 5 Juranic, Z. *et al. Med Hypotheses* The importance of the specific Z-DNA structure and polyamines in carcinogenesis: fact or fiction **35**, 353-357, (1991).
- 6 Johnston, B. H. *J Biomol Struct Dyn* Chemical probing of the B-Z transition in negatively supercoiled DNA **6**, 153-166, (1988).
- 7 Casanovas, J. M. & Azorin, F. *Nucleic Acids Res* Supercoiled induced transition to the Z-DNA conformation affects the ability of a d(CG/GC)<sub>12</sub> sequence to be organized into nucleosome-cores **15**, 8899-8918, (1987).
- 8 Pohl, F. M. *Proc Natl Acad Sci USA* Dynamics of the B-to-Z transition in supercoiled DNA **83**, 4983-4987, (1986).
- 9 Peck, L. J., Wang, J. C., Nordheim, A. & Rich, A. *J Mol Biol* Rate of B to Z structural transition of supercoiled DNA **190**, 125-127, (1986).
- 10 Ellison, M. J., Kelleher, R. J., 3rd, Wang, A. H., Habener, J. F. & Rich, A. *Proc Natl Acad Sci USA* Sequence-dependent energetics of the B-Z transition in supercoiled DNA containing nonalternating purine-pyrimidine sequences **82**, 8320-8324, (1985).
- 11 Xu, Y. C. & Qian, L. *Sci Sin B* Determination of linking number of pBR322 DNA **26**, 602-613, (1983).
- 12 Stockton, J. F. *et al. EMBO J* Left-handed Z-DNA regions are present in negatively supercoiled bacteriophage PM2 DNA **2**, 2123-2128, (1983).
- 13 Rich, A., Nordheim, A. & Azorin, F. *J Biomol Struct Dyn* Stabilization and detection of natural left-handed Z-DNA **1**, 1-19, (1983).
- 14 Nordheim, A. *et al. Cell* Negatively supercoiled plasmids contain left-handed Z-DNA segments as detected by specific antibody binding **31**, 309-318, (1982).
- 15 Bayele, H. K. *et al. Blood* HIF-1 regulates heritable variation and allele expression phenotypes of the macrophage immune response gene SLC11A1 from a Z-DNA forming microsatellite **110**, 3039-3048, (2007).
- 16 Rothenburg, S., Koch-Nolte, F. & Haag, F. *Immunol Rev* DNA methylation and Z-DNA formation as mediators of quantitative differences in the expression of alleles **184**, 286-298, (2001).

- 17 Searle, S. & Blackwell, J. M. *J Med Genet* Evidence for a functional repeat polymorphism in the promoter of the human NRAMP1 gene that correlates with autoimmune versus infectious disease susceptibility **36**, 295-299, (1999).
- 18 Ray, B. K., Dhar, S., Shakya, A. & Ray, A. *Proc Natl Acad Sci USA* Z-DNA-forming silencer in the first exon regulates human ADAM-12 gene expression **108**, 103-108, (2011).
- 19 Taka, S., Gazouli, M., Politis, P. K., Pappa, K. I. & Anagnou, N. P. *Mol Biol Rep* Transcription factor ATF-3 regulates allele variation phenotypes of the human SLC11A1 gene **40**, 2263-2271, (2013).
- 20 Watson, F., Deavall, D. G., Macro, J. A., Kiernan, R. & Dimaline, R. *Biochem J* Transcriptional activation of vesicular monoamine transporter 2 in the pre-B cell line Ea3.123 **337 ( Pt 2)**, 193-199, (1999).
- 21 Lenert, P. *Clin Exp Immunol* Nucleic acid sensing receptors in systemic lupus erythematosus: development of novel DNA- and/or RNA-like analogues for treating lupus **161**, 208-222, (2010).
- 22 Bacolla, A. & Wells, R. D. *Mol Carcinog* Non-B DNA conformations as determinants of mutagenesis and human disease **48**, 273-285, (2009).
- 23 Ray, B. K., Dhar, S., Henry, C., Rich, A. & Ray, A. *Cancer Res* Epigenetic regulation by Z-DNA silencer function controls cancer-associated ADAM-12 expression in breast cancer: cross-talk between MeCP2 and NF1 transcription factor family **73**, 736-744, (2013).
- 24 Temiz, N. A., Donohue, D. E., Bacolla, A., Luke, B. T. & Collins, J. R. *PLoS One* The role of methylation in the intrinsic dynamics of B- and Z-DNA **7**, e35558, (2012).
- 25 Thomas, T. J. & Messner, R. P. *Arthritis Rheum* Effects of lupus-inducing drugs on the B to Z transition of synthetic DNA **29**, 638-645, (1986).
- 26 Van Helden, P. D. *J Immunol* Potential Z-DNA-forming elements in serum DNA from human systemic lupus erythematosus **134**, 177-179, (1985).
- 27 Lafer, E. M. *et al. J Clin Invest* Z-DNA-specific antibodies in human systemic lupus erythematosus **71**, 314-321, (1983).
- 28 Kanai, Y., Moroi, Y., Yokohari, R. & Goto, S. *Princess Takamatsu Symp* Profiles of antibodies to poly(ADP-ribose), left-handed Z-DNA, and other nuclear constituents in systemic lupus erythematosus and progressive systemic sclerosis **13**, 93-100, (1983).
- 29 Blackwell, J. M. & Searle, S. *Immunol Lett* Genetic regulation of macrophage activation: understanding the function of Nramp1 **65**, 73-80, (1999).
- 30 Allinquant, B., Malfoy, B., Schuller, E. & Leng, M. *Clin Exp Immunol* Presence of Z-DNA specific antibodies in Crohn's disease, polyradiculoneuritis and amyotrophic lateral sclerosis **58**, 29-36, (1984).
- 31 Kouzine, F., Sanford, S., Elisha-Feil, Z. & Levens, D. *Nat Struct Mol Biol* The functional response of upstream DNA to dynamic supercoiling in vivo **15**, 146-154, (2008).
- 32 Kouzine, F., Liu, J., Sanford, S., Chung, H. J. & Levens, D. *Nat Struct Mol Biol* The dynamic response of upstream DNA to transcription-generated torsional stress **11**, 1092-1100, (2004).



- 33 Wong, B., Chen, S., Kwon, J. A. & Rich, A. *Proc Natl Acad Sci USA* Characterization of Z-DNA as a nucleosome-boundary element in yeast *Saccharomyces cerevisiae* **104**, 2229-2234, (2007).
- 34 Schroth, G. P., Chou, P. J. & Ho, P. S. *J Biol Chem* Mapping Z-DNA in the human genome. Computer-aided mapping reveals a nonrandom distribution of potential Z-DNA-forming sequences in human genes **267**, 11846-11855, (1992).
- 35 Cer, R. Z. *et al. Nucleic Acids Res* Non-B DB: a database of predicted non-B DNA-forming motifs in mammalian genomes **39**, D383-391, (2011).
- 36 Lange, J. *et al. Cell* Isodicentric Y chromosomes and sex disorders as byproducts of homologous recombination that maintains palindromes **138**, 855-869, (2009).
- 37 Fuertes, M. A., Perez, J. M., Gonzalez, V. M. & Alonso, C. *J Biol Inorg Chem A* kinetic model for the B-Z transition of poly[d(G-C)].poly[d(G-C)] and poly[d(G-m5C)].poly[d(G-m5C)] **6**, 675-682, (2001).
- 38 van de Sande, J. H., McIntosh, L. P. & Jovin, T. M. *EMBO J* Mn<sup>2+</sup> and other transition metals at low concentration induce the right-to-left helical transformation of poly[d(G-C)] **1**, 777-782, (1982).
- 39 van de Sande, J. H. & Jovin, T. M. *EMBO J* Z\* DNA, the left-handed helical form of poly[d(G-C)] in MgCl<sub>2</sub>-ethanol, is biologically active **1**, 115-120, (1982).
- 40 McLean, M. J., Lee, J. W. & Wells, R. D. *J Biol Chem* Characteristics of Z-DNA helices formed by imperfect (purine-pyrimidine) sequences in plasmids **263**, 7378-7385, (1988).
- 41 Xodo, L. E. *et al. J Biomol Struct Dyn* The left-handed Z-DNA conformation in oligodeoxynucleotides containing different amounts of AT base pairs: a far UV circular dichroism study **6**, 1217-1231, (1989).
- 42 Kadalayil, L. P., Majumder, K., Mishra, R. K. & Brahmachari, S. K. *Biochem Int* Sequence specificity of Z-DNA formation in oligonucleotides **17**, 121-131, (1988).
- 43 Mishra, R. K., Latha, P. K. & Brahmachari, S. K. *Nucleic Acids Res* Interruptions of (CG)<sub>n</sub> sequences by GG, TG and CA need not prevent B to Z transition in solution **16**, 4651-4665, (1988).
- 44 Quyen, D. V. *et al. Nucleic Acids Res* Characterization of DNA-binding activity of Z alpha domains from poxviruses and the importance of the beta-wing regions in converting B-DNA to Z-DNA **35**, 7714-7720, (2007).
- 45 Kim, Y. G. *Bull Korean Chem Soc* Protein-Induced Sequence Diversity of Z-DNA **28**, 1063-1066, (2007).
- 46 Fuertes, M. A., Cepeda, V., Alonso, C. & Perez, J. M. *Chem Rev* Molecular mechanisms for the B-Z transition in the example of poly[d(G-C) x d(G-C)] polymers. A critical review **106**, 2045-2064, (2006).
- 47 Dubois, J.-M., Ouanounou, G. & Rouzair-Dubois, B. *Prog Biophys Mol Biol* The Boltzmann equation in molecular biology **99**, 87-93, (2009).
- 48 Zimmer, D. P. & Crothers, D. M. *Proc Nat Acad Sci USA* NMR of enzymatically synthesized uniformly <sup>13</sup>C<sup>15</sup>N-labeled DNA oligonucleotides **92**, 3091-3095, (1995).
- 49 Nikolova, E. N. *et al. Nature* Transient Hoogsteen base pairs in canonical duplex DNA **470**, 498-502, (2011).

- 50 Schwartz, T., Rould, M. A., Lowenhaupt, K., Herbert, A. & Rich, A. *Science* Crystal Structure of the Z $\alpha$  Domain of the Human Editing Enzyme ADAR1 Bound to Left-Handed Z-DNA **284**, 1841-1845, (1999).
- 51 Seo, Y. J. *et al. FEBS Lett* Sequence discrimination of the Z $\alpha$  domain of human ADAR1 during B-Z transition of DNA duplexes **584**, 4344-4350, (2010).
- 52 Kang, Y. M. *et al. J Am Chem Soc* NMR spectroscopic elucidation of the B-Z transition of a DNA double helix induced by the Z  $\alpha$  domain of human ADAR1 **131**, 11485-11491, (2009).
- 53 Bothe, J. R., Lowenhaupt, K. & Al-Hashimi, H. M. *Biochemistry* Incorporation of CC steps into Z-DNA: interplay between B-Z junction and Z-DNA helical formation **51**, 6871-6879, (2012).
- 54 Hodgkin, A. L. & Huxley, A. F. *J Physiol A* A quantitative description of membrane current and its application to conduction and excitation in nerve **117**, 500-544, (1952).
- 55 Weiss, J. N. *FASEB J* The Hill equation revisited: uses and misuses **11**, 835-841, (1997).
- 56 Kypr, J., Kejnovska, I., Renciuk, D. & Vorlickova, M. *Nucleic Acids Res* Circular dichroism and conformational polymorphism of DNA **37**, 1713-1725, (2009).
- 57 Miyahara, T., Nakatsuji, H. & Sugiyama, H. *J Phys Chem A* Helical structure and circular dichroism spectra of DNA: a theoretical study **117**, 42-55, (2013).
- 58 Vorlickova, M., Kejnovska, I., Bednarova, K., Renciuk, D. & Kypr, J. *Chirality* Circular dichroism spectroscopy of DNA: from duplexes to quadruplexes **24**, 691-698, (2012).
- 59 Randazzo, A., Spada, G. P. & da Silva, M. W. *Top Curr Chem* Circular dichroism of quadruplex structures **330**, 67-86, (2013).
- 60 Chang, Y. M., Chen, C. K. & Hou, M. H. *Int J Mol Sci* Conformational Changes in DNA upon Ligand Binding Monitored by Circular Dichroism **13**, 3394-3413, (2012).
- 61 Smith, S. B., Cui, Y. & Bustamante, C. *Science* Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules **271**, 795-799, (1996).
- 62 Motulsky, H. J. & Ransnas, L. A. *FASEB J* Fitting curves to data using nonlinear regression: a practical and nonmathematical review **1**, 365-374, (1987).
- 63 Wands, A. M. *et al. J Biol Chem* Transient-state kinetic analysis of transcriptional activator DNA complexes interacting with a key coactivator **286**, 16238-16245, (2011).
- 64 Walker, G. T. & Aboul-ela, F. *J Biomol Struct Dyn* B-Z cooperativity and kinetics of poly(dG-m5dC) are controlled by an unfavorable B-Z interface energy **5**, 1209-1219, (1988).
- 65 SantaLucia, J., Jr., Allawi, H. T. & Seneviratne, P. A. *Biochemistry* Improved nearest-neighbor parameters for predicting DNA duplex stability **35**, 3555-3562, (1996).
- 66 SantaLucia, J. *Proc Natl Acad Sci USA* A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics **95**, 1460-1465, (1998).

- 67 Ha, S. C., Lowenhaupt, K., Rich, A., Kim, Y. G. & Kim, K. K. *Nature* Crystal structure of a junction between B-DNA and Z-DNA reveals two extruded bases **437**, 1183-1186, (2005).
- 68 Bothe, J. R., Lowenhaupt, K. & Al-Hashimi, H. M. *J Am Chem Soc* Sequence-specific B-DNA flexibility modulates Z-DNA formation **133**, 2016-2018, (2011).

## Chapter 4

### Kinetic Analysis of the B-DNA to Z-DNA Transition in Short DNA Sequences; Examining the Consequences of DNA Methylation\*

#### 4.1 Introduction

Z-DNA exists in the cell in dynamic equilibrium with B-DNA<sup>1</sup> and, in the case of shorter CpG repeats as shown in Chapter 3, potentially at least one additional species is present in this equilibrium. Z-DNA formation is elevated upstream of genes involved in fatal diseases and cancer.<sup>2</sup> Methylation at cytosine C5 (<sup>5m</sup>C) of CpG hot spots in the human genome up-regulates certain oncogenes related to tumor growth<sup>3</sup> and <sup>5m</sup>C modification of CpG repeats increases the propensity to undergo the B-to-Z-DNA transition. Furthermore, some forms of damage are not repaired in Z-DNA.<sup>4</sup> Understanding the mechanism of the B-to-Z-DNA transition will help shed light on how this peculiar conformational change occurs and potentially present Z-DNA as a new drug and cancer therapeutic target. To gain structural insights into the characteristics of the third species observed in the B-to-Z-DNA transition, we seek to perturb the kinetics using chemical modifications. Here, we use <sup>5m</sup>C for its aforementioned role in Z-DNA formation *in vivo*.

Methylation machinery modifies CpG repeats with <sup>5m</sup>C in an epigenetic manner and pushes the thermodynamic B-to-Z-DNA equilibrium sufficiently toward Z-DNA relative to unmodified CpG repeats.<sup>3</sup> Methylation of the Hoogsteen (HG) face of guanine

---

\*HSA and HMA conceived the idea with help from PJO and CF; HSA prepared samples and carried out the experiments and data analysis with help from HMA.

stabilizes Z-DNA thermodynamically *via* hydrophobic contacts<sup>5-7</sup> as well as kinetically.<sup>8</sup> <sup>5m</sup>C modification and Z-DNA formation *in vivo* are thought to be linked.<sup>3,9</sup> Since Z-DNA forms transiently on the timescale of biological functions<sup>10-12</sup> the kinetics of B-to-Z-DNA transition in <sup>5m</sup>C-modified Z-DNA forming sequences are essential to gain a deeper understanding of how epigenetic <sup>5m</sup>C modification affects Z-DNA formation and potentially gene expression. While proposed structural changes for the B-to-Z-DNA transition are abundant in the literature, little emphasis has been placed on how these proposed structural changes might give rise to unique observables in transient kinetic experiments. We therefore seek to describe the B-to-Z-DNA transition using a kinetic model to examine <sup>5m</sup>C perturbations to the underlying mechanism of the B-to-Z-DNA transition.

## **4.2 Methods**

### 4.2.1 DNA Sample Preparation

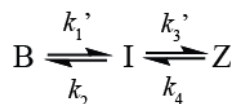
DNA samples were prepared as described in Section 3.2.1.

### 4.2.2 Circular Dichroism Thermodynamics and Kinetics

Thermodynamic and kinetic measurements were performed as described in Sections 3.2.2 and 3.2.3, respectively.

### 4.2.3 Simulations of the Observed Circular Dichroism Signal

Kinetic measurements presented in Chapter 3 were subjected to simulations to determine representative microscopic rate constants of a unimolecular two-step reversible kinetic model, as given by



where B is B-DNA, I is intermediate, Z is Z-DNA,  $k_1' = k_1[NaCl]$  and  $k_3' = k_3[NaCl]$ .

The model was input into KinSim PLUS (JP Consulting) and microscopic rate constants  $k_1$ ,  $k_2$ ,  $k_3$  and  $k_4$  were varied for a simulation from 0.001- 360 sec (for  $({}^5\text{meCG})_3$  and  $\text{CG}_3$ ) with a time step of 0.8 sec and concentration time dependence of B-DNA, intermediate and Z-DNA were calculated. Observed circular Dichroism (CD) signal was calculated from species concentrations using a linear combination of B, I and Z as

$$\theta_{254nm} = \varepsilon_{254nm}^B B + \varepsilon_{254nm}^I I + \varepsilon_{254nm}^Z Z \quad (4.1)$$

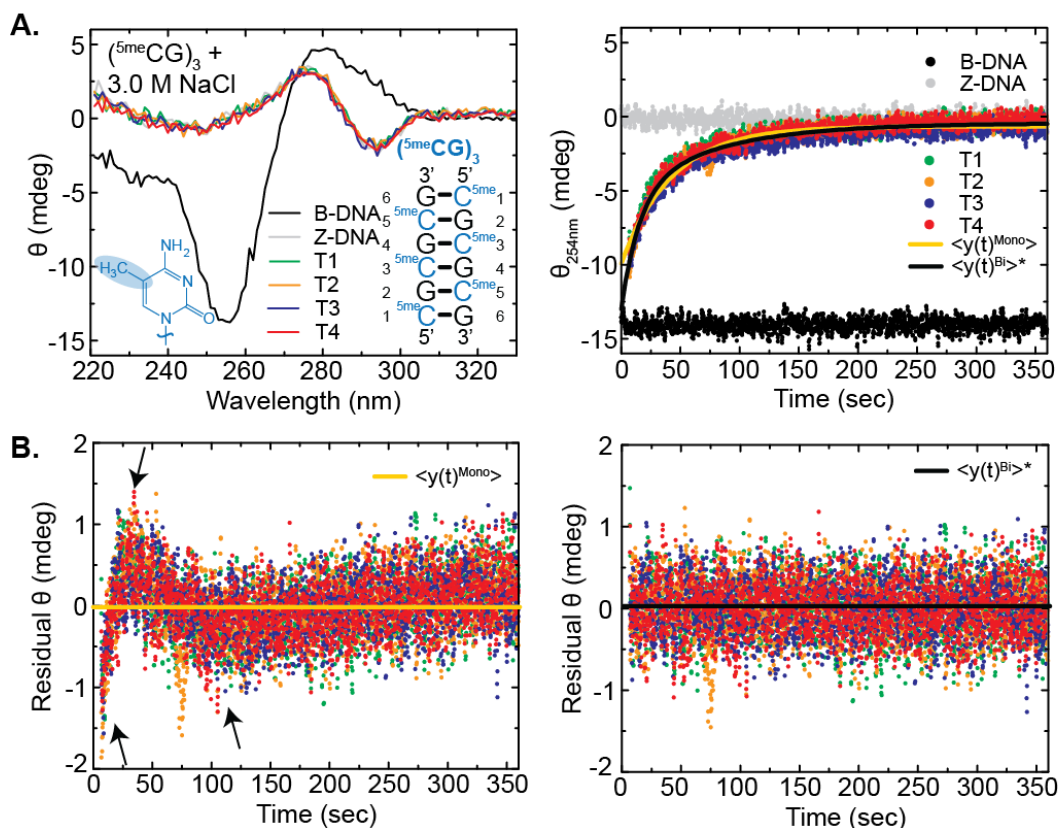
where B, I and Z represent their respective concentrations and  $\varepsilon_{254nm}^B$ ,  $\varepsilon_{254nm}^I$  and  $\varepsilon_{254nm}^Z$  their respective molar extinction coefficients. Molar extinction coefficients of B-DNA and Z-DNA were determined from beginning and endpoints of titrations and that of the intermediate was determined from residual signal at 254 nm calculated from Eq. 3.4. Microscopic rate constants were chosen based on fit of Eq. 4.1 to observed signal. Great care was taken to vary microscopic rate constants by 2 orders of magnitude individually and iteratively to best ensure arrival at a fit representing global minimum of microscopic rate constants.

