Spring 5-14-2016

The Influence of Osmolytes on Nucleic Acid Helical Conformations

Pakinee Phromsiri
The College at Brockport

The Influence of Osmolytes on Nucleic Acid Helical Conformations

Master's Thesis

Presented to the Department of Biology

and the

Faculty of the Graduate College

State University of New York: The College at Brockport

In Partial Fulfillment of the Requirements for the Degree

Master of Science in Biology

by

Pakinee Phromsiri

June 2016

Supervisory Committee

Dr. Joshua Blose, Advisor (Department of Chemistry and Biochemistry)

Dr. Rey Sia, Advisor (Biology Department)

Dr. Rongkun Shen

Acknowledgements

I would like to thank The College at Brockport, The Department of Biology, as well as The Department of Chemistry and Biochemistry for giving me the opportunity to study and conduct research under the guidance of such wonderful professors. I would like to thank all of my committee members: Dr. Joshua Blose, Dr. Rey Sia, and Dr. Rongkun Shen for their help and patience in completing my thesis. I would like to especially thank Dr. Joshua Blose for putting up with me through all the successful and unsuccessful experiments conducted these past 2 (plus) years. It was one of my most rewarding experiences and it was lot of fun. ☺ Lastly, and most of all, I would like to thank my family. I could not have accomplished any of this without their love and support.

Dedication

This work is dedicated to my parents, to whom I am most grateful—for their endless love, support, and encouragement. They inspire me to better myself every day.

Table of Contents

Abstract:

In the cell, nearly 40% of the volume is occupied by macromolecular crowding agents and smaller osmolytes that accumulate in response to environmental stresses. The effects of these cosolutes were observed on the transition of helical conformation from B-DNA to Z-DNA. Distinct from the familiar, right-handed B-DNA, Z-DNA is a left-handed double helical structure with its phosphodiester backbone arranged in a pronounced zig-zag pattern. This pattern, unique to Z-DNA is formed from alternating purine-pyrimidine sequences in the DNA. Due to the correlation between Z-DNA formation potential and regions of active transcription, Z-DNA is believed to serve a vital role in the transcription process. Due to the distinctive characteristics of the two types of DNA, the changes in helical conformation may easily be examined using circular dichroism (CD). Spectral analyses revealed that osmolytes (PEG200) promoted the formation of Z-DNA as well as lessened the salt requirement for Z-form stabilization. These results suggest that the formation of Z-DNA is more energetically favorable as the nonpolar, hydrophobic surfaces of the Z-DNA are stabilized in water-poor conditions. The effects of these cosolutes were also observed on the helical conformation of DNA/RNA hybrid duplexes which play important roles in biological processes such as replication, transcription, reverse transcription, and mRNA degradation. Our analyses revealed that the helical conformation of the hybrid duplexes ranged from A-form like to B-form like, depending on the base composition of each strand. In the presence of macromolecular crowding agents, the conformations shifted to more A-

form like while in the presence of osmolytes the conformations shifted to more B-form like. These results suggest that the accessibility of the helical grooves for a given hybrid sequence may be modulated by the cellular environment.

Background and Significance

The Cellular Environment

Introduction

The molecular environment in the cell is very different from the homogeneous, dilute solutions used for *in vitro* environments, typically with biomolecular concentrations of below 10 grams per liter. Living cells contain a large number of macromolecules with a total concentration of 50 to 400 grams of biomolecules per liter in the eukaryotic cytoplasm and 100 to 400 grams per liter in the nuclei [1]. The cytoplasm is a mixture of aqueous solutions of differing macromolecules including proteins, nucleic acids, and carbohydrates. Above a certain concentration, the cellular matrix undergoes phase separation and becomes compartmentalized. Similarly, within the nucleus, macromolecules, along with the large polymers in which form the chromosomes, are confined at high concentrations. The nuclear environment consists of a variety of particles with dimensions ranging from 1 nm and 1 mm dispersed in a continuous phase of varying composition. The interactions of these macromolecules within the cell are strongly influenced by entropic, Van Der Walls, steric, electrostatic, hydrogen bonding and hydrophobic forces [2].

In addition to these particles, the mobility of water molecules in proximity to these intracellular components is also restricted, effecting solvent parameters such as water activity, viscosity, and dielectric properties. This further adds to the differences between the cell and dilute solution. The intracellular environment therefore contains

interactions that may not be anticipated based on the thermodynamic and kinetic properties of a dilute environment. The concentrated and complex mixture of the cell is maintained in equilibrium by physicochemical principles [3].

Osmolytes

When exposed to hyperosmolality, such a high salt or urea, the cell tries to counter their effects by accumulating organic osmolytes [4]. The major systems of these compounds include combinations of urea and methylamines, polyhydric alcohols, as well as free amino acids and their derivatives [5]. Perturbing and compatible solute mechanisms are recognized for protective effects of cells exposed to hypertonicity. Perturbing solutes refer to those of inorganic ions that once reaching a certain concentration threshold will disturb protein function. Compatible solutes however, are organic osmolytes which do not have as much of an effect compared to perturbing solutes and are thus, compatible. Hypertonicity results in the outward osmotic flux of water from inside the cell, subsequently increasing concentration cellular components [6]. The volume of the cell is quickly reestablished via influx of inorganic salts then osmotic uptake of water. Consequently, the concentration of intracellular inorganic ions is elevated; thus, the cell must replace perturbing ions by accumulating compatible organic osmolytes [4].

In mammals, the osmolality of blood is kept at around 290 milliosmol per kg through equilibrium of thirst and urinary concentrations. Specific cells contain significant concentrations of organic osmolytes, including those of the renal medullary which contain the highest due to continuous exposure to exceedingly

elevated concentrations of NaCl and urea. Other cells known to contain concentrations of organic osmolytes include the brain and liver, although they experience hyperosmolality to a much lesser degree relative to the renal medullary [4].

There are several different organic osmolytes in the cell, all of which have interchangeable, nonspecific protective effects, and are not dependent upon specific chemical interactions. The rapid decrease in cell volume due to hypertonicity causes the reorganization of the cytoskeleton and the increase of intracellular concentration. These changes trigger osmoregulatory response signaling. Initial responses include normalization of cell volume and accumulation of intracellular organic osmolytes, leading to additional normalization of intracellular ionic strength [6].

Nucleic Acids

Introduction

In 1953, Watson and Crick proposed the double helix of DNA structure. With the contributions of Erwin Chargaff's discovery and X-ray crystallography work done by Rosalind Franklin and Maurice Wilkins, a three-dimensional structure of the right-handed, double helical structure was derived. In the proposed model, complementary pairs are held together via hydrogen bonding, congruent/ consistent to Chargaff's Rule and adjoined by a phosphate-sugar backbone [7]. From Watson and Crick's original model, scientists further identified two other conformations of the DNA double helix [7].

DNA Structure

In the majority of living cells, the B-DNA is the predominant helical conformation. Other conformations adopted by DNA in the course of carrying out biological reactions include A-DNA and Z-DNA. Cononical B-DNA is a double helix consisting of two antiparallel strands with A-T and G-C hydrogen bonded base pairs [2]. It has 10.5 residues and an axial rise of 3.4 Å per helical turn, with a helical diameter of 20 Å (Fig 1. A, B). A-DNA is wider and more compact relative to B-DNA with 11 residues and an axial rise of 2.55 Å per helical turn, with a helical diameter of 23 Å (Fig 1.A, B). One of the most dramatic changes in helical structure occurs when going from the familiar B-DNA to a narrower, more elongated Z-DNA. It has 12 residues per helical turn, has an axial rise of 3.7 Å per helical turn, and a helical diameter of 18 Å (Fig 1. A, B).

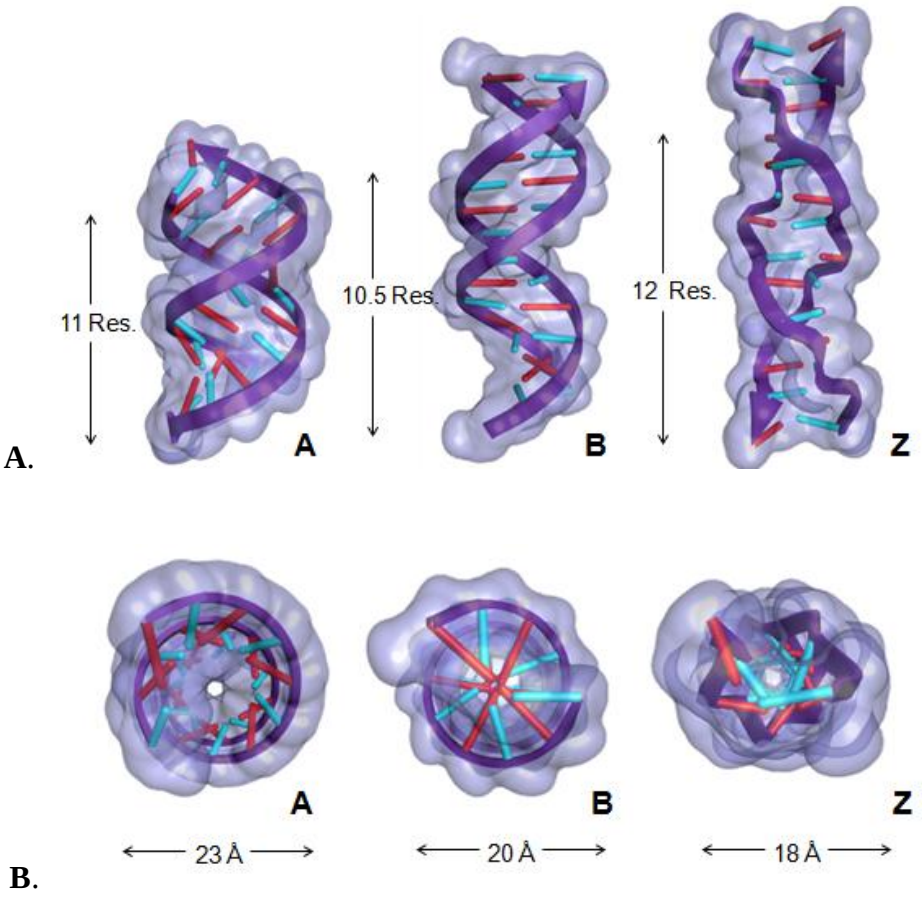


Figure 1. The structural conformations of A-, B-, and Z-DNA. (A) *This is a side view of CG7 A-DNA(left), B-DNA(middle), and Z-DNA (right) strands. The sugar phosphate backbone is in purple, dG bases in red, and dC bases in cyan.* **(B)** *This is a top view of a CG7 A-DNA (left), B-DNA(middle), and Z-DNA (right) strand. The sugar phosphate backbone is in purple, dG bases in red, and dC bases in cyan. Models of A, B, and Z were generated using DS Visualizer with a solvent molecule with a 1.4 Å radius* [8].

Unique to the other two right-handed helices, Z-DNA has a left-handed conformation with a distinct zigzagged backbone [9]. This change in conformation occurs most readily in sequences of alternating purines and pyrimidines, favoring alternating deoxyguanosine and deoxycytidine residues [1011]. This conformation is transient and forms only upon certain biological activity [7]. As previously stated, B-form is the most common form of DNA. A-DNA, although rarely found under standard physiological conditions, may be a vital player in numerous biological processes. The rest of this introduction to nucleic acid structure will focus primarily on Z-DNA as well as the functional roles of hybrids whose conformation range from A to B form.

Functional Roles of A-DNA

Introduction

A-DNA can be found in dehydrated DNA samples, but rarely under standard physiological conditions [7]. Studies of the A-DNA using computer simulation, crystallographic, and biochemical analyses of its structure and complexes with protein however, revealed that A-DNA may be induced upon binding of certain proteins or a significant step in the formation of the distorted DNA conformation seen within several protein-DNA complexes. These investigations indentified numerous structural roles of the A-form in biological processes [121314].

TATA-box Binding Protein (TBP) Binding

It was shown via nanosecond scale molecular dynamics simulations that the GCGTATATAAAACGC DNA oligomer containing the TATA-box binding protein

(TBP) target site adopted the A-form conformation within the TBP-TATA box complex region [15]. This structural distortion of DNA seen in the complex with TBP supports the importance of A-DNA formation in protein-DNA complexes [16].

Cyclic AMP Receptor Protein (CAP) Binding

Studies using the *Escherichia coli* cyclic AMP receptor protein (CAP) showed CAP binding induced bending of the DNA. The recognition elements of the protein are separated by a spacer with either a length of 6bp or 8bp long [17]. When the spacer is 6 bp, the DNA remains in the B-form during bending. When the spacer is 8 bp in length however, a transition into the A-form is required for proper CAP binding [18].

Polymerase Complexes

In cases in which the complexed enzyme cuts or seals at the O3'-P phosphodiester linkage, a transition from B to A- DNA may be necessary to expose otherwise buried atoms of the sugar-phosphate backbone for enzymatic attack [12]. Additionally, crystallographic studies of DNA bound HIV reverse transcriptase revealed a polymerase-induced A-DNA conformation [19]. The conformational change from B-form to A-form at the active site of the polymerase may also increase the fidelity of DNA synthesis [20] by providing bias between correct and incorrect base pairing [21]. Due to a lower dependence on sequence of structural variability in A-DNA relative to B-DNA, presence of A-DNA around the active site of DNA polymerase may improve the fitting of base pairs in the primer-DNA template duplex [21].

Protection from DNA Damage

A study on *Bacillus subtilis*, a sporulating bacteria, described the stabilization of A-DNA by a group of proteins [22]. Relative to the nucleobases in B-DNA, nucleobases in A-DNA are less susceptible to UV damage by several orders of magnitude [23]. This protein-induced change in DNA conformation therefore, may be accountable for spore UV resistance [22].

Z-DNA Conformation

Introduction

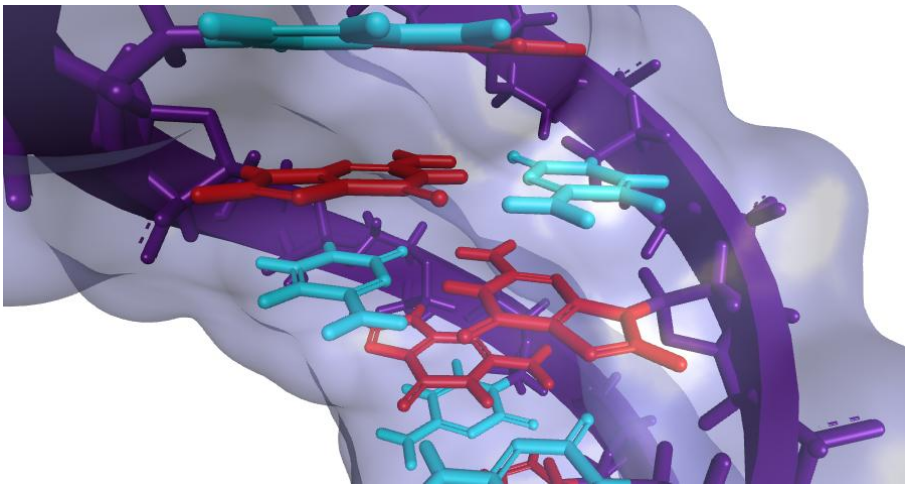
Nucleic acids have the ability to adopt various types of noncanonical structures including the left-handed double helix known as Z-DNA. Formation of this structure is dependent upon both sequence and solution conditions including pH, temperature, and ion concentrations. Regions rich in alternating G-C bases may adopt the Z-form. Potential Z-DNA forming sequences are found in regulatory protein binding sites and control regions in the genome [24]. Due to its relative stability, a Watson-Crick duplex requires a high energy cost to transition to the noncanonical Z-DNA structure in genomic DNA. This energy barrier is lowered in the course of multiple cellular processes. Z-DNA may be formed during gene transcription, DNA supercoiling, nucleosome exclusion, and recombination. Several proteins also bind specifically and induce the formation of Z-DNA [25]. Because of its unique structure and participation in cellular function, Z-DNA is viewed as a promising drug target [26].

Z-DNA Structural Properties

DNA synthesis was developed in the late 1970s, enabling the ability to define structures via single crystal x-ray diffraction. The first atomic view of the double helix however, did not yield the familiar B-DNA structure, but a left-handed helix with a molecular organization completely different from what was anticipated.

Syn and *anti* conformations of the bases are relative to the sugars in the nucleotides. Opposed to the pure *anti* conformation seen in B-DNA, the CG3 DNA hexamer showed an alternation between the *syn*- and *anti*-conformation of bases along the chain [27]. The sugar and glycosidic bond conformation alternated with the C2' endo in *anti* dT/dC and C3' in *syn* dA/dG [2]. As depicted in the following figure, B-DNA has all of its sugars in the C2' endo conformation and all of its bases in the *anti* conformation (Figure 2. A), while Z-DNA possesses the mentioned alternation between C2' endo *anti* dC and C3' endo *syn* dG (Figure 2. B). Due to this, the Z-form helix adopts a zigzag backbone; hence, the name Z-DNA [27].

A.



B.

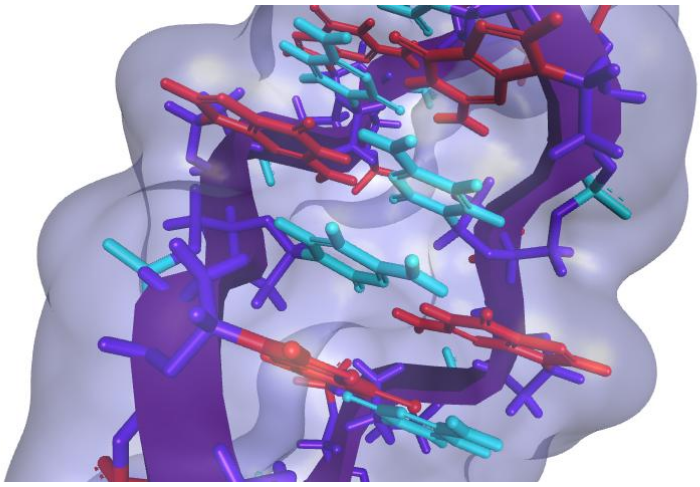


Figure 2. A structural comparison of CG7 B-DNA and Z-DNA. (A) *Model of a cross-section of the CG7 B-DNA strand. The sugar phosphate backbone is in purple, dG bases in red, and dC bases in cyan.* **(B)** *Model of a cross-section of the CG7 Z-DNA strand. The sugar phosphate backbone is in purple, dG bases in red, and dC bases in cyan. Models were generated using DS Visualizer with a solvent molecule with a 1.4 Å radius* [8].

The altered conformation of the base pairs and change in the dioxyribose ring pucker in alternate bases in Z-DNA result in a structure consisting of only one groove, analogous to the -minor groove of B-DNA and a convex surface where the major groove would have been. The phosphate groups in Z-DNA are closer together, relative to B-DNA. Under physiological conditions, the electrostatic repulsion between these charged groups drive the helix into the B-form. Under standard cellular concentrations, Z-DNA is at a higher energy state than B-DNA. Purines have the ability to adopt the syn-conformation without an energy penalty; therefore, the sequence of the strand plays a significant role in B- to Z-form transition. It was found that strands that adopted the Z-form with relatively low energy requirements had alternations of purines and pyrimidines.

Z-DNA Location

In human chromosome 22, 80% of its genes were determined to have sequences favoring the formation of Z-DNA in the vicinity of transcription start sites. The human genome is estimated to contain approximately 100,000 copies of potential Z-DNA forming sequences. Although they are found at various positions within the genome, they are seen at a much higher frequency around transcription start sites [28]. Many studies indicate a strong correlation between transcription and Z-DNA formation. Work on the *Drosophila melanogaster* model showed that antibodies to Z-DNA bound specifically to areas of enhanced transcriptional activity known as puff regions [29].

Z-DNA forming sequences are ten times more frequent in the 5' regions of genes compared to the 3' regions which reflects the overlap between CpG rich islands and the first exon of genes [30]. The distribution of these sequences is congruent with expected association of Z-DNA formation with actively transcribed genes *in vivo*. A computer-aided mapping of over one million base pairs of human DNA, containing 137 genes identified 329 potential Z-forming sequences. Analysis revealed a distinctly nonrandom distribution of these sequences. High potential Z-forming sequences were found more commonly in close proximity to the 5' ends of the genes. It was found that 35% were located upstream of the first expresses exon compared to only 3% downstream of the last expressed exon. The remaining 62% with 47.1% located in introns and 14.9% in exons, were distributed with a high concentration in locations near transcription initiation sites. The strong correlation between the distribution of these Z-forming sequences and regions of the gene expected to be transiently negatively supercoiled during transcription, suggests a possible role for Z-DNA in transcription regulation [31] .

