EXPLORING THE MOLECULAR MECHANISMS BY WHICH AID RECOMBINASE INTERACTS WITH DNA SECONDARY STRUCTURES INVOLVED IN CANCER

by

Salil Kalarn

A Thesis Submitted to the Faculty of the

DEPARTMENT OF CELLULAR AND MOLECULAR MEDICINE

In Partial Fulfillment of the Requirements

For the Degree of

MASTER OF SCIENCE

In the Graduate College

THE UNIVERSITY OF ARIZONA

2017

STATEMENT BY AUTHOR

The thesis titled Exploring the Molecular Mechanisms by which AID Recombinase Interacts with DNA Secondary Structures involved in Cancer prepared by *Salil Kalarn* has been submitted in partial fulfillment of requirements for a master's degree at the University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this thesis are allowable without special permission, provided that an accurate acknowledgement of the source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: *Salil Kalarn*

APPROVAL BY THESIS DIRECTORS

This thesis has been approved on the date shown below:

Samantha Kendrick — *May 8, 2017*

Research Assistant Professor of Pathology — Date

Laurence Hurley — *May 8, 2017*

Professor of Pharmacology and Toxicology — Date

Table of Contents

1. Abstract

Genomic complexity in non-Hodgkin's Diffuse Large B-cell Lymphoma (DLBCL) leads to a treatment failure in ~40% of patients. Activation-Induced Cytosine Deaminase (AID), one of the enzymes involved in generating antibody diversity via class switching recombination (CSR) and somatic hypermutation (SHM) of immunoglobulin (Ig) genes in activated B-cells is one mechanism for the introduction of genomic lesions. In previous studies, AID was shown to preferentially bind to super-enhancer (SE) regions within the genome, but 26% of AID targets were not within the SE regions. The mechanism by which AID interacts with SE elements and its off-target interactions still remains a mystery. Recent evidence suggests that AID may cause genomic lesions in DLBCL via interaction with oncogenes such as *MYC* and *BCL2* resulting in mutations and translocations. Sequences within the *MYC* promoter contain the four-nucleotide AID target sequence (WRCY) and highly G-rich sequences known to form G-quadruplex DNA secondary structures. We hypothesize that key DNA secondary structures act as recruiting elements for aberrant AID activity at promoters and SEs of key genes involved in the development of DLBCL. Here, we first sought to determine whether known AID DNA targets have the potential to form G-quadruplex DNA secondary structures. The data collected from activated mouse B-cells showed 90% of the AID targets contained sequences that could potentially form G-quadruplexes and the data collected from the human Ramos cell line showed 100% of the sequences had the potential to form G-quadruplexes. To further study our hypothesis we used the techniques circular dichroism (CD) and the electrophoresis motility shift assay (EMSA) to explore the potential interaction between AID and the *BCL2* and *MYC* G-quadruplexes. We observed no significant

interactions between AID and these two G-quadruplexes, however further experimentation with different conditions and molecular techniques may show interaction. Additional studies will not only provide key insight into the genomic instability within DLBCL, but will also provide a potential mechanism by which AID is recruited to its DNA targets.

2. Introduction

Since the discovery of DNA in 1953 and the formation of the central dogma of biology in 1957, knowledge regarding genomic instability in cancer has been highly pursued. Genomic instability results from the high frequency of mutations within the genome of a cellular lineage, which include changes in nucleotide sequence, telomere damage, chromosome amplification, chromosome translocation, and epigenetic modifications (1). All of these genomic changes have the potential to eventually lead to a normal cell acquiring the hallmarks of cancer. These hallmarks of cancer are sustaining proliferative signaling, evading growth suppressors, metastasis, gaining immortality, inducing angiogenesis, and resisting cell death. Genomic instability is a characteristic of all cancers, but at what stage of cancer development it arises and what the molecular basis is for such changes, are questions we are only beginning to answer (1).

There are two ways in which a normal cell gains genomic instability; through hereditary means and through sporadic mutation (2). The difference between these two avenues are hereditary mutations are often found in known DNA repair genes whereas sporadic mutations often first happen in oncogenes that drive DNA replication leading to stress induced mutation. However, in both cases the normal cell eventually gains mutations in both DNA repair pathways and cell cycle control, which drives its progression to cancer. Therefore, in recent years, research has been conducted to better understand mutated DNA repair and cell cycle pathways in order to develop a more targeted approach to reducing the genomic instability that drives tumor

progression. Diffuse Large B-cell Lymphoma (DLBCL) cell lines serve as an excellent model for studying genomic instability due to the high rate of mutation frequency in this malignancy.

2.1. Diffuse Large B-cell Lymphoma

Lymphoma is a type of cancer, which affects lymphocytes, the class of cells involved with the adaptive immune response. Lymphoma consists of many different diverse subtypes arising from different stages of differentiation of the lymphocyte. DLBCL is the most common type of lymphoma in adults and is characterized by an increase in cell size of up to three times in size when viewed by Hematoxilin and Eosin staining. When comparing the normal lymphoid tissue to the DLBCL lymphoid tissue there is complete loss of normal lymphoid follicle architecture and an overgrowth of cancerous B-cells (Figure 1A-B). Standard therapies cure about half of the DLBCL patients and the remaining patients typically die within two years (3). The genomic complexity in DLBCL contributes to a treatment failure in ~40% of patients. Two oncogenes commonly involved in the genomic complexity are the *MYC* and *BCL2* oncogenes, which are most commonly mutated and/or translocated leading to increased expression and an aggressive lymphoma phenotype (4). When the *MYC* oncogene is overexpressed it allows the cancerous cell to proliferate uncontrollably without the appropriate growth signals whereas when the *BCL2* gene is overexpressed it promotes cell survival even when there is extensive DNA damage. These two genes are often co-expressed together at high levels and allows the cancer to not only proliferate and survive, but consequently may facilitate the accumulation of mutations in other genes in these anti-apoptotic, persistent cancer cells. A B-cell specific enzyme called Activation-Induced

Cytidine Deaminase (AID) is thought to be involved in the genomic instability of DLBCL, causing these mutations and translocations of in *MYC* and *BCL2* as well as several other important oncogenes (5).

2.1.1. Figure 1: H & E of Normal Lymphoid tissue verse DLBCL Lymphoid tissue.

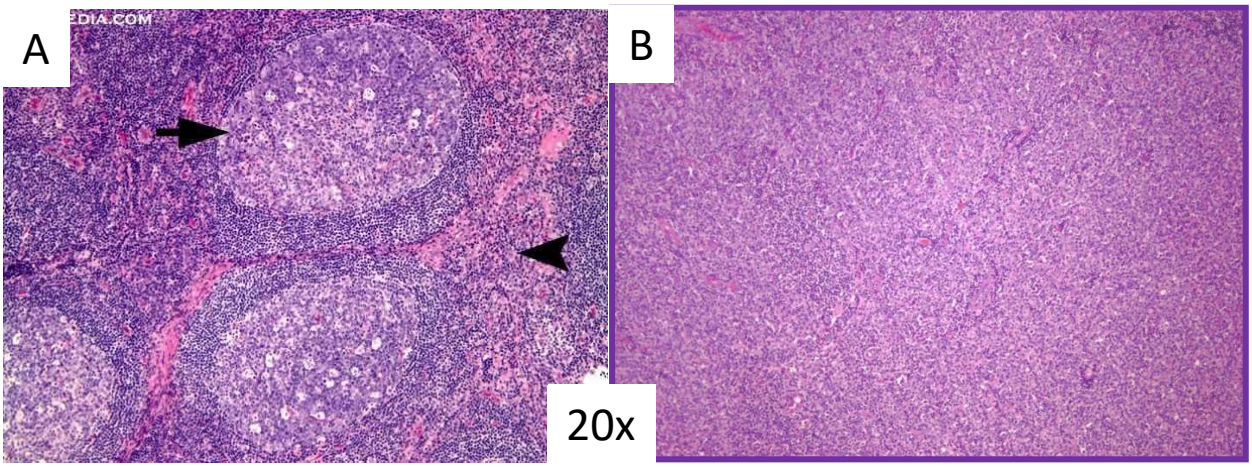


Figure 1: (A) H & E of a normal lymphoid follicle at 20x magnification. (B) H & E of a DLBCL lymphoid follicle at 20x magnification.

2.2. Activation Induced Cytidine Deaminase Normal Function within a B-cell

When a B-cell, a member of the adaptive immune system, is activated during the primary immune response it secretes immunoglobulin M (IgM). IgM binds antigens with low affinity and usually cannot penetrate tissue due to its large pentameric form. After further exposure to the antigen, a B-cell can modify its nucleotide bases within the antibody's variable region, the antigen-binding site. The modified nucleotide bases in the variable region lead to slightly different amino acids thus changing the binding affinity for the antigen, some of which are weaker and some stronger. This process of modifying the bases is called Somatic Hypermutation (SHM) and the process as a whole is called affinity maturation and allows for selection of high affinity antibodies towards an antigen to produce more effective antibodies against the infecting pathogen (5). In addition, an antibody will undergo a process called Class Switching Recombination (CSR) that alters the antibodies fragment crystallizable region (Fc), the part of the antibody that is recognized by immune cells and molecules. The alteration of the Fc portion allows different effector immune cells to bind it, allowing for a specialized, appropriate response to a specific type of foreign antigen. This will change the antibody from the initially made IgM antibody to one of the main classes of antibodies: IgG, the main antibody found to protect the vasculature and tissue; IgA, which protects mucosal surfaces; and IgE, which protects against large parasites. Affinity maturation and CSR involve AID and the mismatch repair (MMR) enzyme (6). In further support of the role of AID in facilitating antibody diversity, B-cells in mice lacking AID are unable to perform SHM or CSR. Specifically, AID is known to deaminate cytosine to uracil in WRCY (W=adenine or thymine, R=adenine or guanine, C=cytosine, and Y=cytosine or thymine) hotspot motifs often leading to incorrect repair of the

immunoglobulin genes, which in turn produces a different antibody (7). The WRCY nucleotide sequence is known to be the preferential site at which AID exerts its deaminase activity.

2.3. Aberrant AID off Target Activity Involved in Mutation and Translocation of Oncogenes

Normally, AID exerts its activity at immunoglobulin gene loci, but recent research has shown that AID introduces genetic instability through the misappropriated recombination and SHM of oncogenes such as *BCL2* and *MYC* (7). Aberrant AID activity still appears to occur at the signature WRCY mutational motif within oncogenes as in Ig genes and recent sequencing studies link SHM at these hotspots in key oncogenes to AID and may promote DLBCL development. These mutation sites are typically within 2 kb of 5′ promoter regions, contain WRCY sequences with preference for AGCT nucleotides, and are in areas of single-stranded DNA (ssDNA) undergoing active transcription (8). Recent work has shown AID targets found in super enhancer regions (SEs) have general features of accessible DNA regions with active transcription and stalled Pol II (8). SEs are areas of the genome with an unusually high density of enhancers that collectively interact with many different transcriptional factors and drive transcription. In this previous study, 1003 SEs were identified in activated mouse embryo B-cells for which only 174 were targeted by AID. In addition, 62 of those AID targets were not located inside these SE regions (9). A follow up study explored these AID off-target sites, which lead to the discovery of AID activity in areas with convergent (ConvT) transcription in GC B-cells from mice and in a Burkitt's lymphoma cell line (Ramos) (10). Convergent transcription is transcription that occurs on both the negative and positive DNA strands at the same time and in the same direction. These sites of ConvT only

occurred in 52% of the SE regions, only correlated with genes highly mutated (70% genes), and did not account for AID off-targeting in genes with low mutation frequency. This suggests that there are other components involved in the mechanism for recruitment of AID to these sites.

Previous studies demonstrated that AID binds to guanine (G)-looped ssDNA within the *MYC* promoter region (13). These G-looped ssDNA regions normally occur in guanine-rich regions and are likely to form G-quadruplex DNA secondary structures. In preliminary in silico analyses, we discovered G-quadruplex forming sequences are common to AID targeted genes whether in overlapping or non-overlapping SE regions in the mouse and the Ramos cell line genomes including the 10 genomic regions with no SE or ConvT associations. The coincidence of these G-quadruplex forming sequences found in AID targeting regions could be the potential mechanism for off-target AID recruitment.

2.4. DNA Secondary Structures

DNA secondary structures formed in G- or cytosine- (C) rich parts of the genome are called the G-quadruplex (G-rich strand) and I-motif (C-rich strand). Normally, most genomic DNA is found in the conformation of a double helix known as B-DNA. Negative supercoiling is known to facilitate the conformational change from B-DNA to non B-DNA structures (Figure 2A). Unwinding of the double helix can occur during cellular events such as transcription, replication, and repair. However, a nuclear enzyme, topoisomerase can relieve negative supercoiling, but it does so randomly leaving portions of ssDNA under torsional stress (14).

These regions could then potentially form G-quadruplexes or I-motifs depending on the sequence composition.