### 4.3 Results and Discussion

#### 4.3.1 Impact of 5-Methylcytosine on Transient Kinetics of the B-to-Z-DNA Transition

The affects of <sup>5m</sup>C modification *in vivo* are, at least in part, due to perturbations to the CpG repeats' structure and dynamic properties.<sup>2,13-15</sup> Understanding the affect of the epigenetic <sup>5m</sup>C modification on CpG repeats' Z-DNA forming capabilities could lay the foundation for providing additional insights into their gene regulating properties. The affects of <sup>5m</sup>C modification as compared to unmodified CpG repeat sequences on the B-to-Z-DNA transition has been probed for a much longer sequence (100 base-pairs) and they find the degree of cooperativity of the transition increases upon <sup>5m</sup>C modification but do not provide a robust investigation of the number of phases observed along the transition.<sup>16</sup>

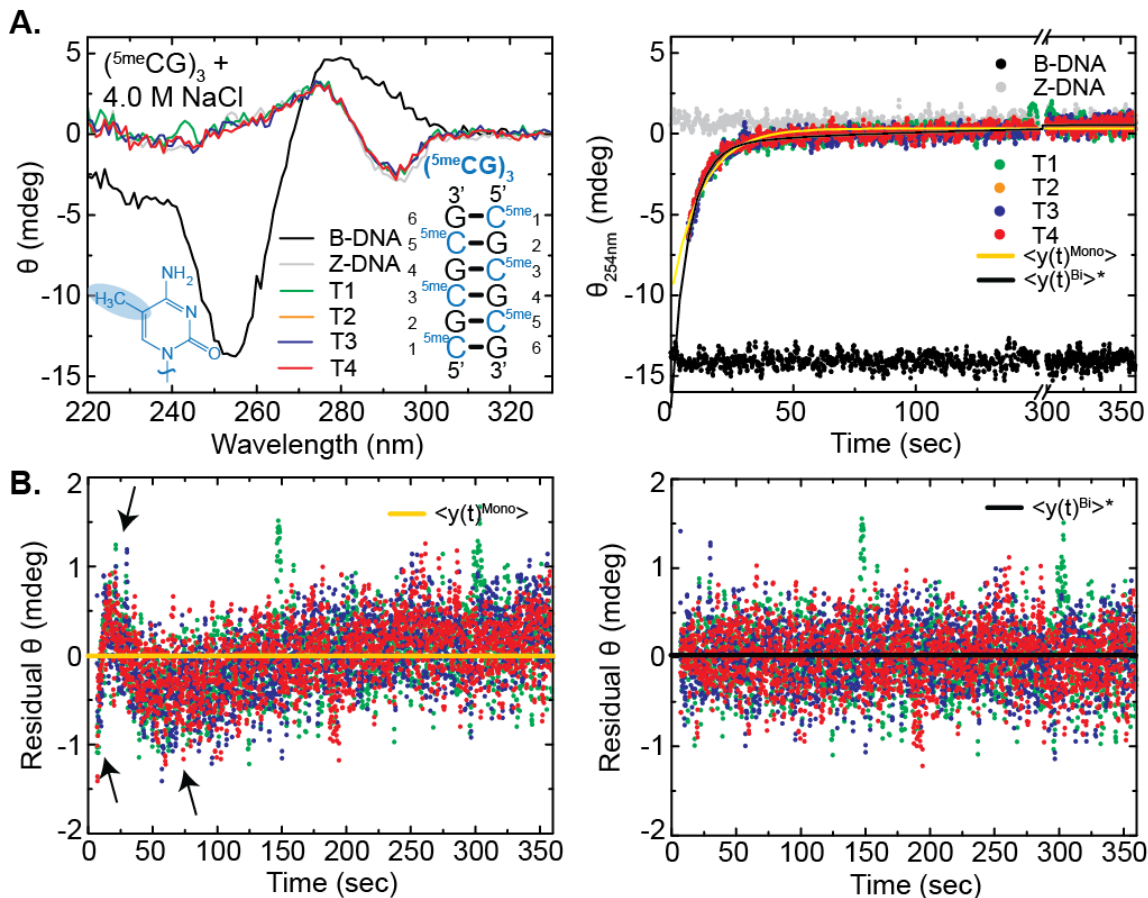
In Chapter 3 we presented a framework for approaching the analysis of monophasic and biphasic fits of transient kinetics. Here, we use the framework presented in Chapter 3 to characterize the differential affects of epigenetic modification in the context of Z-DNA formation of CpG hexamers. We use CpG hexamers to build on the work presented in Chapter 3, and to address the affect of <sup>5m</sup>C modification on such short CpG repeats. The shortest methylated repeats studied in the context of Z-DNA are 14-mers<sup>3</sup> despite the fact that only two CpG repeats are necessary to form Z-DNA under physiologically relevant levels of supercoiling in low salt.<sup>17</sup> Moreover, upon Z-DNA formation in CpG repeats, alternative sequences can be accomodated for further Z-DNA propagation,<sup>18</sup> further extending the efficacy of short Z-DNA forming CpG repeats. To gain a more complete understanding of 5-methylcytosine affects on the B-to-Z-DNA transition, we performed transient kinetic measurements upon installing uniform <sup>5m</sup>C modifications at all C's in CG<sub>3</sub> using 3.0 M (Figure 4.1) and 4.0 M (Figure 4.2) NaCl inducing solutions.



**Figure 4.1 | Transient Kinetics, Controls and Residuals of the NaCl-Induced B-to-Z-DNA Transition in  $(5^{me}CG)_3$  Induced by 3.0 M NaCl.** (A) Wavelength scan controls and corresponding measurements with transients at 254 nm. (B) Residuals of monophasic and biphasic fits (Eq.'s 3.5 and 3.6, respectively). “T” indicates respective trials. \* Indicates significantly better fit as determined by AIC and F-Tests with 95 % confidence level.

We report that the B-to-Z-DNA transition in  $(5^{me}CG)_3$  is biphasic as in  $CG_3$  (Figure 4.1). Inspection of the residuals reveals the two phases are similar relative to one another as compared to the same kinetic conditions upon which  $CG_3$  kinetics were measured (compare Figures 4.1 and 3.7). While this observation does not preclude the mechanism from changing, both results could be described with a single kinetic model, like the one presented in Section 4.2.2.

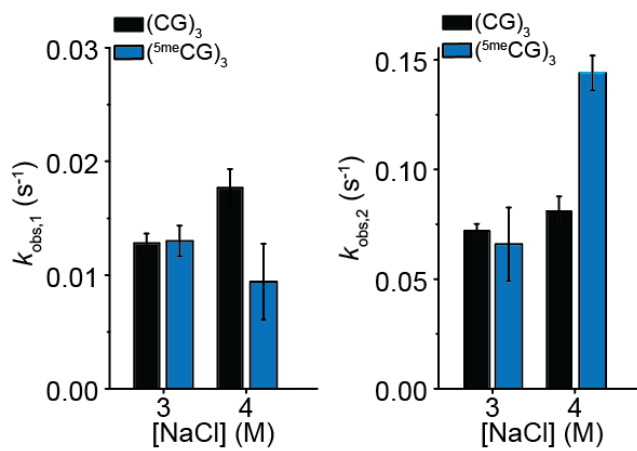




**Figure 4.2 | Transient Kinetics, Controls and Residuals of the B-to-Z-DNA Transition of  $(5^{me}CG)_3$  Induced by 4.0 M NaCl.** (A) Wavelength scan controls and corresponding measurements with transients at 254 nm. (B) Residuals of monophasic and biphasic fits (Eq.'s 3.5 and 3.6, respectively). "T" indicates respective trials. \* Indicates significantly better fit as determined by AIC and F-Tests with 95 % confidence level.

Average and standard deviation of observed rate constants for two different inducing solution concentrations (3.0 M and 4.0 M NaCl) show that the faster phase becomes enhanced upon C5 methylation, albeit two-fold (Figure 4.3), while the slower phase may become slightly slower when comparing standard deviations. This compensatory effect is interesting, but as mentioned in Chapter 3, the relatively poor resolving power of our observed rate constants as a function of salt leaves little room for interpretation of the observed rate constants directly. We also note that no significant

change in observed rate constants upon C5 methylation was observed at 3.0 M NaCl (Figure 4.3).



**Figure 4.3 | Observed Rate Constant Perturbation Upon C5 Methylation.**

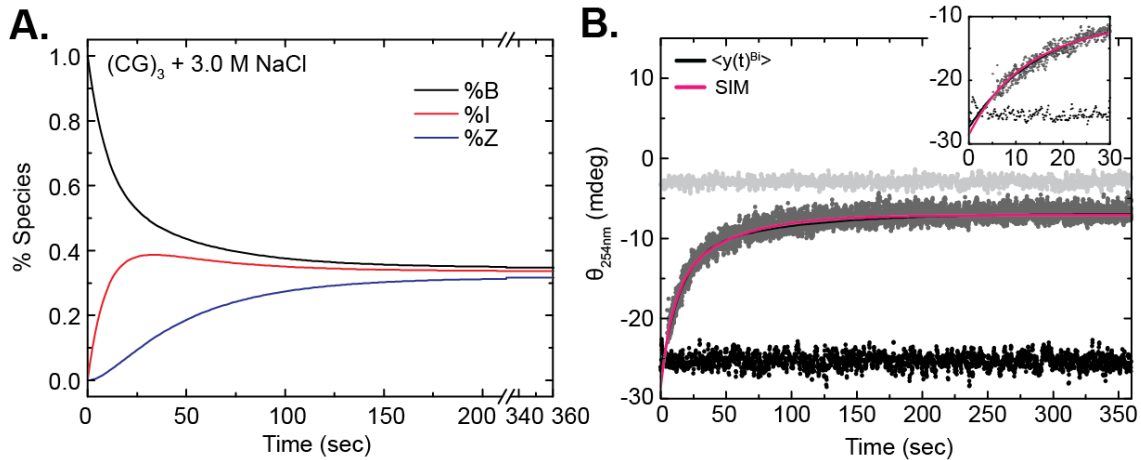
While the observed rate constants remain within the range of reproducibility under 3.0 M NaCl conditions and only the fast observed rate constant is perturbed not within error under 4.0 M NaCl conditions, it is important to note that the observed rate constants are complex convolutions of the microscopic rate constants describing each individual transition within a kinetic model (Eq.'s 1.6-1.8). Assumptions may be made to simplify the full solutions giving rise to  $k_{\text{obs},1}$  (Eq. 1.9) and  $k_{\text{obs},2}$  (Eq. 1.10). To try and gain further insight into how the mechanism is changing as a function of <sup>5me</sup>C, we turn to kinetic simulations, which first involve imposing a model on the B-to-Z-DNA transition.

#### 4.3.2 Simulation of a Kinetic Model for the NaCl-Induced B-to-Z-DNA Transition in CG<sub>3</sub>

In Chapter 2 we demonstrated that the next energetically accessible state of CG<sub>3</sub> is the transient HG base-pair, which occurs at ~0.2 % (Table 2.1),<sup>19</sup> precluding the transition of B- and B-like DNAs to Z-DNA in parallel, competing pathways. In Chapter

3 we presented strong evidence for at least 3 species and two observed phases in the B-to-Z-DNA transition of short CpG repeats, consistent with a unimolecular 2-step reaction. In Chapter 2 we demonstrated the next energetically accessible state of CG<sub>3</sub> on the millisecond timescale happens in ~0.2 % abundance (Table 2.1) which likely rules out parallel pathways as the reason behind the biphasic observations. The unimolecular two-step reversible model is therefore the simplest model to recapitulate our data (shown in Section 4.2.2). Simulations of such a mechanism were implemented using KinSim PLUS. The reversible unimolecular two-step model was expressed and  $k_1$  through  $k_4$  were varied and the resulting concentration time dependence of  $B$ ,  $I$  and  $Z$  was simulated (Figure 4.4 A). Interestingly, simulations of CG<sub>3</sub> + 3.0 M NaCl suggest  $B$ ,  $I$  and  $Z$  are each ~33 % populated at equilibrium. This is consistent with reports that B- and Z-DNA have similar free energies of stabilization.<sup>20</sup>

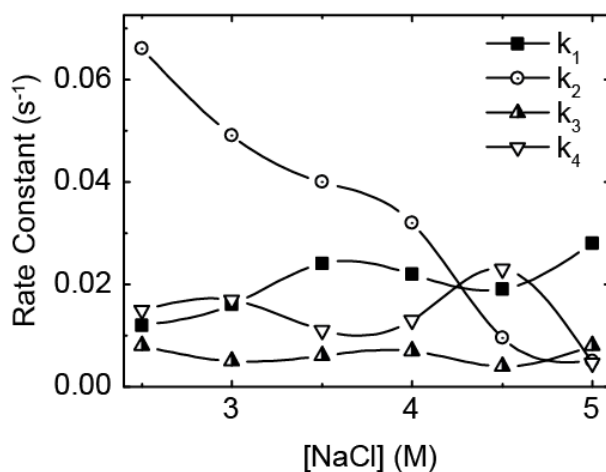
The observe signal is given by a linear combination of species present,  $\theta_{254nm} = \epsilon_{254nm}^B B + \epsilon_{254nm}^I I + \epsilon_{254nm}^Z Z$  (Eq. 4.1) where  $B$ ,  $I$  and  $Z$  represent their respective concentrations and  $\epsilon_{254nm}^B$ ,  $\epsilon_{254nm}^I$  and  $\epsilon_{254nm}^Z$  their respective molar extinction coefficients. For methods used to determine respective extinction coefficients see Section 4.2.3. Eq. 4.1 was used to plot the concentration time dependence of species with transient kinetic traces to evaluate goodness of fit (Figure 4.4 B). Simulations were implemented for all CG<sub>3</sub> and (<sup>5me</sup>CG)<sub>3</sub> conditions used for kinetic measurements.



**Figure 4.4 | Simulations of the B-to-Z-DNA Transition in CG<sub>3</sub>.** (A) Percent species as a function of time calculated in KinSim PLUS using microscopic rate constants that fit observed signal given by Eq. 4.1 (B, yellow line). (B) 4 transients (dark grey) converting from B-DNA (black) to Z-DNA (light grey) biphasic fit to Eq. 3.6 shown in blue. Plot inset shows first 30 sec of transition with corresponding fits.

Microscopic rate constants resulting from the above approach are plotted in Figure 4.5 and presented in Table 4.1. Most importantly, we have demonstrated the successful accommodation of signal due to at least a third species present during the B-to-Z-DNA transition, increasing the merit of the residual signal observed in the NaCl-induced B-to-Z-DNA transition presented in Chapter 3 (Figure 3.6). Interestingly, majority of the microscopic rate constants are very similar to each other, especially at high salt concentrations, and reveal that the increase in observed rate constants with increasing salt is likely due predominantly to a decreasing  $k_2$  (Figure 4.5). This suggests that NaCl concentration is correlated to stability of the third species- the putative intermediate- or correlative changes in B-DNA stability and energy of the first forward barrier. Previous preliminary studies in the lab present evidence for the Watson-Crick (WC) to Hoogsteen (HG) equilibrium shifting slightly more toward WC as the NaCl

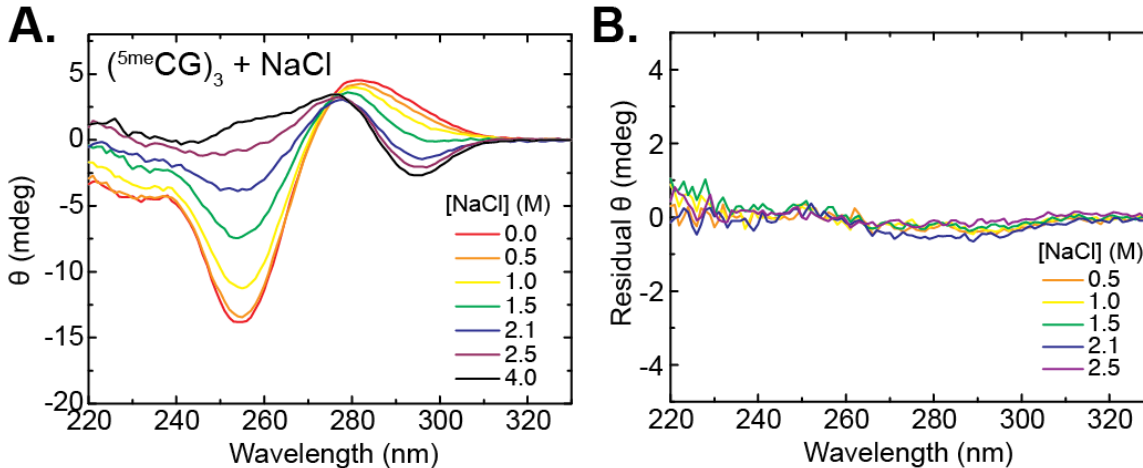
concentration increases from 25 mM to 125 mM (Nikolova, E.N. unpublished results) and it is well known in the literature that  $\text{Na}^+$  ions interact extensively with the DNA phosphate backbone to alleviate phosphate-phosphate repulsion when inter-phosphate distance decreases upon Z-DNA formation.<sup>21,22</sup> This stabilization of the putative intermediate suggests it has a closer phosphate-phosphate distance than that of B-DNA, like that of Z-DNA. Indeed, the residual signal of the NaCl-induced B-Z transition in  $\text{CG}_3$  exhibits residual signal consistent with a left-handed-like structure.



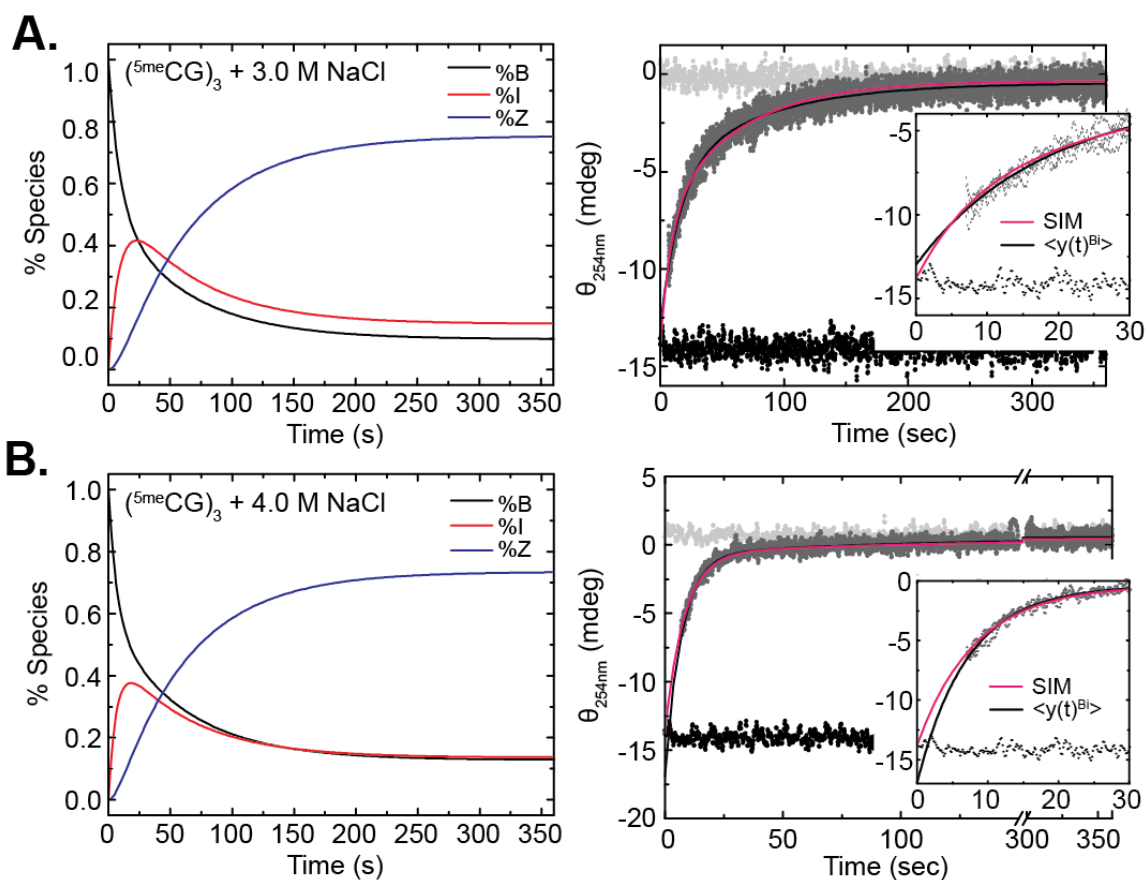
**Figure 4.5 | Microscopic Rate Constants of the B-to-Z-DNA Transition of  $\text{CG}_3$ .**

Next, to understand the potential affect of  $^{5\text{m}}\text{C}$  modification on the microscopic rate constants of the B-to-Z-DNA transition, we simulated the observed signal of the  $(^{5\text{m}}\text{CG})_3$  transition by varying the microscopic rate constants, again calculating residual signal of the equilibrium measurements using Eq. 3.4 (Figure 4.6 and Table A3.1). We note residuals are within noise or non-existent for  $(^{5\text{m}}\text{CG})_3$  so  $\epsilon_I^{254\text{nm}} = 0\text{M}^{-1}\text{cm}^{-1}$  was used in Eq. 4.1. The observation of a concomitant change in observed rate constants and residual signal upon  $^{5\text{m}}\text{C}$  modification (for the case of DNA + 4.0 M NaCl) brings additional merit to the potential relationship between kinetic measurements and residual

signal under equilibrium conditions, at least for the B-Z transition in CpG hexamers. Additional, to control for a change in microscopic rate constants due to accounting for the lower residual signal of  $(^{\delta\text{me}}\text{CG})_3$  as compared to  $\text{CG}_3$ , we repeated the fitting procedure setting the extinction coefficient equal to that of  $\text{CG}_3$  ( $\epsilon_l^{254\text{nm}} = 377,856\text{M}^{-1}\text{cm}^{-1}$ ) and saw between 0- to 5-fold changes in the elementary rate constants (Table A3.2). Care must be taken while interpreting microscopic rate constants from such simulations since more than one solution may exist.



**Figure 4.6 |  $(^{\delta\text{me}}\text{CG})_3$  NaCl Titration and Residual Signal.** (A) Titration and (B) residual signal calculated from Eq. 3.4. 15 mM Phosphate buffer, 0.1 mM EDTA, pH 7.5, 25 °C.



**Figure 4.7 | Simulation of the B-to-Z-DNA Transition in  $(^5\text{meCG})_3$ .** Percent B-DNA (% B), intermediate (% I) and Z-DNA (% Z) as a function of time from simulated reaction using microscopic rate constants in Table 4.1 (left) and corresponding observed signal at 254 nm calculated from Eq. 4.1 for the B-to-Z-DNA transition of  $(^5\text{meCG})_3$  induced by (A) 3.0 M and (B) 4.0 M NaCl. Shown in dark grey are 4 transients converting from B-DNA (black) to Z-DNA (light grey). Biphasic fit to Eq. 3.6 shown in black. Plot inset shows first 30 sec of transition with corresponding fits. 15 mM Phosphate buffer pH 7.5, 0.1 mM EDTA, 5°C.

This demonstrates that changes in elementary rate constants upon  $^5\text{mC}$  modification are not due to change in spectroscopic properties of the intermediate, but rather due to inherent changes in the mechanism itself. Moreover, when compared with the simplistic biphasic fit (Eq. 3.6) as first presented in Chapter 3, the simulated curves are virtually indistinguishable (Figure 4.7). This shows that the method used to simulate the

microscopic rate constants recapitulates the two observed phase changes. Interestingly, unlike the NaCl dependence of CG<sub>3</sub>, that of (<sup>5me</sup>CG)<sub>3</sub> shows that as NaCl concentration increases, the intermediate is actually depleted faster whereas its build-up becomes greater for CG<sub>3</sub>. Indeed, this is reflected in the representative microscopic rate constants (Table 4.1).

**Table 4.1 | Perturbation of Representative Microscopic Rate Constants upon Cytosine C5 Methylation.** Variations in  $k_1$ ,  $k_2$ ,  $k_3$  and  $k_4$  of <5 % produced similar curves, suggesting errors are <5 %.