Z-DNA Stabilization

Negative supercoiling stabilizes Z-DNA with the extent of formation varying with sequence [32]. As RNA polymerase moves through the double helix, negative supercoils are generated behind the moving enzyme [33]. This torsional strain generated by its passage thereby supplies the energy needed to stabilize Z-DNA [34]. Z-DNA formation may also be induced by other enzymes including the SWI-SNF-like BAF complex and the colony-stimulating factor 1 (CSF1) gene. Supercoiling becomes a

major contributor to Z-DNA stability in the case that there is an alternating purine-pyrimidine sequence in a circular DNA molecule. In its formation, the DNA is unwound approximately two supercoils per 12 base pairs of DNA. At the junctions where B-DNA meets Z-DNA, are regions spanning several basepairs of nucleobases that behave as if unpaired. These stretches were also found to be partially reactive to chemicals specific to single-stranded DNA [2]. Thus, Z-DNA form separately from the canonical B-form DNA.

In addition to stabilization by negative supercoiling, increased salt concentration have been shown to stabilize Z-form DNA *in vitro* [35]. As previously mentioned, the phosphate groups are brought closer (relative to B-DNA) due to the structure of the zigzag backbone which increases the electrostatic repulsion between them. Z-DNA is therefore, stabilized by high salt concentration as the presence of cations help shield this repulsion.

Functions of Z-DNA

Introduction

Potential Z-DNA forming sequences are found at various positions in genomes. In eukaryotes, Z-DNA forming sequences were found to accumulate near transcription start sites [30] as well as in transcriptionally active regions implicating numerous functional consequences [2], including transcription, the modulation of supercoiling, nucleosome exclusion, recombination, and protein binding.

The formation of Z-DNA is stabilized by the negative supercoiling following the transcribing RNA polymerase. Furthermore, its occurrence may block the

transcription of that region of the gene from the following RNA polymerase, known as polymerase stalling [36]. This supercoiling of the DNA may remodel the nucleosome and alter the organization of chromosomal domain due to regions that exclude histones and other architectural proteins [37]. The ability of Z-DNA to relieve localized topological train may also facilitate homologous chromosomal recombination by allowing the duplex interactions required for the formation of paranemic joints [38]. Consequently, these changes in DNA topology could alter the recognition sites for protein binding [2].

Transcription

In the process of transcription, as the DNA is unwound, negative torsional strain develops behind the moving RNA polymerase and positive torsional strain develops in front of the moving polymerase. Due to its left handed conformation, Z-DNA is the most under-wound form of DNA; thus formed through the negative torsional strain created by the passing polymerase and drives the formation of Z-DNA upstream the transcribing gene. This formation of Z-DNA may block the following RNA polymerase from transcribing that genetic region, ensuring spatial separation between polymerases and minimizing mis-splicing of messages [36]. Upon cessation of transcription, Z-DNA is quickly reverted back to B-DNA [2].

The formation of an alternative structure such as Z-DNA may impact transcription through modifying levels of supercoiling thus altering the energy cost for protein binding. In addition, due to changes in spatial positioning, interactions between the binding of transcription factors to their respective sites [2].

Modulation of Supercoiling

Because optimal topology of DNA may be required for processes such as transcription, replication, and recombination, it is expected that supercoiling will have an effect on these processes [39]. Z-DNA formation may result in a partial relaxation of excessive superhelicity in the domain. These changes in superhelicity were identified to be depended upon in certain cases of DNA replication and gene expression [40]. Unconstrained negative supercoils that are generated in processes such as transcription stabilize non-B-DNA structures including Z-DNA [33]. The supercoiling of DNA can affect transcription both in prokaryotes and eukaryotes by effecting the open complex formation in prokaryotic genes as well as binding of the TATA-binding protein (TBP) to the TATA element in eukaryotic genes. In eukaryotes, it further impacts transcription through its affect on chromatin remodeling [41].

Nucleosome exclusion

Although the bulk of the eukaryotic genome is negatively supercoiled, it is however, not torsionally stressed due to the fact that a large portion of these supercoilings are accommodated by the packaging of the DNA into nucleosomes [42]. The wrapping of DNA around histones interferes with the binding of promoters and origin of replication. SWI/SNF is a chromatin remodeling system that aids in the unwrapping of DNA from nucleosomes; thereby, turning the gene on. When the nucleosome is unwrapped, the DNA is left in a negatively supercoiled state and adopts the Z-conformation which keeps the site open. This enables the binding of transcription factors and RNA polymerase needed to initiate transcription. This

mechanism of triggering transcription via Z-DNA formation may be extensive due to sequences favoring Z-formation occupying regions near transcription sites [43].

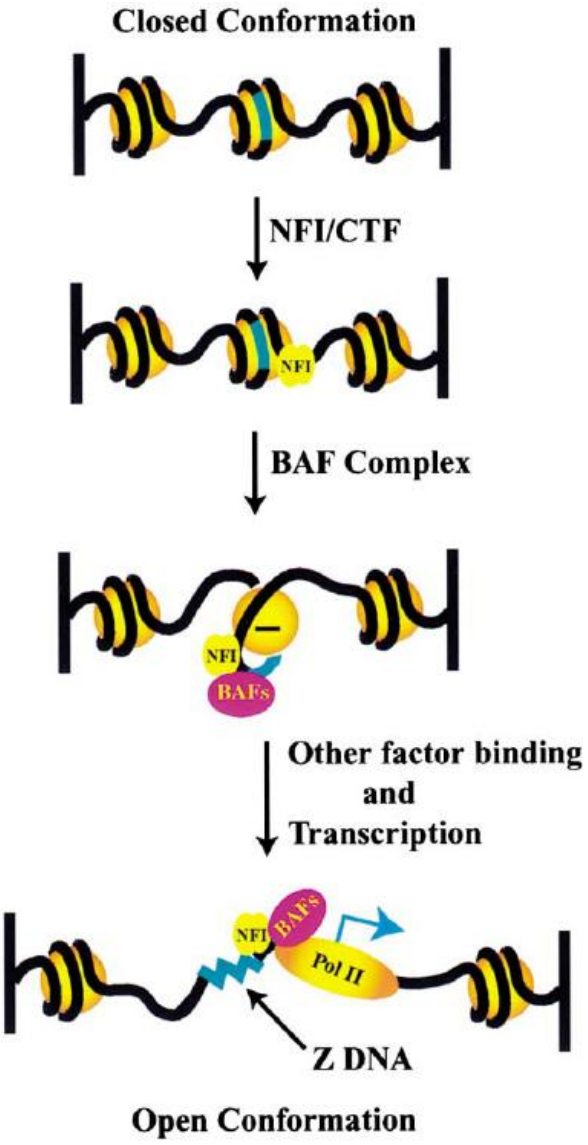


Figure 3. Chromatin remodeling by the BAF complex facilitated by Z-DNA formation. *The formation of Z-DNA necessary is for promoter activation by the complex to remove nucleosome* [43].

The position of the nucleosome as well as how the DNA is organized in chromatin is governed by both the structural and mechanical properties of the DNA itself. Z-DNA in particular, has been reported to play a crucial role in the formation, position, and stability of nucleosomes and thereby, the packaging of DNA in the nuclei [42].

A study identified Z-DNA as a player in gene activation coupled with chromatin remodeling. TG repeats border the promoter of the human colony-stimulating factor 1 (CSF1) gene, which upon activation by BRG1-associated factor (BAF) complex, is converted into Z-DNA *in vivo*. It was also observed *in vitro* that Z-DNA formation in a nucleosomal template was facilitated by the BAF complex. These data suggest that Z-DNA formation induced by BAF at the promoter of the CSF1 gene stabilizes an open chromatin structure, as well as demonstrate why Z-DNA forming sequences are often found in or flanking promoter regions (Figure 3) [43].

Recombination

Numerous relationships are illustrated between DNA recombination and the formation of alternative structures, including several proposed models of Z-DNA assisted recombination [44]. During recombination, the DNA strand exchange requires initial duplex-duplex interaction. This interaction is facilitated by the contact of exposed N7 and C8 guanosines between two Z-DNA duplexes. In the synapsis step of homologous recombination, a paranemic joint may be formed with alternating left-

handed and right-handed turns of the DNA; thus, enabling strands from different DNA molecules to base pair without breaking (Figure 4) [38].

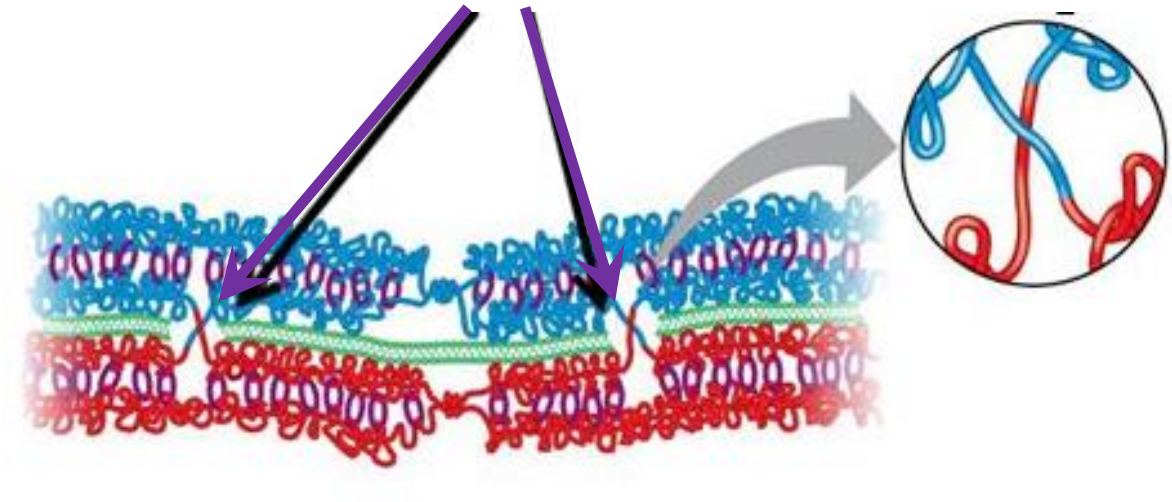


Figure 4 [45]**. Formation of paranemic joint.** *In this interaction (indicated by the arrows), individual complementary strands do not intertwine; thus, a molecule that is base-paired but not topologically linked is produced* [46].

Protein Binding

A major component in understanding the biological roles of Z-DNA comes from properly identifying proteins that specifically interact with it [2]. In identifying Z-DNA binding proteins, the method devised must be able to isolate proteins that bind specifically and with high, unambiguous affinity to Z-DNA. The technique developed was a band shift assay using radioactive labeled Z-DNA in the presence of excess single-stranded DNA and B-DNA. The proteins in which bounded to the Z-DNA

were isolated and identified. Highly specific binders include a Z-DNA-binding nuclear-RNA-editing enzyme called adenosine deaminase acting on RNA-1 (ADAR1), and E3L protein of vaccinia virus [25]. Relatively low specific proteins include HMG proteins, type III intermediate filament proteins, and zeta crystalline [474849].

In addition, the Z-DNA binding domain of the protein, identified as Zα was mapped and expressed in *Escherichia coli*. This enabled the confirmation of its specificity via circular dichroism and Raman spectroscopy; providing direct evidence of Z-DNA in peptide-DNA complexes [50]. These data showed the binding of Zα to Z-DNA composed of different sequences with varying nucleotide composition [51].

Z-DNA Binding Domain: Zα

Zα belongs to the winged helix-turn-helix family of proteins which involve the interaction between conserved residues in α3 and C-terminal β-sheet wing. The wing consists of two conserved prolines and one conserved tryptophan residue that guarantee specificity for the Z-DNA conformation [52]. The prolines interact directly to the Z-DNA backbone while the tryptophan lies in the hydrophobic pocket and interacts indirectly via water mediated DNA contact. In addition, to contributing to the domain's overall rigidity, the tryptophan orientates a conserved tyrosine in α3 for direct DNA contact (Figure 5) [53]. The majority of these Z-DNA interacting residues are prepositioned to bind Z-DNA in solution [52]. This DNA binding domain is also very stable and resistant to thermal denaturation [54]. Interestingly, the Zα domain may

also bind to RNA and induce a Z-form structure, although the biological significance of Z-form RNA is still unknown [55].

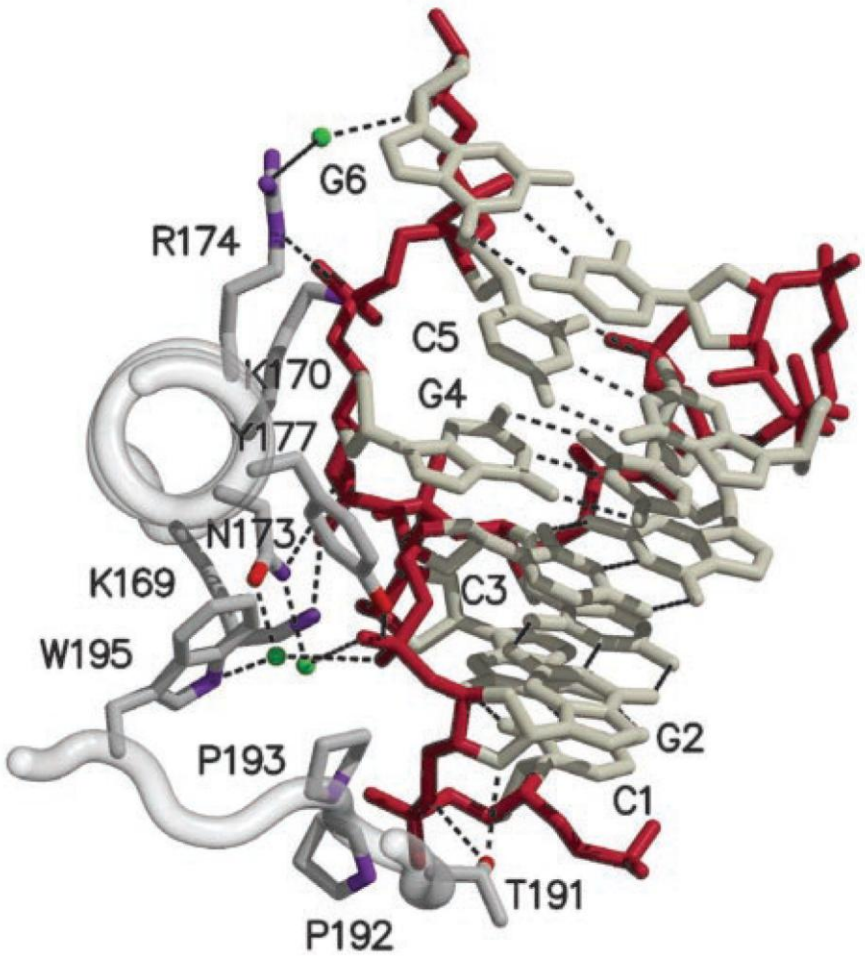


Figure 5. DNA-Protein interaction. *A portion of ADAR1 editing enzyme ($Z\alpha_{ADAR1}$) Z-DNA-binding domain complexed to Z-DNA. Interactions between the amino acids of the protein (left) and the Z-DNA (right) via electrostatic and van der Waals interactions.* [25]

DNA-RNA Hybrid Duplex Structure

Introduction

The DNA-RNA duplex is composed of one strand of DNA and one strand of RNA, thus forming a hybrid of the two. The structural conformations of DNA and RNA duplexes are well characterized; however, studies on hybrid structures of the two often time yield contradictory results. This is due to the fact that the hybrid duplex is not a simple average of the pure structure, but a dynamic structure influenced by multiple factors [56].

Hybridization in the Lab

Hybridization is dependent on the ability of single-stranded nucleic acid sequence to specifically bind to its complementary strand. When double stranded DNA is denatured, it will hybridize when in our case, heat is removed. The annealing process begins with a few bases followed by the zippering of the remaining. Complementary strands of ssDNA and ssRNA hybridize via hydrogen bond formation between the nucleotide bases (A-T, A-U, G-C) (Figure 6. A-D) [57].

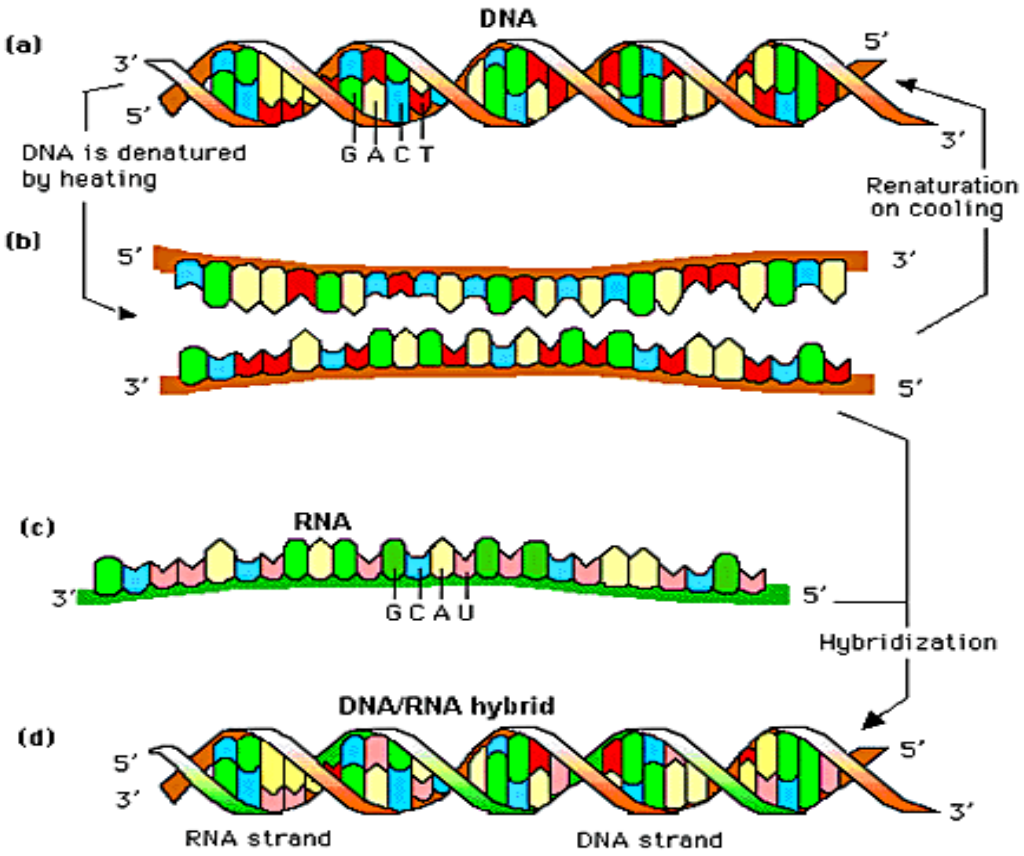


Figure 6. Nucleic acid hybridization. (A) *Double stranded DNA is denatured into* **(B)** *single stranded DNA, and* **(C, D)** *hybridized with its complementary RNA strand.* [57].

DNA-RNA Hybrid Duplex Functions

Transcription and Reverse Transcription

Transcription is a well-known process in which DNA mediates the synthesis of RNA. Using the DNA strand as a template, ribonucleotides are incorporated to the 3' end of the growing strand of RNA, thus forming a hybrid duplex for the duration of their interaction. Mechanistically similar to transcription, reverse transcription is the process in which RNA mediates the synthesis of DNA. This process does not only play an essential part in the life cycle of retroviruses as intermediates in retroviral DNA synthesis, but was found to be required in eukaryotic cells as well—with up to 10% of the genome being a direct product of reverse transcription. Formation of the hybrid duplex may be seen prior to RNase H cleavage of the template strand of RNA and after DNA synthesis which are both mediated by reverse transcriptase [58].

DNA Replication

The DNA replication process begins with the unwinding of the DNA duplex by DNA helicase followed by priming of the DNA strands by proteins and enzymes. The leading strand is replicated upon exposure of the single stranded region of DNA. The lagging strand requires an altered process utilizing Okazaki fragments as the strand is synthesized, forming hybrids every 100 to 200 nucleotides which are then subsequently replaced by DNA [59].

Gene Expression

With the role of DNA-RNA hybrids in the transcription and replication process, it is expected that they will effect gene expression as well. Proteins that dictate cellular function are encoded by genes. The array of genes that are expressed therefore, determine what the cell does. Of all the genes the cell may contain, only a fraction is expressed at any given point in time. The flow of information from DNA to protein in the cell provides potential control points for the regulation of protein manufacturing. Each cell has a set of transcription regulators, unique to that cell type. These regulators may either increase or suppress transcription. In prokaryotes, nutrient availability dictates what is expressed by the regulatory protein. In eukaryotes, this regulation involves a combination of proteins working together in response to environmental change. Transcription and the processes following are therefore, critical for adaptability [60].

Osmolytes

The stability of nucleic acid structures is reliant upon the several factors, including base pairing, electrostatic interactions, and the cellular environment [61]. The hydration of the nucleic acid, along with the osmotic pressure experienced due to the presence of biomolecular crowders in the cell will influence its structure [62]. Hence, understanding the effects of both enthalpic and entropic forces of the cell is critical in nucleic acid study.