Recent research has shown that these structures can form in the promoter regions of many oncogenes and can act as molecular switches for transcriptional activation and repression (14). However, molecular evidence suggests that the G-quadruplex and I-motif cannot form at the same time (14). In addition, in the context of the *MYC* and *BCL2* G-quadruplex and I-motif, the G-quadruplex acts as a transcriptional repressor while the I-motif serves as a transcriptional activator (Figure 2B; 15). Using this knowledge small molecule compounds have been developed to target either the G-quadruplex or I-motif for selective destabilization or stabilization of these structures with the goal of being able to harness this switch mechanism for therapeutic benefit. Furthermore, out of the many G-quadruplexes studied each structure exhibits slightly different folding conformations, which allows for greater specificity when targeting specific oncogenes. Proof of concept has come from Quarfloxin, the first G-quadruplex targeting compound to make it to Phase II clinical trials for treatment of neuroendocrine tumors (16).

2.4.1. Figure 2: Formation of DNA secondary structures from B-DNA

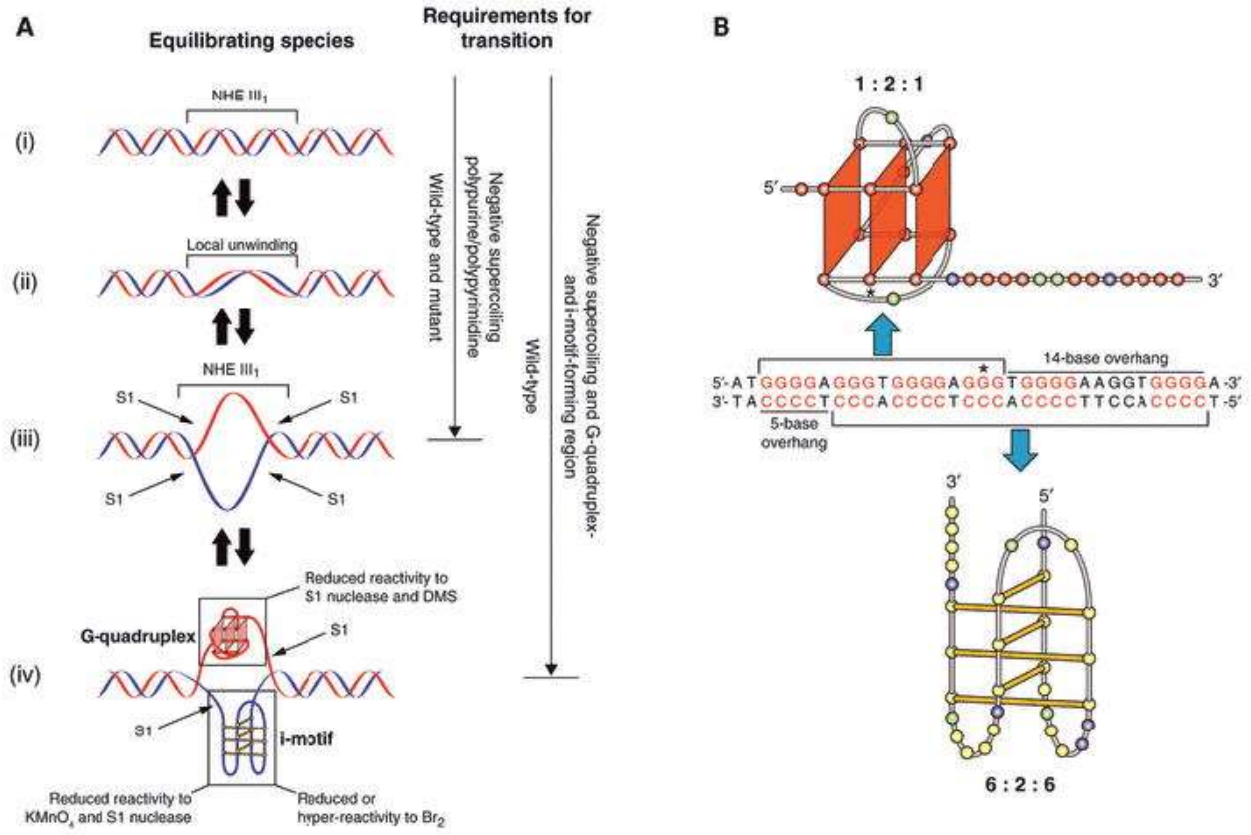


Figure 2: (A) Schematic of B-DNA's transformation to G-quadruplex and I-motif DNA secondary structures. (B) Example of the *MYC* G-quadruplex and I-motif sequences that form DNA secondary structures.

2.5. G-rich Nucleic Acid DNA Secondary Structures

G-quadruplexes form from single-stranded guanine-rich DNA that can fold into a globular structure found under physiological conditions. The G-quadruplexes can be unimolecular or multimolecular, although unimolecular G-quadruplex formation appear to be more prevalent in biological systems (16). These structures form when four guanine bases in the syn or anti conformation engage in Hoogsteen-hydrogen bonding to form a tetrad that can then stack upon other G-tetrads (Figure 3A-B). The reported G-quadruplexes demonstrate many diverse conformations, differing in number of G-tetrads, folding topologies, and loop structures (Figure 3C-F). These structures consist of two to four tetrads, with the most common being four tetrads. The common four-tetrad G-quadruplex can have three or more loops connecting the tetrad corners, and two flanking segments, one on either end (16). G-quadruplexes can adopt a parallel or antiparallel conformation, with adjacent strands running in the same or different directions, respectively, as determined by the sugar glycosidic, anti or syn, conformations (16). G-quadruplexes exist in a parallel conformation and usually have the same glycosidic conformations whereas antiparallel G-quadruplexes have different conformations. G-quadruplex folding is determined by the nucleic acid sequence and the presence of monovalent sodium and/or potassium cations (16). These cations stabilize the interactions of one tetrad to the one beneath it or above it by sitting between and centered on the tetrads. Depending on the cation some G-rich sequences can form hybrid structures that are in equilibrium with each other (16).

2.5.1. Figure 3: Introduction to G-quadruplex structure

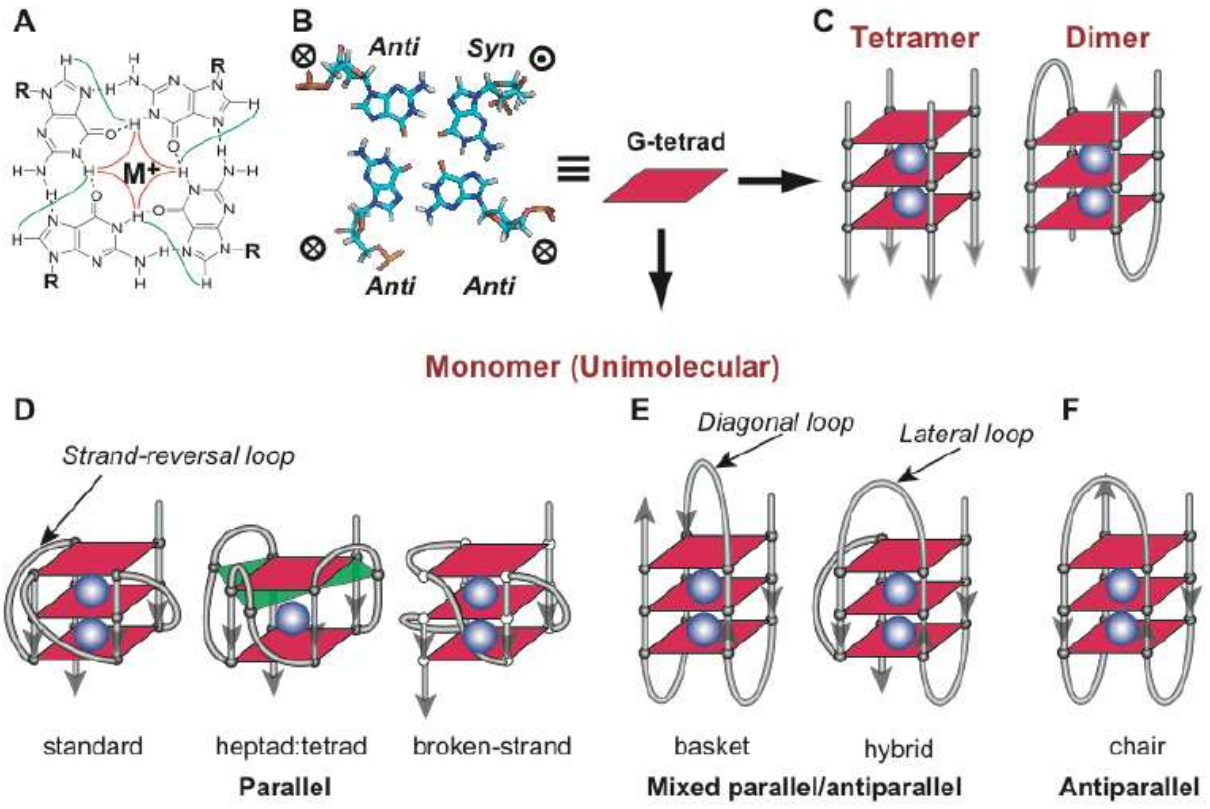


Figure 3: Folding topologies and various glyosidic conformations the G-quadruplex can adopt. (A) Hoogsteen Hydrogen bonding of a G-quadruplex tetrad. (B) anti and syn glyosidic conformations . (C) Tetrameric and Dimeric G-quadruplex formation. (D) Parallel G-quadruplex examples. (E) Mixed G-quadruplex structures. (F) Antiparallel G-quadruplex example (16).

2.6. C-rich Nucleic Acid Forming DNA Secondary Structures

The I-motif is also a non-B DNA structure formed when two parallel C-rich duplexes form intercalated hemiprotonated cytosine-cytosine base pairs. A hemiprotonated cytosine-cytosine base pair involves the sharing of a proton between the N3 of both cytosine bases. Due to the required protonation of the N3 cytosine, this type of bonding requires weak acidic conditions to form (Figure 4A-C; 17). The intercalation of three or more hemiprotonated cytosine-cytosine

base pairs can form via two distinct ways to form the structure (Figure 4D). One uses the outmost cytosine-hemiprotonated cytosine pair at the 3’ end of the DNA and the other uses the cytosine-hemiprotonated cytosine at the 5’ end (17). The 3’end topology is called the 3’E or R form whereas the 5’ end topology is called the 5’E or S form. There is a third topology that is much less stable than the other two, this is called the T-form and is characterized by non-intercalated cytosine-hemiprotonated cytosine interactions.

Stability of the I-motif is determined by sequence length, sequence composition, salt concentration, pH, and phosphate backbone stability. There are two classes of I-motifs, Class I, which comprises shorter loop sequences and class two, which has longer loops (17). Class II is more stable because there are more interactions between non-cytosine base pairs. In addition, longer sequences can form hairpins that intercalate with each other to form a dimeric I-motif. Sequence composition of the non-cytosine bases can also affect I-motif stability through creating adenine-thymine base pairing that stack at both or either the 5’ and 3’ ends of the I-motif to increase its stability. In contrast, thymine-thymine base pairing can destabilize the I-motif. Regardless of I-motif nucleotide sequence conformation, an anion is needed to stabilize the I-motif. Similar to G-quadruplex stabilization with a cation the I-motif requires cation to form in solution. While DNA adopts only the duplex conformation within a 100mM sodium ion solution, a 100mM potassium ion solution facilitates the formation of I-motif and G-quadruplex structures (17). In addition, DNA in the presence of 10mM magnesium and 100mM sodium exists in an equilibrium where all three structures can form. With this experimental evidence, it was concluded that low salt conditions reduced the pKa of cytosine’s N3, which favors I-motif

formation. Other studies demonstrate increasing ionic concentration decreases the I-motif stability.