DNA	[NaCl] (M)	$k_1$ (s <sup>-1</sup> )	$k_2$ (s <sup>-1</sup> )	$k_3$ (s <sup>-1</sup> )	$k_4$ (s <sup>-1</sup> )	$K_{1,app}$	$K_{2,app}$
CG <sub>3</sub>	3	0.016	0.049	0.005	0.016	3.1	3.2
( <sup>5me</sup> CG) <sub>3</sub>	3	0.020	0.040	0.008	0.005	2.0	0.6
CG <sub>3</sub>	4	0.022	0.032	0.007	0.013	1.5	2.0
( <sup>5me</sup> CG) <sub>3</sub>	4	0.030	0.005	0.001	0.004	0.2	3.2

It is known that <sup>5m</sup>C increases the WC-to-HG forward rate of transition by ~2-fold (Nikolova, E.N., data not shown) which is similar to our observed increase in  $k_3$  of ~1.6-fold (Table 4.1). We note <sup>5m</sup>C has a different affect on the B-to-Z-DNA transition under 4.0 M NaCl conditions, whereby  $k_3$  decreases ~7-fold (Table 4.1). It is known that NaCl has an anti-cooperative affect on the WC-to-HG transition (Nikolova, E.N., data not shown), which could explain the differential <sup>5m</sup>C affect due to the change in NaCl concentration. <sup>5m</sup>C modification does not significantly alter the first barrier of the B-to-Z-DNA transition, which suggests it is not involved in transition state formation. Interestingly, the intermediate in both CG<sub>3</sub> and (<sup>5me</sup>CG)<sub>3</sub> is not fully committed to forming Z-DNA. The intermediate is therefore sufficiently populated at equilibrium, at least under the conditions studied herein. This is also evident in the simulations performed (Figure 4.4).



#### 4.4 Conclusions

The B-to-Z-DNA transition in (<sup>5me</sup>CG)<sub>3</sub> is biphasic with similar timescales to those observed for CG<sub>3</sub>. The simplest model to describe our data on CG<sub>3</sub> and (<sup>5me</sup>CG)<sub>3</sub> is the unimolecular reversible two-step reaction. Simulations of the microscopic rate constants were performed according to the observed signal from transient kinetic measurements. Recapitulation of observed transients from simulations overlay exceptionally well with the biphasic exponential fits, i.e. the best fit to the transients. We stress that care must be taken when interpreting microscopic rate constants from such simulations since more than one solution may exist. The positive ellipticity near 255 nm of the CG<sub>3</sub> NaCl-induced B-to-Z-DNA transition residual signal (Figure 3.6 A) indirectly points toward the putative intermediate adopting a left-handed helical structure (i.e. Figure A4.1 A, Scheme 2). Overall these experiments lay the groundwork for interpreting kinetic perturbations of the B-to-Z-DNA transition, shed light on <sup>5me</sup>C perturbations and call for future studies to test the hypotheses presented herein for the molecular model of the B-to-Z-DNA transition.

A concurrent change in observed rate constants and residual signal upon <sup>5me</sup>C modification is observed. The latter two points raise additional merit to the potential relationship between kinetic measurements and residual signal under equilibrium conditions for the B-Z transition in CpG hexamers. Most importantly, we have demonstrated that the signal due to a third species observed under equilibrium conditions can be accommodated within the mechanistic framework used to describe the measurement of transient kinetics of the B-to-Z-DNA transition.

## 4.5 References

- 1 Zacharias, W., Jaworski, A., Larson, J. E. & Wells, R. D. *Proc Natl Acad Sci USA* The B- to Z-DNA equilibrium in vivo is perturbed by biological processes **85**, 7069-7073, (1988).
- 2 Ray, B. K., Dhar, S., Henry, C., Rich, A. & Ray, A. *Cancer Res* Epigenetic regulation by Z-DNA silencer function controls cancer-associated ADAM-12 expression in breast cancer: cross-talk between MeCP2 and NF1 transcription factor family **73**, 736-744, (2013).
- 3 Zacharias, W., Jaworski, A. & Wells, R. D. *J Bacteriol* Cytosine methylation enhances Z-DNA formation in vivo **172**, 3278-3283, (1990).
- 4 Lagravere, C., Malfoy, B., Leng, M. & Laval, J. *Nature* Ring-opened alkylated guanine is not repaired in Z-DNA **310**, 798-800, (1984).
- 5 Kawai, K., Sugiyama, H., Fujimoto, K. & Saito, I. *Nucleic Acids Symp Ser* 8-Methylguanine-containing oligonucleotides: useful structural constraint for Z form DNA, 33-34, (1995).
- 6 Van Lier, J. J., Smits, M. T. & Buck, H. M. *Eur J Biochem* B-Z transition in methylated DNA. A quantum-chemical study **132**, 55-62, (1983).
- 7 Fujii, S., Wang, A. H., van der Marel, G., van Boom, J. H. & Rich, A. *Nucleic Acids Res* Molecular structure of (m5 dC-dG)<sub>3</sub>: the role of the methyl group on 5-methyl cytosine in stabilizing Z-DNA **10**, 7879-7892, (1982).
- 8 Moller, A., Nordheim, A., Nichols, S. R. & Rich, A. *Proc Natl Acad Sci USA* 7-Methylguanine in poly(dG-dC).poly(dG-dC) facilitates z-DNA formation. **78**, 4777-4781, (1981).
- 9 Rothenburg, S., Koch-Nolte, F. & Haag, F. *Immunol Rev* DNA methylation and Z-DNA formation as mediators of quantitative differences in the expression of alleles **184**, 286-298, (2001).
- 10 Kouzine, F., Sanford, S., Elisha-Feil, Z. & Levens, D. *Nat Struct Mol Biol* The functional response of upstream DNA to dynamic supercoiling in vivo **15**, 146-154, (2008).
- 11 Kouzine, F. & Levens, D. *Front Biosci* Supercoil-driven DNA structures regulate genetic transactions **12**, 4409-4423, (2007).
- 12 Kouzine, F., Liu, J., Sanford, S., Chung, H. J. & Levens, D. *Nat Struct Mol Biol* The dynamic response of upstream DNA to transcription-generated torsional stress **11**, 1092-1100, (2004).
- 13 Klysik, J., Stirdivant, S. M., Larson, J. E., Hart, P. A. & Wells, R. D. *Nature* Left-handed DNA in restriction fragments and a recombinant plasmid **290**, 672-677, (1981).
- 14 Singleton, C. K., Klysik, J., Stirdivant, S. M. & Wells, R. D. *Nature* Left-handed Z-DNA is induced by supercoiling in physiological ionic conditions **299**, 312-316, (1982).
- 15 Klysik, J., Stirdivant, S. M., Singleton, C. K., Zacharias, W. & Wells, R. D. *J Mol Biol* Effects of 5 cytosine methylation on the B-Z transition in DNA restriction fragments and recombinant plasmids **168**, 51-71, (1983).

- 16 Fuertes, M. A., Perez, J. M., Gonzalez, V. M. & Alonso, C. *J Biol Inorg Chem* A kinetic model for the B-Z transition of poly[d(G-C)].poly[d(G-C)] and poly[d(G-m5C)].poly[d(G-m5C)] **6**, 675-682, (2001).
- 17 Zacharias, W., Larson, J. E., Kilpatrick, M. W. & Wells, R. D. *Nucleic Acids Res* HhaI methylase and restriction endonuclease as probes for B to Z DNA conformational changes in d(GCGC) sequences **12**, 7677-7692, (1984).
- 18 Bothe, J. R., Lowenhaupt, K. & Al-Hashimi, H. M. *Biochemistry* Incorporation of CC steps into Z-DNA: interplay between B-Z junction and Z-DNA helical formation **51**, 6871-6879, (2012).
- 19 Alvey, H. S., Gottardo, F. L., Nikolova, E. N. & Al-Hashimi, H. M. *Nat Struct Mol Biol* Transient Hoogsteen Base-Pairs Occur Robustly in Duplex DNA **In Submission**, (2014).
- 20 Pohl, F. M. & Jovin, T. M. *J Mol Biol* Salt-induced Co-operative Conformational Change of a Synthetic DNA: Equilibrium and Kinetic Studies with Poly(dG-dC) **67**, 375-396, (1972).
- 21 Misra, V. K. & Honig, B. *Biochemistry* The electrostatic contribution to the B to Z transition of DNA **35**, 1115-1124, (1996).
- 22 Gueron, M. & Demaret, J. P. *Proc Natl Acad Sci USA* A simple explanation of the electrostatics of the B-to-Z transition of DNA **89**, 5740-5743, (1992).

## Chapter 5

### The Impact of C7-Deazaguanine and N1-Methylguanine Modifications on DNA

#### Structure and the Equilibrium Between B-DNA and Z-DNA\*

##### 5.1 Introduction

###### *5.1.1 Importance of C7-Deazaguanine and N1-Methylguanine*

The structural perturbations necessary to undergo the B-to-Z-DNA transition highlight the possibility of a Hoogsteen (HG) or reverse Hoogsteen (rHG) base-pairing pattern during the transition as depicted in Figure A4.1 In Chapter 5 we originally sought to test the hypothesis that HG or rHG base-pairs are formed during the B-to-Z-DNA transition in CG<sub>3</sub> by use of chemical modification. C7-Deazaguanine (<sup>7d</sup>G) contains a single atom mutation whereby N7 of guanine is substituted to C-H, maintaining aromaticity, causing a minimal steric perturbation, and most importantly removing the free electron pair on N7 that is readily donated to form a hydrogen bond.<sup>1,2</sup> G•C<sup>+</sup> HG base-pairs form two hydrogen bonds between the carbonyl at guanine C6 and the exocyclic amino group at cytosine C4 as well as guanine N7 with protonated cytosine N3 (Figure 1.1). Removing the N7 donor electron pair thus prohibits the protonated guanine N7-cytosine N3 base-pair from forming and destabilizes transient HG base-pairs such that their population diminishes to nearly 0 %, <sup>3</sup> and if they still occur in canonical DNA duplexes, they evade detection by NMR R<sub>1ρ</sub> relaxation dispersion. Moreover, a study by Seela and coworkers showed that a uniformly G-<sup>7d</sup>G modified CpG repeat did not

---

\*HSA and HMA conceived the idea; HSA prepared samples and carried out the experiments and data analysis with help from HMA.

undergo the B-to-Z-DNA transition in 4.0 M NaCl,<sup>4</sup> although sample treatment was ambiguous, leading to discrepancies in interpretations. By installing <sup>7d</sup>G into CG<sub>3</sub> we aim to test the hypothesis that the guanine N7 is involved in hydrogen bonding in the B-to-Z-DNA transition and gain insight into the propensity for <sup>7d</sup>G-modified CpG hexamers to adopt Z-DNA under equilibrium conditions.

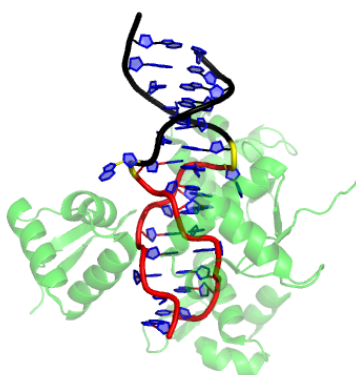
To further probe for the possibility of HG or base-pair formation during the B-to-Z-DNA transition, we installed the mutagenic damaged base N1-methylguanine (<sup>N1me</sup>G) to ‘trap’ the modified base-pairs in the HG conformation.<sup>3,5,6</sup> Trapping a HG base-pair in CG<sub>3</sub> would significantly stabilize any structure for which such base-pairs are involved and potentially manifest in the kinetic observations. The caveat of installing modifications that trap HG base-pairs to study the B-to-Z-DNA transition is that they destroy the ability to form WC base-pairs, which are essential in both B- and Z-DNA. Nevertheless, we tested the affect of <sup>N1me</sup>G on the B-to-Z-DNA transition in a kinetic and thermodynamic manner in an effort to further understand the role of HG base-pairs in Z-DNA formation.

### 5.1.2 NMR Spectroscopy of Z-DNA

To date, the only comprehensive view of Z-DNA on the atomic scale has been obtained from X-ray crystallography<sup>4,7-12</sup> which typically selects for the most rigid conformation and has difficulty detecting flexibility and resolving degeneracy in fits of electron density.<sup>13</sup> NMR spectroscopy offers the unique opportunity to characterize biomolecular structure and dynamics at atomic resolution,<sup>14,15</sup> yet very few heteronuclear (<sup>1</sup>H-<sup>13</sup>C and <sup>1</sup>H-<sup>15</sup>N) NMR studies of Z-DNA resonances have been implemented.<sup>16</sup> 2D

$^1\text{H}$ - $^1\text{H}$  NOESY<sup>17-27</sup> and imino proton exchange<sup>20</sup> data have been collected on Z-DNA and even Z-RNA,<sup>28</sup> but due to severe limitations in NMR sensitivity under high salt conditions ( $\sim 300$  mM NaCl and above),<sup>29</sup> researchers have performed modifications to stabilize Z-DNA which in turn alters the dynamics in question. NMR experiments at molar salt concentrations are possible but require extensive and costly instrumentation upgrades in order to use a cryogenic probe, or significant sensitivity must be sacrificed to use a room temperature probe, in which case transiently sampled species can evade detection.<sup>30</sup> Z-DNA dynamics by NMR have therefore mostly remained elusive.

To this end, we benchmark aromatic carbon chemical shifts of Z-DNA for a system that was previously demonstrated to adopt Z-DNA under a given set of conditions by  $^1\text{H}$  NMR.<sup>25</sup> Next, we employ the Z-DNA binding protein  $Z\alpha$  (Section 1.2.2) to induce Z-DNA to bypass the complications of using high salt for NMR experiments. A crystal structure of a B/Z junction was solved with high resolution ( $<2$  Å) where Z-DNA was induced by  $Z\alpha$ .<sup>31</sup> Amazingly, under crystallization conditions the extrusion of a single base-pair was sufficient for the double helix to adapt from a left-handed to a right-handed configuration (Figure 5.1).



**Figure 5.1 | Crystal Structure of the  $Z\alpha$ -Stabilized B/Z Junction.**<sup>4</sup> Shown with black and red backbones are B- and Z-DNA, respectively. Extruded A•T base-pair junction is highlighted in yellow. Four  $Z\alpha$  domains are shown in green. PDB ID: 2ACJ.

Moreover, the CpG hexameric sequence studied in this thesis is embedded within the Z-DNA forming region, making this construct a desirable candidate for studying Z-DNA using NMR spectroscopy. We seek to obtain atomic resolution dynamics and structural information in solution on Z-DNA using the B/Z junction sequence (ZJXN) presented in Figure 5.1 as a model system for Z-DNA formation. The appendage of B-DNA adjacent to the Z-DNA forming sequences can potentially aid in deconvoluting affects of protein binding versus those of Z-DNA formation. Moreover, junction and Z-DNA formation may alter B-DNA properties (i.e. chemical shift changes due to weakened stacking, etc.), providing insights into the potential long-range affects of Z-DNA formation. Moreover, it is possible crystal packing forces make the complete extrusion of one base-pair comprising the B/Z junction more favorable than in solution. Indeed, solution experiments using the fluorescent probe 2-aminopurine (2AP) in place of the junctional adenine show a large (~5-fold) fluorescence increase upon addition of  $Z\alpha$ .<sup>16,32</sup> However, 2AP fluorescence intensity changes can be sufficient from changes in local electronic environment without complete base unstacking (i.e. helix stretching, etc.).<sup>16</sup> Single-stranded DNA has unique chemical shift perturbation as compared to duplex DNA, making heteronuclear NMR experiments ideal for probing the extent of extrusion of the junctional A•T base-pair. Finally, the chemical shift perturbation of the *anti-to-syn* conformational change is highly uniform (Figure 2.3 C) which could aid in further understanding its contribution to the B-to-Z-DNA transition. We therefore seek to characterize the affects of  $^{7d}G$  and  $^{N1me}G$  modifications as well as the nature of Z-DNA  $^{13}C$  chemical shifts as a function of  $Z\alpha$  titration.

## 5.2 Experimental Methods

### 5.2.1 DNA Sample Preparation

DNA samples for thermodynamic and kinetic measurements were prepared as described in Sections 3.2.2 and 3.2.3, respectively. Covalently modified DNAs were purchased from Midland Certified Reagent Co., including (<sup>5me</sup>CG)<sub>3</sub>, CG<sub>3</sub> with N1-methylguanine modification at G4 (WT<sup>N1meG</sup>) or concomitantly at G2, G4 and G6 (WT<sup>N1meG(2,4,6)</sup>), a single atom C7-deazaguanine modification at G4 ((CG)<sub>3</sub><sup>G47dG</sup>) or all G's ((C<sup>7d</sup>G)<sub>3</sub>). (<sup>5me</sup>CG)<sub>3</sub> Z-DNA sample was prepared as described previously.<sup>25</sup> Final buffer conditions were 10 mM Phosphate buffer pH 7.0, 100 mM NaCl. <sup>13</sup>C/<sup>15</sup>N ZJXN DNA was prepared as described previously<sup>6,33</sup> and in Section 2.2.1. Za (9 kDa) was prepared at natural abundance as described previously<sup>32</sup> except the his tag cleavage step was omitted.

### 5.2.3 Circular Dichroism Equilibrium Measurements

Wavelength scans were measured on a Jasco Spectropolarimeter using a Peltier temperature control unit connected to a recirculating water bath as a heat dump. Spectra were acquired from 220 nm to 330 nm collecting data every 1 nm. Sensitivity was 100 mdeg, scanning rate was 100 nm/min, integration time was 1 sec. Samples were either 110 μM DNA with 200 μL final volume or 11 μM with 2 mL final volume. 200 μL and 2 mL samples were placed in 0.1 cm or 1 cm path length quartz cuvettes (Starna Cells), respectively. 12 scans were collected and averaged, except for CG<sub>3</sub> in which case one scan was acquired. Background subtraction was performed by scanning equivalent samples omitting the DNA. Temperature was 25 °C unless otherwise specified. Buffer



conditions were 15 mM Phosphate buffer pH, 0.1 mM EDTA. pH= 7.5 unless otherwise specified. In addition, 25 mM NaCl was used for cases where  $Z\alpha$  was present and corresponding controls. Samples without  $Z\alpha$  were brought to thermodynamic equilibrium by heating to 95 °C for 5 min followed by cooling at room temperature for at least 10 min. Samples with  $Z\alpha$  were incubated at room temperature for either 1 hr or 1 day as specified in plots.

#### 5.2.4 NMR Experiments

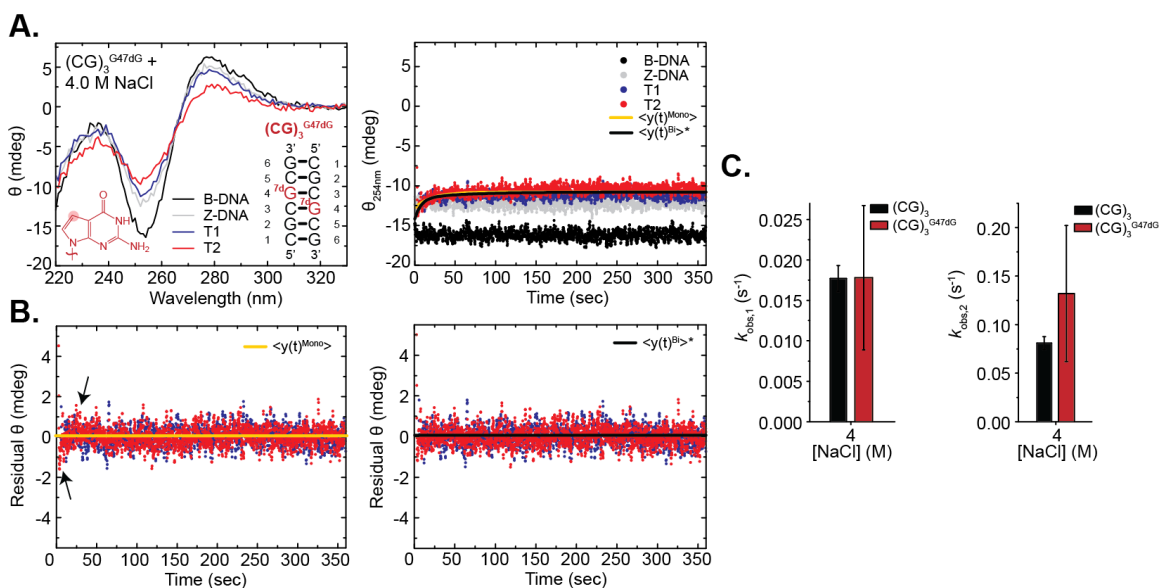
Buffer was 15 mM Phosphate buffer, 0.1 mM EDTA for WT<sup>N1meG</sup> and WT<sup>N1meG(2,4,6)}</sup>, pH= 7.5 unless otherwise specified, 10 mM Phosphate buffer pH 7.0, 100 mM NaCl for (<sup>5me</sup>CG)<sub>3</sub>, and 15 mM Phosphate buffer pH 7.5, 0.1 mM EDTA, 25 mM NaCl for <sup>13</sup>C/<sup>15</sup>N ZJXN in absence and presence (R=2 and R=4) of natural abundance  $Z\alpha$ . Samples were dialyzed or buffer exchanged (Amicon 3 kDa molecular weight cut-off) against RNase/DNase free water, lyophilized and placed in 90 % buffer 10 % D<sub>2</sub>O and transferred to Shigemi tubes. In the case of (<sup>5me</sup>CG)<sub>3</sub>, lyophilized sample was resuspended in 100 % buffer, lyophilized and resuspended in 100 % D<sub>2</sub>O, lyophilized and D<sub>2</sub>O then deuterated methanol (MeOD) was added to a final relative volume of 60 % D<sub>2</sub>O 40 % MeOD. Temperature was 25 °C for covalently modified samples unless otherwise specified and 18 °C in the case of (<sup>5me</sup>CG)<sub>3</sub>. NMR experiments were ran on ~4 mM natural abundance (CG)<sub>3</sub><sup>N4N1meG</sup>, (C<sup>N1meG</sup>)<sub>3</sub> and (<sup>5me</sup>CG)<sub>3</sub> or ~0.4 mM uniformly <sup>13</sup>C/<sup>15</sup>N labeled ZJXN samples. A Mixing time of 320 msec was used for (C<sup>N1meG</sup>)<sub>3</sub> and (CG)<sub>3</sub><sup>N4N1meG</sup>. A mixing time of 375 msec was used for (<sup>5me</sup>CG)<sub>3</sub>. 2D aromatic (<sup>1</sup>H2/6/8-<sup>13</sup>C2/6/8) and sugar (<sup>1</sup>H1'-<sup>13</sup>C1') HSQCs were ran on (CG)<sub>3</sub><sup>N4N1meG</sup>, (C<sup>N1meG</sup>)<sub>3</sub> and

(<sup>5me</sup>CG)<sub>3</sub>. 2D aromatic SOFAST-HMQC experiments were ran on Zα:<sup>13</sup>C/<sup>15</sup>N ZJXN R=2 (~27 kDa) and R=4 (~45 kDa). Samples were ran on a Bruker Avance or Agilent (<sup>13</sup>C/<sup>15</sup>N ZJXN) 600 MHz NMR spectrometer equipped with a 5 mm triple resonance cryogenic probe.