Hydration is an important component of DNA structure, stability, and dynamics [63]. The cell experiences an array of stresses from the environment that may

influence its intracellular water volume. In order to maintain cellular processes, cells overcome the water stresses by accumulating low molecular weight compounds called osmolytes or osmoprotectants that tightly bind to water molecules and thus restricting their mobility [64]. Due to these osmolytes, the dielectric constant of the cell is significantly lower compared to pure water [65]. This decrease results in an increase in electrostatic attraction and repulsion. Because Watson-Crick base pair formation is coupled with water binding, the stability of the duplex depends largely upon water activity. In the presence of small cosolutes, this activity substantially decreases. These small, organic molecules along with macromolecular crowders therefore, greatly impact the molecular environment in a living cell [66].

Crowding Agents and Entropic Forces

Due to the crowded environment within the cell, the macromolecules are enveloped inside are in close proximity to one another. They are therefore, strongly influenced by entropic forces or short-range forces [67]. In a mixture of particles of varying size, these forces favor contact between larger particles. Due to steric effects, the space occupied by macromolecules is inaccessible to other molecules. The volumes excluded by these particles overlap thus decreasing the entropy and accessible volume for smaller particles by the excluded volume effect [68]. The effects of crowding on polymorphic nucleic acid structures suggests a link to transcription control as well as the dense packing of chromatin [69]. In addition, further crowding results from the effective volume exceeding the intrinsic volume of the macromolecule itself. This therefore limits the space available for other

macromolecules of the same size to occupy, enhancing their effective concentrations and thermodynamic activities significantly [70].

Nucleotide Interactions

The stability of nucleic acids is essential for proper function and storage of genetic information. Two major nucleotide interactions that account for its stability include aromatic stacking and hydrogen bonding. In DNA, the purine and pyrimdine bases are position parallel to one another in steps as rungs on a ladder [71]. The aromatic stacking of overlapping p-orbitals, also known as pi stacking, creates weak noncovalent forces between the nucleotides (Figure 7. A), while strong interchain hydrogen bonds hold the two polynucleotide strands together **(**Figure 7. A). In addition, external hydrogen bonds may be formed between polar groups of the nucleotide and surrounding matrix, contributing to overall helical stability **(**Figure 7.B). The strengths of these interactions not only stabilize the structure, but determine sequence-specific efficiencies and helical structure adopted [72].

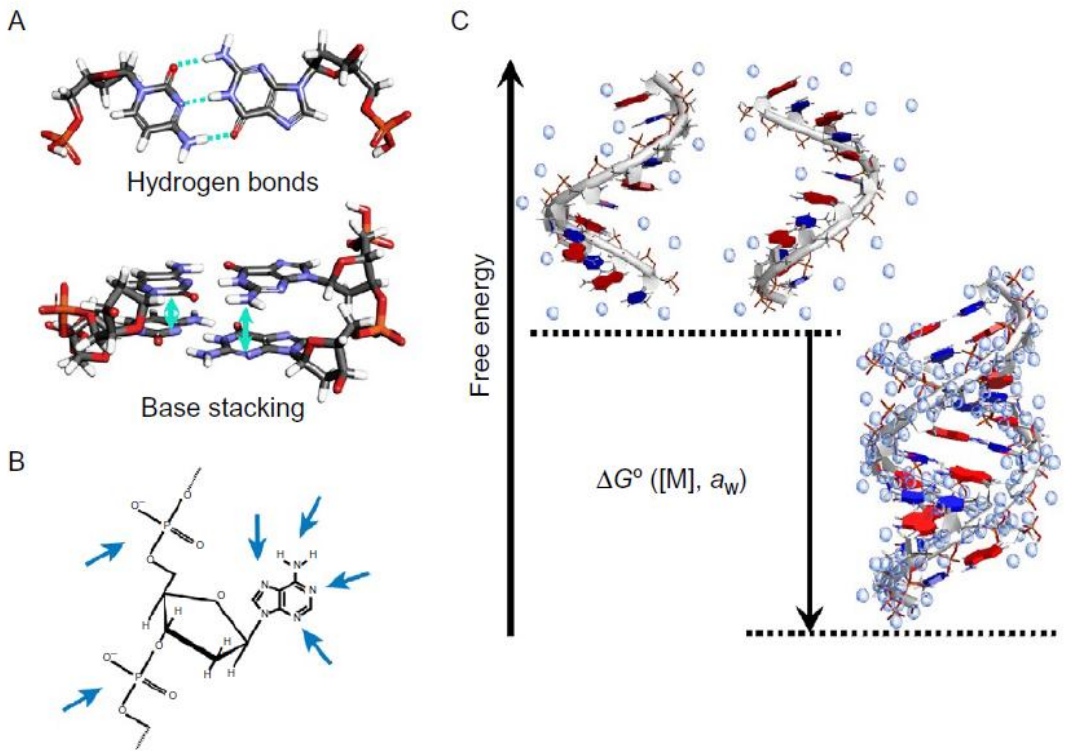


Figure 7. Nucleotide interactions, hybridization, and nucleic acid folding. (A) *Inter-base hydrogen bonding and base stacking in Watson-Crick base pairs.* **(B)** *Potential metal ion and water molecule binding sites on DNA nucleotide indicated by the arrows.* **(C)** *Free energy diagram for the Watson-Crick duplex formation as a function of the cation concentration [M] and the water activity aw.* [72]

Counterions

DNA is a highly charged polyanion due to charged phosphate groups generating a cloud of negative potential. This creates strong intramolecular repulsions within the molecule; thereby, influences its structure, flexibility, and biological function [73]. In the cell, cations and water molecules associate with the

charged sugar-phosphate backbone and polar bases of DNA [74]; effectively reducing these repulsions via screening effect of the solvent [73]. Because high salt concentrations mask the destabilizing repulsion of charge between the backbones, increasing salt concentrations will increase pi-stacking interactions; thereby, resulting in an increase in duplex stability[75]. Counterion interaction and hydration therefore, influence the thermodynamic stability of the nucleic acid molecule. Furthermore, the association and dissociation of ions and water molecules is energetically coupled to the folding nucleic acids **(**Figure 7. C) [72]**.**

Nucleic acids play vital roles in not only the storage of genetic information, but in numerous biological functions as well. The structure, function, and stability of DNA and RNA in the living cell are of great interest in various fields including medicine, pharmacy, chemistry, biology, and biophysics [76].

Specific Aims

The highly crowded media of the cell differs greatly from that of an *in vitro* buffer solution. A large number of macromolecules and organelles occupy a relatively small cellular volume; thus, restricting molecular movements and dynamics. How the cellular environment influences thermodynamic parameters and structural properties of nucleic acid helical conformation; thereby, dictates the conformations adopted within the cellular environment. Using cosolutes as crowding agents to mimic the cellular environment, we quantitatively measured the thermodynamic behavior of nucleic acids as it would behave *in vivo and* studied the influence of osmolytes and crowding agents on nucleic acid conformation. The use of molecular crowding agents

aided in increasing our understanding of the reactions and functions of nucleic acids in the cellular environment. We used Polyethylene Glycol with a molecular weight of 200 g/mol (PEG 200) to mimic the presence of osmolytes in times of stress and Polyethylene Glycol with a molecular weight of 8000 g/mol (PEG 8000) as a crowding agent.

Specific Aim I: The Effects of Osmolytes on the Thermodynamic Parameters of Folding Z-DNA

Our first specific aim was to determine the effect of osmolytes on the structure of Z-DNA and the thermodynamic parameters of B- to Z-DNA transition through qualitative and quantitative analysis of CD spectral data. Z-DNA is a transient structure that appears in multiple biological processes including transcription and specific protein binding [77]. Z-DNA experiences an elevated steric hindrance and repulsion within its structure due to conformation; therefore, requires certain extrinsic conditions for stabilization. It is widely known that the Z-DNA is stabilized by solutions of high salt concentration [78]. Z-conformation is also more favorable in water poor environment due to hydrophobic interactions. Small, uncharged cosolutes have the ability to compete with water for nucleic acid solvation. They are able to preferentially interact with surfaces which may be exposed or buried, making them independent contributors to the m-value or in this case, the cooperativeness of B-to Z-transition [79]. In addition to the m-value, we determined other thermodynamic and structural parameters of the transition, as well as the effects of PEG 200 on the formation of Z-DNA.

Specific Aim II: A Qualitative Analysis of the Influence of Osmolytes on Hybrid Duplex Structures

Our second specific aim was to determine if the presence of osmolytes and macromolecular crowders in solution can act as a helical "conformational switch" for hybrid duplexes. DNA-RNA Hybrid duplexes play significant physiological roles in a number of cellular processes including DNA replication, RNA transcription, and reverse transcription [80 81 56]. DNA adopts the B-form helical conformation and RNA adopts the A-form in solution. The structures possess differences in thermodynamic and chemical stabilities, as well as structural properties. Studies of hybrids indicate a range of helical conformations the duplex may adopt, with some nucleobase-content dependence [80 81 56 82]. The focus of our experiments was on utilizing cosolutes to determine the effects of the cellular environment on hybrids. We believe that certain osmolytes and macromolecular crowders may induce hybrid conformations which may be more A-form or B-form-like. The presence of molecular crowding agents influenced the helical structure to trend towards an A-form conformation because of excluded volume and electrostatic effects. The more compact A-form was favored in a crowded cellular environment, as well as more energetically favorable due to more favorable ion interactions relative to the B-form [83]. These ion interactions with the A-form's major groove were thought to contribute to the chemical preference of one conformation to another in a solution of lowered dielectric constant. Inside the living cell, the dielectric constant is considered to be much less than that of water. This environment may be established in a crowded solution. CD spectroscopy is often used

for secondary protein and nucleic acid structural analysis [84] [80]. Because DNA and RNA have different interaction potentials with increasing salt concentration [83], we were able to assess if a hybrid had more B-or A-form like behavior in the presence or absence of osmolytes and macromolecular crowders.

Materials and Methods

The experimental details for both specific aims are very similar in terms of techniques and solution conditions, and as such this section will serve as a combined Materials and Methods section for both aims.

DNA/ RNA Sequences

All nucleic acid strands were synthesized by Integrated DNA Technologies. The DNA and RNA samples were used without further purification due to efficiency of coupling and the shorter lengths of the strands. Synthesized sequences are as follow:

Specific Aim I Sequences

Label	Sequence
CG3	5'-CG-CG-CG-3'
CG7	5'-CG-CG-CG-CG-CG-CG-CG-3'
CG6CA1	5'-CG-CG-CG-CA-CG-CG-CG-3' (Strand 1)
	3'-GC-GC-GC-GT-GC-GC-GC-5' (Strand 2)
CG4CA3	5'-CG-CA-CG-CA-CG-CA-CG-3' (Strand 1)
	3'-GC-GT-GC-GT-GC-GT-GC-5' (Strand 2)

CA7	5’-CA-CA-CA-CA-CA-CA-CA-3’

Specific Aim II Sequences

Label	**Sequence**
d53	5’-CGC-TAC-ATA-GTG-AGC-3’ (DNA)
d47	5’-GCT-CAC-TAT-GTA-GCG-3’ (DNA)
r53	5’-rCrGrC-rUrArC-rArUrA-rGrUrG-rArGrC-3’ (RNA)
r47	5’-rGrCrU-rCrArC-rUrArU-rGrUrA-rGrCrG-3’ (RNA)

The samples were diluted to 1 mM with 1X TE (Tris-EDTA) pH 7 buffer and kept at room temperature.

Solution Conditions

All samples contained 10 mM sodium phosphate buffer, pH 7. Stock solutions containing 40% w/v cosolutes were prepared to final volumes of 10 or 50 mL with masses of all components recorded for conversion between concentration units as needed. Appropriate amounts of NaCl were included in buffer stocks as necessary to achieve desired final salt concentrations. All chemicals utilized were purchased from Fisher Scientific except for PEG 200 and PEG 8000, which were purchased from Sigma Aldrich. All cosolutes were utilized without further purification.

CD Spectra

The CD (circular dichroism) spectra of the samples were obtained using a Chirascan CD spectrophotometer (Applied Photophysics) flushed with a constant flow of nitrogen gas. Prior to measurement, Z-DNA samples were heated to 60°C and cooled to 3°C at a rate of 3°C/ min to ensure proper folding of the duplexes Spectra were taken for the melted samples and corresponding buffers in a 1 cm path length cuvette at 5°C to compare to previous studies [84] from 320 nm to 230 nm for Z-DNA samples. Spectra were collected from 340 nm to 215 nm, every 1 nm with a 4s averaging time per point. Spectra are the average of 3 scans. Data was repeatable across days. For the hybrid samples, no difference was shown between those that were renatured prior to analysis and those that were not. Spectra were collected from 340 to 200 nm at 20°C, every 1 nm with a 4s averaging time per point, and all data shown are the average of 3 scans. DNA for all samples in both specific aims was approximately 1 μM.

Buffer subtraction and Normalization

The spectra generated by Chirascan were buffer subtracted and smoothed in Pro-Data viewer (Applied Photophysics). The buffer subtracted data was normalized using equation 1:

$$[\theta] = \frac{\theta}{(10\ C \times N \times l)} \quad (1)$$

C is the concentration in M, N, the number of nucleotides, and l, the pathlength in cm. The nucleotide concentration in each sample was determined using Beer's Law (equation 2):

$$A = \varepsilon lc \quad (2)$$

A is the UV absorbance taken at 260 nm, ε, the calculated extinction coefficient of the sequence, *l*, the path length (cm), and *c*, the concentration (mol dm^{-3}). Extinction coefficients for the duplexes were calculated according to reference [85].

Thermodynamic Analysis of Z-DNA Folding

Quantitative analysis of Z-DNA folding was done by fitting CD data at 293 according to a unimolecular folding model as the duplex itself is already hybridized. The equilibrium is considered for the B to Z transition where K is the equilibrium constant for that transition and the fraction in Z-form (f_Z) is defined by equation 3 below, which upon division by [B] yields equation 4:

$$f_Z = \frac{[Z]}{[Z] + [B]} \quad (3)$$

$$f_Z = \frac{K}{K + 1} \quad (4)$$

Equation 5 then provides the relationship between K and free energy (ΔG)

$$K = e^{\left(\frac{\Delta G}{-RT}\right)} \quad (5)$$

The CD signal collected for a given sample is then a combination of some fraction of the molecules in the B-form and an additional fraction in the Z-form. In

fact, this can be expressed such that the signal is equal to the signal of the B-form plus a fractional contribution of the difference between the B and Z-form signals. Rather than assume simply a characteristic value for the B-form and the Z-form, these were modeled as baselines as shown in equation 6 below where m_n and b_n refer to slopes and intercepts of the B or Z-form signal. In this case x refers to the sodium ion concentration but could be used for increasing concentrations of many cosolutes.

$$Signal = m_B x + b_B + \big((m_Z x + b_Z) - (m_B x + b_B)\big) f_Z \quad (6)$$

Equation 7 is obtained by substituting in equation 4 into equation 6

$$Signal = m_B x + b_B + ((m_Z x + b_Z) - (m_B x + b_B)) \frac{K}{K+1} \quad (7)$$

Finally, equation 8 is obtained by substituting equation 5 into equation 7 and expressing the free energy as a function of the free energy of folding without sodium ions present (ΔG_0) and subtracting a linear contribution (m) to the free energy dependent on the sodium ion concentration. The linear contribution is called the m-value [86].

$$Signal = m_B x + b_B + ((m_Z x + b_Z) - (m_B x + b_B)) \frac{e^{\left(\frac{\Delta G_0 - mx}{-RT}\right)}}{e^{\left(\frac{\Delta G_0 - mx}{-RT}\right)} + 1} \quad (8)$$

The m-value is related to the amount of structural change occurring in a folding transition is sometimes related to the amount of surface area changing conformation in a given transition. Based on this interpretation the m-value is a measure of the extent of folding cooperativity.

The CD signal was fit according to equation 8 as a 6 parameter fit using QtiPlot to the baselines of each conformation as well as ΔG_0 and the m-value [87]. For the spectra in which baselines were extremely sloped or have few data points, an alternative to the 6 parameter fit, the linear extrapolation method (LEM) was used. This method utilizes the same general equations as in the nonlinear curve fitting, but it is performed in stepwise fashion. Based on estimates for the folded and unfolded signal, the f_Z can be calculated at every point and used to determine K and ΔG for a given sodium ion concentration. Free energy is plotted vs. $[Na^+]$ and a linear fit is used to obtain ΔG_0 and the m-value which are the intercept and slope of the fit, respectively.

Additional Parameters

After determining ΔG_0 and the m-value for the data additional parameters can be determined. The first is the C_m or midpoint of the transition. This can be calculated by setting ΔG equal to zero and solving for the appropriate sodium ion concentration. This value is somewhat analogous to the T_m for thermal denaturation experiments. At both the C_m and the T_m is assumed that populations of B and Z-form DNA are both 50% of the total molecules in solution.

Also given ΔG_0 and the m-value the ΔG_x for any salt concentration can be calculated and used to determine the K value at that concentration. From K both the fraction of Z at any salt concentration can be determined as well as the relationship between the ln K and the ln $[Na^+]$. A plot of this relationship can be linearly fit to

determine Δn, the number of sodium ions either bound or dissociated in a folding transition [84].

Results and Discussion

Specific Aim I:

The first specific goal of this research was to determine the effects of osmolytes on the structure and thermodynamic parameters of nucleic acids. We specifically studied the transition from B-DNA to Z-DNA. It is widely known that DNA consisting of alternating CG base pairs may adopt the Z-form duplex; therefore, sequences ranging from perfect CG repeats to "broken" sequences containing CA breaks (refer to materials and methods) were used to examine sequence-dependence of Z-DNA formation. Alternating CG sequences adopt the Z-form conformation in the presence of high salt concentrations and B-form in low to moderate concentrations. Polyethylene glycol (PEG) is a neutral polymer which may be used to mimic both osmolyte and crowding agent based on molecular weight. It was chosen as our cosolute due to the fact that it is inert towards nucleotides and relatively silent in terms of CD signal in the region of interest [84]. The conformations of DNA in the presence of varying salt and PEG concentrations were studied at CD wavelengths reflecting base stacking and backbone conformation.

Spectral Region of Interest

CD spectra of the B-form duplex are distinctly different from that of the Z-form duplex due to the forms' opposite chirality [84]. The B-form DNA contains a negative peak around 253 nm and a positive peak around 280 nm, while the Z-form

DNA has a negative peak around 293 nm and a positive peak around 270 nm. The peaks at 253 nm and 293nm both change; however, the peak at 293 nm is more indicative of the Z-form. The peak at 253 nm was found to change in regards to stacking interactions for both B and Z-forms; therefore, we will focus on the changes seen at 293 nm.

Sequence Selection

The alternating GC sequences may adopt the left-handed helical conformation of the Z-form duplex with consecutive G-C base pairs [84]. To verify results and repeatability of past studies, a fit curve was generated to quantitatively identify the C_m or the concentration at which 50% of the DNA is in the B-form and the other 50% is in the Z-form. Our C_m was found to be 2.7+/- 0.2 M which is within error of the literature value of 2.5M [84]**.** Because the data obtained for the CG3 sequence was in good agreement with the literature, we continued our study with the longer CG7 sequence, a sequence often used in protein-DNA binding studies [88]. Due to its increased length, it should form Z-DNA easier as well as give a larger signal at a lower concentration of DNA. In comparing spectra of the CG3 and CG7, CG7 displayed CD signals a magnitude greater than CG3. Additionally, the CG7 sequence is a little more than one full turn of the Z-DNA helix; hence giving us room for the substitutions needed to study the broken sequences.

Influence of PEG 200 on DNA Conformation

Polyethelene glycol (PEG) is a neutral cosolute widely used to mimic the cellular condition. We studied the conformations of the CG7 sequence in the absence

of osmolytes, and the presence of 20 wt% and 40 wt% PEG 200 at varying NaCl concentrations. Figures 8-10 show CD spectra reflecting the ability of the CG7 sequence to transition from B-form DNA to Z-form DNA at increasing cation and PEG 200 concentrations.

Cations are necessary for the creation of ordered nucleotide structures in eliminating electrostatic repulsions of the confined phosphate groups [89]. The folding of nucleotide is accompanied by cationic molecule association. This shields the electronegative potential of the phosphates; thus, enables them to be brought closer in proximity[89]. The region of interest at 293 nm shows the transition from the fully folded B-form (dotted line) to the fully folded Z-form (solid line).