2.6.1 Figure 4: Introduction to I-motif structure

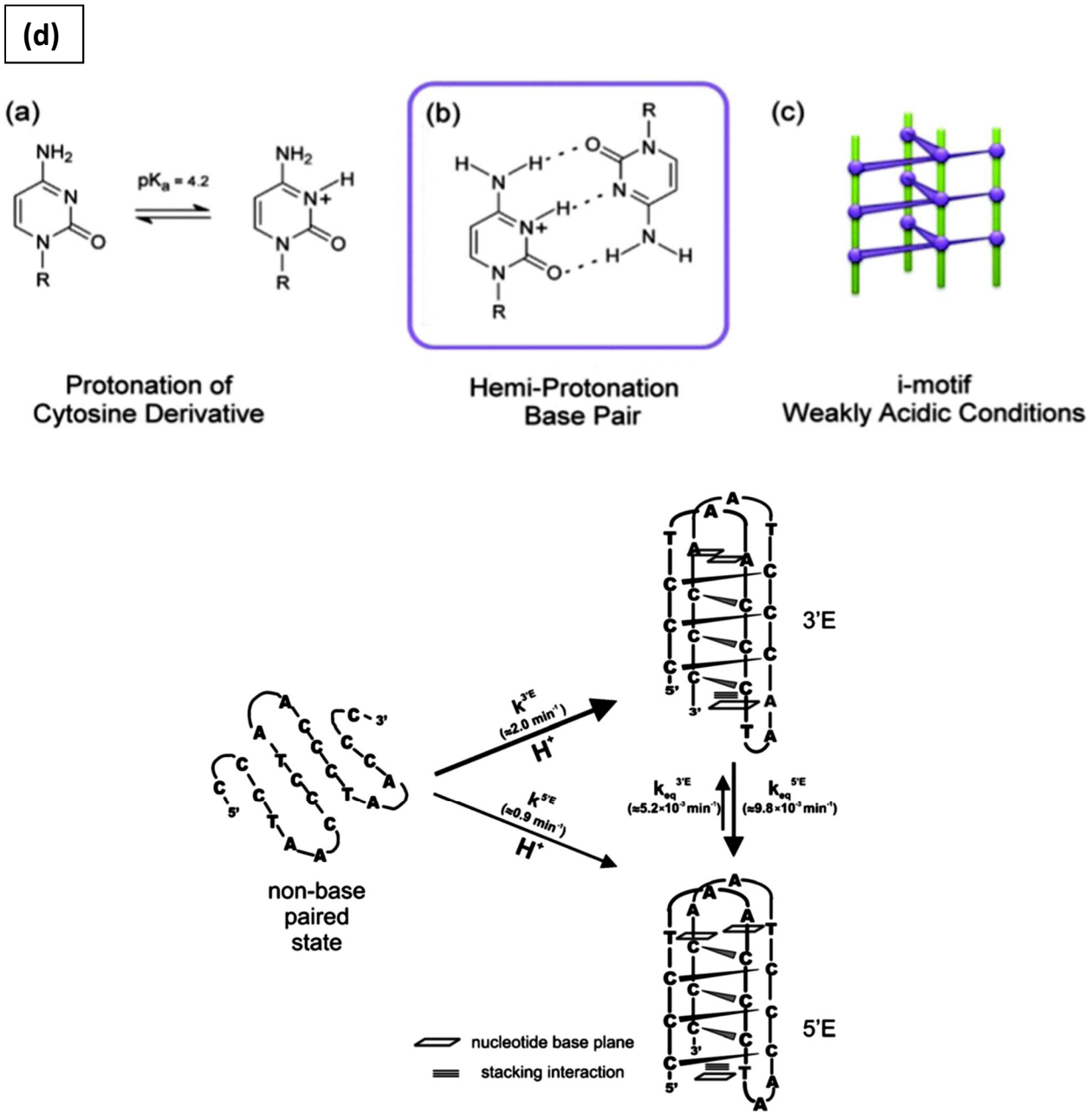


Figure 4: (a) The protonation of N3 of cytosine. (b) The hemiprotonated cytosine-cytosine base pairing needed for I-motif formation. (c) The three dimensional structure of the hemiprotonated cytosine-cytosine basepairing and intercalation. (d) Schematic of single stranded C-rich DNA forming two of the different types of I-motif structures, the 3'E or R form and the 5'E or S form (17).

2.7. MYC G-quadruplex and I-Motif

The two main transcriptional regulatory regions of *MYC* are at the P1 and P2 promoter. However, there is a conserved 27 base pair NHE III in the proximal region of the P1 promoter that can activate transcription regardless of the P1 or P2 promoter (Figure 5a; 18). In the case of the *MYC* promoter, supercoiling occurs 1.8Kb upstream of the P1 promoter, which lies near the NHE III element and supports the potential for this sequence to form a secondary DNA structure (Figure 5b; 18). The NHE III element acts as a silencer element and is regulated by a G-quadruplex structure. When this G-quadruplex is formed the NHE III is silenced, but when the G-quadruplex is destabilized there is transcriptional activity (19). There are twenty-seven base pairs involved in forming this G-quadruplex that can potentially adopt two different conformations. Currently, there are two major *MYC*Pu27 G-quadruplex binding proteins. Nucleolin is recognized to stabilize the G-quadruplex acting as a transcriptional repressor, whereas NM23-H2 destabilizes the G-quadruplex acting as a transcriptional activator (20).

The C-rich strand of the same 27 base pair NHEII just upstream of the *MYC* P1 promotor has a major I-motif that is formed at neutral pH (Figure 5b; 21). The major I-motif is formed from four consecutive runs of cytosines II-VI.

2.7.1. Figure 5: *MYC* G-quadruplex and I-motif

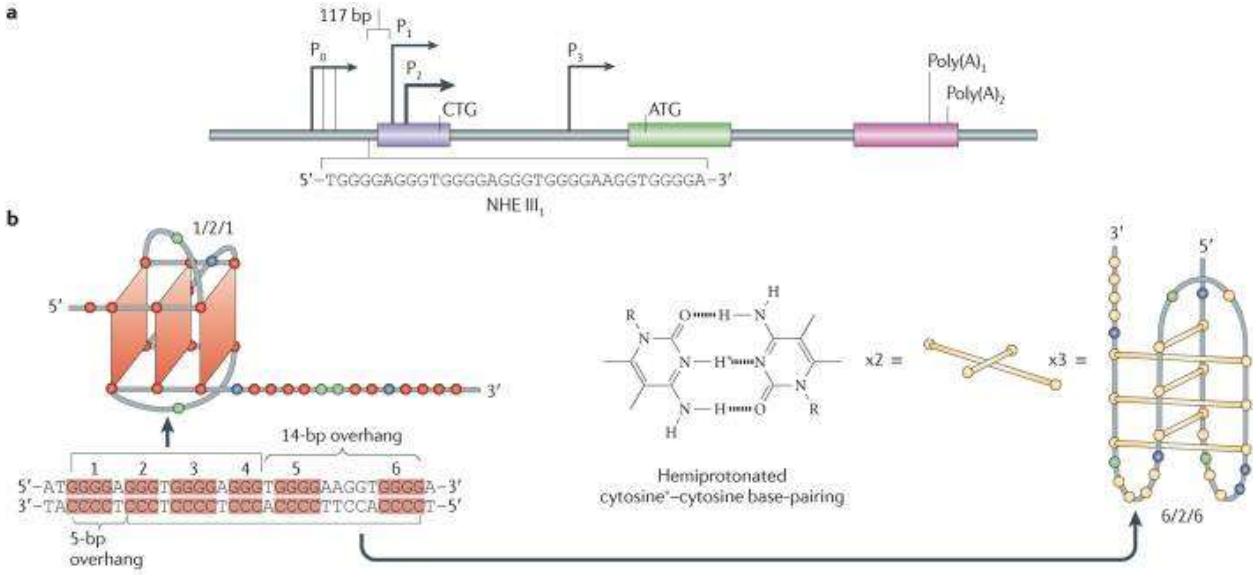


Figure 5: *MYC* G-quadruplex and I-motif (a) The *MYC* gene loci with the 27 base GC-rich sequence within the NHE III just upstream of the P1. (b) The *MYC* G-quadruplex and I-motif forming sequence with the major conformational structures (19).

2.8. The *BCL2* G-quadruplex and I-motif

In the human *BCL2* gene there are two promoters called the P1 and P2 promoters. The major promoter was determined to be the P1 promoter, which is 1,423 base pairs (bp) upstream of the transcriptional start site (Figure 6; 22). This highly GC rich promoter is within a NHE and does not contain a TATA box. Additionally, the 5'end of the P1 promoter is known to be a major site of *BCL2* transcriptional regulation (23). Many of the main transcription factors such as Sp1, Wt1, E2F, and NGF bind the 5' end of the P1 promoter. In this area there is a 39 bp GC rich element located 66 bp upstream of the P1 promoter that is known to form a G-quadruplex (Figure 6) (23).

This *BCL2* G-quadruplex contains six runs of three to five consecutive guaninines, which can potentially form a couple of different quadruplexes (23). The major G-quadruplex uses the first, second, fourth, and fifth G-runs. This G-quadruplex forms a mixed parallel, antiparallel, three-tetrad quadruplex with three loops that contain 3, 7, and 1 nucleotide loops.

The C-rich complimentary strand forms a stable intramolecular I-motif called the 8:5:7 I-motif (25). The structure forms by using the C-runs I-VI and has three loops using 8, 5, and 7 nucleotides per loop.

2.8.1 Figure 6: *BCL2* G-quadruplex and I-motif

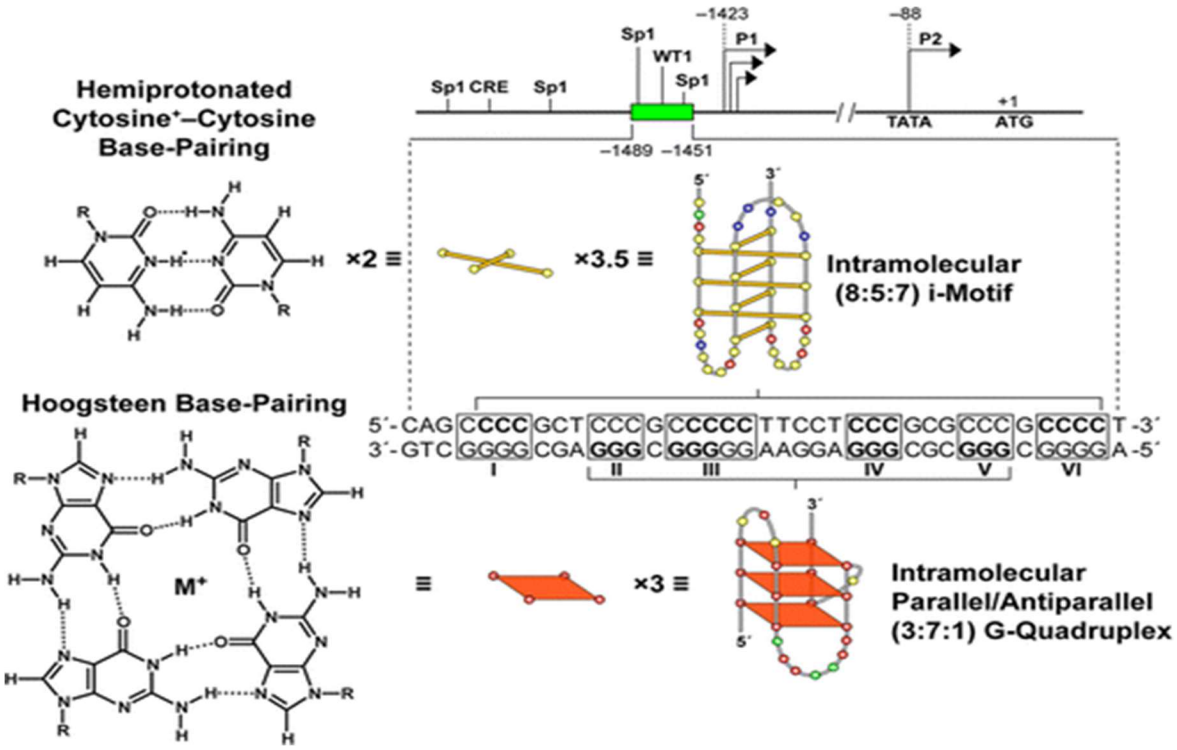


Figure 6: The *BCL2* i-motif (top) and *BCL2* G-quadruplex (bottom) folding topology upstream of the *BCL2* P1 promoter. The 39 base G-C rich DNA secondary structure forming sequence is shown to be just upstream of the p1 promoter (23).

2.9. Project overview

Due to the fact that *MYC* and *BCL2* are frequently mutated and translocated in DLBCL it is important to understand the mechanism by which this occurs so that a therapy can be developed. AID is implicated via sequencing studies to facilitate mutations in WRCY sequences in these off-target oncogenes. In addition, translocations also occur within DLBCL and could be due to the ability of AID to induce double stranded breaks through mutations on both the positive and negative strand of DNA. Although these off-target effects of AID are known, the mechanism by which AID is recruited to these off-target areas of the genome is unknown.

As previously discussed, AID binds to G-looped ssDNA within the *MYC* promoter region. Additional preliminary in silico analyses of DLBCL genomes demonstrated G-quadruplex forming sequences are located within *BCL2* and *MYC* SE genomic locations and promoter AID hotspots where WRCY sites were identified as mutated in published sequencing studies (Figure 7). Therefore, we hypothesize G-quadruplex structures are involved in the direct recruitment of AID, which induces genomic instability. The objective of this project was to determine whether AID directly binds to the *BCL2* or *MYC* G-quadruplex.