## 5.3 Results and Discussion

### *5.3.1 Impact of the 7-Deazaguanine Modification on the B-to-Z-DNA Transition*

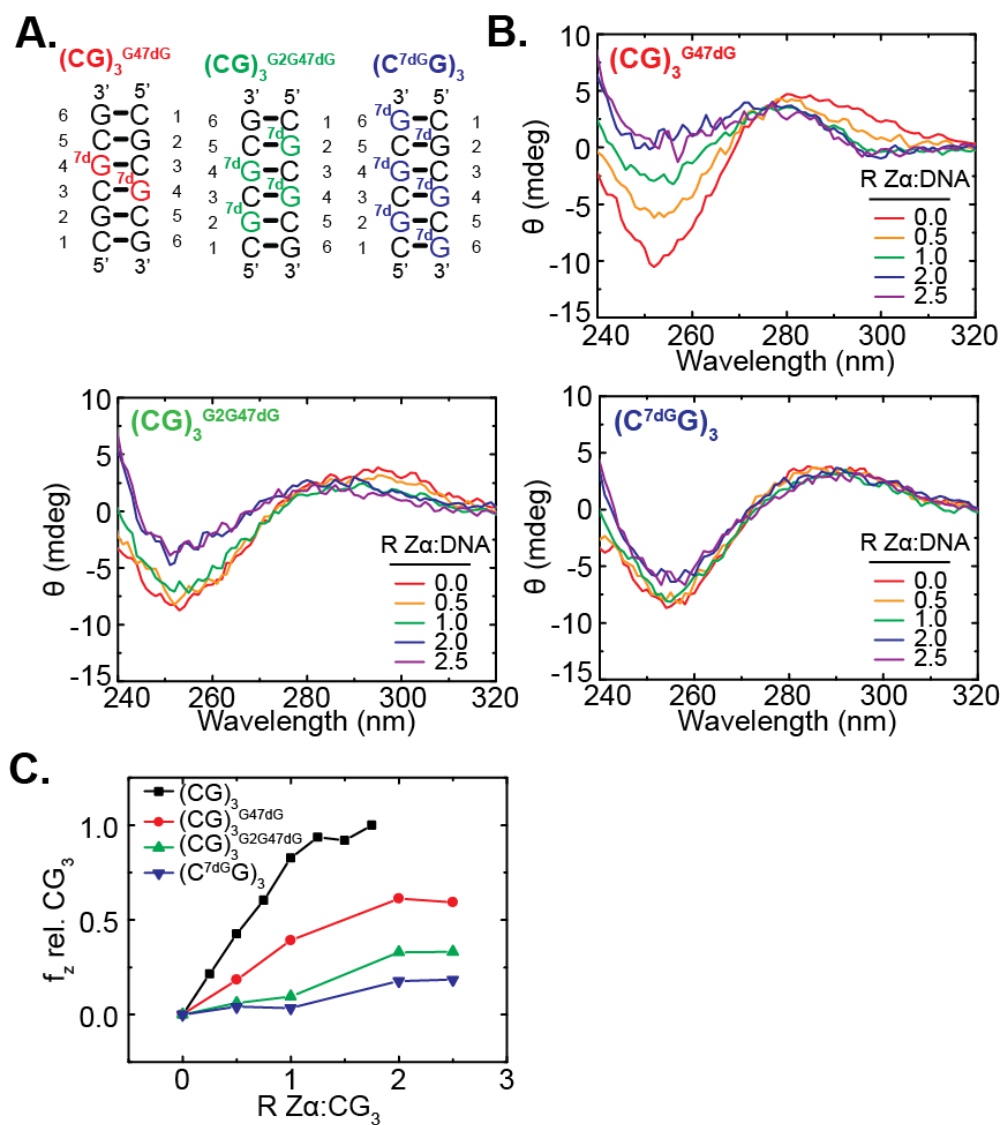
<sup>7d</sup>G was installed in CG<sub>3</sub> using two strategies using one, two or three G-<sup>7d</sup>G substitutions. The single modification causes the two most helical base-pairs to be modified due to the palindromic nature of CG<sub>3</sub> and will be used in kinetic assays. Due to the cooperative nature of the B-to-Z-DNA transition we anticipate an otherwise small perturbation to have an exacerbated affect on the transition. Upon evaluation of the thermodynamic beginning ((CG)<sub>3</sub><sup>G47dG</sup> 0 M NaCl) and ending (4M) points (Figure 5.2 A black and red spectra, respectively) it becomes immediately apparent that there is only ~5 mdeg difference at 254 nm. This severely limits the potential to observe the B-to-Z-DNA transition. Not surprisingly, a small curve is observed in the transients that is difficult to resolve from random signal fluctuations. Nevertheless residual biases for monophasic (Eq. 5.3) fitting do show biases that are moderately accommodated for upon biphasic (Eq. 3.6) fit. F- and AIC tests (Section 1.5) both selected the biphasic model over that of the monoexponential. No change in observed rate constant is observed either. These results suggest that mutation of the two most helical base-pairs from guanine to <sup>7d</sup>G in CG<sub>3</sub> severely limit the thermodynamic propensity to undergo the transition while the observed rate constants are similar to those reported for CG<sub>3</sub> (Section 3.3.4).



**Figure 5.2 | Preliminary Transient Kinetics of the B-to-Z-DNA Transition in  $(CG)_3^{G47dG}$ .** (A) Kinetic traces of the transition (right). Wavelength scan (left) and single wavelength (right) controls, including reference samples at thermodynamic equilibrium (B-DNA: Black, Z-DNA: Grey) and kinetic samples after the B-Z transition has reached equilibrium. (B) Residuals of mono- (left) and bi- (right) exponential fits to Eq. 3.5 and 3.6, respectively, shown in 'A' (right). Arrows highlight poorly distributed residual signal about the fit for monoexponential fit. 'T' represents independent trials. \* Indicates significantly better fit as determined by AIC and F-Tests with 95 % confidence level. (C) Average observed rate constants obtained from fits of Eq. 3.6 to each trial shown in A and B. Error bars represent standard deviation of the observed rate constants.

It is possible the microscopic rate constants change in such a way that the observed rate constants simply vary within the range of reproducibility from independent trials. More experiments are necessary to characterize the kinetic affect of the  $^{7d}G$  mutation, perhaps under conditions that make Z-DNA formation more favorable. Due to the difficulties mentioned above in measuring the kinetics upon  $^{7d}G$  mutation, we turned to thermodynamic investigations as well as exploratory techniques to try and push the B-Z equilibrium more toward Z-DNA. Understanding these affects will provide key clues into the essential structural interactions along the B-to-Z-DNA transition pathway. We note that mutations have the potential to change the transition pathway, so ultimately care

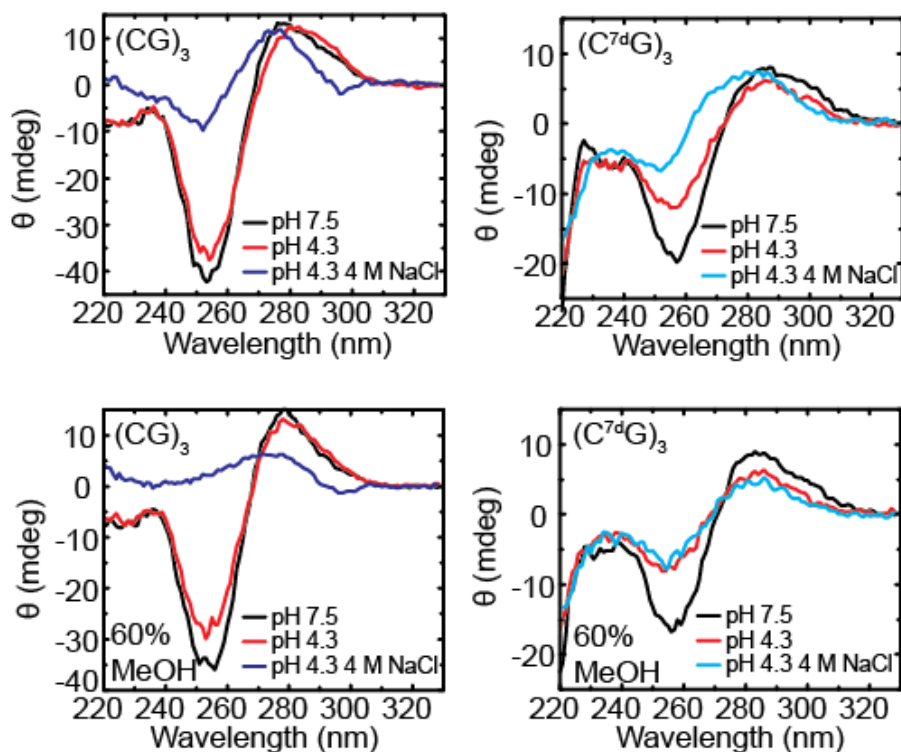
must be taken when making inferences about the mechanistic scheme upon mutation. For the purposes of evaluating the thermodynamic tendencies of  ${}^7\text{dG}$ -containing sequences, we use a single, double and triple  ${}^7\text{dG}$  substitution (Figure 5.3 A) and induce the transition using  $Z\alpha$  for its tendency to stabilize Z-DNA in sequences that do not convert readily to Z-DNA in high salt.<sup>19,34</sup> For all three substitution varieties we titrate  $Z\alpha$  up to saturation as monitored by CD ( $R= 2.5$ ) and observe a change in CD profiles consistent with the B-to-Z-DNA transition (Figure 5.3). Increasing the amount of  ${}^7\text{dG}$  mutations significantly decreases the propensity to adopt Z-DNA. When the fraction of Z-DNA relative to that of  $\text{CG}_3$  is calculated for each titration, it is apparent that the modest extent of B-to-Z-DNA transition is saturated at  $R= 2$  independent of the number of  ${}^7\text{dG}$  mutations present (Figure 5.3 C). This observation could suggest that the  $Z\alpha$  affinity for the DNA substrate is not the main reason for lack of Z-DNA formation, that rather, the inherent instability of  ${}^7\text{dG}$ -containing Z-DNA relative to B-DNA is dominating this affect. Consistent with our observations, it has been shown that substitution at the 7 position of purines stabilize B-DNA.<sup>1,35,36</sup> However, measuring the affinity of  $Z\alpha$  for each  ${}^7\text{dG}$ -containing species relative to  $\text{CG}_3$  using a band shift or standard NMR titration assay would confirm this.



**Figure 5.3 | Thermodynamic Propensity of N7-Deazaguanine Mutations.** (A) Shown are sequences used with increasing number of mutation sites and 7-deazaguanine moiety. Titrations of  $Z\alpha$  into (B)  $(CG)_3^{G47dG}$ ,  $(CG)_3^{G2G47dG}$  and  $(C^{7d}GG)_3$  and (C) fraction of Z-DNA formed relative to  $CG_3$  ( $f_z$  rel.  $CG_3$ ).

We next sought to understand what solution conditions might considerably stabilize Z-DNA in the  $^{7d}G$  modified sequences. Nucleophilic solvents are known to stabilize Z-DNA in non-modified pyr-pur repeats and it is believed this is due to the electrophilic purine N7 becoming more solvent exposed in Z-DNA relative to B-DNA.<sup>37,38</sup> It is known that low pH also favors Z-DNA in unmodified DNAs.<sup>39</sup> If low pH

stabilizes the Z-form of  $^7\text{dG}$ -containing species, it is possible protonation makes possible a distorted Z-form helix in which the  $^7\text{dG}$  base may be accommodated, albeit potentially distorted relative to the canonical Z-DNA WC base-pair. To test these hypotheses we use  $(\text{C}^{7\text{d}}\text{G})_3$  for its poorest Z-DNA favorability in high NaCl and  $Z\alpha$  conditions. We find that pH 4.3 with 4.0M NaCl stabilizes  $(\text{C}^{7\text{d}}\text{G})_3$  in Z-form and that even pH 4.3 in the 25 mM NaCl induces a change in structure consistent with the B-to-Z-DNA transition (Figure 5.4). Moreover, the nucleophilic solvent methanol (MeOH) can be used (60 % v/v) in place of 4.0M NaCl for the case of  $(\text{C}^{7\text{d}}\text{G})_3$  pH 4.3. These observations demonstrate that both nucleophilic solvent and low pH markedly stabilize Z-DNA even upon the unfavorable  $^7\text{dG}$  substitution. With all electrophilic N7's removed in  $(\text{C}^{7\text{d}}\text{G})_3$ , it is possible other nucleophilic interactions and protonation events stabilize the Z-form of such chemically modified DNA sequences (i.e. other tertiary nitrogens as well as oxygens and phosphates may be involved). Lastly, since conditions were found that push the B-to-Z-DNA equilibrium of  $(\text{C}^{7\text{d}}\text{G})_3$  sufficiently toward Z-DNA (4.0 M NaCl, pH 4.3, 25 °C), kinetic studies of the corresponding transition can be measured under identical conditions. Kinetic measurements of  $^7\text{dG}$ -containing DNAs would directly probe the impact of HG base-pairs on the B-to-Z-DNA transition.

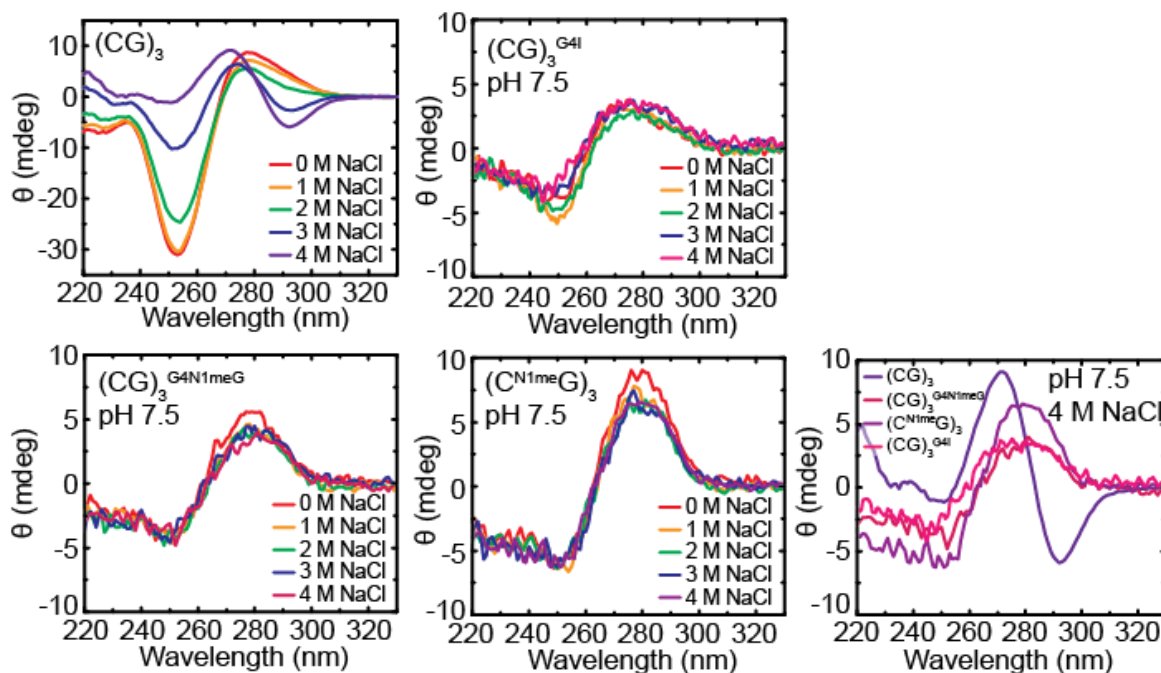


**Figure 5.4 | pH and Solvent Affects on the B-Z-DNA Equilibrium in  $(C^{7d}G)_3$ .** NaCl, pH and methanol (MeOH) affect on  $CG_3$  and  $(C^{7d}G)_3$  B-to-Z-DNA transition. 15 mM Phosphate buffer, 0.1 mM EDTA, 25 mM NaCl unless otherwise specified, 25 °C, pH specified in plot insets.

### 5.3.2 Impact of the N1-Methylguanine Modification on DNA Structure

$N^{1me}G$  is a naturally occurring form of damage in mammalian cells that we use here to trap the HG form of C•G base-pairs. Interestingly, installing one  $N^{1me}G$  at the most helical  $CG_3$  base-pair causes a dramatic structural change in low-salt conditions. Due to the loss of roughly 85 % of the CD signal with one or three  $N^{1me}G$  substitutions, we thought the modification might be causing the double helix to adopt an unstructured helix. Since even single-stranded DNAs are chiral, it is also conceivable for such structures to give rise to this nominal signal. As a control we modified G4 to the purine derivative inosine (I) that is formed upon deamination of adenosine by  $Z\alpha$ .<sup>40</sup> Inosine

weakens WC base-pairs but does not destroy them. A purely IpC repeat helix is less stable in the Z-form as compared to a pure CpG repeat,<sup>41</sup> consistent with the notion of having weaker WC base-pairs in Z-form as well. The structure adopted by the <sup>N1meG</sup> and I modifications is relatively independent of salt and is not consistent with Z-DNA (Figure 5.5). Interestingly, however, the sequences undergo a slow *Zα*-induced conformational change (Figure A4.2).

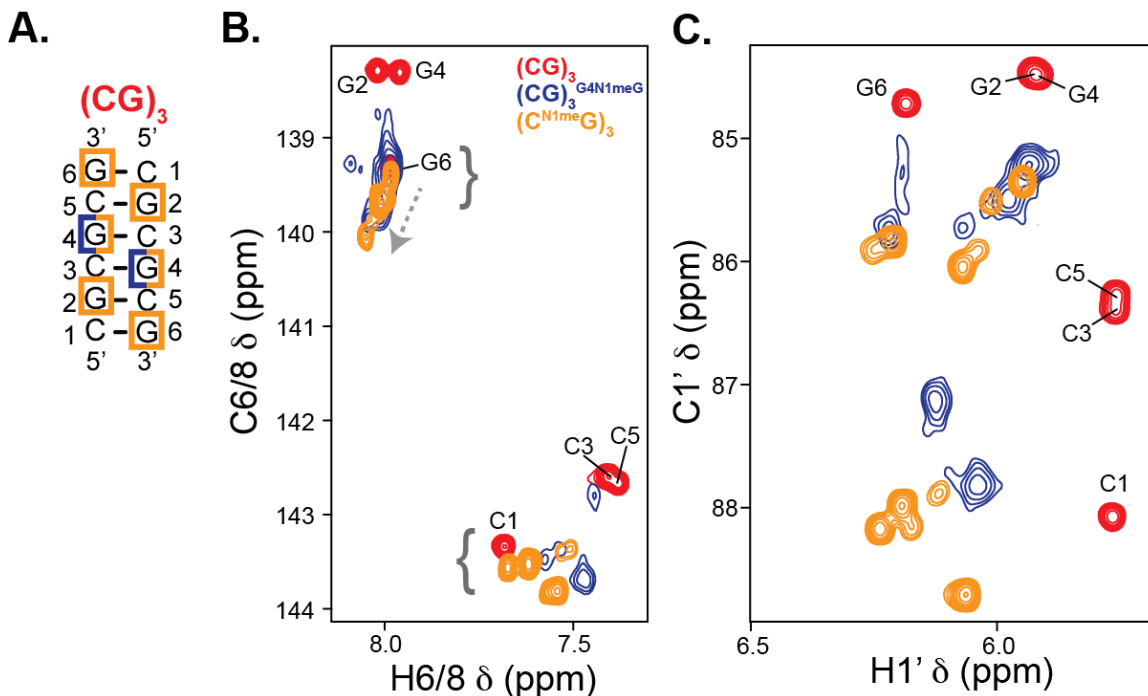


**Figure 5.5 | Structural Impact of N1-Methylguanine on Short DNAs.** CD spectra of NaCl titrations of (top)  $CG_3$  and  $(CG)_3^{G4I}$  for comparison to (bottom)  $(CG)_3^{G4N1meG}$  and  $(C^{N1meG})_3$ . (Bottom right) Shown are CD Spectra for samples in 4.0M NaCl for comparison. 15 mM Phosphate buffer pH 7.5, 0.1 mM EDTA, 25 °C.

To better understand the structural features of the <sup>N1meG</sup> and I modifications we turn to NMR heteronuclear correlation experiments (here, HSQCs)<sup>42</sup> since the chemical shifts of helical, terminal, single-stranded and HG conformations are generally distinct.<sup>6,43</sup> To our surprise, we found the <sup>N1meG</sup> substituted sequences to possess chemical shifts strikingly similar to those of terminal residues that undergo end-fraying



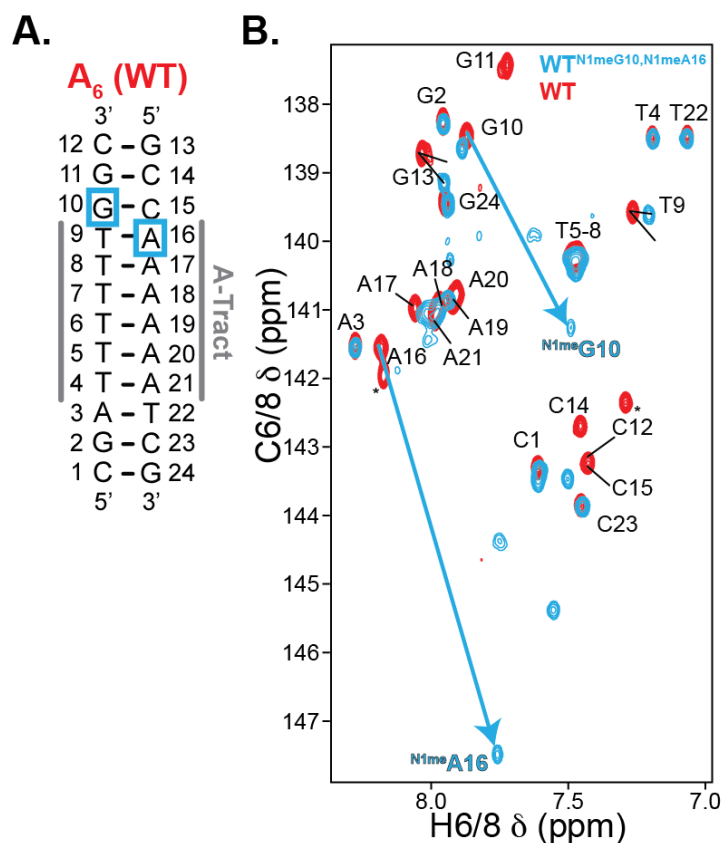
affects (Figure 5.6). The  $G^{N1me}G$  aromatic chemical shifts also exhibit increasing single-stranded nature as indicated by a grey dotted arrow in Figure 5.6 B. Furthermore, imino proton peaks were not detected in 1D  $^1H$  NMR experiments at 5 °C for  $(CG)_3^{G4N1meG}$  (data not shown). These data strongly suggest the  $N1meG$  modified sequences do not form a double helix but rather remain single-stranded.