Spectral Analysis for CG7 Sequence

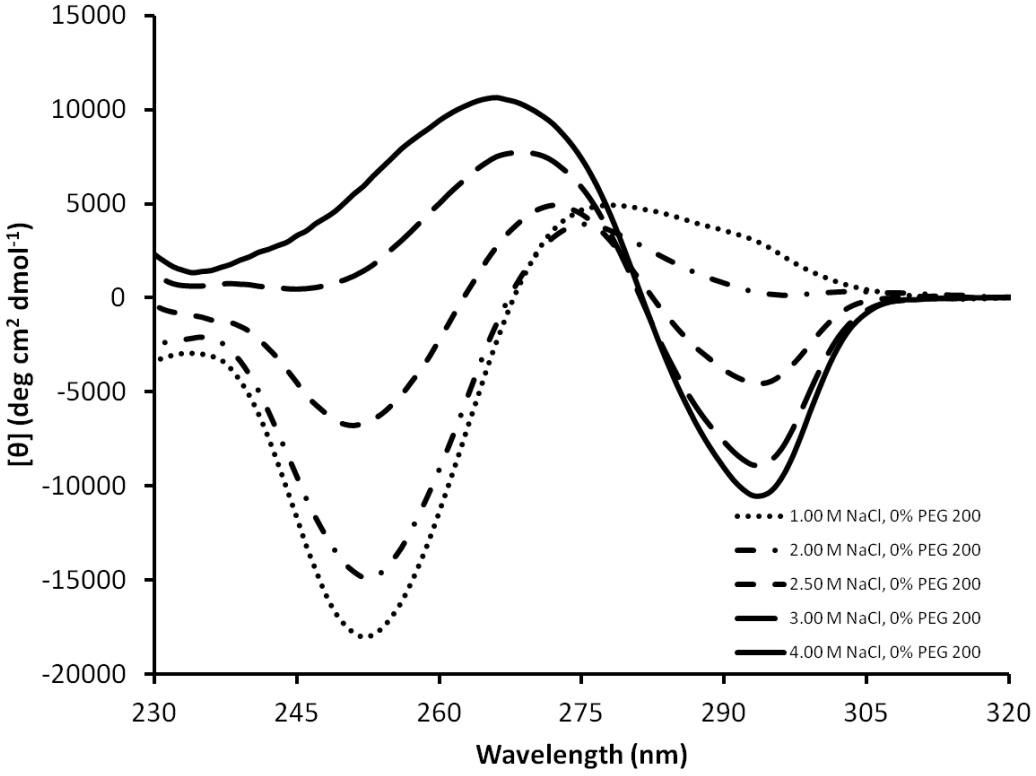


Figure 8. CD spectra of the CG7 sequence at varying salt concentrations in the absence of osmolytes at 5°C and pH 7. *Spectra are the average of three scans and shown from 340 nm to 230 nm. The addition of salt was found to favor Z-DNA formation as can be seen by the increase in magnitude of the negative peak at 293 nm and positive peak at 253 nm.*

The midpoint of the transition for CG7 in the absence of osmolyte was found to be approximately 2.5M NaCl and the complete transition required a minimum concentration of 4.0M NaCl, which is also consistent with studies on CG3[84] (Figure

8). This suggests that the length of the Z-DNA is not a contributing factor to the overall dependence of the data on sodium ion concentration. The obtained information was in agreement with the literature; therefore, we further studied sequences in the presence of PEG 200. [84].

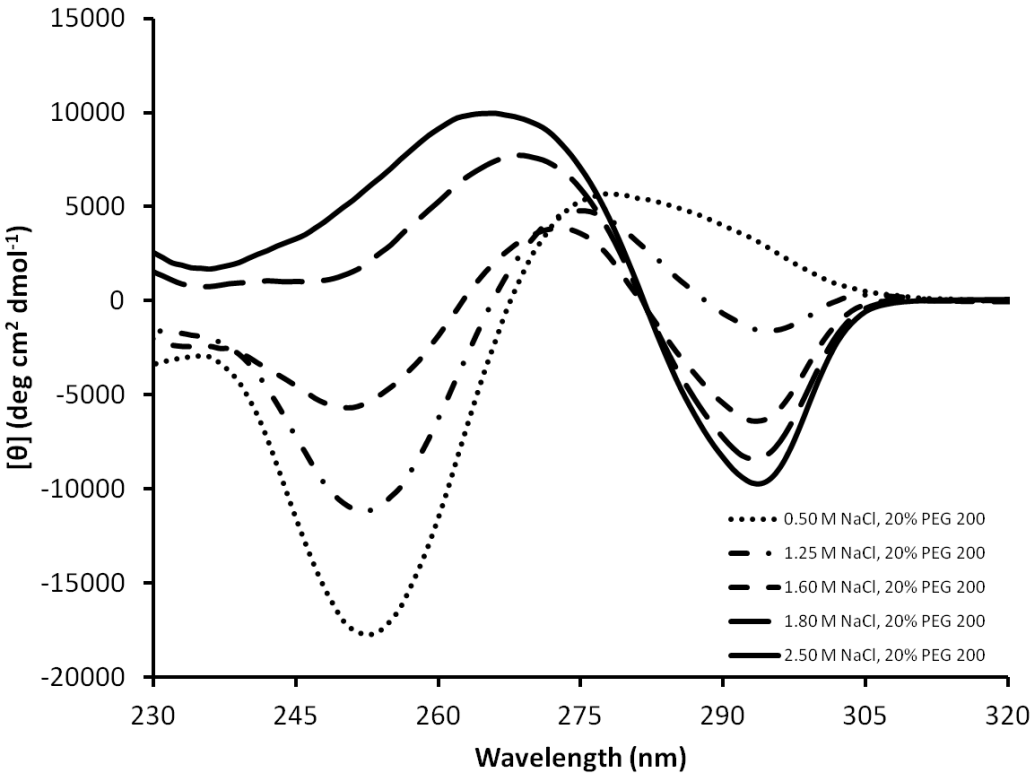


Figure 9. CD spectra of the CG7 sequence at varying salt concentrations in the presence of 20% PEG 200 at 5°C and pH 7. *Spectra are the average of three scans and shown from 340 nm to 230 nm. The addition of salt was found to favor Z-DNA formation as can be seen by the increase in magnitude of the negative peak at 293 nm and positive peak at 253 nm.*

According to the spectra obtained, the addition of 20% PEG 200 reduced the NaCl concentration requirement for complete Z-form stabilization by about half (Figure 9). This indicated that the presence of osmolyte greatly stabilized Z-DNA.

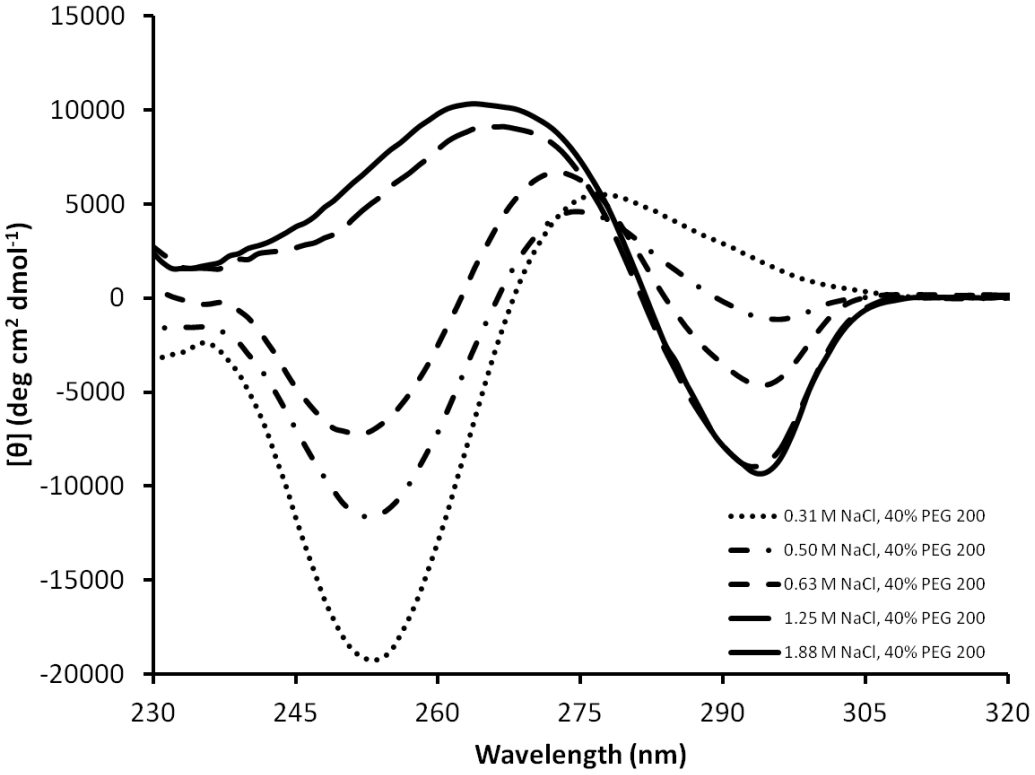


Figure 10. CD spectra of the CG7 sequences at varying salt concentrations in the presence of 40% PEG 200 at 5°C and pH 7. *Spectra are the average of three scans and shown from 340 nm to 230 nm. The addition of salt was found to favor Z-DNA formation as can be seen by the increase in magnitude of the negative peak at 293 nm and positive peak at 253 nm.*

An additional increase in PEG concentration to from 20% to 40% PEG 200 in solution showed even further reductions in NaCl requirement for full B- to Z-form transition (Figure 10). From the CD spectral data obtained of the CG7 sequence in varying concentrations of salt and PEG 200, it was deduced that 1) sequences of pure/ long stretches of CG repeats have the ability to adopt complete Z-form relatively easily with salt and 2) PEG 200 substantially reduces the salt requirement for B- to Z-form transition.

The Z-form DNA has narrower grooves and phosphate groups closer to one another relative to the B-form. This increases steric hindrance and repulsion within the structure. Extrinsic conditions such as high salt concentrations, were therefore, needed to overcome these limitations. Quantitative analysis of Z-DNA folding was done by fitting CD data at 293 according to a unimolecular folding model. The CD signal was fit according to equation 8 to the baselines, ΔG_0, and m-value as a 6 parameter fit using QtiPlot. Additional parameters were then able to be calculated: the C_m or midpoint of the transition, where 50% of the DNA population adopts the B-and 50% adopts the Z- form, the ΔG_x for any salt concentration, as well as Δn, the number of ions bound or dissociated in the transition. The ΔG_0 value indicates the energy required to form Z-DNA at 0 M NaCl. As seen in the Table 1, ΔG_0 decreases by 1.5 kcal/ mol in the presence of 20% PEG 200 and by an additional 1.6 kcal mol^{-1} in the presence of 40% PEG 200. Therefore, increasing the concentration of PEG, even without the presence of cations in solution, lowers the energy requirement for Z-DNA formation a considerable amount. The value of $\Delta\Delta G_0$ indicates the difference between

the ΔG_0 of solutions containing PEG and the ΔG_0 of the solution without osmolyte. The data suggests a direct, linear relationship between the concentration of PEG in solution and the $\Delta\Delta G_0$ value. Comparison of the sample containing osmolyte to the one containing none, the value of $\Delta\Delta G_0$ is linear within error.

The conformational change from B-form to Z-form is coupled with nucleotide dehydration. Due to the release of water molecules during this transition, the Z-form structure is more energetically favored in the presence of osmolyte [90]. The C_m value, calculated using the sum of squares, indicates the concentration of salt required for the B- to Z-form transition. C_m was found to decrease by a factor of two with every 20% increase in PEG concentration. PEG acts as an osmolyte; thus, reduces water activity of the solution itself. It therefore, makes sense that a lower concentration of salt will be needed for transition [91]. Again, ΔC_m shows a predictable, linear relationship.

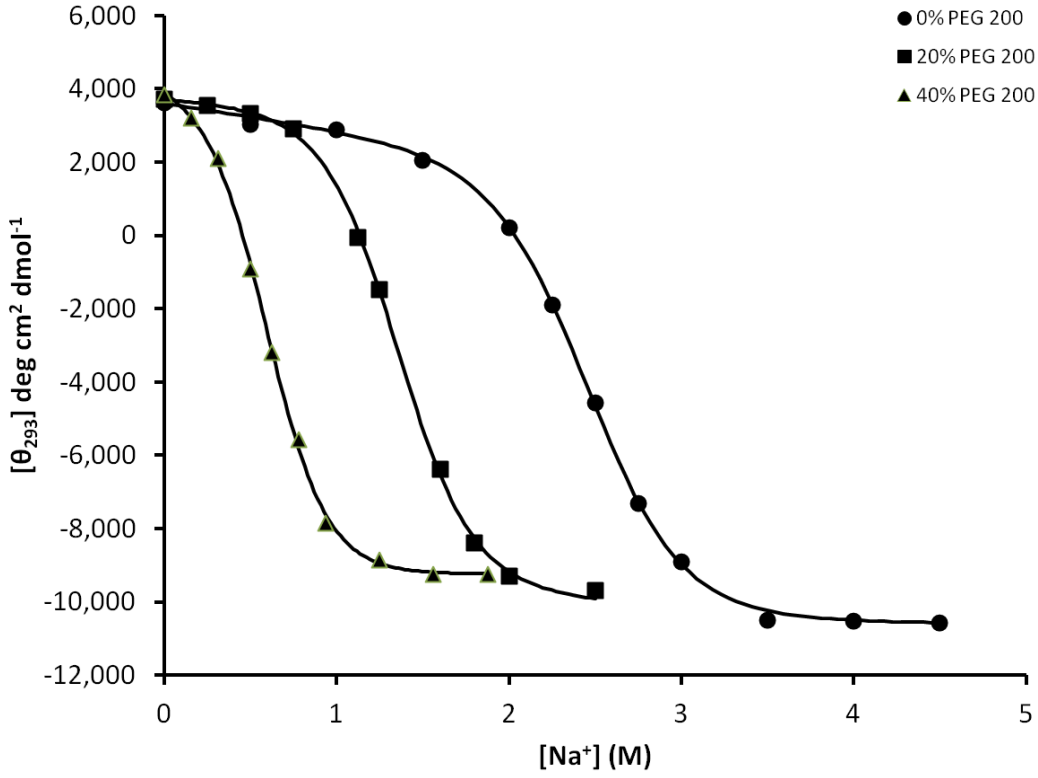


Figure 11. A 6 parameter fit using QtiPlot to the baselines of the CG7 sequence in 0%, 20%, and 40% PEG 200. *Fit done as detailed in Materials and Methods. Points were taken from CG7 scans at 293 nm. Shift in folding curve to lower sodium concentration shows that PEG stabilizes Z-form as well as increases the cooperatively in folding.*

The m value increases from 2.0 ± 0.1 kcal/ mol*M in the presence of no osmolyte to 2.6 ± 0.1 kcal/ mol*M in 20% PEG 200, and 3.2 ± 0.1 kcal/ mol*M in 40% PEG 200 (Table 1). The data shows more cooperative folding over a small range of cationic concentration which means fewer sodium ions are needed for the

transition. Δm also indicates a linear relationship between the presence of PEG and the m-value. This may be qualitatively seen in Figure 11. The cooperativeness in folding increased significantly and is evident in the complete change of conformation from B-to Z- in a reduced range of $[Na^+]$ with increasing PEG concentration.

Based on the ΔG_0 and m values obtained from the fitting of the salt dependent CD data, the ΔG at any salt concentration can be calculated and the equilibrium constant determined from that free energy. The number of sodium ions (Δn) bound or released in a given transition is related to the slope of the plot of ln K vs. ln $[Na^+]$ [84]. For the B to Z transition Δn should be a positive value as the additional sodium is required to shield repulsion of the phosphate groups which are on average closer together in Z-DNA than in B-DNA. It was found previously for the CG3 duplex that an additional 4.3 sodium ions were bound to Z-DNA in the B to Z transition [84]**.** The CG7 contains 2.3 times as many basepairs, and based on this value the number of ions bound in the transition from B to Z DNA should be 9.9 (~10) sodium ions. From the linear fit of the 0% PEG data in Figure 12 the Δn value returned is 9.7, which is in good agreement with the predicted value. It was also determined for CG3 that there was no significant change in Δn upon addition of up to 50% PEG 200, suggesting that PEG 200 does not influence sodium ion binding of Z-DNA as it does in duplex formation from single strands, where upon addition of PEG 200 less sodium is taken up by the duplex. However, our data for 20% and 40% PEG 200 are not in agreement with this observation. We find ΔΔn values of 2.9 and 5.7 for 20% and 40% PEG 200 respectively, which are significant and outside the error of Δn (Table 1). Our findings

suggest that PEG 200 does in fact lower the number of additional sodium ions required for Z-DNA formation and this is one contributing factor for the decreasing C_m with increasing PEG 200.

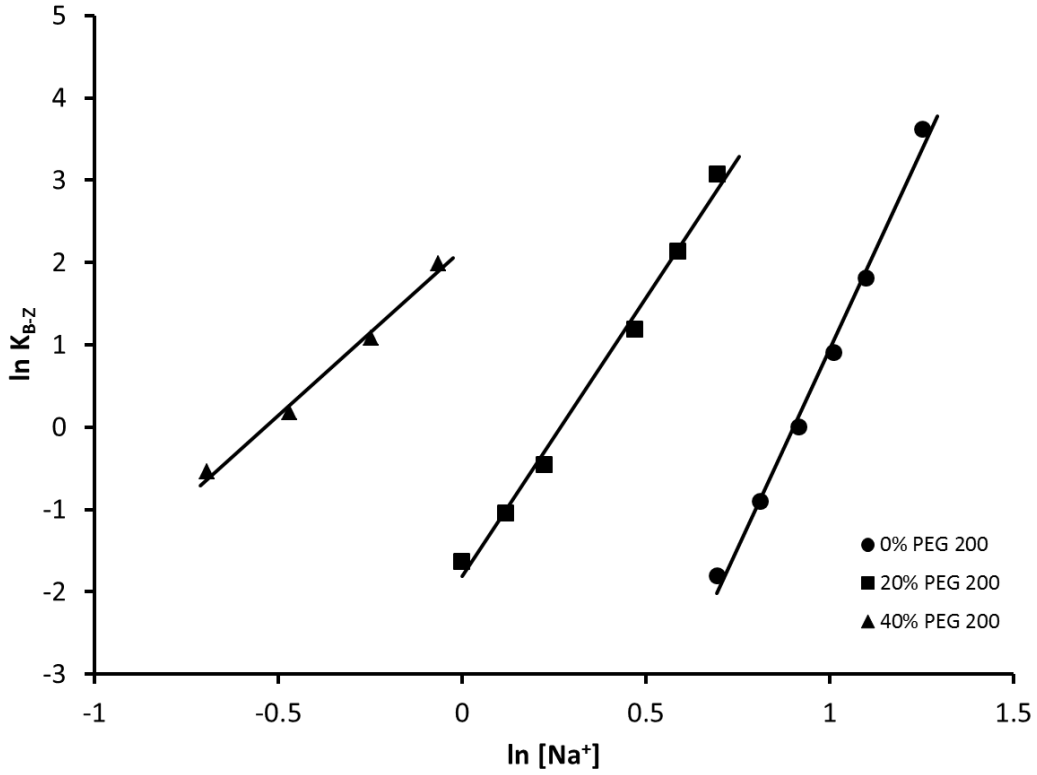


Figure 12. Linear fit of the CG7 at 0% PEG 200, 20% PEG 200, and 40% 200 spectral data. *The change in slope of the linear fits suggests Δn decreases with increasing PEG 200.*

The differences in our observations are primarily due to which regions of the ln K vs. ln [Na^+] curves that were analyzed rather than differences in ΔG or K values

due to comparing thermal versus chemical folding data as evidenced by our agreement of the 0% osmolyte data. For both our CG7 data and the CG3 folding data from the literature, the ln K values utilized in the fitting correspond to sodium ion concentrations that include the folding transition region as determined by our isothermal CD data. However, the same sodium ion concentration range was utilized for the samples containing PEG 200[84], whereas we adjusted the sodium ion concentration to include the folding transition region. From the perspective of chemical folding of the Z-DNA, the prior approach to keep the sodium ion concentration range the same, utilized fewer values from the actual folding transition and more values from the folded baseline as the concentration of PEG is increased. If in our analysis we utilize the approach from the literature and included increasing numbers of folded baseline points and fewer transition points our $\Delta\Delta n$ values approach 0 and our observation matches that from the literature. For example for 40% PEG 200 the $\Delta\Delta n$ is 5.7 whereas if we use values from the baseline as was done in the literature, $\Delta\Delta n$ is 0.6, which is effectively 0 within the error and constraints of the measurement. Thus, we believe that by using ln K values that correspond to the transition region is a more direct chemical comparison and that our data reveal an important mechanism by which osmolytes may influence formation of Z-DNA *in vivo*.