2.9.1. Figure 7: *BCL2* and *MYC* G-quadruplex sequences with AID target elements

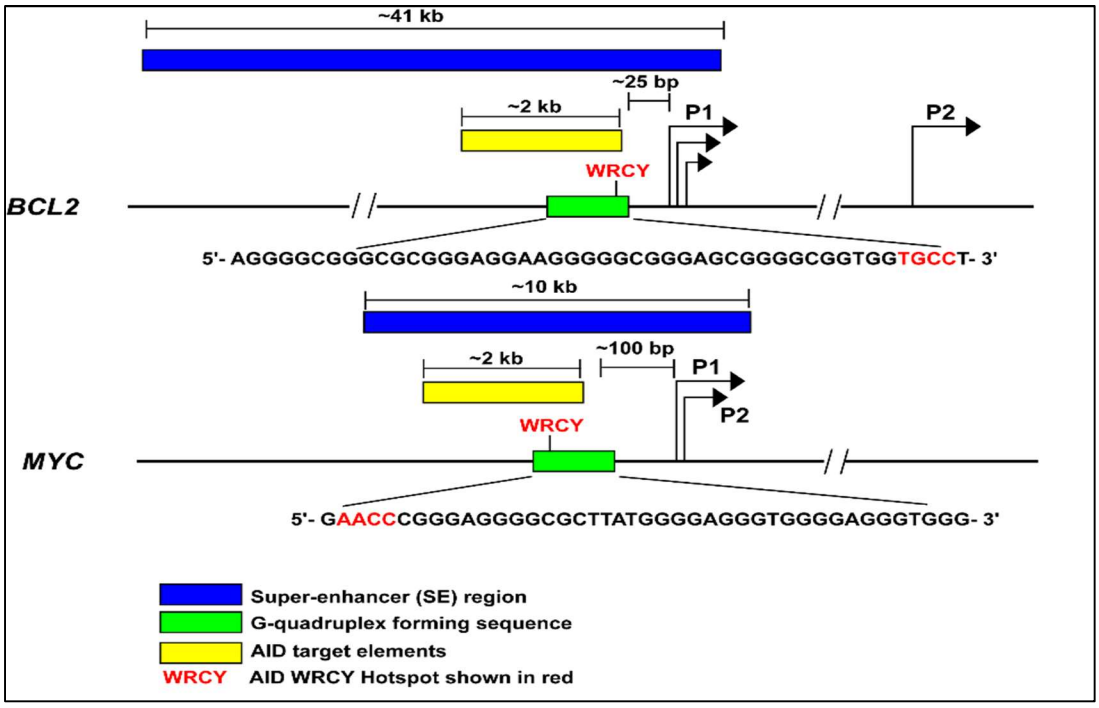


Figure 7: Schematic of *BCL2* and *MYC* genomic regions with AID target elements. Alignment of the SE regions identified for the *BCL2* and *MYC* genes in human DLBCL cell lines with the promoter regions detected as AID targeted elements, G-quadruplex forming sequence, and AID hotspot (44).

3. Methods

3.1 In silico

The paper from Qian J *et al* (9) was used to determine how many of the previously identified AID targeting sequences contain putative/known G-quadruplex forming sequences. The sequences discovered in this paper were shown to be AID targets in mouse activated B-cells. NCBI's, "Genomes and Maps" was used to acquire the positive and negative strand nucleotide sequence. This sequence was then inputted into the QGRS Mapper program hosted by the Ramapo College of Bioinformatics (http://bioinformatics.ramapo.edu). The sequences were analyzed for their ability to form G-quadruplex structures using the following parameters: runs of 3Gs; 40 nucleotide maximum length; and 17 nucleotide maximum loop size. In addition to looking at mouse activated B-cells, the Human Ramos Burkitt lymphoma cell line was also characterized using the same methods in a follow-up publication from Meng FL *et al* (10).

3.2 Circular Dichroism

Circular dichroism (CD) spectroscopic studies of the oligonucleotides were performed on a Jasco J-810 spectropolarimeter equipped with a thermoelectrically controlled cell holder. A quartz cell of 1 mm optical path length was used. A blank sample consisting of the appropriate buffer was used for baseline correction for the experiment to eliminate buffer imperfections in the data CD spectroscopic measurements were the averages of three scans collected between 200 and 350 nm. The scanning speed of the CD instrument was 100 nm/min, and the response time was 1 s. Tm values were measured by CD melting and annealing experiments performed at 264 nm for three repeats, with a heating or cooling rate of 1 C/min. All CD Tm data was fitted

using Sigma plot 47 13's curve fitting software to determine the point at which the G-quadruplex is fifty percent denatured. All CD samples contained 10uM DNA in 10mM Li-phosphate buffer and 5mM KCl (pH. 7.0) with a DNA to compound ratio of 1:3. By saturating the complex with compound to determine the thermal stability of the complex we can compare it to the free G-quadruplex to determine if the compound stabilizes the G-quadruplex and by how much the compound stabilizes the G-quadruplex.

3.3 Electrophoresis Motility Shift Assay

All oligomers were purchased from Eurofin MWG Operon and were PAGE-purified. The oligomer concentration was 5uM and was annealed from 95 C to room temperature. The EMSA binding buffer contained; 20mM Hepes, 2mM $MgCl_2$, 1mM EDTA, 1mM DTT, 0.1ug/uL BSA, 0.1% Tween 20, 10% Glycerol. The protein and DNA were allowed to incubate for 30 minutes on ice before addition into the gel. The gels used were a 5% TBE pre-cast mini-protean gel with a 30uL well size. Each gel was pre-ran at 75V for 15minutes before addition of the experimental contents. The gel was then ran for 1 hour using the 75V setting.

4 Results

4.1 G-quadruplex forming sequences enriched in AID off-target areas of the genome

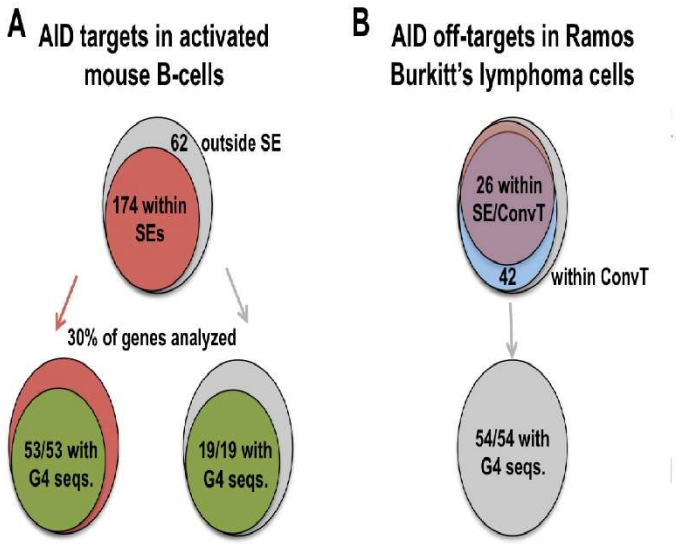
We explored the relationship between the occurrence of AID targeting and G-quadruplex forming sequences using previous research from Qian J *et al* mouse activated B-cell sequencing data. In the mouse activated B-cell there were 1003 SEs discovered (10). The data showed 824 SE regions as non-AID targets, 174 SE regions were AID targets, and 62 AID targets were not in SE regions. Sequences from the study were inputted into the QGRS mapper to determine whether these sequences had the potential to form G-quadruplex structures. Out of the 236 AID targets we determined that 90% (223/236) had the potential to form G-quadruplex structures suggesting an enrichment of these structures are present in AID targeted regions of the genome (Figure 8A).

Using the follow-up study by Meng FL *et al*, we looked into the interaction of AID between SE regions and ConvT areas (11). Out of the 54 sequences targeted by AID, 26 were areas with SE regions and ConvT areas, 16 were in ConvT areas only, 2 were in SE regions, and 10 areas were in neither SE regions or ConvT areas (Figure 8B). When the sequences from this data were analyzed for the presence of G-quadruplex forming sequences, 100% (54/54) had the potential to form G-quadruplex structures.

Although it is known that AID exerts off target activity in both SE regions and ConvT areas of the genome, there are other areas of the genome it can target that have neither. This indicates that there is another mechanism by which AID is recruited to these off target sites. Our data

supports the possibility that AID could be recruited to its off target sites by DNA G-quadruplex secondary structures.

4.1.1. Figure 8: In silico data of known AID targets



C

Genomic Overlap	Genomic Region*	G-quadruplex forming sequence (5'-3')
SE	chr1:36365700-36368700; *Arid5a*	GGGGCGGGTCCATGGGCGGGTCCGCGGGGG
ConvT	chr3:186500475-186503615; *EIF4A2*	GGGGGAGGGTCCTAGGCTTTACTTGGGGAAGGG
SE and ConvT	chr3:187457660-187464090; *BCL6*	GGGGGGGAGGGTGGGGGTGGCGGGG
None	chr6:32496717-32498380; *HLADRB5*	GGGCTCCCTGGGCGGGGTGCGGG

Figure 8: Venn diagrams depicting the fraction of AID targets with G-quadruplex (G4) forming sequences. (A) In the 236 genomic regions identified as AID targets in activated mouse B-cells, 174 overlap with SE's and 62 do not. (B) All 54 AID targets within the Ramos cell line genome contain G4 forming sequences regardless of association with ConvT and/or SE regions . (C) Example of genomic data obtained from in silico analysis.

4.2 Introduction to Circular Dichroism in DNA Secondary Structure Research

CD spectroscopy in the DNA secondary structure field is often used to show the folding topology of the structure and the thermal stability of the structure. CD is a technique that can depict the difference in the absorption of left-handed and right-handed circularly polarized light. This can be used to explore changes in absorption of the DNA secondary structure upon changing the ex vivo environment, including the addition of different cations, small molecules, or binding proteins. The characteristic parallel G-quadruplex structure has a maxima at 264nm and a minima at 240nm. An example of this parallel G-qaudruplex is the *MYC* G-qaudruplex. The *BCL2* G-quadruplex is a mixed parallel, antiparallel structure, which will have a characteristic maxima at 264nm, an additional positive peak at 290nm, and a minima at 240nm.

Generally a shift downstream in the maxima towards 250nm is characteristic of a change in structure to an unordered-coiled structure whereas a decrease in the maxima means there is less G-quadruplex folded. Therefore, a significant increase shows that there is an increase in the amount of folded structures in solution, which infers that this sequence is slightly more stable at the same DNA concentration. A decrease or increase of 300,000 molar ellipticity from the baseline is generally considered a substantial change (15).

CD is also able to characterize the thermal stability of the G-quadruplex through evaluating the difference in molar ellipticity from 20 °C to 90 °C. The midpoint on the curve between the maximum molar ellipticity and the minimum molar ellipticity is the melting temperature of the structure. An increase or decrease by >5 °C is considered a substantial change to the thermal stability of the structure. As well, these thermal stability studies require a melting temperature

between 50 ^{o}C and 70 ^{o}C for accurate melting temperature determination. This provides an adequate amount of data points in the upper molar ellipticity and the lower molar ellipticity and offers a window for destabilization of stabilization of the structure. Since the G-quadruplex requires cations for optimal stabilization ex vivo, mainly potassium to form its structure, altering the buffer concentrations can achieve a desired melting temperature range. Normally decreasing the cation concentration destabilizes the G-quadruplexes and lowers the melting temperature. Modifying cation size also plays a role in how stable the G-quadruplex. The concentration of potassium was then addressed in order to determine the effect on thermal stability for both sequences.

4.3 Modified *MYC* and *BCL2* G-quadruplex sequences have similar characteristics to previously studied sequences

For our experimentation, we extended the natural nucleotide promoter sequence of the established *MYC* and *BCL2* G-quadruplex sequences in order to provide slight overhang in anticipation of AID binding (Table 1). The previously studied G-quadruplex sequences did not contain the WRCY hotspot motif, which lead to the addition of 15 nucleotides to the 27 nucleotide long *MYC* G-quadruplex and 9 nucleotides to the 39 nucleotide long *BCL2* G-quadruplex sequences.

We did not observe any shifts in spectra peak wavelength between the known and extended G-quadruplex forming sequences (Figure 9A). The *MYC* CD spectra showed a slight increase in molar ellipticity for the longer *MYC* sequence with the AID WRCY motif either in the most stable

buffer conditions of 100 mM potassium (Figure 9A orange vs yellow) or in a less stable buffer of 20 mM potassium (Figure 9A blue vs grey). When comparing the CD melting temperature spectra of the previously studied *MYC* sequence and the *MYC* sequence with the AID WRCY motif at 20 mM (Figure 9B blue and grey) versus 100 mM (Figure 9B orange and yellow) potassium ion concentration, the 100 mM potassium ion concentration was more thermally stable for both sequences. For further experimentation to test the effect of AID on these structures, the 20 mM potassium ion concentration in the TrisHCl buffer was used. In addition, the purpose of this project is to address AID interaction in respect to its effect on thermal stability, which means a slightly less stable G-quadruplex would be favorable to show this effect.

Similar to the *MYC* sequences, the *BCL2* CD spectra data showed no real significant change between the previously studied *BCL2* sequence and the longer *BCL2* sequence in the more stable buffer conditions with 100 mM potassium (Figure 10A blue vs grey) or in the less stable buffer conditions with 20 mM potassium (Figure 10A orange vs yellow). Both sequences contained the characteristic *BCL2* spectra positive peak at approximately 290nm. This spectrum confirms the known mixed parallel (maxima 264nm) and antiparallel (positive peak ~290nm) folding topology. The potassium ion concentration was also addressed in the same melting temperature experimentation as the *MYC* G-quadruplex with the 100 mM potassium (Figure 10B orange and yellow) buffer conditions being more stable when compared to the buffer with 20 mM potassium (Figure 10B blue and grey). The *BCL2* sequences had a similar result with the 100 mM potassium ion concentration being the most stable. Similarly, 20 mM potassium ion concentration was used to assess interaction of AID with the *BCL2* G-quadruplex.