**Figure 5.6 | Changes in  $^{13}C$  and  $^1H$  NMR Chemical Shifts due to the N1-Methylguanine Modification in  $CG_3$ .** (A) Wild type  $(CG)_3$  sequence shown in black with blue and yellow indicating positions of N1-methylguanine mutations in singly and triply substituted sequences, respectively.  $^1H$ - $^{13}C$  HSQCs showing (B) H6-C6 and H8-C8 (aromatic) and (C) H1'-C1' (sugar) resonances for sequences shown in 'A'. Grey brackets and arrow in aromatic HSQC represent typical chemical shifts for end fraying and single-stranded DNA, respectively.

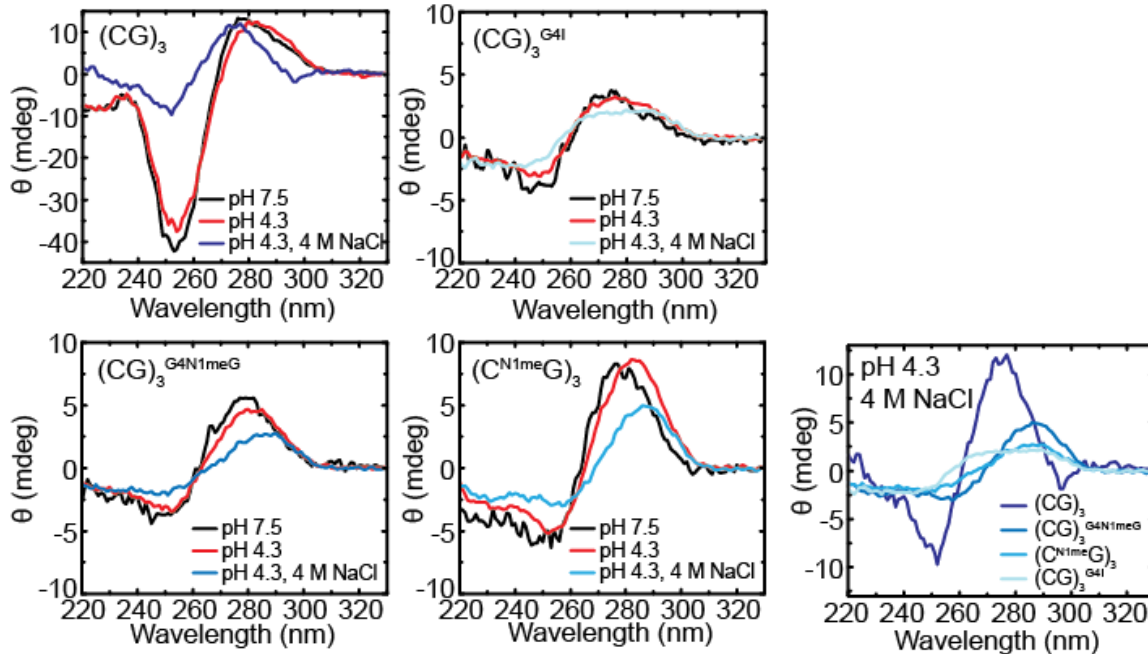
For  $(C^{N1me}G)_3$  HSQCs we see double the number of peaks expected indicating two conformations are present that are in slow exchange with one another. Interestingly, the guanine H8-C8 resonances do not exhibit conformational heterogeneity. They may be undergoing chemical exchange faster than the NMR timescale (psec-low msec) or perhaps they are conformationally rigid. Using  $R_{1\rho}$  relaxation dispersion experiments to

probe for  $\mu\text{sec}$ -  $\text{msec}$  dynamics would help alleviate this conundrum. For  $(\text{CG})_3^{\text{G4N1meG}}$  HSQCs we observe approximately 6 aromatic and sugar peaks as expected for a non- or fast-exchanging system. Interestingly, we do not observe this affect in  $\text{A}_6$  with two neighbors N1-methylated (Figure 5.6) despite its lower melting temperature (Table A.12). It is also possible the single and triple  $\text{N1meG}$  modified sequences adopt different conformations. More experiments are necessary to fully understand the structures adopted by the  $\text{N1meG}$  substituted sequences and the equilibrium between those structures. The experiments presented in this section provide strong evidence for the lack of Z-DNA conversion in  $\text{N1meG}$  modified sequences being attributed to their single-stranded nature.



**Figure 5.7 | Changes in  $^{13}\text{C}$  and  $^1\text{H}$  NMR Chemical Shifts due to the N1-Methylguanine Modification in  $\text{A}_6$ .** (A) Wild type ( $\text{A}_6$ ) sequence shown in black with blue indicating positions of N1-methylguanine mutations.  $^1\text{H}$ - $^{13}\text{C}$  HSQCs showing (B) H6-C6 and H8-C8 resonances for sequences shown in (A). Sample preparation and data acquisition performed by ENN.

Next, we hypothesized that the reason the <sup>N1me</sup>G mutation is not accommodated in the CG<sub>3</sub> duplex is because of the poor stacking efficiency of HG base-pairs relative to WC base-pairs that is suggested by our findings in Chapter 2. We use the combinatorial low pH and high salt strategy presented in Section 5.3.3 in an attempt to stabilize duplex formation in these <sup>N1me</sup>G modified sequences. We find that a change in pH alone does not alter the global structure, but low pH and high NaCl concentrations together do, in fact, induce a conformational change. This conformational change is actually consistent with Z-DNA formation as opposed to B-DNA formation, but the overall magnitudes of pH 4.3 4.0M NaCl spectra suggest the samples are still single-stranded in nature. Interestingly, this structural perturbation occurs to an increasing extent upon comparison of (CG)<sub>3</sub><sup>G4I</sup>, (CG)<sub>3</sub><sup>G4N1meG</sup> and (C<sup>N1me</sup>G)<sub>3</sub>, which correlates directly with each sample's propensity to form HG base-pairs. Due to the conserved magnitude of signal comparing pH 7.5 to pH 4.3 4.0 M NaCl profiles, it is likely the conformational change did not merit duplex formation.



**Figure 5.8 | pH Affects on Modified DNAs.** CD spectra of pH and NaCl change of (top)  $(CG)_3$  and  $(CG)_3^{G41}$  for comparison to (bottom)  $(CG)_3^{G4N1meG}$  and  $(C^{N1meG})_3$ . (Bottom right) Shown are CD Spectra for samples in 4.0M NaCl pH 4.3 for comparison. 15 mM Phosphate buffer, 0.1 mM EDTA, 25 °C, pH and [NaCl] specified in plots.

The CD profiles do, however, change in a fashion consistent with the B-to-Z-DNA transition. This suggests the conformational change may involve a single-stranded right-to-left-handed transition and subsequently that HG base-pairs can be accommodated within a left-handed framework. These data also raise the concept that, when a HG base-pair must be accommodated within a left-handed framework, that the collective HG stability of the overall structure (i.e. the sum of HG affects along at each base-pair to arrive at a global description) will contribute to the stability of Z-DNA. HG helices of pure A•T base-pairs have been studied by X-ray crystallography and solution NMR, both indicating such structures can be accommodated within the structural framework of a deformed right-handed helix.<sup>44,45</sup> Lastly, we note that the correlation of  $N^{1meG}$ -modified

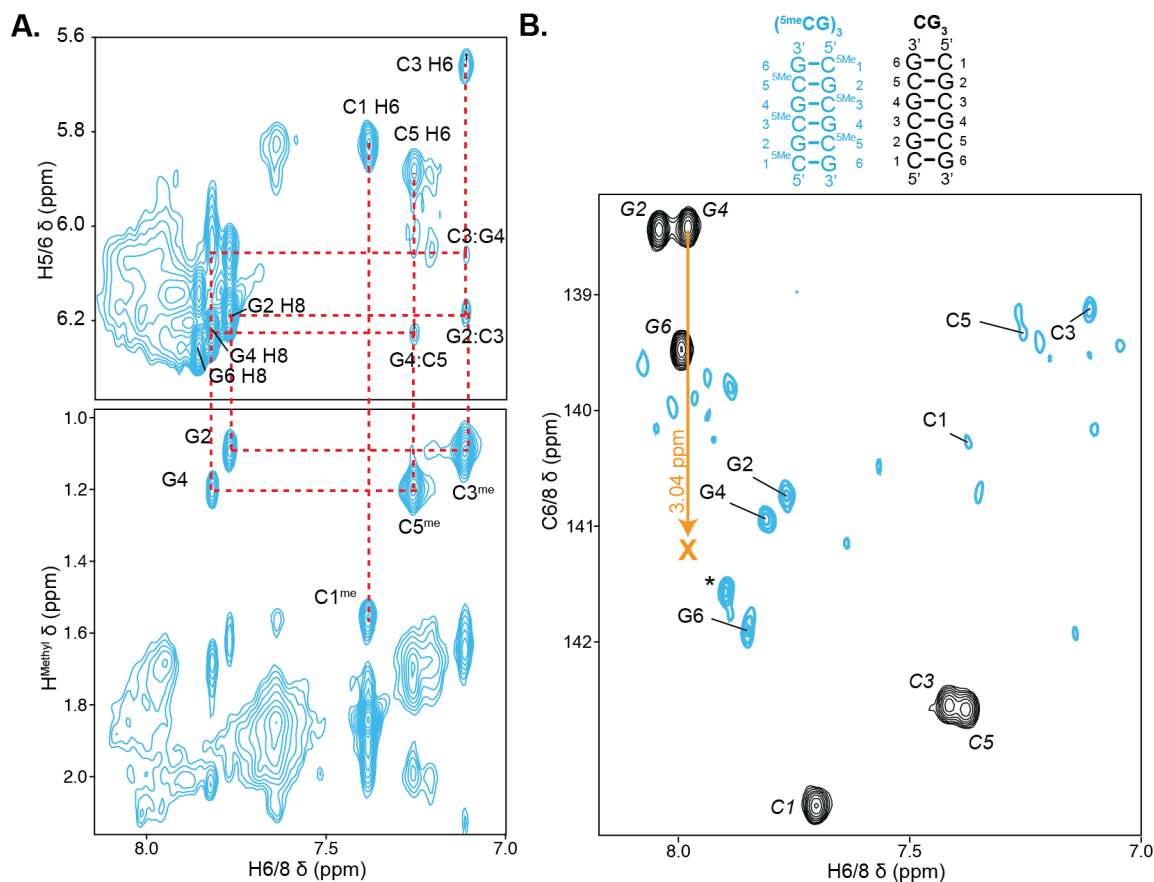
strand annealing of CpG repeats with HG forming propensity could reveal a new relationship between WC-HG equilibrium and larger conformational changes that could more readily be recognized *in vivo*. Future directions include testing the hypothesis that sequence specificity of the WC-HG equilibrium in N1-methylated DNA is correlated with duplex annealing propensity and potentially even investigating the sequence specificity of the cooperativity of neighboring HG formation.

### 5.3.3 Preliminary NMR Spectroscopic Investigations of Z-DNA: ( $^{\delta me}CG$ )<sub>3</sub> and a (CG)<sub>3</sub>:Z $\alpha$ Complex

We have prepared a sample that was previously demonstrated to adopt Z-DNA using a natural abundance ( $^{\delta me}CG$ )<sub>3</sub> construct<sup>25</sup> to investigate the behavior of the system before preparing a  $^{13}C/^{15}N$  labeled sample. The previous study did not report carbon chemical shifts, so we plan to benchmark our proton assignments against theirs, then subsequently measure  $^1H$ - $^{13}C$  correlations to gain insight into aromatic carbon chemical shifts of Z-DNA for the first time.

As demonstrated in Chapter 4,  $^{\delta me}C$  modifications and methanolic solvent increase the propensity to form Z-DNA in a mutually exclusive manner. The benchmark system utilized herein employs these two strategies to form ~80 % Z-DNA and ~20 % B-DNA. Standard  $^1H$ - $^1H$  NOESY assignments were implemented and are consistent with those previously published (Figure 5.9 A).<sup>25</sup> Corresponding HSQC reveals aromatic carbon chemical shifts of Z-DNA for the first time (Figure 5.9 B). Guanine C8 resonances are shifted downfield ~ 2 ppm, consistent with the *anti*-to-*syn* transition as monitored *via* the WC-to-HG transition and demonstrated in Chapter 2<sup>43</sup> and previously

published studies.<sup>3,6</sup> We note that the proton chemical shift of the WC-to-HG transition can be as large as  $\sim 0.5$  ppm,<sup>6</sup> while those of guanine and cytosine presented here do not change markedly from B- to Z-DNA, consistent with previous findings.<sup>25,26,37,38</sup> Interestingly, cytosine aromatic chemical shifts move upfield  $\sim 2$  ppm, indicating they are deshielded (Figure 1.8) relative to their B-DNA structure. This is conceivable since cytosine loses one of its neighboring base-pair stacking interactions upon Z-DNA formation due to the extension of the helix and increased step size between base-pair neighbors (Table 1.2). The strong intraresidue crosspeaks in the  $^1\text{H}$ - $^1\text{H}$  NOESY suggest that the weak cytosine C6-H6 resonances in the aromatic  $^1\text{H}$ - $^{13}\text{C}$  HSQC are not predominantly due to chemical exchange. Further optimization of sample conditions to minimize precipitation upon addition of methanol to high salt solutions<sup>25</sup> as well as potential adjustments to the HSQC experiment, such as optimizing the carbon carrier frequency and excitation pulse bandwidth could improve spectral quality.



**Figure 5.9 | NMR Spectroscopy of  $(^5\text{meCG})_3$  in Z-Form.** (A)  $^1\text{H}$ - $^1\text{H}$  NOESY and (B) Sequence and  $^1\text{H}$ - $^{13}\text{C}$  aromatic HSQC of  $(^5\text{meCG})_3$  (light blue) overlaid with B-form  $\text{CG}_3$  (black). Experimental conditions for  $(^5\text{meCG})_3$  are 10 mM Phosphate buffer, 100 mM NaCl, 60 % MeOD (deuterated methanol), pD 6.6 (pH 7.0), 18 °C. Experimental conditions for  $\text{CG}_3$  are 15 mM Phosphate buffer, 0.1 mM EDTA, 25 mM NaCl, pH 5.4. Orange highlights the transient state Hoogsteen chemical shift of  $\text{CG}_3$  G4 C8 measured in Chapter 2 as a reference for *syn* guanine carbon chemical shifts.

Current NMR studies have provided some insight into features of the B-to-Z-DNA transition in the B/Z junction sequence first presented in Chapter 2 (ZJXN) and similar sequences, relying predominantly on imino proton exchange.<sup>17,46,47</sup> These reports demonstrate that when B-DNA is adjacent to a CpG repeat, the B-DNA lifetime of the base open state increases in the transition. If the junctional base-pair goes from base-paired in B-DNA to extruded upon Z-DNA formation, were completely flipped out in

solution, the exchange rate between the closed and open state ( $k_{\text{cl-op}}$ ) would increase, then become immeasurable once the base open state has a long enough lifetime to where the thymine (T) imino proton exchange with water and broadens out the T imino peak. To the contrary, imino proton exchange measurements reveal the base pair that is extruded in the crystal structure has a low  $k_{\text{cl-op}}$  that does not change from R= 0 to R= 1.5 ( $k_{\text{cl-op}} \sim 8 \text{ s}^{-1}$ ).<sup>17</sup> This suggests the junctional base-pair that is extruded in the crystal structure is not extruded in solution. In an attempt to discern between the extruded or base-paired junction, we again turn to heteronuclear HSQCs to probe for chemical shift perturbations of the junctional A•T base-pair in a similar fashion to that in Section 5.3.4.

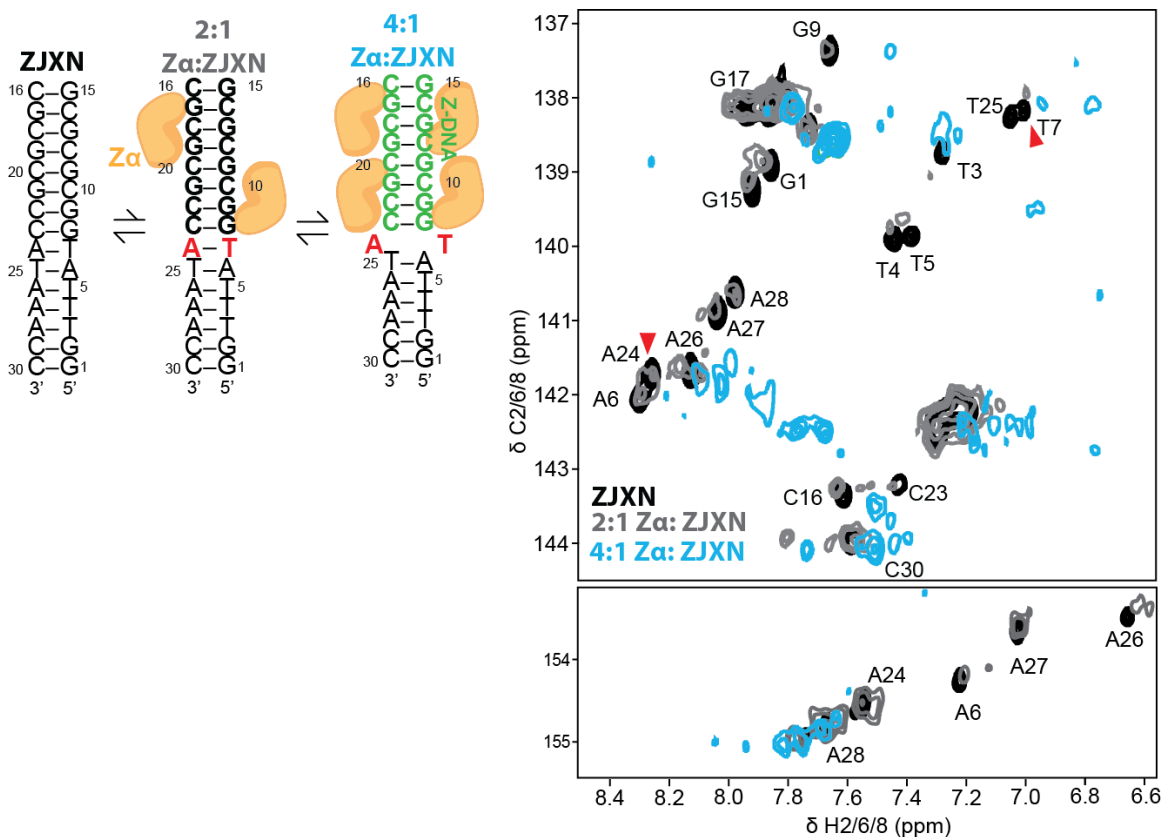
In Figure 5.10 we present a titration of natural abundance  $Z\alpha$  into  $^{13}\text{C}/^{15}\text{N}$  labeled ZJXN. From these experiments dynamics can only be inferred within each measurement since overall molecular weight, and subsequently tumbling time and line width, changes from one titration point to the next. In the absence of protein, all ZJXN resonances are relatively sharp (Figure 5.10 black). At R= 2:1 nearly all B-DNA resonances are still observable but have undergone significant line broadening due to an increase in effective molecular weight from  $\sim 9$  kDa to  $\sim 27$  kDa. We note the R= 2:1 complex has less than half the amount of protein required to saturate the B-Z conversion (saturation is at R= 6).<sup>48</sup> Since there are more peaks than C-H resonances under R=2:1 conditions, it is likely B-DNA is in slow exchange with another conformation. These additional chemical shifts are likely not due to protein interaction with the C-H resonances as  $Z\alpha$  only makes contacts with the DNA phosphate backbone.<sup>4</sup> We note with respect to extra peaks in the regions of Z-DNA forming sequence (guanine and cytosine), the cytosine region contains markedly more extra peaks, albeit with low intensity, as compared to the guanine region.



At R= 4 the CG-rich region adopts Z-DNA and the A•T junctional base-pair (A24 and T7) is putatively flipped out into solution. It is possible addition of more protein would further stabilize the Z-form, but at R= 4 under the conditions used herein for NMR, the protein is at its solubility limit (~1 mM). Most importantly, the junctional base-pair resonances have completely shifted from their respective B-DNA chemical shifts (Figure 5.10 red arrow heads). Peaks in the R= 4 spectrum are not detectable in the region where A24 and T7 are expected to shift if they are extruded (~1 ppm downfield in carbon and proton). This does not preclude base-pair extrusion at the B/Z junction since the final titration point is not at saturating protein concentration. To bypass this limitation in the future we plan to isotopically label ZJXN at specific residues, allowing for immediate assignment and precise probing of a given resonance. We note that with simply these HSQC experiments we cannot rule out a number of other scenarios causing disappearance of A24 and T7 chemical shifts, including line broadening due to slower tumbling of the ~45 kDa 4:1 complex compared with the ~9 kDa free DNA or intermediate exchange. It is also possible A24 and T7 chemical shifts have changed in an unexpected direction. Chemical shift assignments will help aid in the identification of signatures of the B/Z junction.

Remarkably, despite previous studies demonstrating that the CD spectrum of the  $Z\alpha$ :ZJXN complex can, in fact, be fit to a linear combination of B-DNA and Z-DNA spectra,<sup>16</sup> nearly all B-DNA resonances shift or disappear upon Z-DNA formation under the conditions used herein. This observation suggests that even if the extrusion of one base-pair is sufficient to accommodate a left-handed structure next to a right-handed structure, that any potential dynamic affects due to Z-DNA do propagate past the

junction, into the neighboring B-DNA. If this were true, the dynamic affects on Z-DNA formation *in vivo* could spread even further than the Z-DNA location and perhaps act as a long-range signaling mechanism to recruit DNA binding proteins.



**Figure 5.10 | Titration of Z $\alpha$  into  $^{13}\text{C}/^{15}\text{N}$  ZJXN.** Aromatic SOFAST-HMQC spectra of ZJXN (black), a Z $\alpha$ :ZJXN 2:1 complex (grey) and a Z $\alpha$ :ZJXN 4:1 complex (blue). 15 mM Phosphate buffer pH 7.5, 0.1 mM EDTA, 25 mM NaCl, 25 °C. B-DNA chemical shifts are labeled according to previously published assignments.<sup>16</sup> Red arrowheads denote B-DNA chemical shifts of junctional base-pair.