In addition to providing information concerning the number of sodium ions bound in the B to Z transition the equilibrium constants generated using ΔG_0 and m-values can be used to calculate the fraction of molecules in the Z-DNA conformation

at a given [Na^+] according to equation 8. Using 150 mM as a model for *in vivo* monovalent salt conditions, almost no Z-DNA is formed in the absence of PEG 200. However, at 40% PEG 200, 7% of the CG7 molecules adopt the Z conformation without being induced into the conformation by negative supercoiling or being bound to proteins. This result suggests that increase of osmolyte concentrations within the cell may significantly facilitate the formation of functional Z-DNA structure and may make more Z-DNA available to participate in regulation of cellular process and promote stronger interactions with its binding partners. Moreover, while not examined here, a mixed solution of cosolutes may in fact promote even higher percentages of Z-DNA formation. It has been noted that mixed cosolute solutions are more effective at promoting ribozyme function over high concentrations of one osmolyte alone [92].

Substituted Sequences

B-form to Z-form transitions of duplexes of different sequence compositions of d(CG) repeats and d(CA) repeats were studied using circular dichroism (CD) spectroscopy. Substituted sequences selected contained one, three, or complete CG→CA substitutions. The CD spectra gathered for these sequences enabled us to investigate and quantify the effects of PEG 200 on B-to Z-transition.

Spectral Analysis for Substituted Sequences: CG6CA1

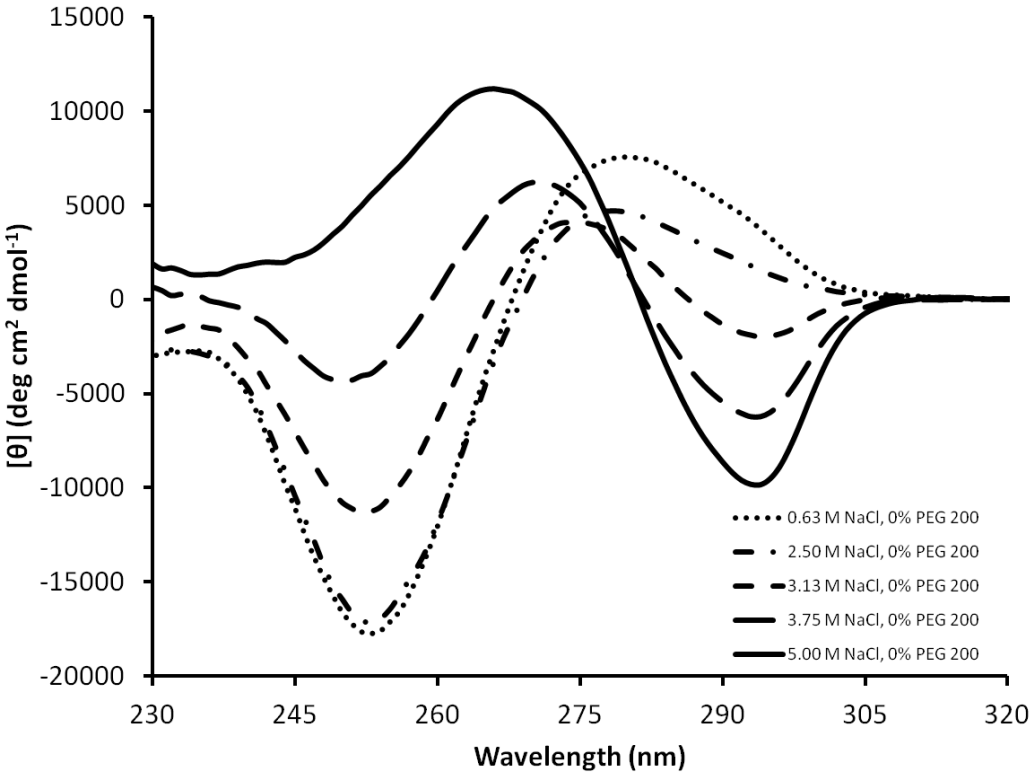


Figure 13. CD spectra of the CG6CA1 sequence at varying salt concentrations in the absence of osmolytes at 5°C and pH 7. *Spectra are the average of three scans and shown from 340 nm to 230 nm. The addition of salt was found to favor Z-DNA formation as can be seen by the increase in magnitude of the negative peak at 293 nm and positive peak at 253 nm.*

In the absence of osmolyte, the formation of the Z form in the CG6CA1 substituted sequence was only seen at large concentrations of NaCl. At 0.63 M NaCl,

the DNA adopted full B-form conformation. After the addition of 5.00 M NaCl to the solution media, full Z-conformation was achieved (Figure 13).

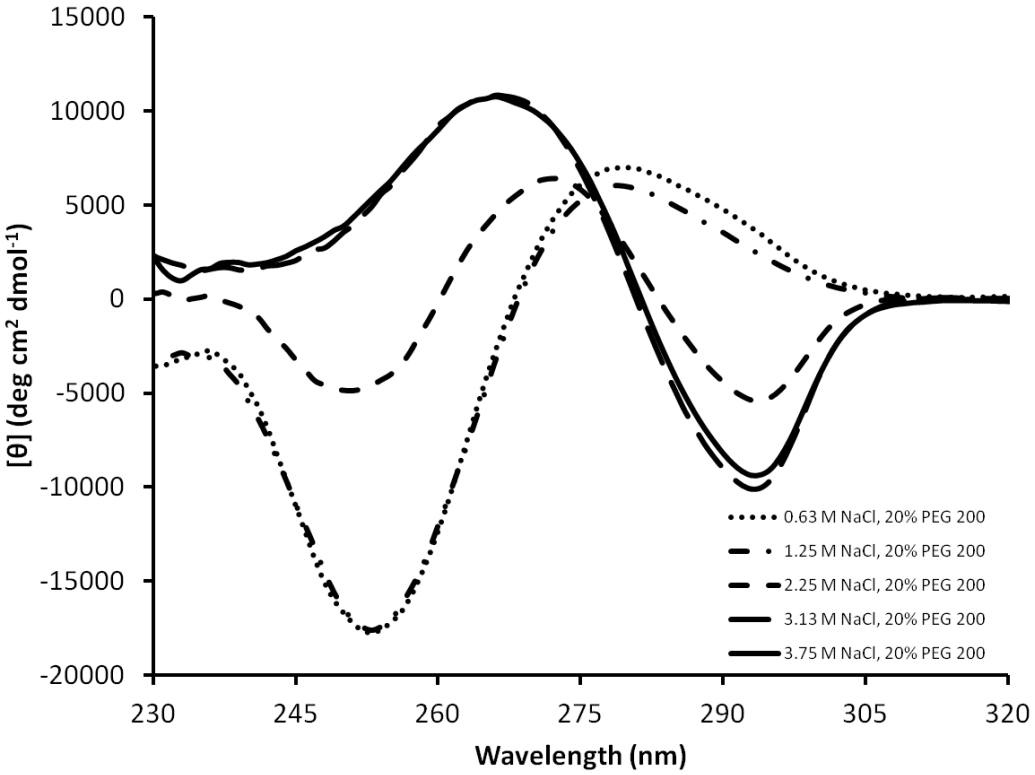


Figure 14. CD spectra of the CG6CA1 sequence at varying salt concentrations in the presence of 20% PEG 200 at 5°C and pH 7. *Spectra are the average of three scans and shown from 340 nm to 230 nm. The addition of salt was found to favor Z-DNA formation as can be seen by the increase in magnitude of the negative peak at 293 nm and positive peak at 253 nm.*

With the introduction of 20% PEG 200 into the solution, the transition from B- to Z-form became much more favorable. By at most 3.13 M NaCl, full Z-

conformation was stabilized, significantly reducing the salt concentration needed for the transition (Figure 14).

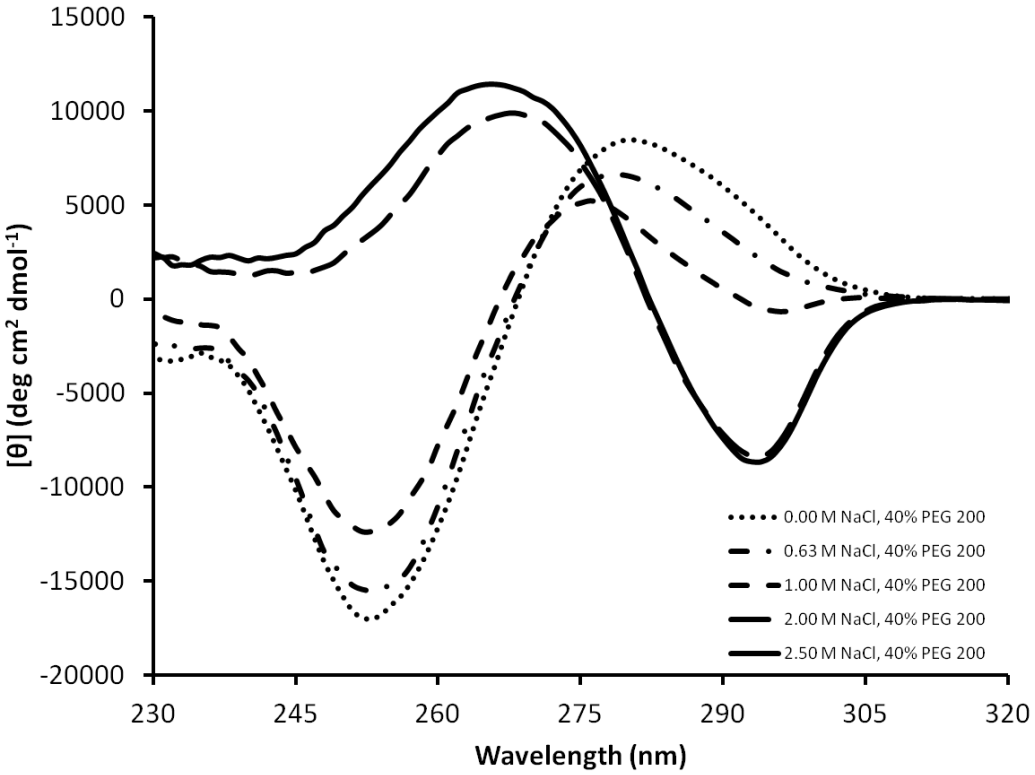


Figure 15. CD spectra of the CG6CA1 sequence at varying salt concentrations in the presence of 40% PEG 200 at 5°C and pH 7. *Spectra are the average of three scans and shown from 340 nm to 230 nm. The addition of salt was found to favor Z-DNA formation as can be seen by the increase in magnitude of the negative peak at 293 nm and positive peak at 253 nm.*

In the presence of 40% PEG 200, this transition became substantially more favorable. The duplex adopted full Z-form conformation at no more than 2.00 M

NaCl, further demonstrating the role of osmolytes in Z-DNA stabilization (Figure 15).

Spectral Analysis for Substituted Sequences: CG4CA3

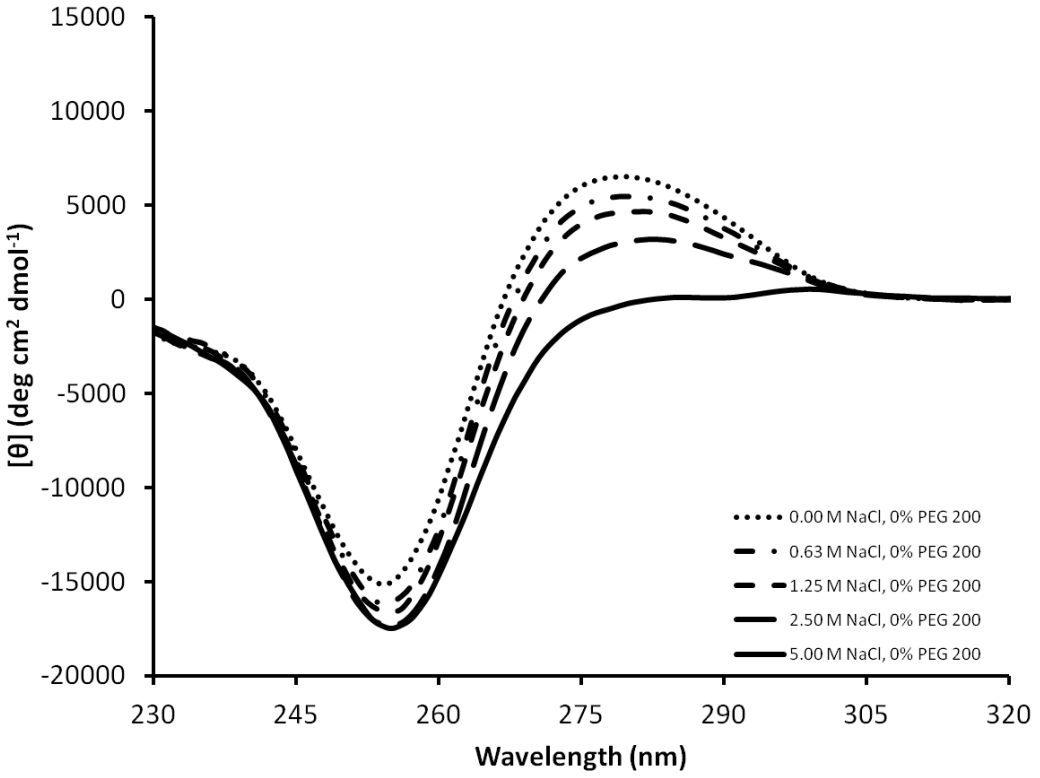


Figure 16. CD spectra of the CG4CA3 sequence at varying salt concentrations in the absence of osmolytes at 5°C and pH 7. *Spectra are the average of three scans and shown from 340 nm to 230 nm. The addition of salt for this sequence was found to favor Z-DNA formation; however, full Z-conformation was not stabilized even in the presence of 5 M NaCl.*

CG4CA3 spectra revealed that the presence of 5M NaCl was not enough to push the duplex to fully adopt the Z-form; however, we did observe a steady trend towards Z-conformation at increasing salt concentration (Figure 16).

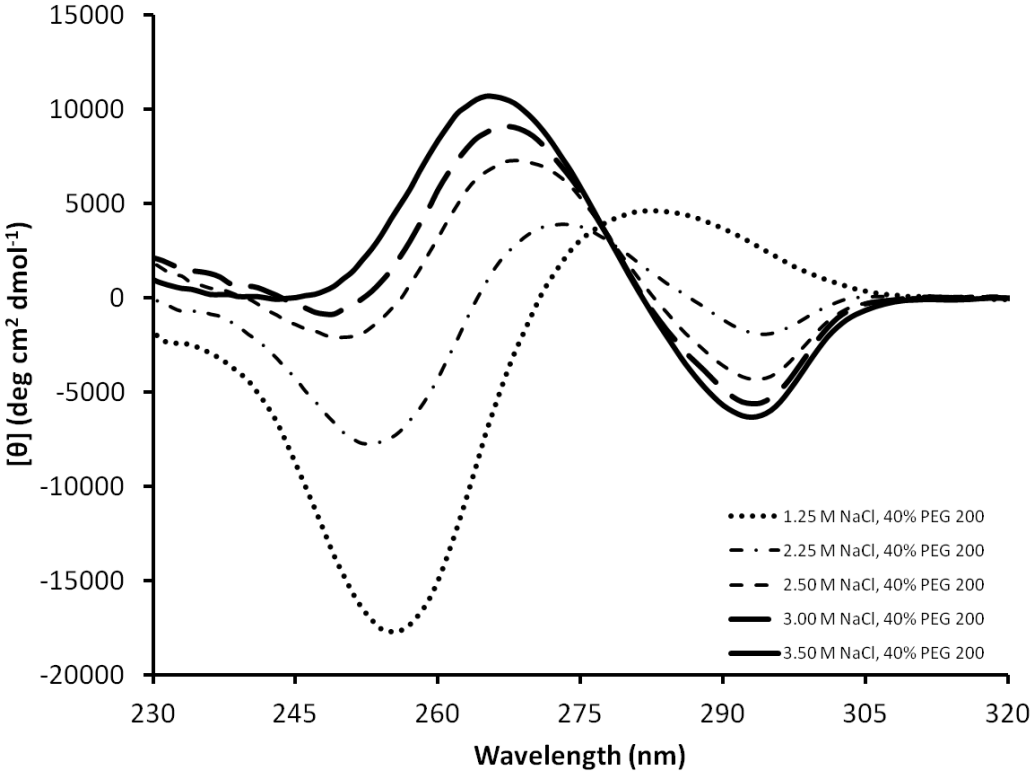


Figure 17. CD spectra of the CG4CG3 sequence at varying salt concentrations in the presence of 40% PEG 200 at 5°C and pH 7. *Spectra are the average of three scans and shown from 340 nm to 230 nm. The addition of salt was found to favor Z-DNA formation as can be seen by the increase in magnitude of the negative peak at 293 nm and positive peak at 253 nm.*

This trend is amplified but complete Z-conformation was only observed at 3.50 M salt in the presence of 40% PEG 200 (Figure 17). At 0 M NaCl, the expected complete B-form spectrum is seen. The sequence is about halfway folded around 2.25 M (relative to spectrum at 3.50 M NaCl), which is considerably high especially when compared to the CG7 sequence which was already 50% Z-conformation at around 2.50 M even without the presence of PEG 200.

Spectral Analysis for Substituted Sequences: CA7

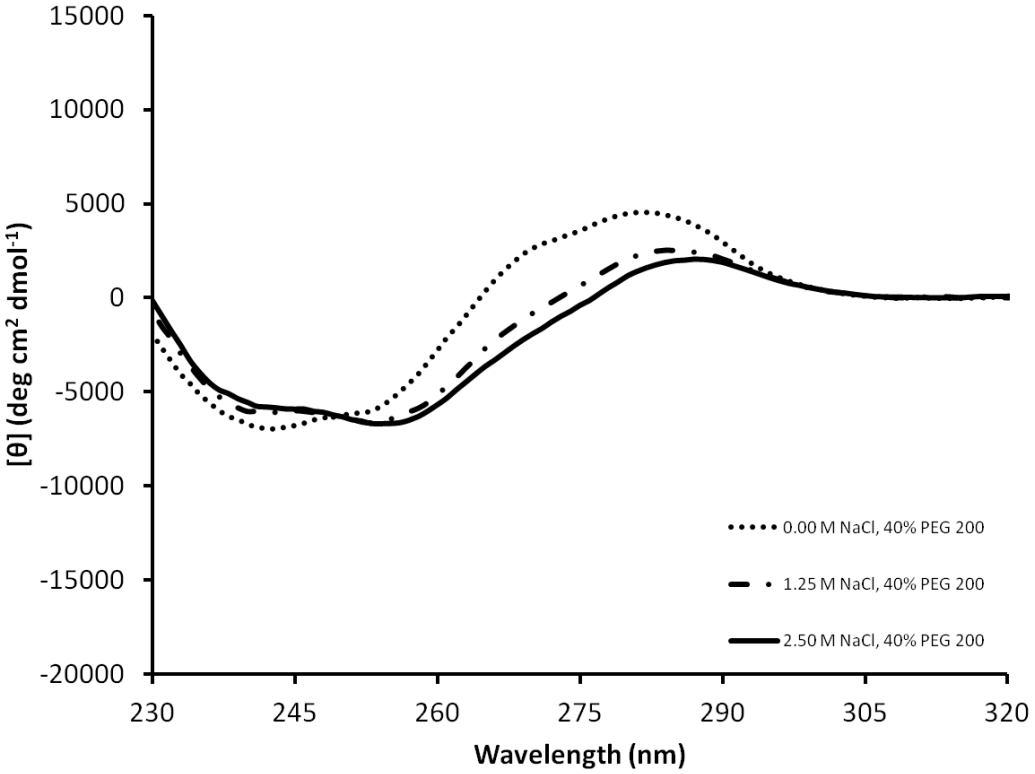


Figure 18. CD spectra of the CA7 sequence at varying salt concentrations in the presence of 40% PEG 200 at 5°C and pH 7. *Spectra are the average of three scans and shown from 340 nm to 230 nm. The addition of salt for this sequence was found to have very little to no effect on Z-DNA formation even at 2.50 M NaCl in 40% PEG 200.*

Compared to the CG7 sequence (Figure 10), spectral data of the CA7 duplex revealed it was virtually unaffected by PEG (Figure 18). Even at high salt (2.5 M

NaCl) conditions in addition to the presence of 40% PEG 200, there was very little shift towards Z-conformation, indicating that Z-DNA formation is also largely dependent on sequence composition. Study on this sequence was abandoned due to these negligible effects. Qualitative analysis of the "broken" sequences revealed that with enough substitutions, a solution media containing osmolytes alone is not sufficient to induce Z-formation without the aid of other elements. We suspect that the presence of divalent cations such as Mg^{+} may aid in transition to allow B-to Z folding.

Folding Parameters for CA-substituted Sequences

The association of cationic molecules is required to aid in the folding of otherwise fairly unfavorable sequences. The understanding of the energetic contribution of cationic interaction in nucleotide folding events is largely reliant on their binding affinities. Cations, such as sodium stabilize nucleotides in the folded conformation [89]. The presence of NaCl, as seen in the substituted sequence study above, may not be enough to fully fold the DNA into the Z-conformation; thus requiring the aid of osmolytes to induce the transition. We examined the thermodynamic and structural parameters of folding of these sequences to further understand and quantify sequence dependence in B to Z-transitions.