4.3.1. Table 1: *MYC* and *BCL2* G-quadruplex sequences

MYC Sequences:
5’-**GAACCCGGGAGGGGCGCTTA**TGGGGAGGGTGGGGAGGGTGGG-3’ (Pu27+ AID sequence)
5’- TGGGGAGGGTGGGGAGGGTGGGGAAGG-3’ (Pu27)
BCL2 Sequences:
5’- AGGGGCGGGCGCGGGAGGAAGGGGGCGGGAGCGGGGCTG**TGGTGCCT**-3’ (Pu39 + AID sequence)
5’- AGGGGCGGGCGCGGGAGGAAGGGGGCGGGAGCGGGGCTG-3’ (Pu39)

4.3.2. Figure 9: CD data comparison of previously studied *MYC* sequence and *MYC* sequence with AID WRCY motif

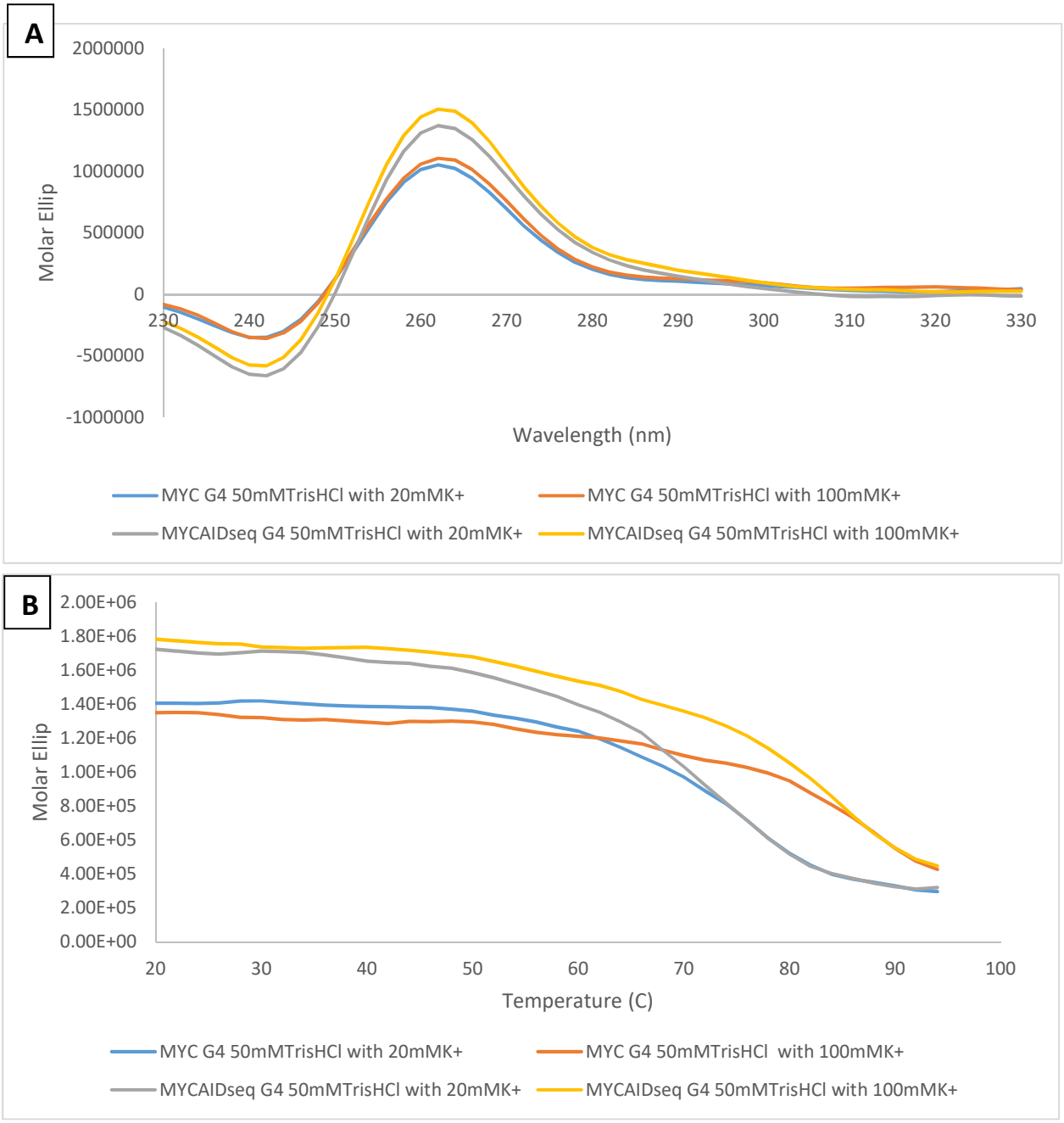


Figure 9: (A) CD spectrum of 5uM *MYC* G-quadruplex DNA with AID WRCY sequence in 50mM TrisHCl with 20mM or 100mM potassium ion concentrations. (B) CD T-melt spectra of 5uM *MYC* DNA with AID WRCY sequence in 50mM TrisHCl with 20mM or 100mM potassium ion.

4.3.3. Figure 10: CD data comparison of previously studied *BCL2* sequence and *BCL2* sequence with AID WRCY motif

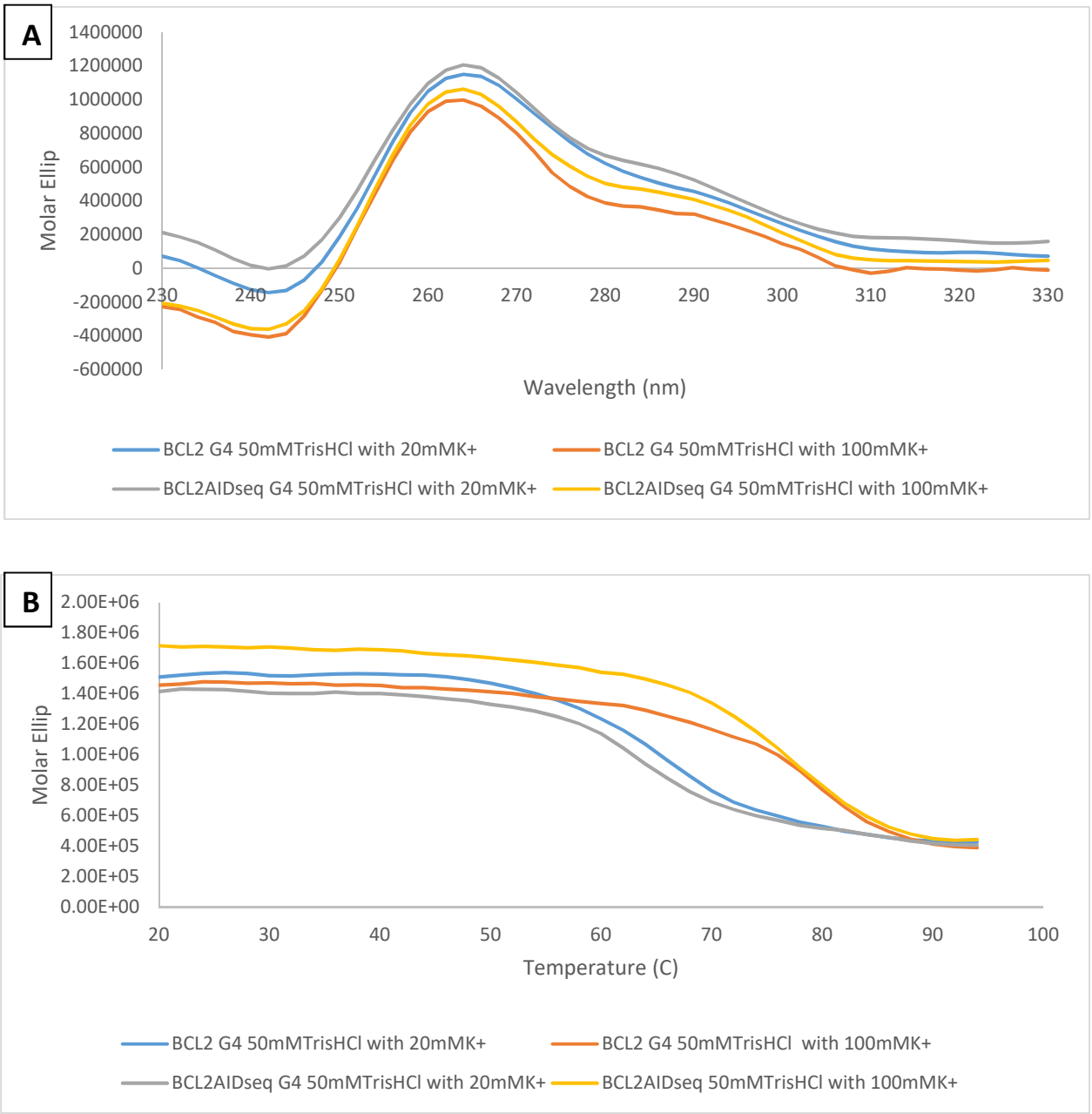
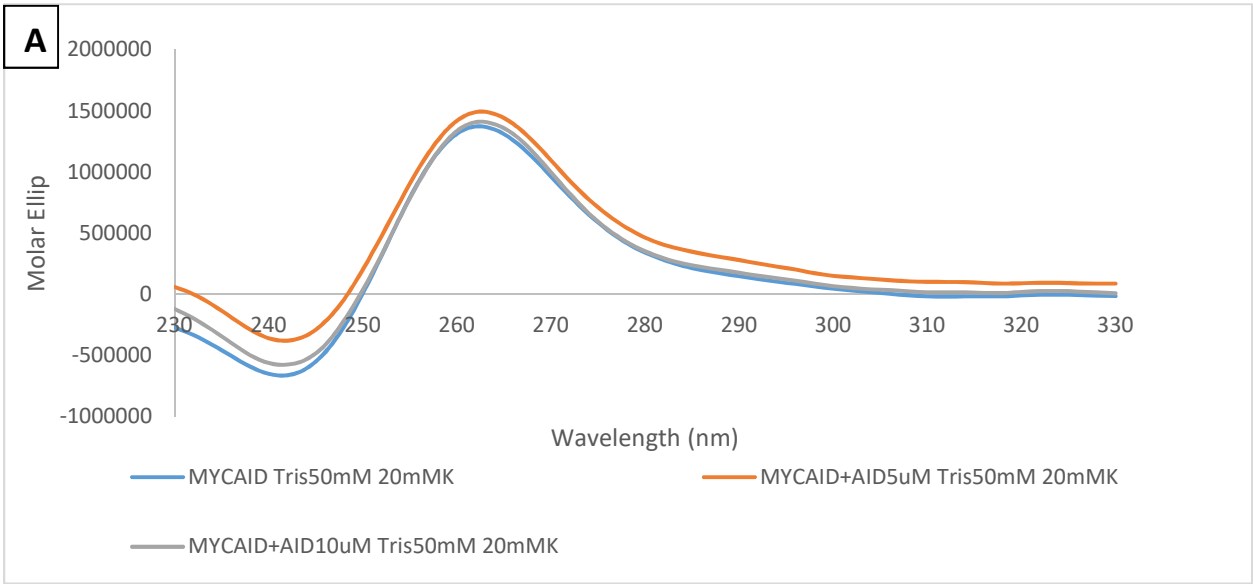


Figure 10: (A) CD spectrum of 5uM *BCL2* G-quadruplex DNA with AID WRCY sequence in 50mM TrisHCl with 20mM or 100mM potassium ion concentrations. (B) CD T-melt spectra of 5uM *BCL2* DNA with AID WRCY sequence in 50mM TrisHCl with 20mM or 100mM potassium ion concentration.

4.4 AID does not alter *BCL2* and *MYC* G-quadruplex folding or stability

With experimental conditions optimized, we titrated in increasing concentrations of commercially available recombinant AID protein to the *MYC* or *BCL2* modified AID sequence G-quadruplex in a TrisHCl buffer with 20mM potassium ion concentration. Comparison of the melting temperatures of DNA only to DNA plus AID protein demonstrated there was no effect of the presence of AID on the thermal stability for either the *MYC* (Figure 11) or *BCL2* (Figure 12) G-quadruplexes. While we would expect an interaction of AID with the G-quadruplex to either increase or decrease its thermal stability by at least 5 °C, our data suggests that the protein may not be interacting to stabilize or destabilize the G-quadruplex, respectively.