Additional studies are needed to explore this hypothesis in more detail, including Z-DNA chemical shift assignments. Assignments of the Z $\alpha$ :ZJXN complex *via*  $^1\text{H}$ - $^1\text{H}$  NOESY are non-trivial due to the excess of protein relative to DNA. A conventional assignment experiment will detect  $^1\text{H}$ - $^1\text{H}$  correlations of all protons in the system. To detect  $^1\text{H}$ - $^1\text{H}$  correlations of the DNA only, one must select for  $^1\text{H}$  spins connected to  $^{13}\text{C}$ ,

which requires implementation of a new pulse sequence, optimization for DNA resonances, and finally a well-behaved model system.

## 5.4 Conclusions

$^{7d}\text{G}$  mutation significantly diminishes propensity for  $\text{CG}_3$  to form Z-DNA thermodynamically. As such, our kinetic experiments cannot discern between any distinct changes in observed rate constants. We have presented a detailed analysis of the Z-DNA forming capabilities of CpG repeats containing  $^{7d}\text{G}$  and  $^{\text{N1me}}\text{G}$  mutations. We also demonstrate that a single methylation at guanine N1 on each strand is sufficient to denature the duplex. To a first approximation, the propensity for  $^{\text{N1me}}\text{G}$  modified sequences to adopt Z-DNA correlates with their HG forming propensity.  $\text{Z}\alpha$  binds single-stranded  $^{\text{N1me}}\text{G}$  modified sequences and slowly induces a conformational change consistent with the B-to-Z-DNA transition.

For the first time, to our knowledge, we have presented aromatic carbon chemical shifts using a benchmark system first presented by Rich and coworkers.<sup>25</sup> Guanine C8 chemical shifts are consistent with the *syn* orientation while, surprisingly, those of cytosine C6 shift upfield, consistent with being deshielded. This could be explained by the loss of stacking interaction with one neighboring base-pair upon Z-DNA formation.

Titration of  $\text{Z}\alpha$  into isotopically enriched ZJXN as monitored by aromatic SOFAST-HMQC shows many changes in chemical shifts, including the junctional A•T base-pair. The preliminary investigations into Z-DNA aromatic carbon chemical shifts call for further NMR studies of the  $\text{Z}\alpha$ : $^{13}\text{C}/^{15}\text{N}$  ZJXN complex to enhance signal-to-noise, assign chemical shifts, calculate changes in chemical shifts upon Z-DNA formation using

Density Functional Theory (DFT)<sup>49</sup> and ultimately probe the dynamics of Z-DNA at atomic resolution. Using residue-specific isotopic labeling strategies will propel these efforts greatly.

## 5.5 References

- 1 Seela, F., Peng, X. & Ming, X. *Nucleosides Nucleotides Nucleic Acids* 7-deazapurin-2,6-diamine and 7-deazaguanine: synthesis and property of 7-substituted nucleosides and oligonucleotides **24**, 839-841, (2005).
- 2 Seela, F. & Driller, H. *Nucleic Acids Res* Alternating d(G-C)<sub>3</sub> and d(C-G)<sub>3</sub> hexanucleotides containing 7-deaza-2'-deoxyguanosine or 8-aza-7-deaza-2'-deoxyguanosine in place of dG **17**, 901-910, (1989).
- 3 Nikolova, E. N., Gottardo, F. L. & Al-Hashimi, H. M. *J Am Chem Soc* Probing Transient Hoogsteen Hydrogen Bonds in Canonical Duplex DNA Using NMR Relaxation Dispersion and Single-Atom Substitution **134**, 3667-3670, (2012).
- 4 Ha, S. C., Lowenhaupt, K., Rich, A., Kim, Y. G. & Kim, K. K. *Nature* Crystal structure of a junction between B-DNA and Z-DNA reveals two extruded bases **437**, 1183-1186, (2005).
- 5 Delaney, J. C. & Essigmann, J. M. *Proc Natl Acad Sci USA* Mutagenesis, genotoxicity, and repair of 1-methyladenine, 3-alkylcytosines, 1-methylguanine, and 3-methylthymine in alkB Escherichia coli **101**, 14051-14056, (2004).
- 6 Nikolova, E. N. *et al. Nature* Transient Hoogsteen base pairs in canonical duplex DNA **470**, 498-502, (2011).
- 7 Ha, S. C. *et al. Proc Natl Acad Sci USA* A poxvirus protein forms a complex with left-handed Z-DNA: crystal structure of a Yatapoxvirus Zalpha bound to DNA **101**, 14367-14372, (2004).
- 8 Herbert, A. & Rich, A. *Genetica* Left-handed Z-DNA: structure and function **106**, 37-47, (1999).
- 9 Egli, M., Williams, L. D., Gao, Q. & Rich, A. *Biochemistry* Structure of the pure-spermine form of Z-DNA (magnesium free) at 1-A resolution **30**, 11388-11402, (1991).
- 10 Coll, M., Wang, A. H., van der Marel, G. A., van Boom, J. H. & Rich, A. *J Biomol Struct Dyn* Crystal structure of a Z-DNA fragment containing thymine/2-aminoadenine base pairs **4**, 157-172, (1986).
- 11 Ho, P. S. *et al. EMBO J* G.T wobble base-pairing in Z-DNA at 1.0 A atomic resolution: the crystal structure of d(CGCGTG) **4**, 3617-3623, (1985).
- 12 Fujii, S., Wang, A. H., van der Marel, G., van Boom, J. H. & Rich, A. *Nucleic Acids Res* Molecular structure of (m<sup>5</sup> dC-dG)<sub>3</sub>: the role of the methyl group on 5-methyl cytosine in stabilizing Z-DNA **10**, 7879-7892, (1982).
- 13 Nikolova, E. N. *et al. Biopolymers* A historical account of hoogsteen base-pairs in duplex DNA **99**, 955-968, (2013).

- 14 Wijmenga, S. S. & van Buuren, N. M. *Prog NMR Spectrosc* The use of NMR  
methods for conformational studies of nucleic acids **32**, 287-387, (1998).
- 15 Al-Hashimi, H. M. *J Magn Reson* NMR studies of nucleic acid dynamics **237**,  
191-204, (2013).
- 16 Bothe, J. R., Lowenhaupt, K. & Al-Hashimi, H. M. *J Am Chem Soc* Sequence-  
specific B-DNA flexibility modulates Z-DNA formation **133**, 2016-2018, (2011).
- 17 Lee, Y. M. *et al. Biophys Chem* NMR investigation on the DNA binding and B-Z  
transition pathway of the Zalpha domain of human ADAR1 **172**, 18-25, (2013).
- 18 Kim, H. E. *et al. FEBS Lett* The Zbeta domain of human DAI binds to Z-DNA via  
a novel B-Z transition pathway **585**, 772-778, (2011).
- 19 Seo, Y. J. *et al. FEBS Lett* Sequence discrimination of the Zalpha domain of  
human ADAR1 during B-Z transition of DNA duplexes **584**, 4344-4350, (2010).
- 20 Lee, E. H. *et al. FEBS Lett* NMR study of hydrogen exchange during the B-Z  
transition of a DNA duplex induced by the Zalpha domains of yatapoxvirus E3L  
**584**, 4453-4457, (2010).
- 21 Kang, Y. M. *et al. J Am Chem Soc* NMR spectroscopic elucidation of the B-Z  
transition of a DNA double helix induced by the Z alpha domain of human  
ADAR1 **131**, 11485-11491, (2009).
- 22 Yang, X. L. & Wang, A. H. *Biochemistry* Structural analysis of Z-Z DNA  
junctions with A:A and T:T mismatched base pairs by NMR **36**, 4258-4267,  
(1997).
- 23 Sugiyama, H. *et al. Nucleic Acids Res* Synthesis, structure and thermodynamic  
properties of 8-methylguanine-containing oligonucleotides: Z-DNA under  
physiological salt conditions **24**, 1272-1278, (1996).
- 24 Bendel, P. *Biochem Biophys Res Commun* Measurement of the helix opening rate  
in Z-DNA by <sup>1</sup>H nuclear magnetic resonance relaxation spectroscopy **128**, 352-  
359, (1985).
- 25 Feigon, J., Wang, A. H., van der Marel, G. A., Van Boom, J. H. & Rich, A.  
*Nucleic Acids Res* A one- and two-dimensional NMR study of the B to Z  
transition of (m5dC-dG)<sub>3</sub> in methanolic solution **12**, 1243-1263, (1984).
- 26 Patel, D. J., Kozlowski, S. A., Nordheim, A. & Rich, A. *Proc Natl Acad Sci USA*  
Right-handed and left-handed DNA: studies of B- and Z-DNA by using proton  
nuclear Overhauser effect and P NMR **79**, 1413-1417, (1982).
- 27 Cohen, J. S., Wooten, J. B. & Chatterjee, C. L. *Biochemistry* Characterization of  
alternating deoxyribonucleic acid conformations in solution by phosphorus-31  
nuclear magnetic resonance spectroscopy **20**, 3049-3055, (1981).
- 28 Popena, M., Milecki, J. & Adamiak, R. W. *Nucleic Acids Res* High salt solution  
structure of a left-handed RNA double helix **32**, 4044-4054, (2004).
- 29 Voehler, M. W., Collier, G., Young, J. K., Stone, M. P. & Germann, M. W. *J Mag  
Res* Performance of cryogenic probes as a function of ionic strength and sample  
tube geometry **183**, 102-109, (2006).
- 30 Binbuga, B., Boroujerdi, A. F. B. & Young, J. K. *Protein Science* Structure in an  
extreme environment: NMR at high salt **16**, 1783-1787, (2007).
- 31 Kim, D. *et al. Nucleic Acids Res* Base extrusion is found at helical junctions  
between right- and left-handed forms of DNA and RNA **37**, 4353-4359, (2009).

- 32 Schwartz, T., Rould, M. A., Lowenhaupt, K., Herbert, A. & Rich, A. *Science* Crystal Structure of the Za Domain of the Human Editing Enzyme ADAR1 Bound to Left-Handed Z-DNA **284**, 1841-1845, (1999).
- 33 Zimmer, D. P. & Crothers, D. M. *Proc Nat Acad Sci USA* NMR of enzymatically synthesized uniformly <sup>13</sup>C<sup>15</sup>N-labeled DNA oligonucleotides **92**, 3091-3095, (1995).
- 34 Ho, P. S., Ellison, M. J., Quigley, G. J. & Rich, A. *EMBO J* A computer aided thermodynamic approach for predicting the formation of Z-DNA in naturally occurring sequences **5**, 2737-2744, (1986).
- 35 Seela, F. & Peng, X. *Curr Protoc Nucleic Acid Chem* Synthesis and properties of 7-substituted 7-deazapurine (pyrrolo[2,3-d]pyrimidine) 2'-deoxyribonucleosides **Chapter 1**, Unit 1.10, (2005).
- 36 Seela, F., Budow, S., Shaikh, K. I. & Jawalekar, A. M. *Org Biomol Chem* Stabilization of tandem dG-dA base pairs in DNA-hairpins: replacement of the canonical bases by 7-deaza-7-propynylpurines **3**, 4221-4226, (2005).
- 37 Hua, N., van der Marel, G. A., van Boom, J. H. & Feigon, J. *Nucleic Acids Res* Non-contiguous regions of Z-DNA in a DNA dodecamer **17**, 7923-7944, (1989).
- 38 Feigon, J., Wang, A. H., van der Marel, G. A., van Boom, J. H. & Rich, A. *Science* Z-DNA forms without an alternating purine-pyrimidine sequence in solution **230**, 82-84, (1985).
- 39 Chen, F. M. *Biochemistry* Base protonation facilitates B-Z interconversions of poly(dG-dC) X poly(dG-dC) **23**, 6159-6165, (1984).
- 40 Herbert, A. *et al. Proc Natl Acad Sci USA* A Z-DNA binding domain present in the human editing enzyme, double-stranded RNA adenosine deaminase **94**, 8421-8426, (1997).
- 41 Vorlickova, M. & Sagi, J. *Nucleic Acids Res* Transitions of poly(dI-dC), poly(dI-methyl5dC) and poly(dI-bromo5dC) among and within the B-, Z-, A- and X-DNA families of conformations **19**, 2343-2347, (1991).
- 42 Mandal, P. K. & Majumdar, A. *Concepts in Magnetic Resonance Part A* A comprehensive discussion of HSQC and HMQC pulse sequences **20A**, 1-23, (2004).
- 43 Alvey, H. S., Gottardo, F. L., Nikolova, E. N. & Al-Hashimi, H. M. *Nat Struct Mol Biol* Transient Hoogsteen Base-Pairs Occur Robustly in Duplex DNA **In Submission**, (2014).
- 44 Abrescia, N. G., Gonzalez, C., Gouyette, C. & Subirana, J. A. *Biochemistry* X-ray and NMR studies of the DNA oligomer d(ATATAT): Hoogsteen base pairing in duplex DNA **43**, 4092-4100, (2004).
- 45 Abrescia, N. G., Thompson, A., Huynh-Dinh, T. & Subirana, J. A. *Proc Natl Acad Sci USA* Crystal structure of an antiparallel DNA fragment with Hoogsteen base pairing **99**, 2806-2811, (2002).
- 46 Lee, Y. M. *et al. J Am Chem Soc* NMR study on the B-Z junction formation of DNA duplexes induced by Z-DNA binding domain of human ADAR1 **134**, 5276-5283, (2012).
- 47 Lee, A. R. *et al. Biochem Biophys Res Commun* NMR dynamics study of the Z-DNA binding domain of human ADAR1 bound to various DNA duplexes **428**, 137-141, (2012).

- 48 Bothe, J. R., Lowenhaupt, K. & Al-Hashimi, H. M. *Biochemistry* Incorporation of CC steps into Z-DNA: interplay between B-Z junction and Z-DNA helical formation **51**, 6871-6879, (2012).
- 49 Xu, X. & Au-Yeung, S. *J Phys Chem B* Investigation of chemical shift and structure relationships in nucleic acids using NMR and density functional theory methods **104**, 5641-5650, (2000).

## Chapter 6

### Conclusions and Future Perspectives

#### 6.1 Transient Watson-Crick to Hoogsteen Transitions in Canonical Duplex DNA

The DNA double helix must maintain structure plasticity to be packaged efficiently into a cell nucleus while transcriptionally active regions remain unpacked and even un-paired during gene expression. DNA must simultaneously maintain structural integrity to protect the genetic code from damage and serve as a substrate for many DNA binding proteins. The sequence-specific recognition of DNA is, in part, a conundrum since there are only four different types of nucleotide building blocks to comprise the 3 billion base-pairs in the human genome. Alternative structures formed by DNA are thought to help increase specificity for particular sequences. Our lab recently showed that DNA WC base-pairs transiently morph into HG base-pairs at flexible CA/TG and TA/TA steps on the microsecond to millisecond timescale.<sup>1,2</sup> Chapter 2 of this thesis unequivocally demonstrates the robust occurrence of transient excursion to the HG base-pair form across a wide variety of sequences with an ~30-fold and 20-fold range in populations and lifetimes, respectively. Remarkably, every base-pair probed exhibited a transient state consistent with the HG base-pair. This work implies an appreciable presence of HG base-pairs in the context of duplex DNA and provides a broader implication for such cases within the genome as well.



## 6.2 The B-DNA to Z-DNA Transition and Consequences of Sequence Variations

Another such structural transition owing to DNA's flexibility is its inversion of handedness, discovered accidentally.<sup>3</sup> Studies of this B-to-Z-DNA transition have mainly elucidated where Z-DNA occurs and what sequences can be accommodated within the structural framework. Despite Z-DNA's transient nature *in vivo*, little emphasis has been placed on the kinetic nature of how Z-DNA forms. In Chapter 3 we established the B-to-Z-DNA transition is biphasic for CpG repeats of 3 and 6 dinucleotide repeats, likely involving an intermediate species that resembles a left-handed structure.

Z-DNA formation was postulated as a gene expression regulatory mechanism in 1982.<sup>4</sup> Recently, overexpression of the cancer metastasis-promoting ADAM-12 was shown to likely be due to the loss in negative regulation by an upstream epigenetic Z-DNA silencer.<sup>5</sup> In Chapter 4 we present a kinetic and molecular model to describe our observations of B-Z transition perturbations upon introduction of the epigenetic modification <sup>5me</sup>C on transient Z-DNA formation and observe perturbations in the observed phases. Together Chapters 3 and 4 present a comprehensive conceptual framework describing perturbations to B-to-Z-DNA transition observables. In an attempt to test the hypothesis that the guanine *anti-to-syn* transition manifests in the observed transient kinetics, we employed <sup>N1me</sup>G modifications and learned N1 methyl damage at two guanines in CG<sub>3</sub> sufficiently destabilized the duplex such that single-stranded signatures were observed by CD and NMR. This was not the case for such modifications within A<sub>6</sub> despite the lower melting temperature of A<sub>6</sub> as compared to CG<sub>3</sub>, suggesting a more complex mechanism owing to duplex stability.

High salt conditions have largely impeded atomic resolution dynamics studies of Z-DNA using modern NMR experiments. The discovery of Z-DNA binding proteins has aided in such efforts but the crux of Z-DNA being a highly repetitive sequence element will preclude atomic resolution dynamics of ~10 base-pairs or more, even at high magnetic fields. A combinatorial approach of benchmarking (<sup>5me</sup>CG)<sub>3</sub> Z-DNA proton assignments with those previously reported<sup>6</sup> was used, followed by Z-DNA aromatic carbon chemical shift determination. Z-DNA guanine carbon chemical shifts (C8) are consistent with those observed for the *syn* transient HG state while those of cytosine (C6) shift upfield, potentially due to the decrease in electron density around the cytosine base due to the loss of stacking with its 3' neighbor. Finally, the titration of Z $\alpha$  into a B/Z junction forming sequence shows the same evidence for Z-DNA *syn* guanines and potentially Z-form cytosines. The signal-to-noise of cytosine C6 is lower as compared to guanine C8 in the Z-form. These NMR experiments of Z-DNA present the first visualization of Z-DNA aromatic carbon chemical shifts and call for further investigation of structure and dynamics of Z-DNA using NMR. Two such options for high molecular weight samples are site-specific labeling strategies to employ 1D NMR experiments, or the methyl TROSY experiment whose power was recently demonstrated on nucleosomes.<sup>7</sup> Overall, in this thesis I have highlighted the complexities of DNA dynamics and flexibility, with emphasis on transient excursions away from WC base-paired B-DNA to HG base-paired B-DNA and WC base-paired Z-DNA.

### 6.3 Future Perspective

The central dogma of biology explains the flow of genetic information, whereby DNA is the template for RNA synthesis, RNA is the template for protein synthesis, and proteins are in turn used to make DNA. In this cyclical process the exciting and often surprising sequence-specific structural and dynamic features of DNA are quickly overlooked. The pinnacle of this thesis is to highlight the dynamic nature of naked duplex DNA. The sequence and positional specificity of HG base-pair formation is complex and calls for characterization of all 64 trinucleotide sequences. The work presented herein also calls for the understanding of Z-DNA's impact on transient HG base-pair formation, since together they may have a cooperative affect on *in vivo* signaling. Distance from helix termini and B/Z junctions affects the structure and dynamics of canonical B-DNA, HG base-pairs and Z-DNA. While the discovery of Z-DNA in 1979 and transiently forming HG base-pairs in duplex DNA in 2011 were both met with skepticism for their seemingly small impact on biology, this thesis lays the ground work for their increased presence in nature.

### 6.4 References

- 1 Nikolova, E. N. *et al. Nature* Transient Hoogsteen base pairs in canonical duplex DNA **470**, 498-502, (2011).
- 2 Nikolova, E. N., Gottardo, F. L. & Al-Hashimi, H. M. *J Am Chem Soc* Probing Transient Hoogsteen Hydrogen Bonds in Canonical Duplex DNA Using NMR Relaxation Dispersion and Single-Atom Substitution **134**, 3667-3670, (2012).
- 3 Wang, A. H.-J. *et al. Nature* Molecular structure of a left-handed double helical DNA fragment at atomic resolution **282**, 680-686, (1979).
- 4 Check, W. A. *JAMA* Z-DNA: a new twist in gene regulation? **247**, 3175-3176, (1982).

- 5 Ray, B. K., Dhar, S., Henry, C., Rich, A. & Ray, A. *Cancer Res* Epigenetic regulation by Z-DNA silencer function controls cancer-associated ADAM-12 expression in breast cancer: cross-talk between MeCP2 and NF1 transcription factor family **73**, 736-744, (2013).
- 6 Feigon, J., Wang, A. H., van der Marel, G. A., Van Boom, J. H. & Rich, A. *Nucleic Acids Res* A one- and two-dimensional NMR study of the B to Z transition of (m5dC-dG)<sub>3</sub> in methanolic solution **12**, 1243-1263, (1984).
- 7 Kato, H. *et al. Proc Natl Acad Sci USA* Architecture of the high mobility group nucleosomal protein 2-nucleosome complex as revealed by methyl-based NMR **108**, 12283-12288, (2011).