Table 1. Thermodynamic and Structural Parameters for Folding of Z-DNA Duplexes

Sequence	Solution Conditions	ΔG_0 (kcal mol^{-1})	C_m (M)	m (kcal mol^{-1} M^{-1})	Δn
CG7[a]	0% PEG 200	5.0 ± 0.1	2.5 ± 0.1	2.0 ± 0.1	9.7 ± 0.4
CG7	20% PEG 200	3.5 ± 0.2	1.3 ± 0.1	2.6 ± 0.1	6.8 ± 0.3
CG7	40% PEG 200	1.9 ± 0.1	0.6 ± 0.1	3.2 ± 0.1	4.0 ± 0.3
CG6CA1[b]	0% PEG 200	5.8 ± 0.3	3.6 ± 0.3	1.5 ± 0.1	9.3 ± 0.4
CG6CA1	20% PEG 200	4.8 ± 0.4	2.2 ± 0.5	2.2 ± 0.2	7.4 ± 0.3
CG6CA1	40% PEG 200	2.9 ± 0.2	1.2 ± 0.3	2.5 ± 0.2	5.4 ± 0.3
CG4CA3[a]	40% PEG 200	6.5 ± 0.4	2.2 ± 0.4	2.9 ± 0.2	10.5 ± 0.4

Sequence	Solution Conditions	$\Delta\Delta G_0$[c] (kcal mol^{-1})	ΔC_m (M)	Δm (kcal mol^{-1} M^{-1})	$\Delta\Delta n$
CG7	0% PEG 200				
CG7	20% PEG 200	-1.5 ± 0.2	-1.2 ± 0.1	0.6 ± 0.1	-2.9 ± 0.5
CG7	40% PEG 200	-3.1 ± 0.1	-1.9 ± 0.1	1.2 ± 0.1	-5.7 ± 0.5
CG6CA1[a]	0% PEG 200				
CG6CA1	20% PEG 200	-1.0 ± 0.5	-1.4 ± 0.6	0.7 ± 0.2	1.9 ± 0.3
CG6CA1	40% PEG 200	-2.9 ± 0.4	-2.4 ± 0.4	1.0 ± 0.2	3.9± 0.3

[a] Values are from nonlinear curve fitting, [b] Values are from using LEM, [c] Values are relative to those of 0% PEG 200

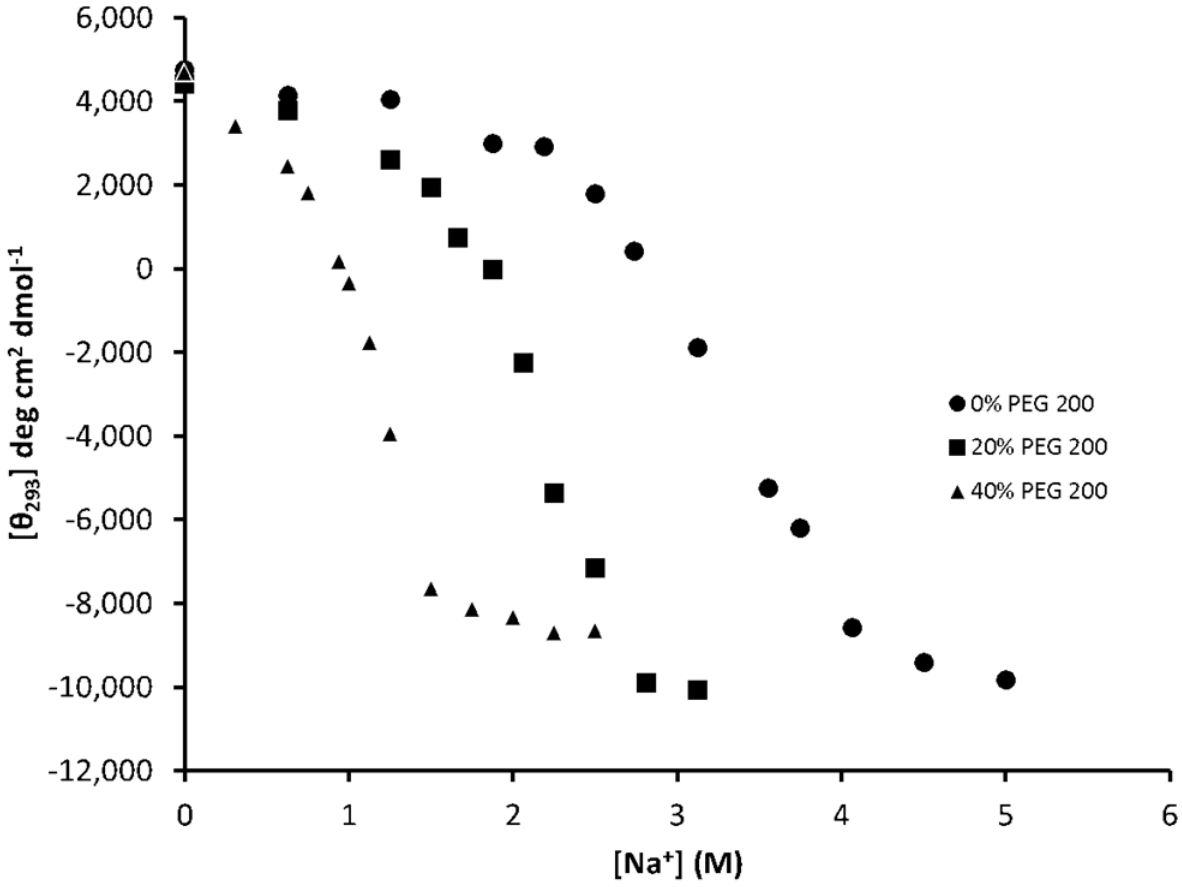


Figure 19. Compiled CD spectral data at 293 nm for CG6CA1 substituted sequence at varying salt concentration in the presence of 0%, 20%, and 40% NaCl concentration. *The plot qualitatively indicates that B- to Z-DNA transition of the CG6CA1 substituted sequence is much less cooperative compared to the CG7 sequence.*

Qualitative analysis of the compiled data at 293 nm reveal that although the structure does adopt the Z-form, the transition is significantly less cooperative than its CG7 counterpart as it occurs over a broader range of sodium ion concentrations

(Figure19). Quantitatively, this was supported by the smaller m-value yielded by the substituted sequence (2.5± 0.2 in 40% PEG 200, relative to 3.2± 0.1 at same conditions for CG6CA1 and CG7, respectively) as well as a larger C_m (1.2 ± 0.3 for CG6CA1 at 40% PEG 200, relative to 0.6 ± 0.1 at same conditions for CGCA1 and CG7, respectively) in comparison to the parameters calculated for the CG7 sequence (Table 1). This observed lower m-value in addition to the sodium ion concentrations with 1 M greater C_M value suggested a decreased propensity for forming Z-DNA.

Table 2. Folding Parameters for CG7 and CG4CA3 from Nonlinear Curve Fitting and LEM

Nonlinear Curve Fitting					
Sequence	**Solution Conditions**	**ΔG_0 (kcal mol^{-1})**	**C_m (M)**	**m (kcal mol^{-1} M^{-1})**	**Δn**
CG7	0% PEG 200	5.0 ± 0.1	2.5 ± 0.1	2.0 ± 0.1	9.7 ± 0.4
CG7	20% PEG 200	3.5 ± 0.2	1.3 ± 0.1	2.6 ± 0.1	6.8 ± 0.3
CG7	40% PEG 200	1.9 ± 0.1	0.6 ± 0.1	3.2 ± 0.1	4.0 ± 0.3
CG4CA3	40% PEG 200	6.5 ± 0.1	2.2 ± 0.1	2.9 ± 0.2	10.5 ± 0.4
LEM					
Sequence	**Solution Conditions**	**ΔG^a (kcal mol^{-1})**	**C_m (M)**	**m^a (kcal mol^{-1} M^{-1})**	**Δn**
CG7	0% PEG 200	4.7 ± 0.1	2.5 ± 0.1	1.9 ± 0.1	9.9 ± 0.4
CG7	20% PEG 200	3.4 ± 0.2	1.3 ± 0.1	2.7 ± 0.1	6.9 ± 0.3
CG7	40% PEG 200	2.2 ± 0.1	0.6 ± 0.1	3.4 ± 0.2	4.0 ± 0.2
CG4CA3	40% PEG 200	6.9 ± 0.4	2.2 ± 0.4	3.1 ± 0.2	11.1 ± 0.4

[a] Values are from nonlinear curve fitting

It can be observed in Figure 19 that addition of PEG 200 shifts the folding transitions to lower sodium ion values. Obtaining direct nonlinear fits to this data is problematic; however, as the low sodium baseline becomes increasing sloped and the high sodium baselines being short or sloped as well. It is notable that this large increase in slope of the lower baseline was not observed for CG7 and was also not

observed in CG4CA3 in 40% PEG 200. This does permit for the possibility of the presence of an additional state whereby the terminal CG repeats are attempting to fold into the Z-form, but the nucleotides surrounding the only CA need additional sodium to adopt the full Z-form structure. As PEG 200 is added this could sufficiently stabilize the Z-form, but preferentially assist the CG repeats over the CA. A 3 state fit was applied to the data but due to a lack of sufficient parameters for the potential intermediate, the fits were unsuccessful in modeling the data. In the discussion of the CG6CA1 the focus will be on the steeper of the two transitions out of which the Z-DNA does fully form. In order to mitigate the fitting issues of the data to analyze this transition LEM was used to obtain folding parameters for CG6CA1. To ensure that LEM is a reliable alternative, it was used to analyze the CG7 data, and the analysis yielded the same parameters as did the nonlinear curve fitting (Table 2). The linear fits from LEM analysis of the CG6CA1 are shown in Figure 20 for reference, and the lnK vs. ln $[Na^+]$ plot resulting from this analysis is shown in Figure 21.

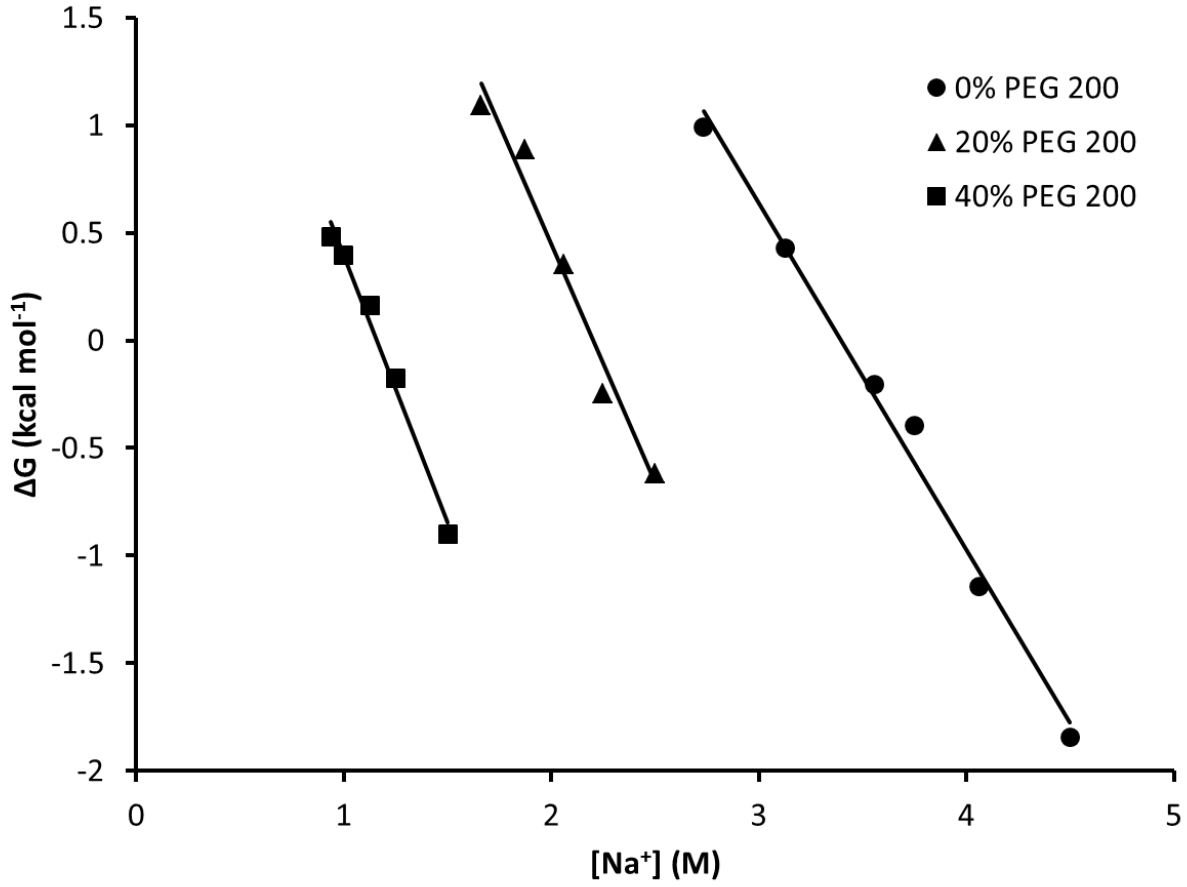


Figure 20. Fitting of CG6CA1 CD data at 293 nm with varying salt concentration using LEM (Linear Extrapolation Method). The linear fit provides both the free energy of folding at no osmolyte as well as the associated m-value.

The ΔG_0 for CG6CA1 are less favorable than their CG7 counterparts (Table 1) under all conditions, as more sodium is required to fold the structure (C_m) and the transitions are less cooperative (m-value). However, the associated changes in these parameters with the addition of PEG 200 are the same for CG7 and CG6CA1. For

example, the $\Delta\Delta G_0$ values upon addition of 40% PEG 200 are -3.1 ± 0.1 and -2.9 ± 0.4 kcal mol^{-1} for CG7 and CG6CA1, respectively. This result suggests that the CG and CA repeats are stabilized to a similar extent by PEG 200 in the major transition observed in the folding curves.

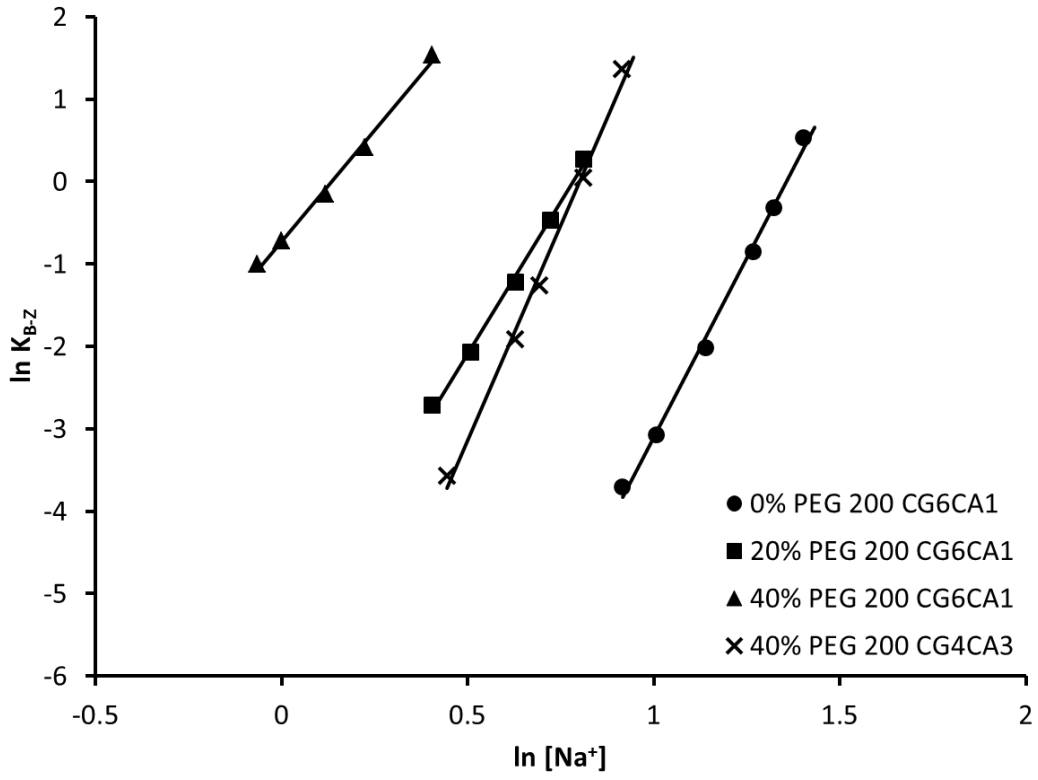


Figure 21. Plot of $\ln K_{B-Z}$ vs $\ln[Na^+]$. *The change in slope of the linear fits suggests Δn decreases with increasing PEG 200*

The number of sodium ions Δn bound in this transition though differs between CG7 and CG6CA1 as PEG 200 is added with the mutant requiring more ions be

bound (4.0 vs. 5.4, Table 1). This is evidenced in the ΔΔn values for increasing PEG 200 (Table 1). The ΔΔn for an increase in 20% PEG is closer to 3 ions for CG7 where it is closer to 2 for the CG6CA1. An increased number of CA repeats destabilizes the Z-DNA structure and requires an increase in bound sodium as discussed for the CG4CA3 mutant below.

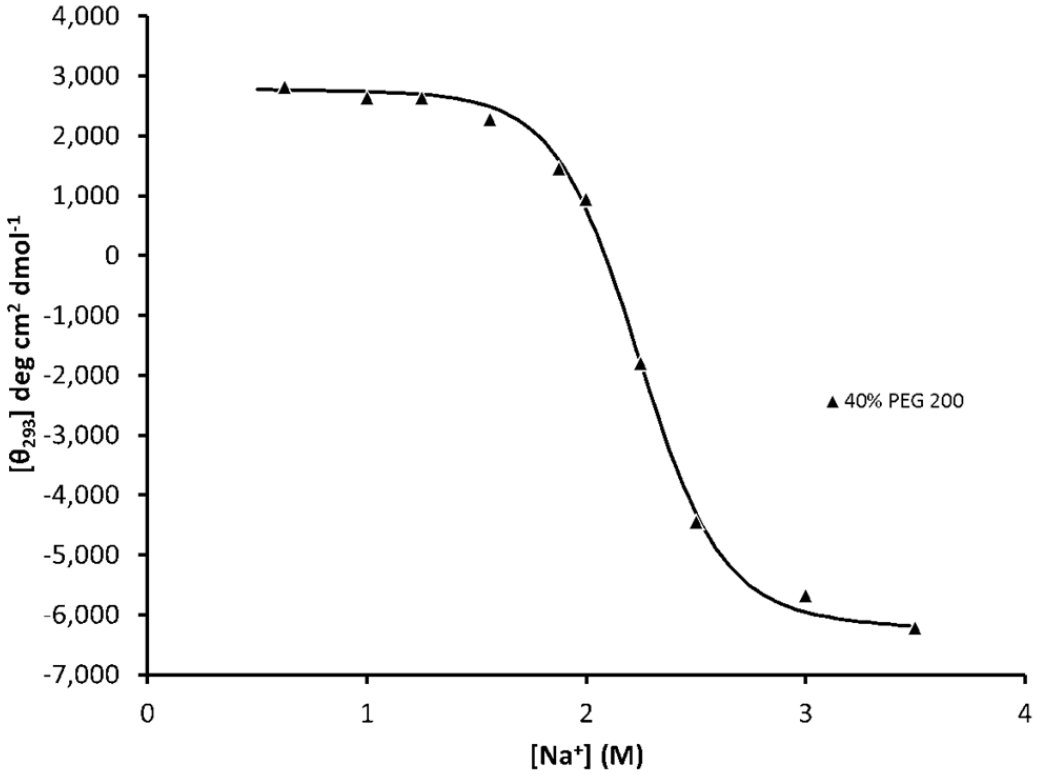


Figure 22. 6 parameter fit of CD spectra at 293 nm for the CG4CA3 substituted sequence at varying salt concentrations in the presence of 40% PEG 200. *Fit done as detailed in Materials and Methods. This plot shows the salt dependence of the CG4CA3 substituted sequence in the transition from B- to Z-DNA.*

As discussed previously the CG4CA3 sequence does not fold without osmolytes even in 5 M NaCl, and to our knowledge this is the first parameterization of its folding *in vitro* with osmolytes, and without proteins present to fold the structure. While some folding was induced by the presence of 20% folding we determined that a fully folded population would require a higher sodium concentration than feasible to remain soluble in 20% PEG 200 so we utilized the 40% PEG condition. Interestingly, our CD data for this sequence was able to be fit using both nonlinear curve fitting (Figure 22) and LEM, giving similar results within error (Table 2).