4.4.1. Figure 11: CD data of *MYC* G-quadruplex with AID



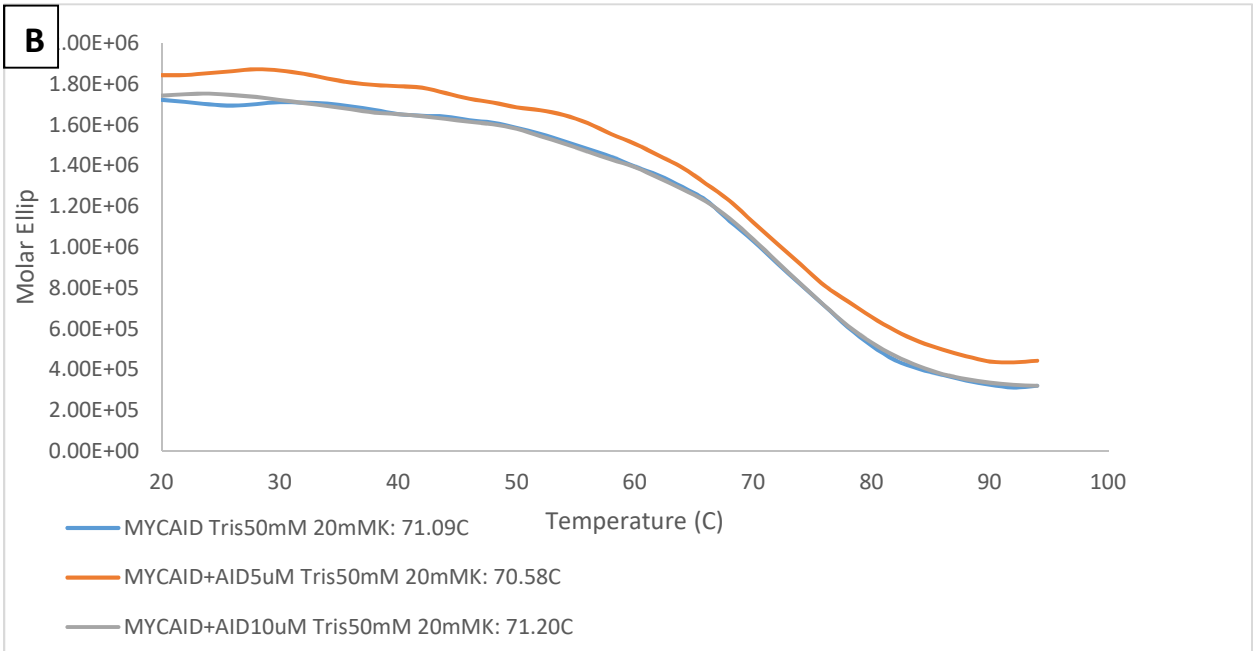


Figure 11: (A) CD spectrum of 5uM *MYC* G-quadruplex DNA with AID WRCY sequence in 50mM TrisHCl with 20mM potassium ion with varying protein concentrations. (B) CD T-melt spectra of 5uM c-*MYC* DNA with AID WRCY sequence in 50mM TrisHCl with 20mM potassium ion with varying protein concentrations.

4.4.2. Figure 12: CD data of *BCL2* G-quadruplex with AID

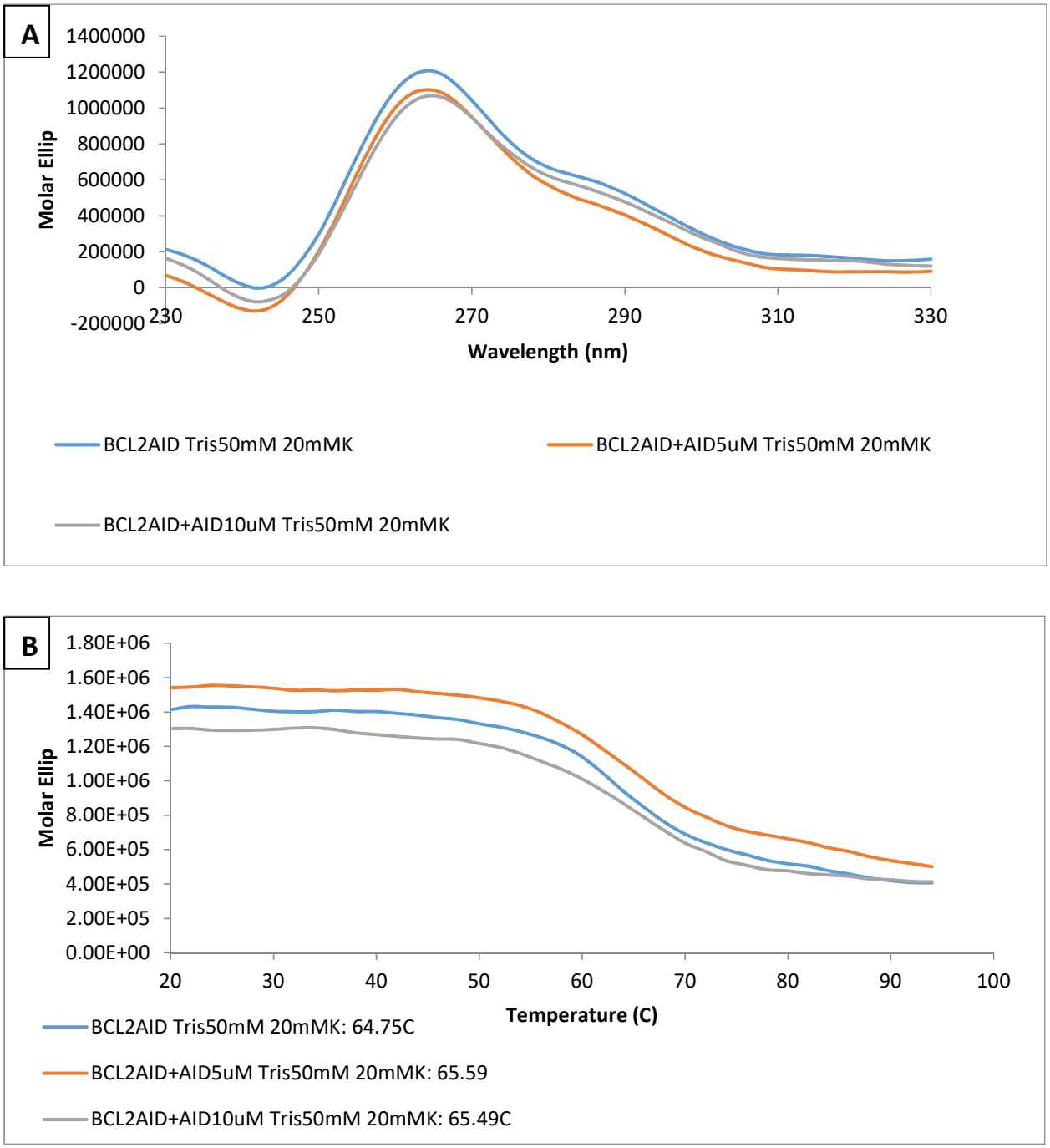


Figure 12: (A) CD spectrum of 5uM *BCL2* G-quadruplex DNA with AID WRCY sequence in 50mM TrisHCl with 20mM potassium ion with varying protein concentrations. (B) CD T-melt spectra of 5uM *BCL2* with AID WRCY sequence in 50mM TrisHCl with 20mM potassium ion with varying protein concentrations.

4.5. The presence of AID recombinase decreases the thermal stability of the *MYC* and *BCL2* I-motif

With the absence in any detectable interaction of AID with the G-quadruplex using CD, we then proposed AID may interact and bind with the opposite, complementary strand of DNA that has potential to form the I-motif. To explore this possibility, we similarly modified the *BCL2* and *MYC* I-motif sequences as previously discussed for the G-quadruplex sequences (Table 2) and tested the effect of AID on spectra and melting temperature at the same increasing concentration. We also decided to test an optimized EMSA buffer. An EMSA buffer was used instead of the normal sodium cacodylate buffer because previous CD studies demonstrated interaction between an I-motif and a nuclear protein under these conditions. Furthermore, the G-quadruplex forms readily in a TrisHCl buffer whereas the I-motif requires a sodium cacodylate buffer and an acidic pH to form in solution. However, both the *MYC* and *BCL2* I-motif are quite stable and can form readily at a pH of 6.6.

The I-motif structure has a different maxima than the G-quadruplex making it easy to distinguish between the two structures. A characteristic I-motif spectrum has a maxima at 286nm and a minima at 264nm.

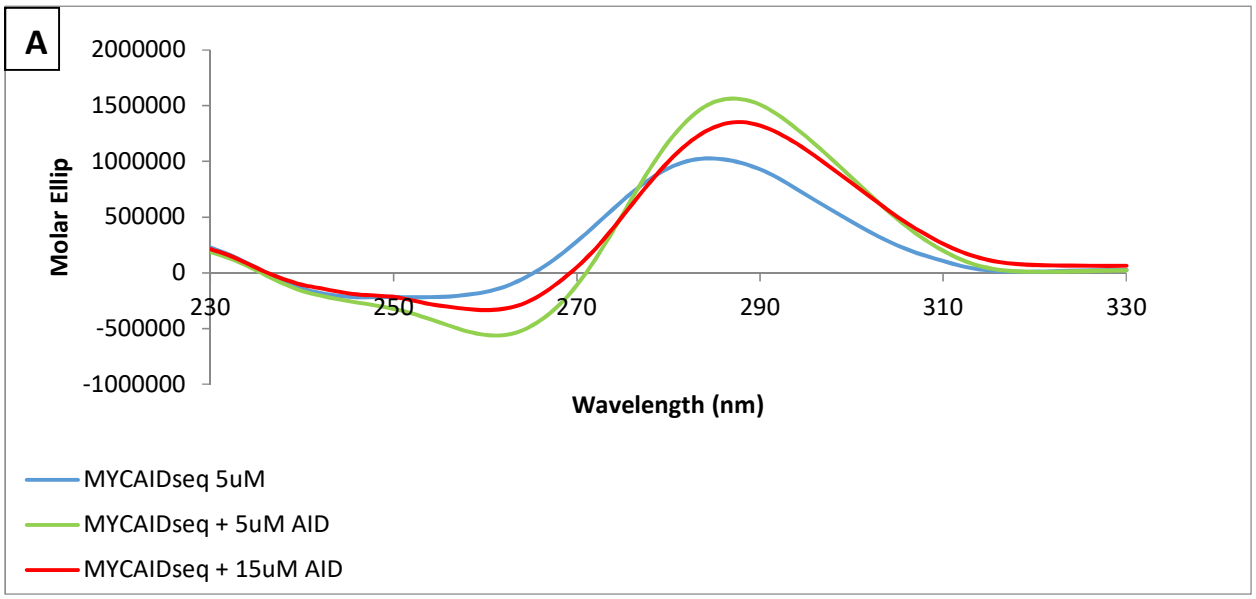
The *MYC* I-motif CD spectra data showed a shift to the right towards 290nm and an increase in maxima (Figure 13A). This indicates there is a change in structure upon protein addition and a shift in the equilibrium towards I-motif formation at the same DNA concentration. However, this is in contrast to the melting temperature studies where there was a protein dependent decrease in thermal stability as more protein was added and indicates that AID is binding and destabilizing the I-motif structure (Figure 13B).

Unlike our observations with the *MYC* I-motif in the presence of AID, the *BCL2* I-motif CD spectra data did not show any shift in peak maxima (Figure 14A). With no shift in the maxima or a significant increase in peak of the maxima, we presume there is no change in structure or relative amount of I-motif folding in solution. The melting temperature data shows there is a destabilizing shift with a lower melting temperature at the 1:1 and 1:3 DNA: protein ratio when compared to the free DNA (Figure 14B). However, this was not a concentration dependent decrease in thermal stability as was seen in the *MYC* I-motif data. The 1:1 and 1:3 ratios showed very similar melting temperatures, which could mean that less protein is needed in order to see an effect on thermal stability. With that said, the decrease in thermal stability was only 2°C, which is not considered a significant difference when compared to the *MYC* data which had a 6 °C difference between the free DNA and the 1:3 DNA to protein ratio.