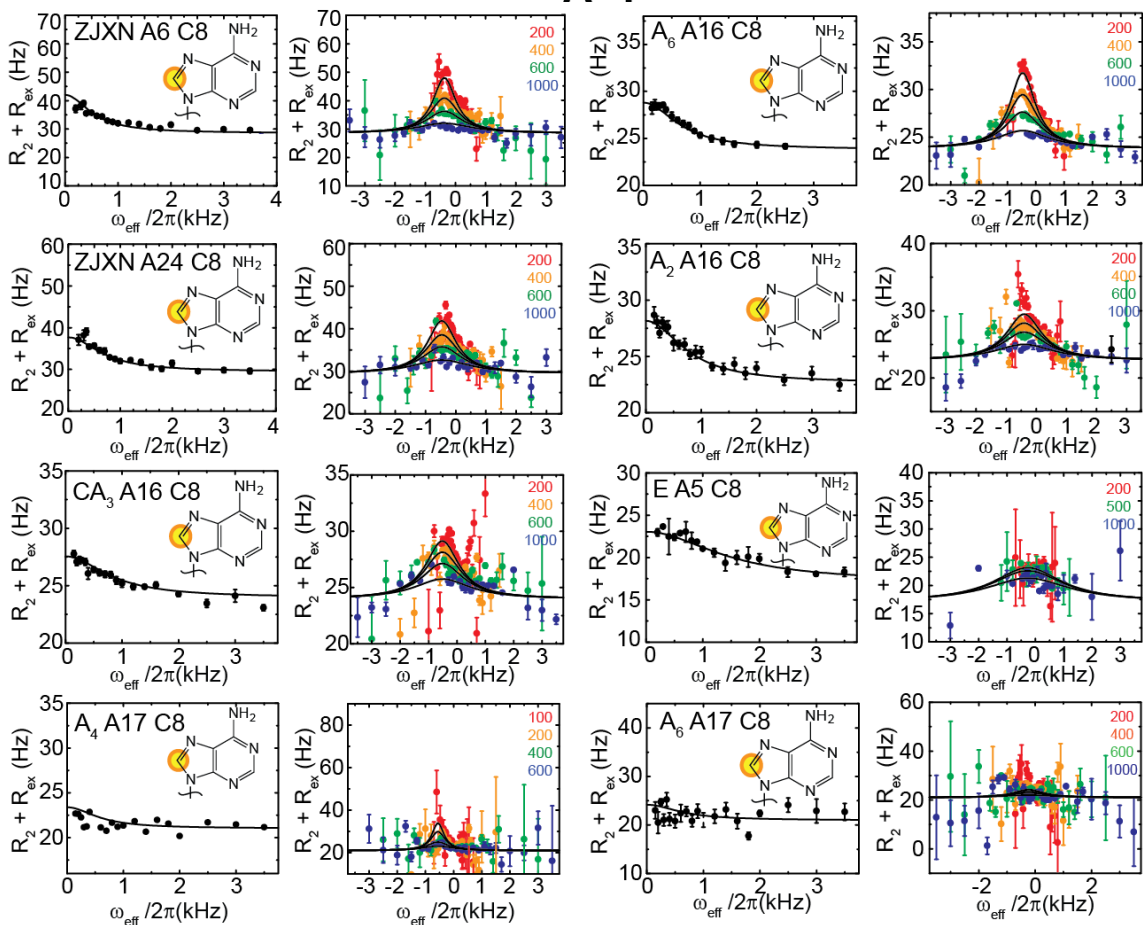
## Appendix 1

### $R_{1\rho}$ Relaxation Dispersion Experimental Parameters

**Table A1.1 |  $R_{1\rho}$  RD experimental parameters. Probes measured with variable spin lock power and offset at 14.1 T.**

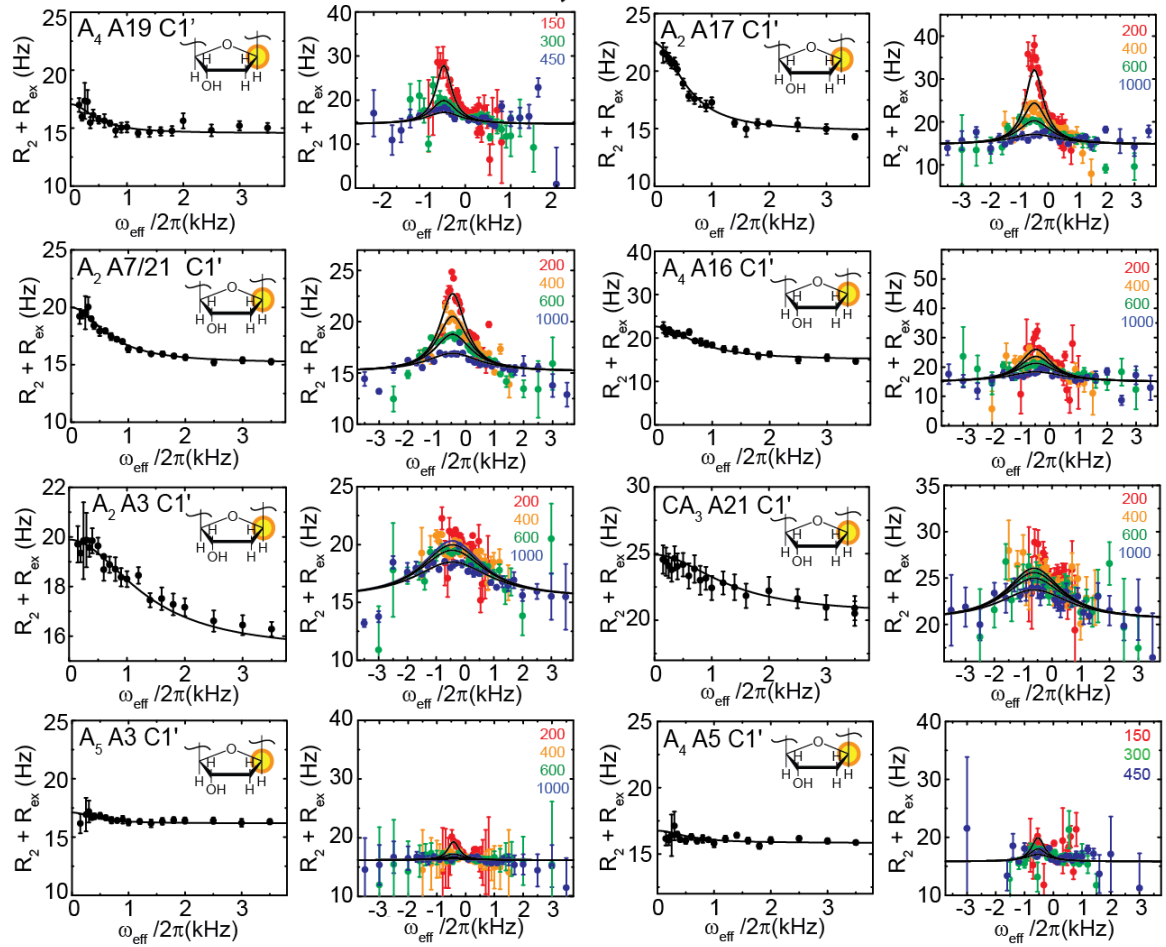
Probe
A <sub>2</sub> C2 C6, A <sub>2</sub> A3 C1', A <sub>2</sub> T8 N3, A <sub>2</sub> T9 C6, A <sub>2</sub> G10 C1', A <sub>2</sub> A16 C8, A <sub>2</sub> A17 C1', A <sub>2</sub> A17 C2, A <sub>2</sub> A7/21 C1', A <sub>4</sub> A16 C1', A <sub>6</sub> A16 C8, A <sub>6</sub> A17 C8, A <sub>6</sub> G10 N1, CA <sub>3</sub> A16 C8, CA <sub>3</sub> A21 C1', ZJXN A6 C8, ZJXN A24 C8
<b>On-resonance spin lock power (<math>\omega</math>) / Off-resonance spin lock power (<math>\omega</math>) &amp; <math>\pm</math> {offset (<math>\Omega</math>)}</b>
150, 200, 250, 300, 350, 400, 500, 600, 700, 800, 900, 1000, 1200, 1400, 1600, 1800, 2000, 2500, 3000, 3500 Hz/ 200 Hz & $\pm$ {30, 60, 90, 120, 150, 180, 210, 240, 270, 300, 360, 420, 480, 540, 600, 700, 800, 330 Hz} 400 Hz & $\pm$ {50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, 1000, 1200, 1500, 550 Hz} 600 Hz & $\pm$ {75, 150, 225, 300, 375, 450, 550, 650, 800, 1000, 1200, 1400, 1600, 2000, 2500, 3000, 3500, 4000, 5000 Hz} 1000 Hz & $\pm$ {100, 200, 300, 400, 500, 600, 700, 900, 1100, 1300, 1500, 1700, 2000, 2500, 3000, 3500, 4000, 5000, 7500 Hz}
A <sub>2</sub> G11 C8, A <sub>4</sub> G11 C8, A <sub>4</sub> C15 C6, A <sub>4</sub> A17 C8, A <sub>5</sub> C9 C6, A <sub>5</sub> G10 C8, A <sub>5</sub> G11 C8, A <sub>6</sub> G10 C8, CA <sub>3</sub> C17 C6, CA <sub>3</sub> C19 C6
150, 200, 250, 300, 350, 400, 500, 600, 700, 800, 900, 1000, 1200, 1400, 1600, 1800, 2000, 2500, 3000, 3500 Hz/ 100 Hz & $\pm$ {30, 60, 90, 120, 150, 180, 210, 240, 270, 300, 360, 420, 480, 540, 600, 700, 800, 330 Hz} 200 Hz & $\pm$ {50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, 1000, 1200, 1500, 550 Hz} 400 Hz & $\pm$ {75, 150, 225, 300, 375, 450, 550, 650, 800, 1000, 1200, 1400, 1600, 2000, 2500, 3000, 3500, 4000, 5000 Hz} 600 Hz & $\pm$ {100, 200, 300, 400, 500, 600, 700, 900, 1100, 1300, 1500, 1700, 2000, 2500, 3000, 3500, 4000, 5000, 7500 Hz}
A <sub>4</sub> A19 C1', A <sub>4</sub> C5 C1'
150, 200, 250, 300, 350, 400, 500, 600, 700, 800, 900, 1000, 1200, 1400, 1600, 1800, 2000, 2500, 3000, 3500 Hz/ 150 Hz & $\pm$ {30, 60, 90, 120, 150, 180, 210, 240, 270, 300, 360, 420, 480, 540, 600, 700, 800, 330 Hz} 300 Hz & $\pm$ {50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, 1000, 1200, 1500, 550 Hz} 450 Hz & $\pm$ {75, 150, 225, 300, 375, 450, 550, 650, 800, 1000, 1200, 1400, 1600, 2000, 2500, 3000, 3500, 4000, 5000 Hz}
A <sub>4</sub> G10 C1', CA <sub>3</sub> G10 C1'
150, 200, 250, 300, 350, 400, 500, 600, 700, 800, 900, 1000, 1200, 1400, 1600, 1800, 2000, 2500, 3000, 3500 Hz/ 150 Hz & $\pm$ {30, 60, 90, 120, 150, 180, 210, 240, 270, 300, 360, 420, 480, 540, 600, 700, 800, 330 Hz} 250 Hz & $\pm$ {50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, 1000, 1200, 1500, 550 Hz} 400 Hz & $\pm$ {75, 150, 225, 300, 375, 450, 550, 650, 800, 1000, 1200, 1400, 1600, 2000, 2500, 3000, 3500, 4000, 5000 Hz} 700 Hz & $\pm$ {100, 200, 300, 400, 500, 600, 700, 900, 1100, 1300, 1500, 1700, 2000, 2500, 3000, 3500, 4000, 5000, 7500 Hz}
CG <sub>3</sub> G4 C8
150, 200, 250, 300, 350, 400, 500, 600, 700, 800, 900, 1000, 1200, 1400, 1600, 1800, 2000, 2500, 3000, 3500 Hz/ 200 Hz & $\pm$ {30, 60, 90, 120, 150, 180, 210, 240, 270, 300, 360, 420, 480, 540, 600, 700, 800, 330 Hz} 400 Hz & $\pm$ {50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, 1000, 1200, 1500, 550 Hz} 800 Hz & $\pm$ {75, 150, 225, 300, 375, 450, 550, 650, 800, 1000, 1200, 1400, 1600, 2000, 2500, 3000, 3500, 4000, 5000 Hz}
A <sub>2</sub> G23 N1, A <sub>5</sub> T4 N3*, A <sub>5</sub> T5 N3, A <sub>5</sub> T6 N3, A <sub>5</sub> T7 N3, A <sub>5</sub> T8 N3, A <sub>6</sub> G10 N1
100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 700, 800, 900, 1000, 1200, 1400, 1600, 1800, 2000 Hz/ 200 Hz & $\pm$ {20, 40, 50, 60, 70, 80, 100, 120, 140, 150, 170, 180, 210, 220, 250, 300, 350, 400, 500, 600, 700, 800, 1000 Hz} 400 Hz & $\pm$ {30, 60, 90, 110, 120, 140, 150, 170, 180, 210, 220, 250, 300, 350, 400, 500, 600, 700, 800, 1000 Hz} *500 Hz & $\pm$ {30, 60, 90, 120, 150, 180, 210, 250, 300, 350, 400, 500, 600, 700, 800, 1000 Hz} 600 Hz & $\pm$ {40, 80, 120, 150, 160, 200, 240, 250, 300, 350, 400, 450, 500, 600, 800 Hz} *800 Hz & $\pm$ {40, 80, 120, 160, 200, 240, 300, 400, 500, 600, 800, 1000, 1200, 1400, 1600, 2000} 1000 Hz & $\pm$ {100, 200, 300, 400, 500, 600, 700, 900, 1100, 1300, 1500, 1700, 2000, 2500, 3000, 3500 Hz}
A <sub>2</sub> G10 N1
100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 700, 800, 900, 1000, 1200, 1400, 1600, 1800, 2000 Hz/ 200 Hz & $\pm$ {20, 40, 50, 60, 70, 80, 100, 120, 140, 150, 170, 180, 210, 220, 250, 300, 350, 400, 500 Hz} 400 Hz & $\pm$ {30, 60, 90, 110, 120, 140, 150, 170, 180, 210, 220, 250, 300, 350, 400, 500, 600, 700, 800, 1000 Hz} 600 Hz & $\pm$ {40, 80, 120, 150, 160, 200, 240, 250, 300, 350, 400, 450, 500, 600, 800, 1000, 1200, 1400, 1600, 2000 Hz}
E A5 C8
200, 300, 400, 500, 600, 700, 800, 900, 1000, 1200, 1400, 1600, 1800, 2000, 2500, 3000, 3500 Hz 200 Hz & $\pm$ {30, 60, 90, 120, 180, 240, 300, 420, 480, 540, 600, 700 Hz} 500 Hz & $\pm$ {50, 100, 150, 200, 250, 300, 400, 500, 600, 800, 1000, 1200 Hz} 1000 Hz & $\pm$ {75, 150, 225, 300, 375, 450, 600, 800, 1000, 1400, 1600, 200, 2500, 3000 Hz}

# A•T



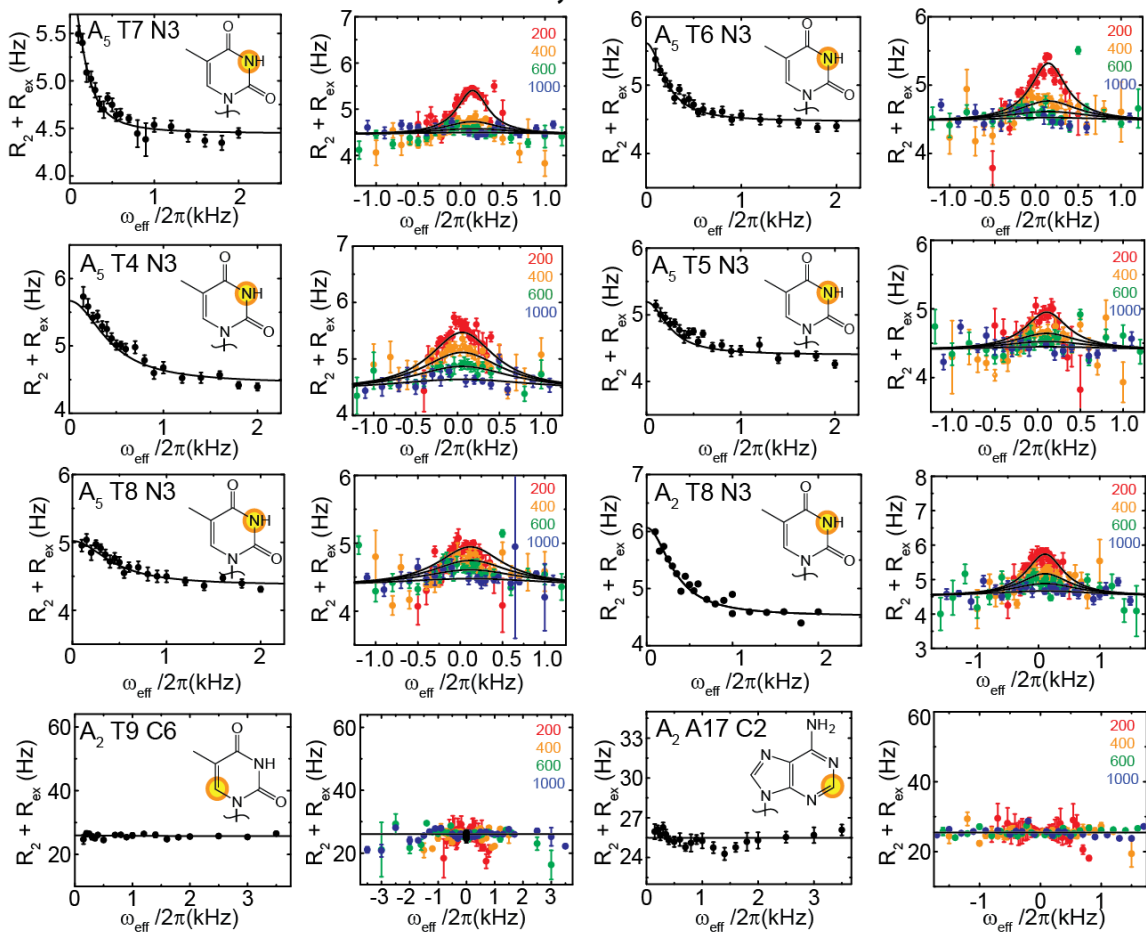
cont'd

# A•T, cont'd



cont'd

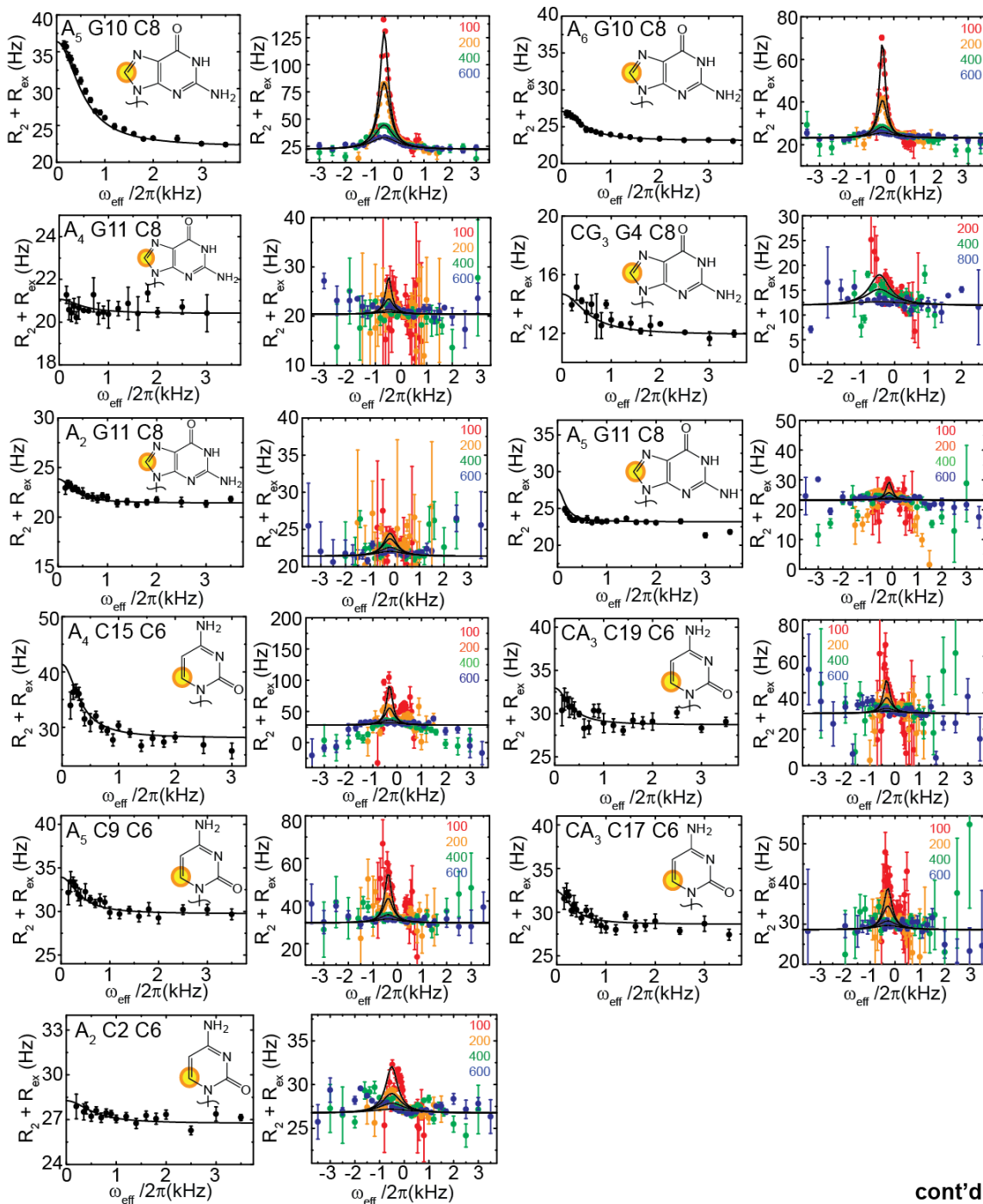
# A•T, cont'd



cont'd

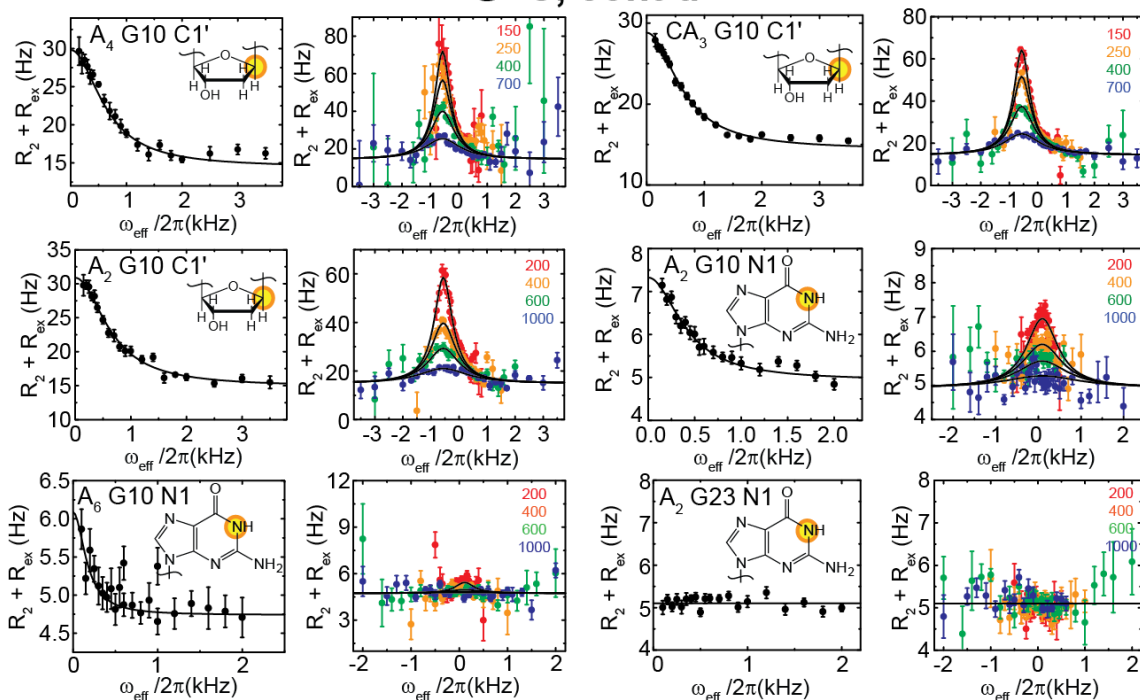


# G•C



cont'd

## G•C, cont'd



**Figure A1.1 | Characterizing transient Hoogsteen base-pairs using on- and off-resonance  $^{13}\text{C}$  and  $^{15}\text{N}$   $R_{1\rho}$  relaxation dispersion.** Shown are representative examples of relaxation dispersion profiles for different sequence and positional contexts. The various spin lock powers used (Hz) are indicated in the insets of the off-resonance relaxation dispersion profiles. Data are fitted to Eq. 1. Error bars represent experimental uncertainty (one standard deviation) estimated from monoexponential fitting of duplicate  $R_{1\rho}$  data. Buffer conditions are 15 mM Sodium phosphate, 25 mM NaCl, 0.1 mM EDTA, 10 %  $\text{D}_2\text{O}$  pH 5.2 ( $A_6$  C15 C6,  $A_5$ ,  $A_4$ : C15 C6, A16 C1', G10 C1'), 5.4 ( $A_2$ ,  $A_4$ ,  $A_6$ ,  $CA_3$ , CG<sub>3</sub>), 6.8 (E) or 7.5 (ZJXN), 26 °C ( $A_2$ ,  $A_4$ ,  $A_5$ ,  $A_6$ ,  $CA_3$ , E) or 25 °C ( $A_2$  A3 C1', CG<sub>3</sub>, ZJXN).