We note that the lower baseline is significantly less sloped for this sequence, behaving more like CG7 than CG6CA1. This result does lend support to the idea that there is an additional state present in CG6CA1, because inclusion of the additional CA's in the CG4CA3 in alternating fashion restores a level of homogeneity to the structure, such that a cooperative folding transition is restored. In 40% PEG 200, the CG4CA3 has a stability and C_m value similar to that of CG7 without PEG, but an elevated m-value similar to that of CG7 with 40% PEG 200. This indicates that PEG 200 interacts with CG and CA in a similar fashion to induce Z-DNA structure related changes, but cannot quite offset the higher Na^+ requirements needed by the CA repeats. In fact the Δn for CG4CA3 is similar to that for CG7 in the absence of PEG, and is 2.5X that needed by CG7 in 40% PEG 200.

In spite of the folding deficiencies of the CA containing mutants, PEG 200 does promote folding of the sequences, and for CG4CA3 induces complete Z-DNA

folding when the sequence fails to induce at best 30% of the molecules to adopt Z-form even in 5 M NaCl. This result shows not only that changes in osmolyte concentration in response to stress can influence the folding of Z-DNA from typical CG rich sequences, but that it could rescue the folding of sequence that include a significant percentage of CA repeats.

Summary I

The conformational flexibility and stability of nucleic acids are largely dependent upon interactions with the aqueous surrounding [93 94]. DNA possesses a highly favorable free energy of solvation for its polar groups on the surface of its structure. Fundamentally, hydrophobic groups are preferred on the interior and more hydrophilic on the exterior of the DNA and thus, become a major driving force for folding. The structure of the Z-form DNA however, exposes these hydrophobic regions and therefore, will be stabilized in water poor environments. That is why the transition from B-to Z-conformation was much more favorable and cooperative with increasing salt or osmolyte concentrations, as seen in the spectral data as well as the thermodynamic parameters generated. We determined from our study that the presence of PEG 200, will increases the cooperativeness of folding as well as lessens the salt requirement for transition to Z-form. Sequence; however, may have a critical role in ease of transitioning. It was seen that high cytosine and guanine content may not be the only determining factor in stabilizing the Z-form as data yielded support significant sequence dependence.

Specific Aim II:

Although the helical structures of DNA and RNA are well characterized, hybrid duplexes are not as straight forward. As stated earlier, it is not a simple average of the two strands but is dependent upon sequence and environmental conditions. Small cosolutes such as osmolytes, affect nucleic acid processes by introducing chemically favorable or unfavorable interactions with exposed or buried surfaces of the structure. Large cosolutes like crowding agents exclude volume and affect processes of conformational change. We will use PEG 200 for the osmolyte and PEG 8000 for the crowder to investigate how they affect hybrid duplex conformation. This section will provide qualitative analyses of structural changes undergone by the duplex in varying osmolyte conditions.

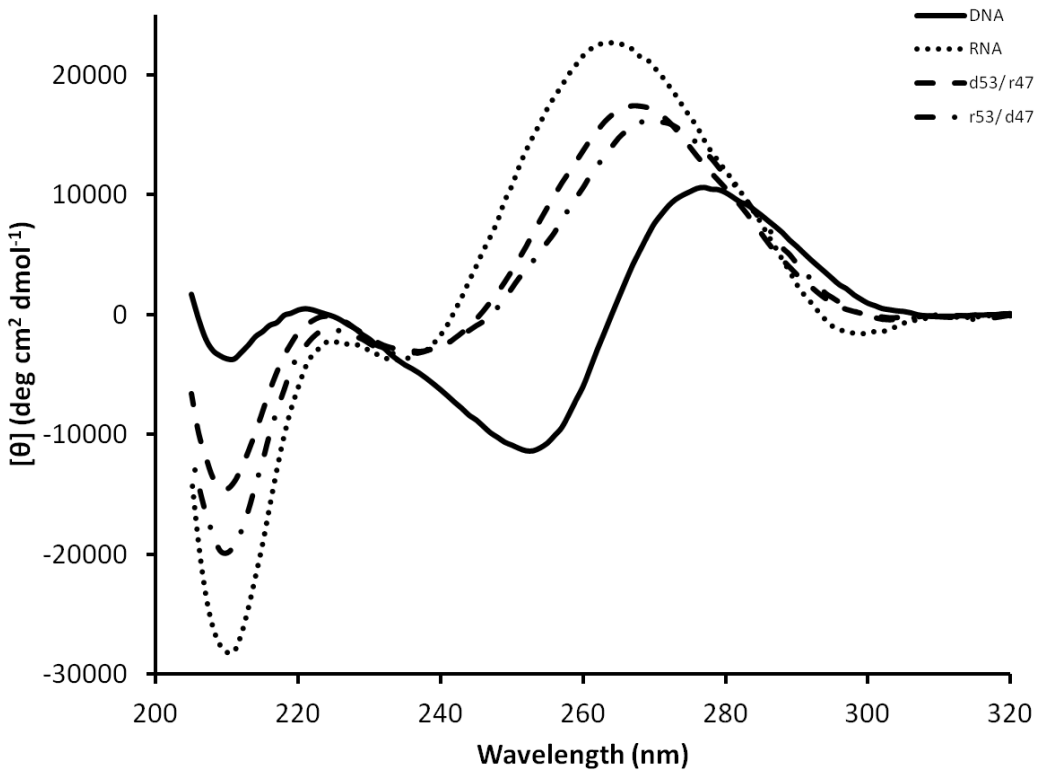


Figure23. Spectra of control and hybrid duplexes in the absence of osmolytes at 20°C and pH 7. *Spectra are the average of three scans and shown from 320 nm to 200 nm. Information on the stacking interactions and helical conformation of the nucleic acids may be determined by the signal at 260-280 nm and 200-260 nm, respectively. The characteristic peaks of each nucleic acid form may be seen within these regions.*

The signal at 260-280 nm provides information about stacking interactions and the signal at 200-260 nm provides information about helical conformation. Characteristic peaks for the B form (DNA) include a positive peak at ~280 nm, a

moderately negative peak at ~250 nm, and a Less negative peak at ~210 nm. Characteristic peaks of the A-form (RNA) include a positive peak at ~260 nm, a slightly negative peak at ~235 nm, and a significantly negative peak at ~210 nm. The hybrids fall somewhere in between the two pure DNA and RNA spectra with the d53/r47 hybrid spectra closer to the DNA, and the r53/d47 closer to the RNA at the 200-260 nm region of the graph. This indicates that the hybrid containing a 53% purine RNA strand favors A-form, while the hybrid containing a 53% purine DNA strand favors B-form.

Effect of Osmolytes on Hybrid Helical Structure

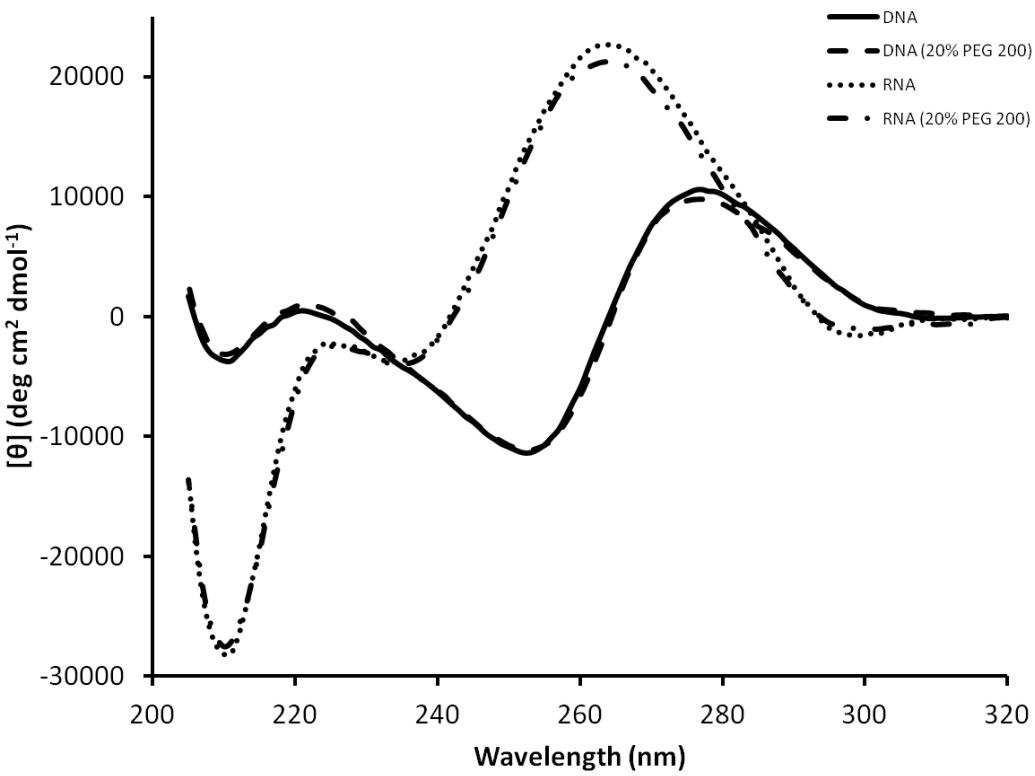


Figure 24. Spectra of DNA and RNA in the presence of PEG 200 at 20°C and pH 7. *Spectra are the average of three scans and shown from 320 nm to 200 nm. DNA and RNA was found to be unaffected by the presence of PEG 200.*

In the presence of PEG 200, the DNA and RNA spectra remain unchanged indicating that osmolytes have little to no effect on pure A- and B-DNA conformation (Figure 24).

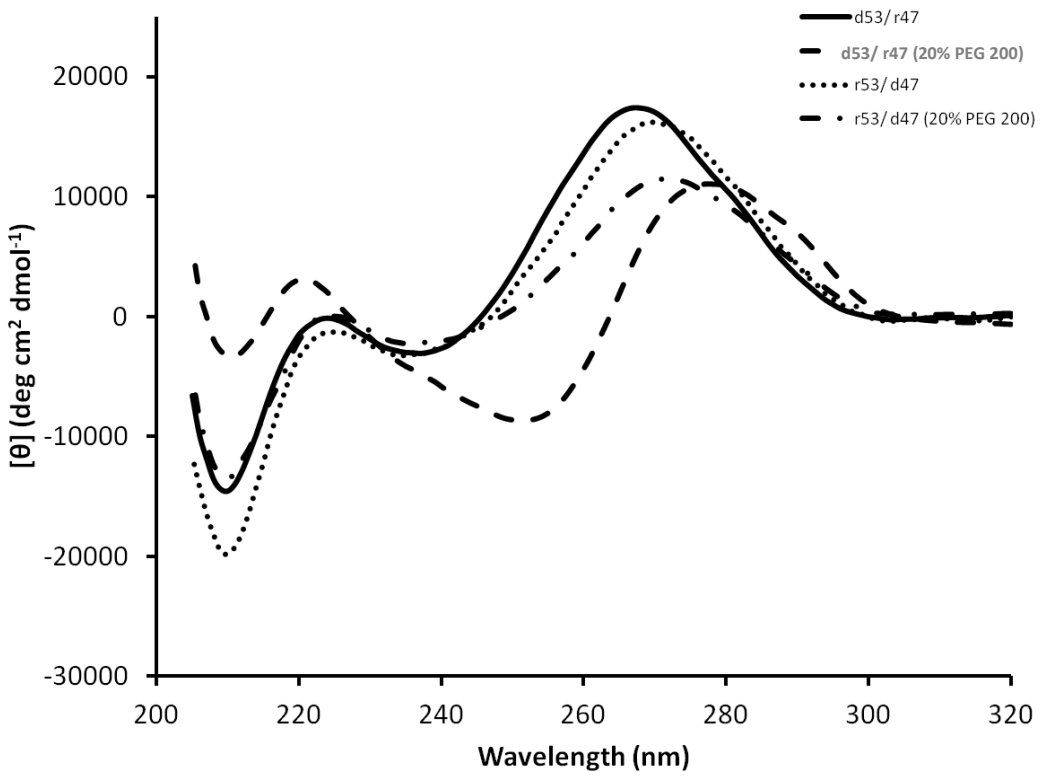


Figure 25. Spectra of hybrids in the absence and presence of PEG 200 at 20°C and pH 7. *Spectra are the average of three scans and shown from 320 nm to 200 nm. In the presence of PEG 200, hybrids were found to favor B-DNA conformation as can be seen by the upward shift at 210 nm and downward shift in the 260-280 nm region.*

This figure of combined spectra was generated to determine the effects of osmolytes on the hybrid duplex (Figure 25). The spectrum for the d53/ r47 hybrid showed a pronounced upward shift at 210 nm and a downward shift in its 260-280 nm region. The r53/ d47 hybrid also showed an upward shift at 210 nm and a downward shift in its 260-280 nm region, although not as strong of a change relative to the d53/

r47 hybrid. Overall, the hybrid spectra of the hybrids shifted toward B-form, especially the d53/ r47 spectrum when PEG 200 was added. This conformational switch to more B-form like structures is suspected to be due to requirement of fewer water molecules for solvation in hybridizing the B-form helix; therefore, may be favored in the water poor environment of 20% osmolyte [76].

Effect of Crowding Agents on Hybrid Helical Structure

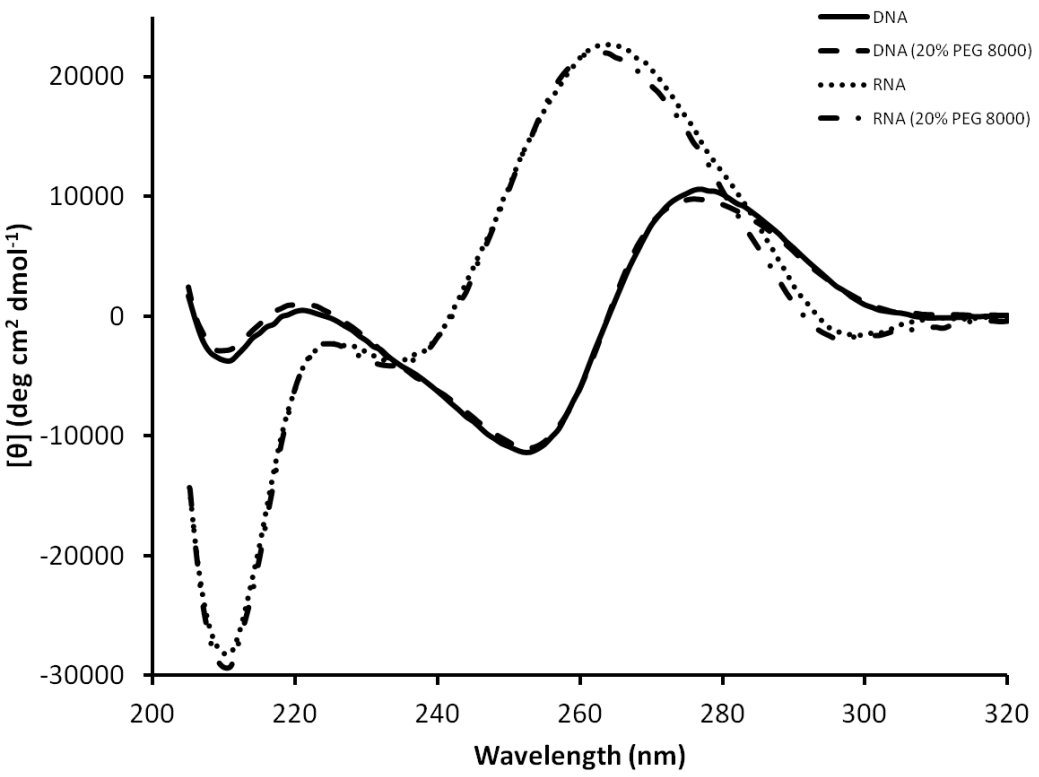


Figure 26. Spectra of DNA and RNA in the presence of PEG 8000 at 20°C and pH 7. *Spectra are the average of three scans and shown from 320 nm to 200 nm. DNA and RNA was found to be unaffected by the presence of PEG 8000.*

Similar to what was seen in the PEG 200 spectra, the DNA and RNA spectra remain unchanged in the presence of PEG 8000 indicating that crowding agents have little to no effect on pure A- and B-form conformation (Figure 26).

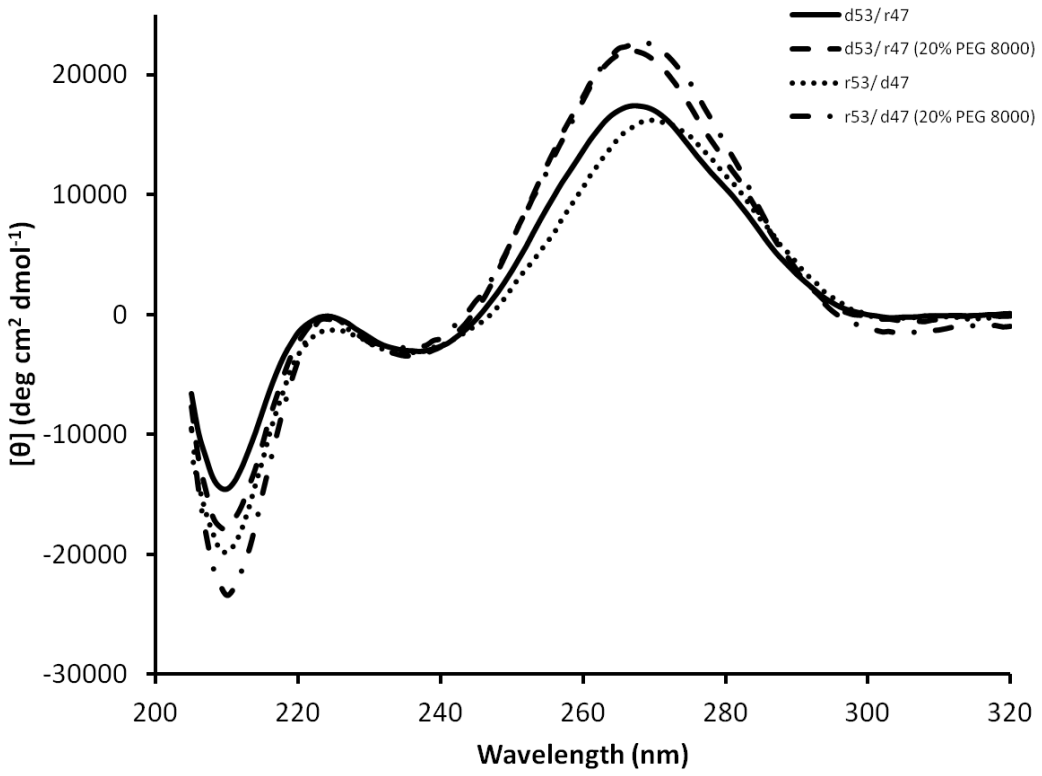


Figure 27. Spectra of hybrids in the absence and presence of PEG 8000 at 20°C and pH 7. *Spectra are the average of three scans and shown from 320 nm to 200 nm. In the presence of PEG 8000, the hybrids were found to favor A-DNA conformation as can be seen by the downward shift at 210 nm and upward shift in the 260-280 nm region.*

In the presence of PEG 8000, the hybrid spectra shifted toward the more compact A-form. Spectrum of the d53/ r47 hybrid reveals a downward shift at 210 nm and an upward shift in its 260-280 nm region. For the r53/ d47 hybrid, there is also a downward shift at 210 nm and an upward shift in its 260-280 nm region

(Figure 27), indicative of the formation of a more A-form like structure. Because large cosolutes such as PEG 8000 occupy volume and reduce the amount of available space in solution, the increased concentration of this physically large species contributes to the excluded volume effect. Due to this, the nucleic acid is forced to adopt a more compact state [79].

Summary II

From the data collected, there is evidence that macromolecular crowders influence the conformation of hybrid duplexes in solution and thus, also may affects its structure, stability, and biological functions in the living cell. CD spectra of the hybrids gathered at various solution conditions of osmolyte and crowding agents revealed that these duplexes trend towards the B-form in the presence of osmolytes, while the presence of crowders swayed the hybrids towards the A-form. This finding was in agreement with studies using the hybrid duplex which showed destabilization upon formation in the presence of osmolytes, which appeared to be due to decreased water activity [76]. Because folding of the B-form requires less solvation for formation, it understandable why this form would be favored in the water poor environment of 20% PEG 200. The presence of molecular crowders that occupy larger volumes of the already constricted cellular environment will encourage the more compact A-form due to excluded volume effects. The form adopted by the hybrids is what is most enthalpically and entropically favored in the specific solution condition is in. That is why further experiments must be conducted to fully understand the effects of the dynamic cellular environment on hybrid duplex conformation.