4.5.1. Table 2: MYC and BCL2 I-motif Sequences

MYC I-motif:
5’- CCCACCCTCCCCACCCTCCCCA**TAAGCGCCCCTCCCGGGTTC**-3’ (Py27+ AID sequence)
5’- CCTTCCCCACCCTCCCCACCCTCCCCA -3’ (Py27)
BCL2 I-motif:
5’- **AGGCACCA**CCGCCCCGCTCCCGCCCCCTTCCTCCCGCGCCCGCCCCT-3’ (Py39+ AID sequence)
5’- CCGCCCCGCTCCCGCCCCCTTCCTCCCGCGCCCGCCCCT-3’ (Py39)

4.5.2. Figure 13: CD data of *MYC* I-motif with AID



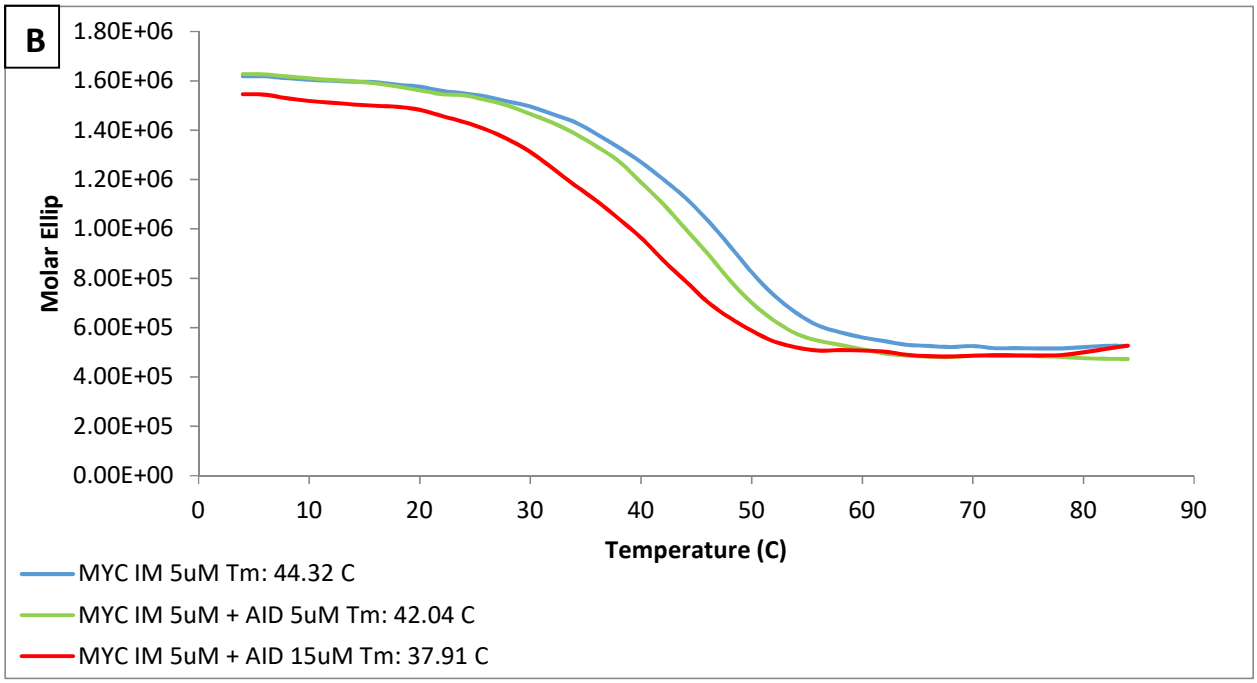
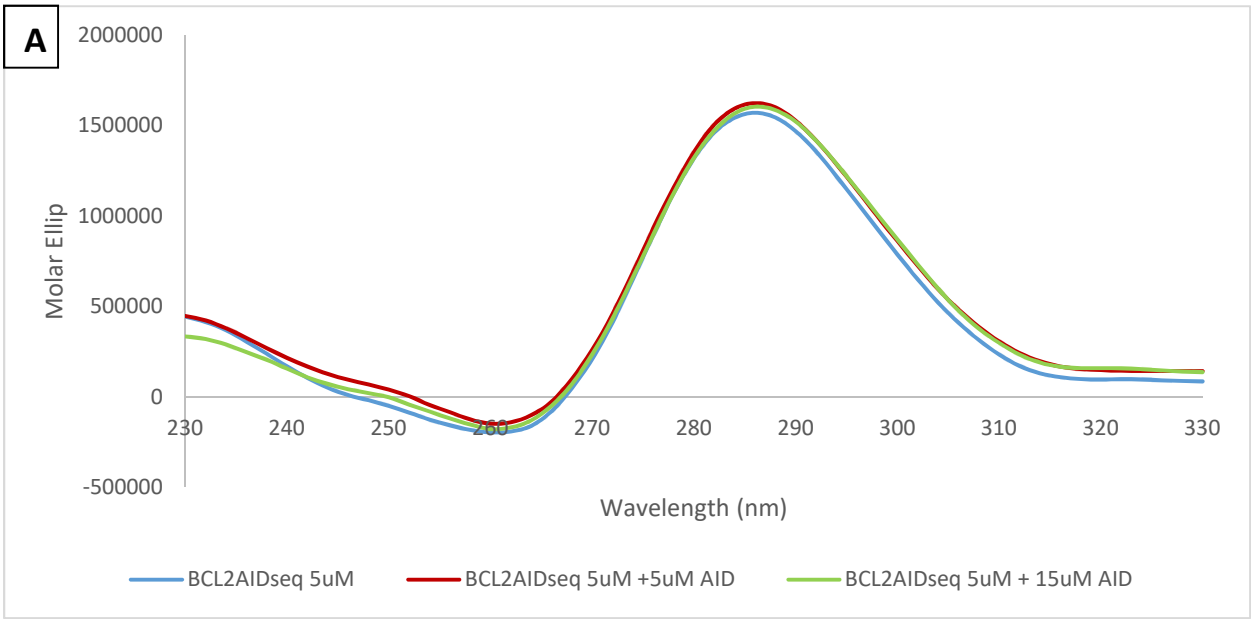


Figure 13: (A) CD spectrum of 5uM *MYC* I-motif DNA with AID WRCY sequence in EMSA Buffer with 25mM potassium ion with varying protein concentrations. (B) CD T-melt spectra of 5uM *MYC* i-motif DNA with AID WRCY sequence in EMSA buffer with 75mM potassium ion with varying protein concentrations. EMSA buffer contains 20mM Hepes, 2mM $MgCl_2$, 1mM EDTA, 1mM DTT, 0.1ug/uL BSA, 0.1% Tween 20, 10% Glycerol.

4.5.3. Figure 14: CD data of *BCL2* I-motif with AID



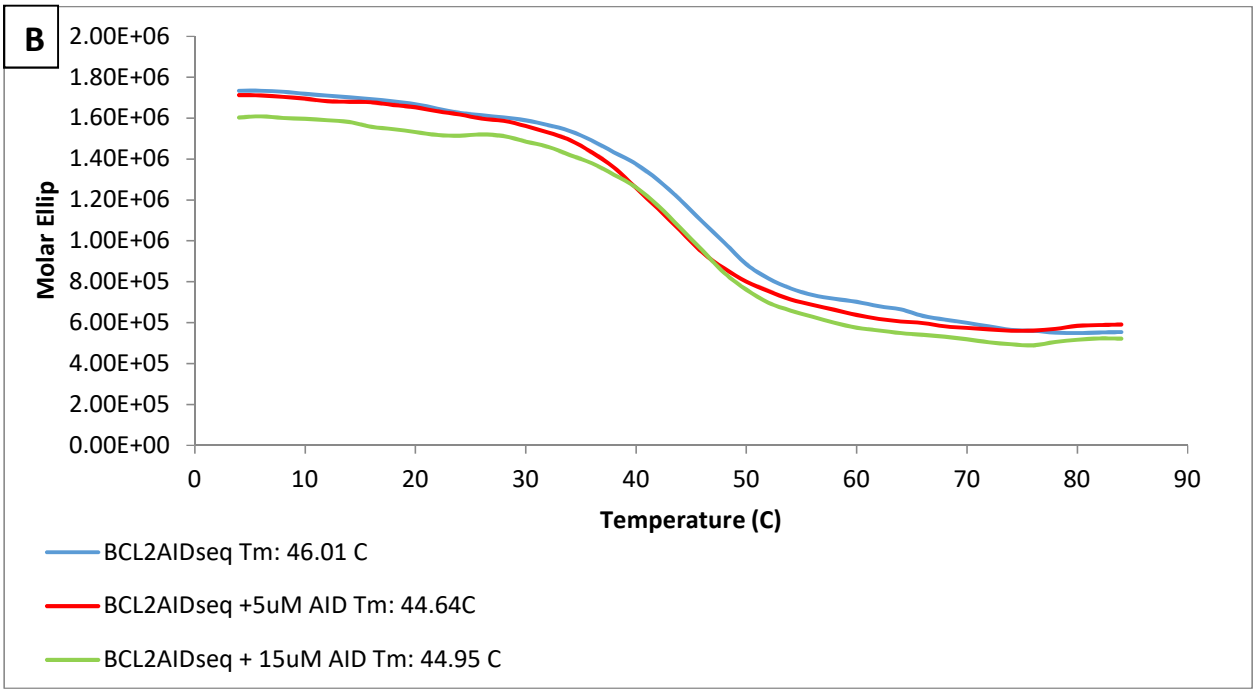


Figure 14: (A) CD spectrum of 5uM *BCL2* i-motif DNA with AID WRCY sequence in EMSA Buffer with 75mM potassium ion with varying protein concentrations. (B) CD T-melt spectra of 5uM *BCL2* i-motif DNA with AID WRCY sequence in EMSA buffer with 75mM potassium ion with varying protein concentrations. EMSA buffer contains 20mM Hepes, 2mM $MgCl_2$, 1mM EDTA, 1mM DTT, 0.1ug/uL BSA, 0.1% Tween 20, 10% Glycerol.

4.6 AID does not directly bind the MYC or BCL2 G-quadruplex and I-motif as shown with EMSA

The use of the Electrophoresis Motility Shift Assay (EMSA) native non-denaturing gel allows for the visualization of AID interaction with the G-quadruplex and I-motif. When a protein interacts with a specific element of DNA a shift occurs in which the complex formed between DNA and protein runs with less mobility than the free DNA on the gel. The optimal binding buffer is one of the most important components to producing accurate results for EMSA. Different buffers were tested for the G-quadruplex and I-motif to determine optimal conditions. The G-quadruplex buffer was the same buffer as previously used during the I-motif CD experiments, a TrisHCl at pH 7.4 buffer with 20mM potassium ion concentration. The second buffer tested was based on previous studies performed with the *BCL2* I-motif and I-motif interactive protein hnRNP LL (24). This EMSA buffer contains 20mM Hepes, 2mM 2mM MgCl2, 1mM EDTA, 1mM DTT, 0.1ug/uL BSA, 0.1% Tween 20, and 10% Glycerol.

If AID interacts with the DNA, but does not alter the conformation stability as we did not observe any noteworthy changes in CD analyses, there should be two bands upon concentration dependent addition of AID: a lower band representing un-complexed DNA and a slower migrating band above representing AID complexed with the DNA. In addition, at higher concentrations of AID, the lower DNA only band should disappear because all of the DNA should be bound to AID. These bands were visualized using SYBR green to stain the DNA. From the data obtained, there was only one band on the gel even after addition of AID to the *BCL2* and *MYC* G-quadruplex and I-motif (Figures 15 and 16). This means that with the current

experimental conditions used here, there is no binding of AID to the *BCL2* and *MYC* G-quadruplex and I-motif.

4.6.1. Figure 15: EMSA data of *MYC* G-quadruplex and I-motif with AID

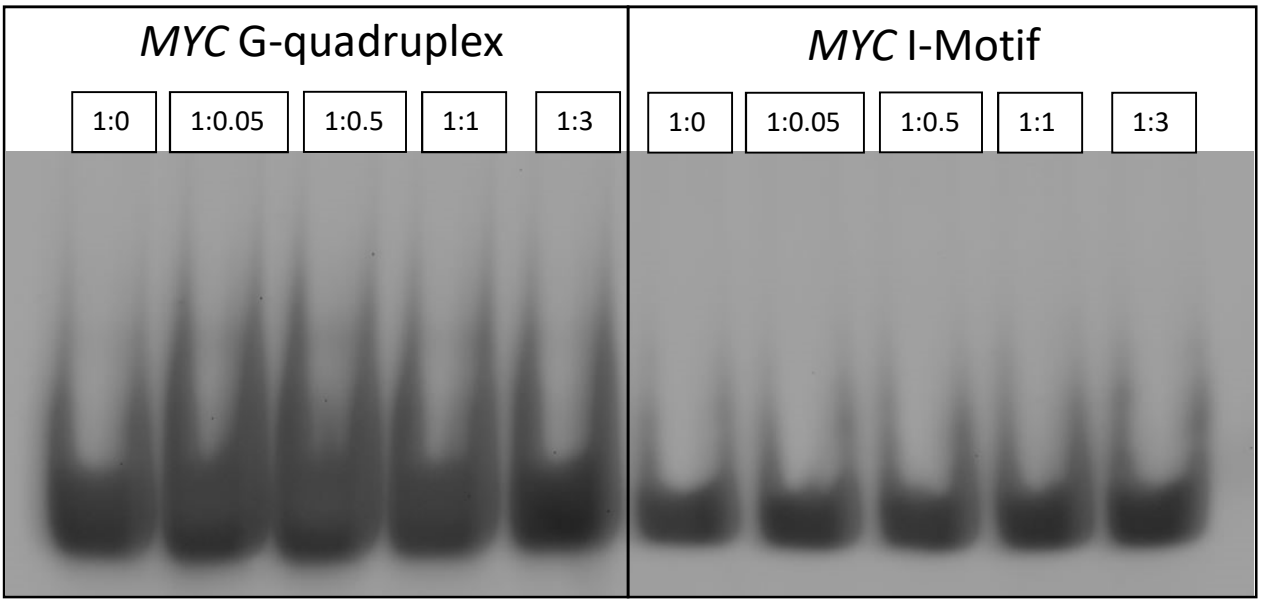


Figure 15: The ratio above each lane is DNA: AID. The *MYC* G-quadruplex is shown on the left whereas the *MYC* I-motif is shown on the right.