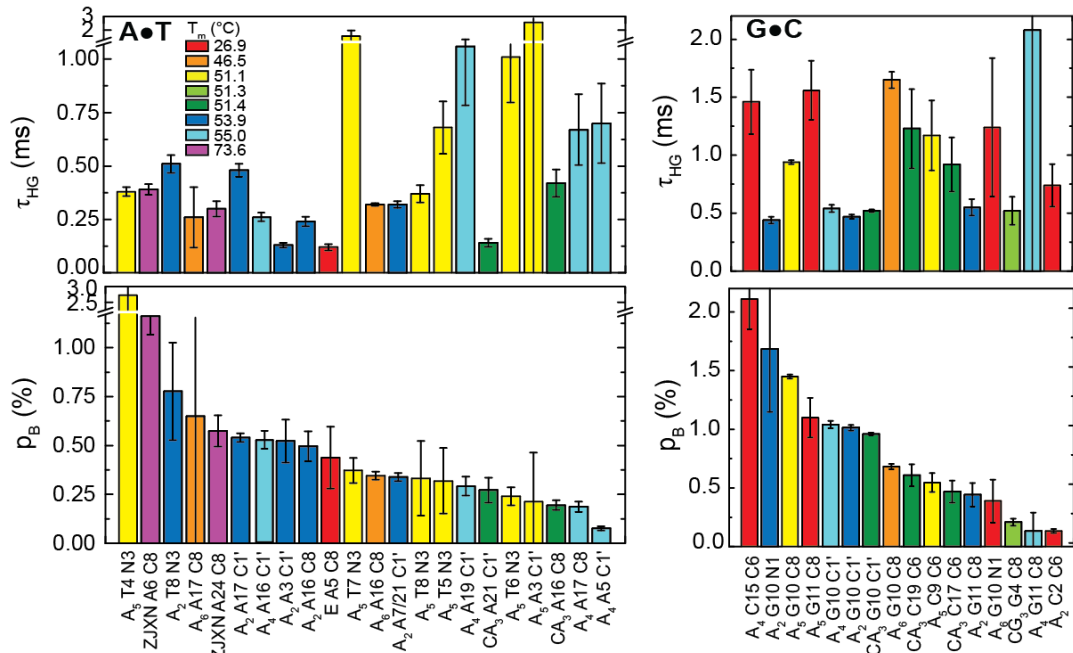


Figure A1.2 | Melting Temperature Compared to Population and Lifetime of Hoogsteen Transient State.

## Appendix 2

### CG<sub>3</sub> B-to-Z-DNA Controls and Transition Fits to Sum of Exponentials

‘Mono’ and ‘Bi’ fits were performed with Eq.’s 3.5 and 3.6, respectively. Experimental conditions for data from which the following fit parameters were obtained are as follows: 15 mM Phosphate buffer pH 7.5, 0.1 mM EDTA, 25 °C, NaCl concentration specified in each table for NaCl-induced transitions. 25 mM NaCl for Z $\alpha$ -induced transition.

**Table A2.1 | CG<sub>3</sub> + 2.5 M NaCl B-to-Z-DNA Transition Fitting Values**

Equation	Parameter, $p$	$\langle p \rangle$	$\sigma$	$\langle \chi^2 \rangle$	$\langle R^2 \rangle$	n
Mono	$k_{\text{obs}}$ (s <sup>-1</sup> )	0.0241	0.0005			
	A (mdeg)	11.87	0.28	0.44	0.91	
	$y_{\infty}$ (mdeg)	-13.0	0.2			
Bi	$k_{\text{obs},1}$ (s <sup>-1</sup> )	0.0129	0.0037			4
	$k_{\text{obs},2}$ (s <sup>-1</sup> )	0.0743	0.0283			
	A <sub>1</sub> (mdeg)	6.03	1.84	0.37	0.92	
	A <sub>2</sub> (mdeg)	11.4	1.2			
	$y_{\infty}$ (mdeg)	-12.7	0.2			

**Table A2.2 | CG<sub>3</sub> + 3.0 M NaCl B-to-Z-DNA Transition Fitting Values**

Equation	Parameter, $p$	$\langle p \rangle$	$\sigma$	$\langle \chi^2 \rangle$	$\langle R^2 \rangle$	n
Mono	$k_{\text{obs}}$ (s <sup>-1</sup> )	0.0318	0.0014			
	A (mdeg)	15.2	0.7	0.49	0.92	
	$y_{\infty}$ (mdeg)	-6.70	1.12			
Bi	$k_{\text{obs},1}$ (s <sup>-1</sup> )	0.0128	0.0009			4
	$k_{\text{obs},2}$ (s <sup>-1</sup> )	0.0720	0.0032			
	A <sub>1</sub> (mdeg)	5.62	1.08	0.32	0.95	
	A <sub>2</sub> (mdeg)	14.9	1.0			
	$y_{\infty}$ (mdeg)	-6.87	0.13			

**Table A2.4 | CG<sub>3</sub> + 4.0 M NaCl B-to-Z-DNA Transition Fitting Values**

Equation	Parameter, $p$	$\langle p \rangle$	$\sigma$	$\langle \chi^2 \rangle$	$\langle R^2 \rangle$	n
Mono	$k_{\text{obs}}$ (s <sup>-1</sup> )	0.0539	0.0002			
	A (mdeg)	25.5	0.2	0.41	0.96	
	$y_{\infty}$ (mdeg)	1.00	0.001			
Bi	$k_{\text{obs},1}$ (s <sup>-1</sup> )	0.0177	0.0016			3
	$k_{\text{obs},2}$ (s <sup>-1</sup> )	0.0810	0.0068			
	A <sub>1</sub> (mdeg)	5.76	1.33	0.34	0.97	
	A <sub>2</sub> (mdeg)	26.6	0.8			
	$y_{\infty}$ (mdeg)	1.91	0.32			

**Table A2.5 | CG<sub>3</sub> + 4.5 M NaCl B-to-Z-DNA Transition Fitting Values**

Equation	Parameter, $p$	$\langle p \rangle$	$\sigma$	$\langle \chi^2 \rangle$	$\langle R^2 \rangle$	n
Mono	$k_{\text{obs}}$ (s <sup>-1</sup> )	0.0793	0.0003			
	A (mdeg)	30.6	0.1	0.36	0.96	
	$y_{\infty}$ (mdeg)	1.50	0.002			
Bi	$k_{\text{obs},1}$ (s <sup>-1</sup> )	0.0239	0.0010			4
	$k_{\text{obs},2}$ (s <sup>-1</sup> )	0.113	0.007			
	A <sub>1</sub> (mdeg)	4.37	0.54	0.32	0.96	
	A <sub>2</sub> (mdeg)	34.4	1.3			
	$y_{\infty}$ (mdeg)	1.97	0.33			

**Table A2.6 | CG<sub>3</sub> + 5.0 M NaCl B-to-Z-DNA Transition Fitting Values**

Equation	Parameter, $p$	$\langle p \rangle$	$\sigma$	$\langle \chi^2 \rangle$	$\langle R^2 \rangle$	n
Mono	$k_{\text{obs}} (\text{s}^{-1})$	0.0684	0.0067			
	A (mdeg)	19.1	3.3	0.40	0.77	
	$y_{\infty}$ (mdeg)	5.96	0.37			
Bi	$k_{\text{obs},1} (\text{s}^{-1})$	0.0139	0.0039			3
	$k_{\text{obs},2} (\text{s}^{-1})$	0.120	0.017			
	A <sub>1</sub> (mdeg)	2.16	0.71	0.35	0.80	
	A <sub>2</sub> (mdeg)	32.6	3.6			
	$y_{\infty}$ (mdeg)	6.16	0.01			

**Table A2.7 | 5.5  $\mu\text{M}$  (0.5 X) CG<sub>3</sub> + 3.0 M NaCl B-to-Z-DNA Transition Fitting Values**

Equation	Parameter, $p$	$\langle p \rangle$	$\sigma$	$\langle \chi^2 \rangle$	$\langle R^2 \rangle$	n
Mono	$k_{\text{obs}} (\text{s}^{-1})$	0.029	0.004			
	A (mdeg)	8.32	0.43	0.17	0.91	
	$y_{\infty}$ (mdeg)	-3.90	0.19			
Bi	$k_{\text{obs},1} (\text{s}^{-1})$	0.015	0.0004			3
	$k_{\text{obs},2} (\text{s}^{-1})$	0.069	0.005			
	A <sub>1</sub> (mdeg)	3.71	0.90	0.15	0.93	
	A <sub>2</sub> (mdeg)	7.14	1.21			
	$y_{\infty}$ (mdeg)	-3.76	0.20			

**Table A2.8 | 22  $\mu\text{M}$  (2 X) CG<sub>3</sub> + 3.0 M NaCl B-to-Z-DNA Transition Fitting Values**

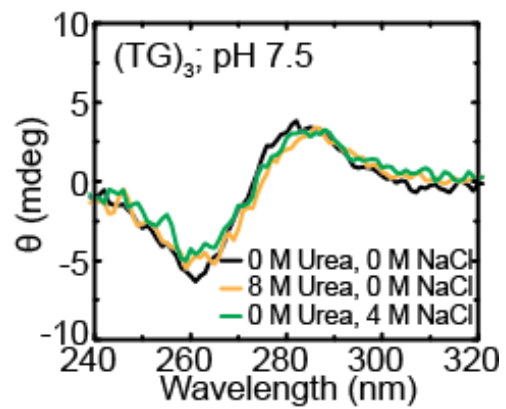
Equation	Parameter, $p$	$\langle p \rangle$	$\sigma$	$\langle \chi^2 \rangle$	$\langle R^2 \rangle$	n
Mono	$k_{\text{obs}} (\text{s}^{-1})$	0.0232	0.0043			
	A (mdeg)	29.6	3.4	4.73	0.84	
	$y_{\infty}$ (mdeg)	-14.3	0.6			
Bi	$k_{\text{obs},1} (\text{s}^{-1})$	0.0134	0.0017			3
	$k_{\text{obs},2} (\text{s}^{-1})$	0.0927	0.0547			
	A <sub>1</sub> (mdeg)	16.1	5.0	4.41	0.85	
	A <sub>2</sub> (mdeg)	33.1	2.8			
	$y_{\infty}$ (mdeg)	-13.7	0.6			

**Table A2.9 | CG<sub>6</sub> + 4.0 M NaCl B-to-Z-DNA Transition Fitting Values**

Equation	Parameter, $p$	$\langle p \rangle$	$\sigma$	$\langle \chi^2 \rangle$	$\langle R^2 \rangle$	n
Mono	$k_{\text{obs}} (\text{s}^{-1})$	$7.30 \times 10^{-4}$	$1.4 \times 10^{-5}$	0.65	1.00	
	A (mdeg)	68.4	0.8			
	$y_{\infty}$ (mdeg)	19.5	1.0			
Bi	$k_{\text{obs},1} (\text{s}^{-1})$	$7.17 \times 10^{-4}$	$1.0 \times 10^{-5}$	0.57	1.00	2
	$k_{\text{obs},2} (\text{s}^{-1})$	0.0253	0.0135			
	A <sub>1</sub> (mdeg)	67.8	1.0			
	A <sub>2</sub> (mdeg)	6.44	1.97			
	$y_{\infty}$ (mdeg)	19.7	1.1			

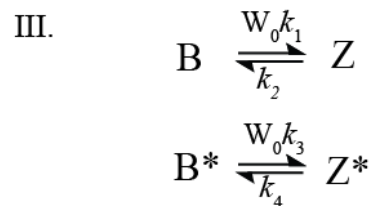
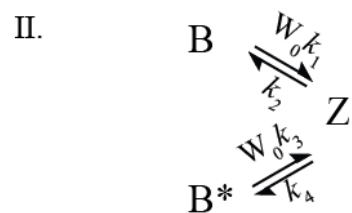
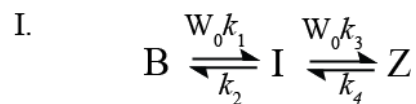
**Table A2.10 | CG<sub>3</sub> + 22  $\mu\text{M}$  Z $\alpha$  B-to-Z-DNA Transition Fitting Values**

Equation	Parameter, $p$	$\langle p \rangle$	$\sigma$	$\langle \chi^2 \rangle$	$\langle R^2 \rangle$	n
Mono	$k_{\text{obs}} (\text{s}^{-1})$	$1.40 \times 10^{-3}$	$1.9 \times 10^{-4}$	1.12	0.95	
	A (mdeg)	20.3	1.6			
	$y_{\infty}$ (mdeg)	3.47	0.09			
Bi	$k_{\text{obs},1} (\text{s}^{-1})$	$9.3 \times 10^{-4}$	$1.2 \times 10^{-4}$	0.90	0.96	2
	$k_{\text{obs},2} (\text{s}^{-1})$	$9.41 \times 10^{-3}$	$2.98 \times 10^{-3}$			
	A <sub>1</sub> (mdeg)	18.5	0.3			
	A <sub>2</sub> (mdeg)	8.94	1.91			
	$y_{\infty}$ (mdeg)	5.38	0.32			



**Figure A2.1 | Single-Stranded DNA CD Spectrum Benchmark.**





**Figure A2.2 | Simplest Kinetic Models Giving Rise to Two Observed Phases.** Shown are three schemes of B-DNA transitioning to Z-DNA, involving a(n) (Scheme I) intermediate, a species structurally similar to (Scheme II) B-DNA (B\*) or (Scheme III) Z-DNA (Z\*). In Schemes II and III, B and B\* are interconverting at a rate orders of magnitude slower than the conversion to Z-DNA. W<sub>0</sub> represents the NaCl concentrations, which is orders of magnitude larger than the DNA concentration.

### Appendix 3

#### Kinetic Simulation Parameters for (<sup>5me</sup>CG)<sub>3</sub>

**Table A3.1 | Molar Extinction Coefficients Used for Eq. 4.1.**

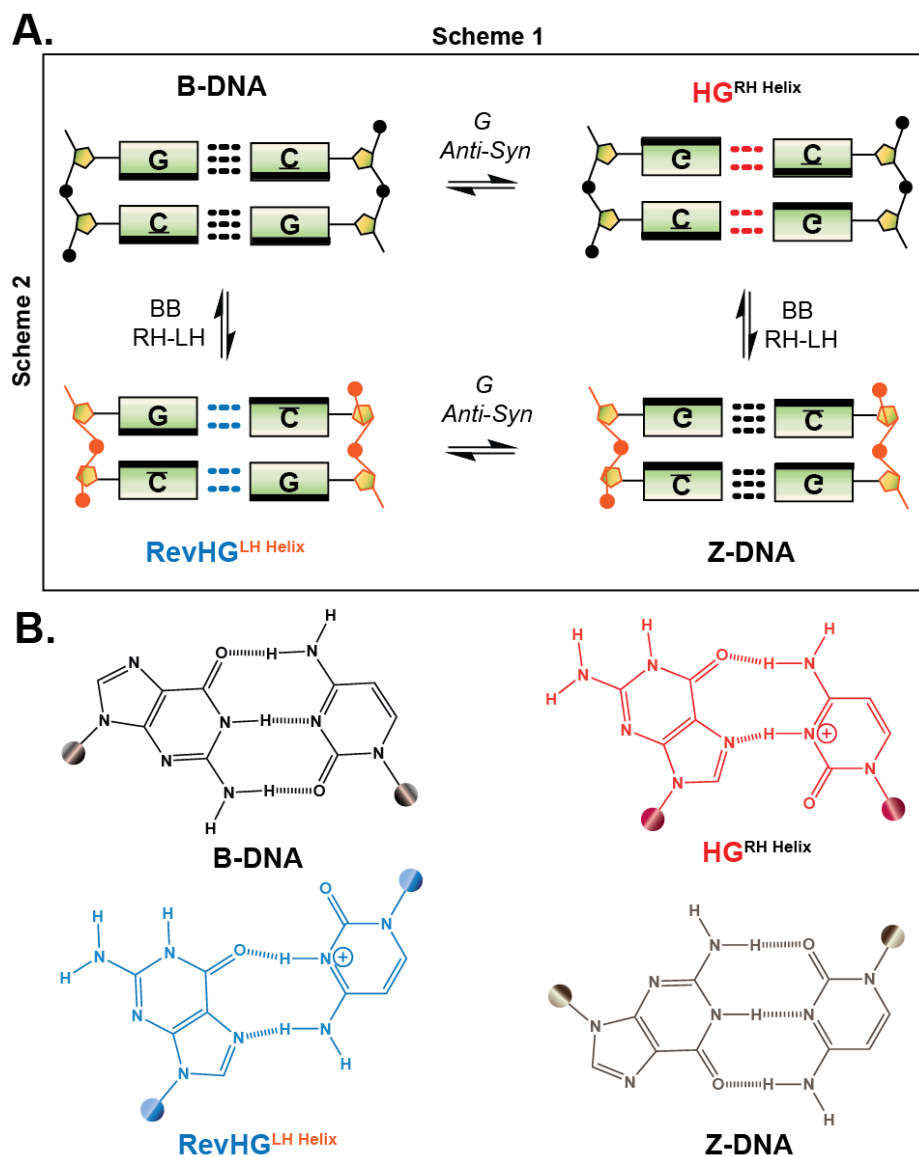
DNA	$\epsilon_{254nm}^B$ (M <sup>-1</sup> cm <sup>-1</sup> )	$\epsilon_{254nm}^I$ (M <sup>-1</sup> cm <sup>-1</sup> )	$\epsilon_{254nm}^Z$ (M <sup>-1</sup> cm <sup>-1</sup> )
CG <sub>3</sub>	-2.6 x10 <sup>6</sup>	3.8 x10 <sup>5</sup>	5.6 x10 <sup>5</sup>
( <sup>5me</sup> CG) <sub>3</sub>	-1.3 x10 <sup>6</sup>	0	1.2 x10 <sup>5</sup>
CG <sub>6</sub>	-4.7 x10 <sup>6</sup>	0	1.8 x10 <sup>5</sup>

**Table A3.2 | Microscopic Rate Comparison for 3.0 M NaCl-Induced (<sup>5me</sup>CG)<sub>3</sub>**

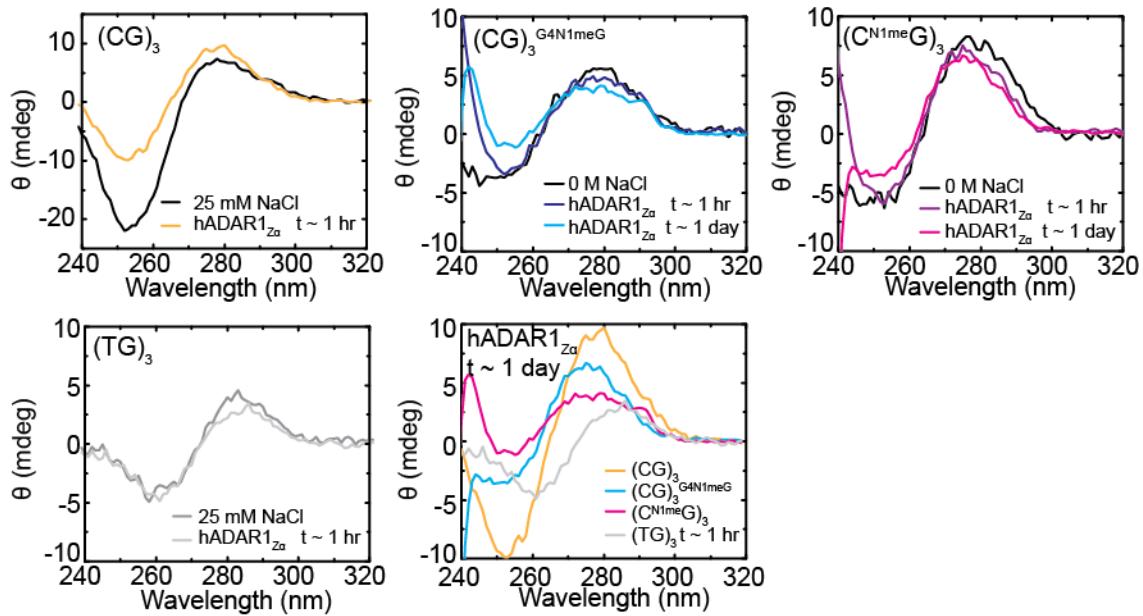
$\epsilon_{254nm}^I$ (M <sup>-1</sup> cm <sup>-1</sup> )	$k_1$ (s <sup>-1</sup> )	$k_2$ (s <sup>-1</sup> )	$k_3$ (s <sup>-1</sup> )	$k_4$ (s <sup>-1</sup> )
0	0.020	0.040	0.008	0.005
3.8 x10 <sup>5</sup>	0.018	0.068	0.011	0.006

## Appendix 4

### Single-Stranded DNA Circular Dichroism Controls



**Figure A4.1 | Two Most Plausible Mechanistic Schemes for the B-to-Z-DNA Transition.** (A) Two schema of the B-to-Z-DNA transition. Orange represents left-handed backbone (BB), upside-down G's represent syn orientation. Corresponding base-pairing pattern for each putative intermediate is indicated. (B) Corresponding base-pairs for helices in (A).



**Figure A4.2 | Recognition of  $(CG)_3$ ,  $(CG)_3^{G4N1meG}$  and  $(C^{N1meG})_3$  but not  $(TG)_3$  by  $Z\alpha$ .**