Conclusion

Several thermodynamic studies have been done on the formation of DNA duplexes in the presence of osmolyte [95]. It was found that polytheylene glycol 200 (PEG 200), a neutral osmolyte, decreased nucleic acid thermodynamic stability [96 97]. Studies on the effects of osmolytes on Z-DNA formation found that the interplay of neutral solute and NaCl in modulating the transition from B- to Z-form suggested that the Z-form may be stabilized through osmotic stress [63]. We have confirmed through our research thus far that the addition of osmolytes to the solution matrix does indeed increase favorability of Z-form transition. For the future goals of this project, we hope to examine the effects of the position of CA "breaks" in the GC sequence (i.e. CACGCGCG vs CGCACGCG), as well as the addition of divalent cations in stabilizing the Z-conformation. We also hope to study hybrids of Z-DNA and Z-RNA and the effects that the cellular environment may have on its structure and thermodynamic stability.

For our hybrid duplex project, due to large structural sensitivity to slight changes in macromolecular crowding concentrations, as well as diminutive shifts in the region of interest, we decided to focus the majority of our time on the Z-DNA project. Future goals if continuing our hybrid duplex project include examining the effects of larger crowding agents, osmolytes containing amine groups, temperature dependence (5 °C, 20 °C, and 37 °C) of formation, as well as mixed systems of both crowding agents and oxmolytes as they have opposing effects on hybrid structural conformation.

References

(1) Cheung, M. S., Klimov, D., and Thirumalai, D. (2005) Molecular crowding enhances native state stability and refolding rates of globular proteins. *Proc. Natl. Acad. Sci. 102*, 4753–4758.

(2) Ohyama, T. DNA CONFORMATION AND TRANSCRIPTION Molecular Biology Intelligence Unit Library of Congress Cataloging-in-Publication Data.

(3) Dix, J. A., and Verkman, A. S. (2008) Crowding Effects on Diffusion in Solutions and Cells. *http://dx.doi.org/10.1146/annurev.biophys.37.032807.125824*.

(4) Burg, M. B., and Ferraris, J. D. (2008) Intracellular organic osmolytes: function and regulation. *J. Biol. Chem. 283*, 7309–13.

(5) Yancey, P. H., Clark, M. E., Hand, S. C., Bowlus, R. D., and Somero, G. N. (1982) Living with water stress: evolution of osmolyte systems. *Science 217*, 1214–22.

(6) Burg, M. B., Ferraris, J. D., and Dmitrieva, N. I. (2007) Cellular response to hyperosmotic stresses. *Physiol. Rev. 87*, 1441–74.

(7) Pray, L. A. (2008) Discovery of DNA Structure and Function: Watson and Crick. *Nat. Educ.*

(8) (2016) Discovery Studio Modeling Environment. Dassault Systemes BIOVIA, San Diego.

(9) Wang, A. H.-J., Quigley, G. J., Kolpak, F. J., Crawford, J. L., van Boom, J. H., van der Marel, G., and Rich, A. (1979) Molecular structure of a left-handed double helical DNA fragment at atomic resolution. *Nature 282*, 680–686.

(10) Kłysik, J., Stirdivant, S. M., Larson, J. E., Hart, P. A., and Wells, R. D. (1981) Left-handed DNA in restriction fragments and a recombinant plasmid. *Nature 290*, 672–677.

(11) Haniford, D., and Pulleyblank, D. (1983) Facile transition of poly into a left-handed helix in physiological conditions.

(12) Lu, X.-J., Shakked, Z., and Olson, W. K. (2000) A-form Conformational Motifs in Ligand-bound DNA Structures. *J. Mol. Biol. 300*, 819–840.

(13) Timsit, Y. (2000) DNA Structure and Polymerase Fidelity: A New Role for A-DNA. *J. Biomol. Struct. Dyn. 17 Suppl 1*, 169–76.

(14) Nekludova, L., and Pabo, C. O. (1994) Distinctive DNA conformation with enlarged major groove is found in Zn-finger-DNA and other protein-DNA complexes. *Proc. Natl. Acad. Sci. U. S. A. 91*, 6948–52.

(15) Flatters, D., and Lavery, R. (1998) Sequence-dependent dynamics of TATA-Box binding sites. *Biophys. J. 75*, 372–81.

(16) Guzikevich-Guerstein, G., and Shakked, Z. (1996) A novel form of the DNA double helix imposed on the TATA-box by the TATA-binding protein. *Nat. Struct. Biol. 3*, 32–7.

(17) Barber, A. M., Zhurkin, V. B., and Adhya, S. (1993) CRP-binding sites: evidence for two structural classes with 6-bp and 8-bp spacers. *Gene 130*, 1–8.

(18) Ivanov, V. I., Minchenkova, L. E., Chernov, B. K., McPhie, P., Ryu, S., Garges, S., Barber, A. M., Zhurkin, V. B., and Adhya, S. (1995) CRP-DNA Complexes: Inducing theA-likeForm in the Binding Sites with an Extended Central Spacer. *J. Mol. Biol. 245*, 228–240.

(19) Jacobo-Molina, A., Ding, J., Nanni, R. G., Clark, A. D., Lu, X., Tantillo, C., Williams, R. L., Kamer, G., Ferris, A. L., and Clark, P. (1993) Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 A resolution shows bent DNA. *Proc. Natl. Acad. Sci. U. S. A. 90*, 6320–4.

(20) Timsit, Y. (2013) DNA self-assembly: from chirality to evolution. *Int. J. Mol. Sci. 14*, 8252–70.

(21) Kiefer, J. R., Mao, C., Braman, J. C., and Beese, L. S. (1998) Visualizing DNA replication in a catalytically active Bacillus DNA polymerase crystal. *Nature 391*, 304–7.

(22) Mohr, S. C., Sokolov, N. V, He, C. M., and Setlow, P. (1991) Binding of small acid-soluble spore proteins from Bacillus subtilis changes the conformation of DNA from B to A. *Proc. Natl. Acad. Sci. U. S. A. 88*, 77–81.

(23) Becker, M. M., and Wang, Z. (1989) THE JOURNAL OF BIOLOGICAL CHEMISTRY B-A Transitions within a 5 S Ribosomal RNA Gene Are Highly Sequence-specific *264*, 4163–4167.

(24) Wang, G., and Vasquez, K. M. (2006) Non-B DNA structure-induced genetic instability. *Mutat. Res. Mol. Mech. Mutagen. 598*, 103–119.

(25) Kim, Y.-G., Muralinath, M., Brandt, T., Pearcy, M., Hauns, K., Lowenhaupt, K., Jacobs, B. L., and Rich, A. A role for Z-DNA binding in vaccinia virus pathogenesis.

(26) Oganesian, L., and Bryan, T. M. (2007) Physiological relevance of telomeric G-quadruplex formation: a potential drug target. *Bioessays 29*, 155–65.

(27) Rich, A., and Zhang, S. (2003) Timeline: Z-DNA: the long road to biological function. *Nat. Rev. Genet. 4*, 566–572.

(28) Hamada, H., and Kakunaga, T. (1982) Potential Z-DNA forming sequences are

highly dispersed in the human genome. *Nature 298*, 396–398.

(29) Kobayashi, M., Aita, N., Hayashi, S., Okada, K., Ohta, T., and Hirose, S. (1998) DNA supercoiling factor localizes to puffs on polytene chromosomes in Drosophila melanogaster. *Mol. Cell. Biol. 18*, 6737–44.

(30) Schroth, G. P., Chou, P. J., and Ho, P. S. (1992) Mapping Z-DNA in the human genome. Computer-aided mapping reveals a nonrandom distribution of potential Z-DNA-forming sequences in human genes. *J. Biol. Chem. 267*, 11846–55.

(31) Schroth, G. P., Chou, P. J., and Ho, P. S. (1992) Mapping Z-DNA in the human genome. Computer-aided mapping reveals a nonrandom distribution of potential Z-DNA-forming sequences in human genes. *J. Biol. Chem. 267*, 11846–55.

(32) Herbert, A., and Rich, A. (1996) The Biology of Left-handed Z-DNA. *J. Biol. Chem. 271*, 11595–11598.

(33) Liu, L. F., and Wang, J. C. (1987) Supercoiling of the DNA template during transcription. *Proc. Natl. Acad. Sci. U. S. A. 84*, 7024–7.

(34) Dröge, P. (1994) Protein tracking-induced supercoiling of DNA: A tool to regulate DNA transactions *in vivo* ? *BioEssays 16*, 91–99.

(35) Sugimoto, N. (2014) Noncanonical Structures and Their Thermodynamics of DNA and RNA Under Molecular Crowding, in *International review of cell and molecular biology*, pp 205–273.

(36) Peck, L. J., and Wang, J. C. (1985) Transcriptional block caused by a negative supercoiling induced structural change in an alternating CG sequence. *Cell 40*, 129–137.

(37) Garner, M. M., and Felsenfeld, G. (1987) Effect of Z-DNA on nucleosome placement. *J. Mol. Biol. 196*, 581–590.

(38) Iyer, V., and Struhl, K. (1995) Poly(dA:dT), a ubiquitous promoter element that stimulates transcription via its intrinsic DNA structure. *EMBO J. 14*, 2570–9.

(39) Herbert, A., and Rich, A. (1999) RNA processing and the evolution of eukaryotes. *Nat. Genet. 21*, 265–269.

(40) Herbert, A. (2005) Roles for Z-DNA and Double-Stranded RNA in Transcription, in *DNA Conformation and Transcription*, pp 93–104. Springer US, Boston, MA.

(41) Weinreb, a, Katzenberg, D. R., Gilmore, G. L., and Birshtein, B. K. (1988) Site of unequal sister chromatid exchange contains a potential Z-DNA-forming tract. *Proc. Natl. Acad. Sci. U. S. A. 85*, 529–33.

(42) van Holde, K., and Zlatanova, J. (1994) Unusual DNA structures, chromatin and transcription. *Bioessays 16*, 59–68.

(43) Liu, R., Liu, H., Chen, X., Kirby, M., Brown, P. O., and Zhao, K. (2001) Regulation of CSF1 Promoter by the SWI/SNF-like BAF Complex. *Cell 106*, 309–318.

(44) Blaho, J. A., and Wells, R. D. (1989) Left-handed Z-DNA and genetic recombination. *Prog. Nucleic Acid Res. Mol. Biol. 37*, 107–26.

(45) Stafford, B. Meiosis – The spice of life Meiosis – The SPICE of life. *2015*.

(46) Bianco, P. R., Tracy, R. B., and Kowalczykowski, S. C. (1998) 570 DNA STRAND EXCHANGE PROTEINS: A BIOCHEMICAL AND PHYSICAL COMPARISON. *Front. Biosci. 3*, 570–603.

(47) Potaman, V. N., and Sinden, R. R. (2013) DNA: Alternative Conformations and Biology.

(48) Brunel, F., Zakin, M. M., Buc, H., and Buckle, M. (1996) The polypyrimidine tract binding (PTB) protein interacts with single-stranded DNA in a sequence-specific manner. *Nucleic Acids Res. 24*, 1608–15.

(49) Hildebrandt, M., Lacombe, M.-L., Mesnildrey, S., and Véron, M. (1995) A human NDP-kinase B specifically binds single-stranded poly-pyrimidine sequences. *Nucleic Acids Res. 23*, 3858–3864.

(50) Imre Berger, ‡,§, William Winston, ‡, Ramasamy Manoharan, ∥, Thomas Schwartz, ‡, Jens Alfken, ‡, Yang-Gyun Kim, ‡, Ky Lowenhaupt, ‡, Alan Herbert, ‡ and, and Alexander Rich*, ‡. (1998) Spectroscopic Characterization of a DNA-Binding Domain, Zα, from the Editing Enzyme, dsRNA Adenosine Deaminase: Evidence for Left-Handed Z-DNA in the Zα−DNA Complex†.

(51) Lushnikov, A. Y., Brown, B. A., Oussatcheva, E. A., Potaman, V. N., Sinden, R. R., and Lyubchenko, Y. L. (2004) Interaction of the Zalpha domain of human ADAR1 with a negatively supercoiled plasmid visualized by atomic force microscopy. *Nucleic Acids Res. 32*, 4704–12.

(52) Schade, M., Turner, C. J., Kühne, R., Schmieder, P., Lowenhaupt, K., Herbert, A., Rich, A., and Oschkinat, H. (1999) The solution structure of the Zalpha domain of the human RNA editing enzyme ADAR1 reveals a prepositioned binding surface for Z-DNA. *Proc. Natl. Acad. Sci. U. S. A. 96*, 12465–70.

(53) Schwartz, T., Rould, M. A., Lowenhaupt, K., Herbert, A., and Rich, A. (1999) Crystal structure of the Zalpha domain of the human editing enzyme ADAR1 bound to left-handed Z-DNA. *Science 284*, 1841–5.

(54) Herbert, A. G., Spitzner, J. R., Lowenhaupt, K., and Rich, A. (1993) Z-DNA binding protein from chicken blood nuclei. *Proc. Natl. Acad. Sci. 90*, 3339–3342.

(55) Barraud, P., and Allain, F. H.-T. (2012) ADAR proteins: double-stranded RNA and Z-DNA binding domains. *Curr. Top. Microbiol. Immunol. 353*, 35–60.

(56) Romainczyk, O., Endeward, B., Prisner, T. F., and Engels, J. W. (2011) The RNA-DNA hybrid structure determined by EPR, CD and RNase H1. *Mol. Biosyst. 7*, 1050–2.

(57) Howard Hughes Medical Institute. (2006) Nucleic Acid Techniques.

(58) Temin, H. M. Review Reverse Transcription in the Eukaryotic Genome: Retroviruses, Pararetroviruses, Retrotransposons, and Retrotranscripts'.

(59) Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2002) DNA Replication Mechanisms.

(60) Gene Expression.

(61) Miyoshi, D., and Sugimoto, N. (2008) Molecular crowding effects on structure and stability of DNA. *Biochimie 90*, 1040–1051.

(62) S B Zimmerman, and, and Minton, A. P. (2003) Macromolecular Crowding: Biochemical, Biophysical, and Physiological Consequences. *http://dx.doi.org/10.1146/annurev.bb.22.060193.000331*.

(63) Preisler, R. S., Chen, H. H., Colombo, M. F., Choe, Y., Jr., B. J. S., and Rau, D. C. (2002) The B Form to Z Form Transition of Poly(dG-m5dC) Is Sensitive to Neutral Solutes Through an Osmotic Stress.

(64) Gilles, R. (1997) "Compensatory" organic osmolytes in high osmolarity and dehydration stresses: history and perspectives. *Comp. Biochem. Physiol. A. Physiol. 117*, 279–90.

(65) Asami, K., Hanai, T., and Koizumi, N. (1976) Dielectric properties of yeast cells. *J. Membr. Biol. 28*, 169–180.

(66) Marky, L. A., and Kupke, D. W. (2000) Enthalpy-entropy compensations in nucleic acids: Contribution of electrostriction and structural hydration. *Methods Enzymol. 323*, 419–441.

(67) Leckband, D., and Israelachvili, J. (2001) Intermolecular forces in biology. *Q. Rev. Biophys. 34*, 105.

(68) Asakura, S., and Oosawa, F. (1954) On Interaction between Two Bodies Immersed in a Solution of Macromolecules. *J. Chem. Phys. 22*, 1255.

(69) Grosberg, A. Y. (2012) How two meters of DNA fit into a cell nucleus: Polymer models with topological constraints and experimental data. *Polym. Sci. Ser. C 54*, 1–10.

(70) Schnell, S., and Hancock, R. (2008) The intranuclear environment. *Methods Mol. Biol. 463*, 3–19.

(71) Rafael, and Arnold, P. (2009) DNA Structure: Hydrogen Bonds.

(72) Hancock, R., and Kwang, J. W. CELL AND MOLECULAR.

(73) Rueda, M., Cubero, E., Laughton, C. A., and Orozco, M. (2004) Exploring the Counterion Atmosphere around DNA: What Can Be Learned from Molecular Dynamics Simulations? *Biophys. J. 87*, 800–811.

(74) Harding. Nucleotides and Nucleic Acids.

(75) Atdbio. DNA duplex stability.

(76) Pramanik, S., Nagatoishi, S., Saxena, S., Bhattacharyya, J., and Sugimoto, N. (2011) Conformational Flexibility Influences Degree of Hydration of Nucleic Acid Hybrids.

(77) Ha, S. C., Lowenhaupt, K., Rich, A., Kim, Y.-G., and Kim, K. K. (2005) Crystal structure of a junction between B-DNA and Z-DNA reveals two extruded bases. *Nature 437*, 1183–1186.

(78) Chen, F. Y., Park, S., Otomo, H., Sakashita, S., and Sugiyama, H. (2014) Investigation of B-Z transitions with DNA oligonucleotides containing 8-methylguanine 0–6.

(79) Knowles, D. B., LaCroix, A. S., Deines, N. F., Shkel, I., and Record, M. T. (2011) Separation of preferential interaction and excluded volume effects on DNA duplex and hairpin stability. *Proc. Natl. Acad. Sci. U. S. A. 108*, 12699–704.

(80) Pramanik, S., Nagatoishi, S., Saxena, S., Bhattacharyya, J., and Sugimoto, N. (2011) Conformational flexibility influences degree of hydration of nucleic acid hybrids. *J. Phys. Chem. B 115*, 13862–72.

(81) Wheelhouse, R. T., and Chaires, J. B. (2010) Drug binding to DNA x RNA hybrid structures. *Methods Mol. Biol. 613*, 55–70.

(82) Gyi, J. I., Lane, A. N., Conn, G. L., and Brown, T. (1998) Solution structures of DNA.RNA hybrids with purine-rich and pyrimidine-rich strands: comparison with the homologous DNA and RNA duplexes. *Biochemistry 37*, 73–80.

(83) Pabit, S. A., Qiu, X., Lamb, J. S., Li, L., Meisburger, S. P., and Pollack, L. (2009) Both helix topology and counterion distribution contribute to the more effective charge screening in dsRNA compared with dsDNA. *Nucleic Acids Res. 37*, 3887–96.

(84) Nakano, S.-I., Wu, L., Oka, H., Karimata, H. T., Kirihata, T., Sato, Y., Fujii, S., Sakai, H., Kuwahara, M., Kuwahara, M., Sawai, H., and Sugimoto, N. (2008) Conformation and the sodium ion condensation on DNA and RNA structures in the presence of a neutral cosolute as a mimic of the intracellular media. *Mol. Biosyst. 4*, 579–88.

(85) Tataurov, A. V., You, Y., and Owczarzy, R. (2008) Predicting ultraviolet

spectrum of single stranded and double stranded deoxyribonucleic acids. *Biophys. Chem. 133*, 66–70.

(86) Lambert, D., and Draper, D. E. (2012) Denaturation of RNA secondary and tertiary structure by urea: simple unfolded state models and free energy parameters account for measured m-values. *Biochemistry 51*, 9014–26.

(87) Vasilief, I., and Besch, S. The QtiPlot Handbook.

(88) Chen, F. Y.-H., Park, S., Otomo, H., Sakashita, S., and Sugiyama, H. (2014) Investigation of B-Z transitions with DNA oligonucleotides containing 8-methylguanine. *Artif. DNA. PNA XNA 5*, e28226.

(89) Nakano, S., Kirihata, T., and Sugimoto, N. (2008) Capture of cationic ligands bound diffusely to base pairs during DNA refolding. *Chem. Commun. (Camb)*. 700–2.

(90) MacGregor, R. B., and Chen, M. Y. Delta V0 of the Na(+)-induced B-Z transition of poly[d(G-C)] is positive. *Biopolymers 29*, 1069–76.

(91) Preisler, R. S., Chen, H. H., Colombo, M. F., Choe, Y., Short, B. J., and Rau, D. C. (1995) The B form to Z form transition of poly(dG-m5dC) is sensitive to neutral solutes through an osmotic stress. *Biochemistry 34*, 14400–7.

(92) Nakano, S., Kitagawa, Y., Miyoshi, D., and Sugimoto, N. (2014) Hammerhead ribozyme activity and oligonucleotide duplex stability in mixed solutions of water and organic compounds. *FEBS Open Bio 4*, 643–650.

(93) Saenger, W. (2003) Structure and Dynamics of Water Surrounding Biomolecules. *http://dx.doi.org/10.1146/annurev.bb.16.060187.000521*.

(94) Westhof, E. (1988) Water: an integral part of nucleic acid structure. *Annu. Rev. Biophys. Biophys. Chem. 17*, 125–44.

(95) Yancey, P. H. (2005) Organic osmolytes as compatible, metabolic and counteracting cytoprotectants in high osmolarity and other stresses. *J. Exp. Biol. 208*, 2819–30.

(96) Nakano, S., Karimata, H., Ohmichi, T., Kawakami, J., and Sugimoto, N. (2004) The effect of molecular crowding with nucleotide length and cosolute structure on DNA duplex stability. *J. Am. Chem. Soc. 126*, 14330–1.

(97) Miyoshi, D., Nakamura, K., Tateishi-Karimata, H., Ohmichi, T., and Sugimoto, N. (2009) Hydration of Watson-Crick base pairs and dehydration of Hoogsteen base pairs inducing structural polymorphism under molecular crowding conditions. *J. Am. Chem. Soc. 131*, 3522–31.