4.6.2. Figure 16: EMSA data of *BCL2* G-quadruplex and I-motif with AID

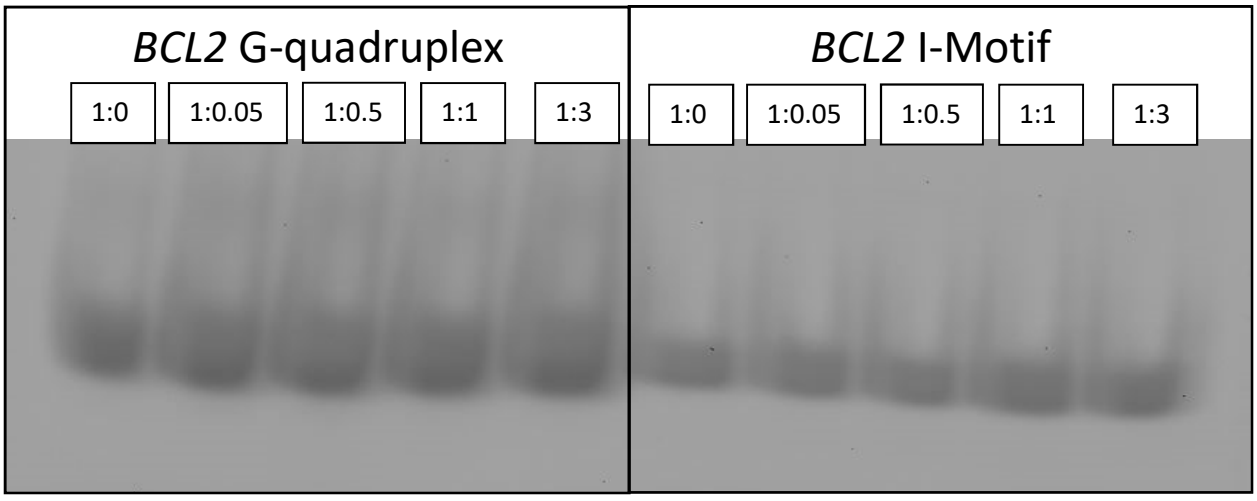


Figure 16: The ratio above each lane is DNA: AID. The *BCL2* G-quadruplex is shown on the left whereas the *BCL2* I-motif is shown on the right.

5. Discussion

The purpose of this project was to determine whether AID recombinase binds to either the *MYC* or *BCL2* G-quadruplex. The in-silico data suggested a commonality between known AID targeting regions of the genome and potential G-quadruplex forming sequences; SE regions, ConvT areas, and complete off target areas. In the mouse activated B-cell, 223 out of 226 of AID targets had the potential to form G-quadruplex structures and the Ramos cell line all of the AID targets demonstrated the potential to form G-quadruplexes. In addition, this data is supported by a study, which showed that AID binds to G-loop secondary structures in the *MYC* promoter region. The fact that all of these AID targeting sequences with and without SE regions or ConvT areas almost all have the potential to form G-quadruplexes provides convincing evidence to support further investigation.

The GC-rich *BCL2* and *MYC* P1 promoters are known to form G-quadruplexes upstream of the transcriptional start site and within regions of SEs, and AID WRCY sequence response elements. Determining this possible interaction between AID and these G-quadruplex structures will not only provide a mechanism by which AID can cause genomic instability in DLBCL, but will provide the scientific community with a molecular mechanism by which AID is recruited to DNA to exert its deaminase activity.

The first portion of the project addressed whether the previously studied *BCL2* and *MYC* G-quaduplex folding topology and thermal stability was similar to the longer nucleotide sequence needed to address the interaction of AID with WRCY hotspots. Our data demonstrates that the BCL2 and MYC sequences that are in the areas of AID WRCY targeting are capable of forming G-

quadruplexes and I-motifs, suggesting these structures can form where AID is present in the genome.

We then determined whether AID had an effect on the structure of the G-quadruplex using the CD spectra and to determine whether AID had an effect on thermal stability of the G-quadruplex using the CD melting temperature spectra. The data from both the CD spectra and the CD melting temperature spectra indicates there may not be a direct interaction between AID and the *BCL2* or *MYC* G-quadruplex structures.

We then considered the possibility that AID, due to its deaminase activity of cytosine could be interacting with the C-rich strand which lead us to expand our study to include the *BCL2* and *MYC* I-motif. The *BCL2* and *MYC* I-motif CD data both showed a destabilizing effect upon AID addition in a concentration depended manor. This result supports the idea that there might be a possible interaction between AID and the I-motif. However, I-motif CD data must be supported with EMSA studies to demonstrate direct interactions with AID.

An additional sample to include in future experiments is a positive control to address whether the buffer conditions and oligonucleotides used in this study confirm previously investigated protein-G-quadruplex interactions. For example, nucleolin is known to bind and stabilize the *MYC* G-quadruplex whereas NM23H2 has been shown to destabilize the *MYC* G-quadruplex. An example of a protein binding to the I-motif is hnRNP LL, which has been shown to interact with the *BCL2* I-motif. A positive control using these proteins with the longer *MYC* or *BCL2* sequences, which contain the AID WRCY motif, would help address the structural nature of these G-quadruplex or I-motif sequences. If the longer sequences are shown not to interact

with these proteins it is possible that the structures formed from these sequences has adopted a slightly different conformation that prevents recognition by the known interactive proteins and possibly AID. If there is binding within the positive controls then it would mean that AID does not bind directly to the structures and if there is no binding the buffer conditions used would not be optimal. Alternatively, there is a possibility that AID requires an even longer sequence in order to bind.

Another aspect to consider, is ensuring the recombinant AID we received is active and has retained its DNA-binding abilities. Although, this protein was purchased from a reliable company and recommended for use in western blot and gel shift assays, we should still rule this out as a contributing factor the negative results we observed.

AID may also require a co-protein in order to bind these secondary structures. Interestingly enough, previous research using co-precipitation has shown that the *BCL2* G-quadruplex co-precipitates with Eukaryotic translation elongation factor 1 alpha (EIF1a), a protein known to interact with AID. This is quite strange due to the fact that EIF1a is normally found in the cytoplasm to interact with RNA whereas the *BCL2* G-quadruplex is found within the nucleus. However, interaction within the nucleus might occur before EIF1a is exported, but thus far no additional functions of EIF1a are known. In order to provide further insight, co-treatment experiments could be run to determine whether AID interaction is enhanced with the *MYC* or *BCL2* secondary structures and address whether AID-associated proteins such as EF1a is able to facilitate AID binding or association to DNA.

Our studies provide preliminary data to support the continued investigation of the role of DNA secondary structures in AID targeting throughout the genome and possibly suggest that I-motifs may be involved. If there is an interaction between AID and these DNA secondary structures, this may indicate the molecular mechanism behind how AID is recruited to DNA. In addition, it could potentially lead to a new, targetable mechanism to prevent genomic instability within DLBCL and other lymphomas. Small molecule compounds identified through high throughput screening of known compounds to interact with the *BCL2* and *MYC* secondary structures and compound libraries could be used to bind to the G-quadruplex in such a way that would inhibit AID from binding and mutagenic activity. Due to the fact that the *MYC*, *BCL2,* and other known G-quadruplexes have distinct structures, this approach could be a selective method for targeting each oncogene specifically without disrupting normal AID function at the immunoglobulin loci.

6. References

1. Duensing, Stefan, and Karl Manger. "Mechanisms of Genomic Instability in Human Cancer: Insights from Studies with Human Papillomavirus Oncoproteins."International Journal of Cancer 109.2 (2003): 157-62.
2. Negrini, Simona, Vassilis G. Gorgoulis, and Thanos D. Halazonetis. "Genomic Instability an Evolving Hallmark of Cancer." Nature Reviews Molecular Cell Biology 11.3 (2010): 220-28. Web.
3. Pasqualucci L, Guglielmino R, Houldsworth J, et al. Expression of the AID protein in normal and neoplastic B cells. Blood 2004; 104:3318-25.
4. KhodaBAKhshi AH, Morin RD, Fejes AP. Recurrent targets of aberrant somatic hypermutation in lymphoma. Oncotarget 2012; 3:1308-19.
5. Morin RD, Mungall K, Pleasance E, et al. Mutational and structural analysis of diffuse large B-cell lymphoma using whole genome sequencing. Blood 2013; 122:2156-65.
6. Lohr JG, Stojanov P, Lawrence MS, et al. Discovery and prioritization of somatic mutations in diffuse large B-cell lymphoma (DLBCL) by whole-exome sequencing. Proc Natl Acad Sci USA 2012; 109:3879-84.
7. Schuetz JM, Johnson NA, Morin RD, et al. *BCL2* mutations in diffuse large B-cell lymphoma. Leukemia 2012; 26:1383-90.
8. Hackney JA, Misaghi S, Sendger K, Garris C, Sun Y, Lorenzo MN, Zarrin AA. DNA targets of AID: Evolutionary link between antibody somatic hypermutation and class switch recombination. Advances in Immunology 2009; 101:163-189.
9. Qian J, Wang Q, Dose M, et al. B cell super-enhancers and regulatory clusters recruit AID tumorigenic activity. Cell 2014; 159:1-14.
10. Meng FL, Du Z, Federation A, et al. Convergent transcription at intragenic super-enhancers targets AID-initiated genomic instability. Cell 2014; 159:1538-48.
11. Alinikula J and Schatz DG. Super-Enhancer Transcription Converges on AID. Cell 2014; 159:1490-92.
12. Duquette ML, Huber MD, Maizels N. G-rich proto-oncogenes are targeted for genomic stability in B-cell lymphomas. Cancer Res 2007; 76:2586-94.
13. Duquette, ML, Pham P, Goodman MF, Maizel N. AID binds to transcription-induced structures in *MYC* that map to regions associated with translocation and hypermutation. Oncogene 2005; 24:5791-8.
14. Johnson NA, Slack GW, Savage KJ, et al. Concurrent expression of *MYC* and *BCL2* in diffuse large B-cell lymphoma treated with rituximab plus cyclophosphamide, doxorubicin, vincristine, and prednisone. J Clin Oncol 2012; 30:3452-9.
15. You, Huijuan, Jingyuan Wu, Fangwei Shao, and Jie Yan. "Stability and Kinetics of c-*MYC* Promoter G-Quadruplexes Studied by Single-Molecule Manipulation." Journal of the American Chemical Society 137.7 (2015): 2424-427. Web.

16. Brooks, T. A., and L. H. Hurley. "Targeting *MYC* Expression through G-Quadruplexes." Genes & Cancer 1.6 (2010): 641-49. Web.
17. Yang, Danzhou, and Laurence H. Hurley. "Structure of the Biologically Relevant G-Quadruplex in The c-*MYC* Promoter." Nucleosides, Nucleotides and Nucleic Acids 25.8 (2006): 951-68. Web.
18. Gonzalez, V., K. Guo, L. Hurley, and D. Sun. "Identification and Characterization of Nucleolin as a c-*MYC* G-quadruplex-binding Protein." Journal of Biological Chemistry 284.35 (2009): 23622-3635. Web.
19. Dai, Jixun, Emmanuel Hatzakis, Laurence H. Hurley, and Danzhou Yang. "I-Motif Structures Formed in the Human c-*MYC* Promoter Are Highly Dynamic–Insights into Sequence Redundancy and I-Motif Stability." PLoS ONE 5.7 (2010): n. pag. Web.
20. Yang, D.; Okamoto, K. Structural Insights Into G-Quadruplexes: Towards New Anticancer Drugs. Future Medicinal Chemistry 2010, 2, 619–646.
21. Dai, Jixun, Thomas S. Dexheimer, Ding Chen, Megan Carver, Attila Ambrus, Roger A. Jones, and Danzhou Yang. "An Intramolecular G-Quadruplex Structure with Mixed Parallel/Antiparallel G-Strands Formed in the Human BCL-2 Promoter Region in Solution." Journal of the American Chemical Society 128.4 (2006): 1096-098. Web.
22. Dexheimer, Thomas S., Daekyu Sun, and Laurence H. Hurley. "Deconvoluting the Structural and Drug-Recognition Complexity of the G-Quadruplex-Forming Region Upstream of thebcl-2P1 Promoter." Journal of the American Chemical Society 128.16 (2006): 5404-415. Web.
23. Dai, Jixun, Ding Chen, Roger A. Jones, Laurence H. Hurley, and Danzhou Yang. "NMR solution structure of the major G-quadruplex structure formed in the human *BCL2* promoter region." Nucleic Acids Research 34.18 (2006): 5133-144. Web.
24. Kendrick, Samantha, Yoshitsugu Akiyama, Sidney M. Hecht, and Laurence H. Hurley. "The i-Motif in thebcl-2P1 Promoter Forms an Unexpectedly Stable Structure with a Unique 8:5:7 Loop Folding Pattern." Journal of the American Chemical Society 131.48 (2009): 17667-7676. Web.
25. Kang, Hyun-Jin, Samantha Kendrick, Sidney M. Hecht, and Laurence H. Hurley. "The Transcriptional Complex Between the*BCL2*i-Motif and hnRNP LL Is a Molecular Switch for Control of Gene Expression That Can Be Modulated by Small Molecules." Journal of the American Chemical Society 136.11 (2014): 4172-185. Web.
26. Woski, Stephen. "The Dynamic Character of the *BCL2* Promoter i-Motif Provides a Mechanism for Modulation of Gene Expression by Compounds That Bind Selectively to the Alternative DNA Hairpin Structure." F1000 - Post-publication peer review of the biomedical literature (2014): n. pag. Web.