Genomic Deoxyuracil: Targeting, Regulation And Repair

Nate Warren Holton
Illinois State University, nwholto@ilstu.edu

Follow this and additional works at: https://ir.library.illinoisstate.edu/etd
Part of the Molecular Biology Commons

Recommended Citation
https://ir.library.illinoisstate.edu/etd/468
Maintaining genomic integrity is critical to avoid sequence mutations and genetic disease. However the immune system requires genomic rearrangements in order to produce novel antibodies and elicit a full immunological response. The expression of DNA modifying enzymes is essential for antibody production but the off target effects are not well understood. The work presented here was performed to elucidate molecular mechanisms responsible for the regulation and targeting of DNA editing enzymes, as well as characterize repair of these lesions at off target loci.

KEYWORDS: Activation Induced Cytosine Deaminase, AID, miRNA, miR155, miR93, G-quadruplex, G4, Uracil DNA Glycosylase, UDG, UNG
Copyright 2015 Nate W. Holton
GENOMIC DEOXYURACIL: TARGETING, REGULATION AND REPAIR

NATE W. HOLTON

COMMITTEE MEMBERS:

Erik Larson, Chair
John Sedbrook
Brian Wilkinson
Kevin Edwards
Laura Vogel
CONTENTS

CONTENTS i

TABLES v

FIGURES vi

CHAPTER

I. URACIL IN THE GENOME: MUTAGENIC THREAT AND IMMUNOLOGICAL NECESSITY 1

References 8
Figure 11

II. HISTONES H2A AND H2B ARE MONOUBIQUITINATED AT AID-TARGETED LOCI 12

Abstract 13
Introduction 14
Results 17

AID and Hypermutation Target Two Discrete Loci in Ramos 17
Chromatin at the Re-VH and c-MYC Gene in Ramos B Cells is Monoubiquitinated 20
The Chromatin of Ig Genes in Activated Primary Murine B Cells is Ubiquitinated 22
AID and MUb-H2B Colocalize in Discrete Nuclear Foci 23

Discussion 24
Material and Methods 29

Cells and Cell Lines 29
Chromatin Immunoprecipitations 30
Western Blotting 30
III. REPRESSION OF HUMAN ACTIVATION INDUCED CYTIDINE DEAMINASE BY MIR-93 AND MIR-155

Abstract 51
Introduction 52
Results 55

The Human AID 3'UTR Contains Sequences Complementary to MiR-93 and MiR-155 55
MiR-93 and MiR-155 Interact with the AID 3'UTR to Inhibit its Expression 56
MiR-93 and MiR-155 can Repress Endogenous AID 58
Endogenous MiR-93 and MiR-155 Restrain AID Protein Translation 59

Discussion 60

Model for AID Repression 60
Disruptions of MiR-155 and MiR-93 are Associated with AID-Induced Oncogenesis 61
MicroRNA Target Sites Obscured by Repetitive Element Insertions 62

Conclusion 63
Materials and Methods 63

Reagents and Cell Lines 63
Luciferase Assays 64
Western Blotting 64
Vector Construction 65

References 67
Figures 73
Table 81
# TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Oligonucleotide Table</td>
<td>48</td>
</tr>
<tr>
<td>2. Oligonucleotide Master List</td>
<td>81</td>
</tr>
<tr>
<td>3. DNA Oligonucleotide Sequence</td>
<td>127</td>
</tr>
</tbody>
</table>
FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Cytosine Deamination and the Outcomes of Uracil DNA Repair</td>
<td>11</td>
</tr>
<tr>
<td>2.</td>
<td>Diagram of the IgH Loci in Ramos Burkitt's Lymphoma</td>
<td>40</td>
</tr>
<tr>
<td>3.</td>
<td>Ramos Burkitt's Lymphoma Cells Constitutively Hypermutate at Distinct Genomic Loci</td>
<td>41</td>
</tr>
<tr>
<td>4.</td>
<td>ChIPs Identify MUb-H2A and MUb-H2B at Ramos Hypermutating Loci</td>
<td>42</td>
</tr>
<tr>
<td>5.</td>
<td>MUb-H2A and MUb-H2B Associate with Hypermutation in Activated Mouse Primary B Cells</td>
<td>43</td>
</tr>
<tr>
<td>6.</td>
<td>AID and Monoubiquitinated H2B Colocalize in Discrete Nuclear Foci</td>
<td>44</td>
</tr>
<tr>
<td>7.</td>
<td>PCNA and Polymerase Eta are Associated with the Ramos Re-VH Region</td>
<td>45</td>
</tr>
<tr>
<td>8.</td>
<td>HEK293 and A549 ChIPs Using AID, MUb-H2A and MUb-H2B Antibodies Find No Enrichment of Sequences Undergoing Hypermutation in Ramos cells</td>
<td>46</td>
</tr>
<tr>
<td>9.</td>
<td>AID Antibody Used for ChIPs is Specific for AID</td>
<td>46</td>
</tr>
<tr>
<td>10.</td>
<td>AID IP Demonstrating AID Antibody Used for ChIPs is Detecting AID</td>
<td>47</td>
</tr>
<tr>
<td>11.</td>
<td>MUb-H2b Colocalizes with AID in Multiple Discrete Foci</td>
<td>47</td>
</tr>
<tr>
<td>12.</td>
<td>The AID 3'UTR</td>
<td>73</td>
</tr>
<tr>
<td>13.</td>
<td>The AID 3’ UTR is Recognized by MiRs-93 and -155</td>
<td>74</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
<td></td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>14. MiR-93 and MiR-155 Can Repress Endogenous AID</td>
<td>75</td>
<td></td>
</tr>
<tr>
<td>15. Endogenous MiR-93 and MiR-155 Restrain AID Protein Expression</td>
<td>76</td>
<td></td>
</tr>
<tr>
<td>16. Model of MiR-93 and MiR-155 Regulation of AID Mediated Genomic Instability</td>
<td>77</td>
<td></td>
</tr>
<tr>
<td>17. The AID 3'UTR is Recognized by MiRs -93 and -155</td>
<td>78</td>
<td></td>
</tr>
<tr>
<td>18. MicroRNAs MiR-93 and MiR-155 are Widely Expressed</td>
<td>79</td>
<td></td>
</tr>
<tr>
<td>19. MiR-93 Represses Endogenous AID in a Dose Responsive Manner</td>
<td>80</td>
<td></td>
</tr>
<tr>
<td>20. Sequences Used and Diagrams of G4 Structures</td>
<td>106</td>
<td></td>
</tr>
<tr>
<td>21. Detection of G4 Formation</td>
<td>107</td>
<td></td>
</tr>
<tr>
<td>22. G-Tetrads in TCF3 G4 Interfere with UDG Activity</td>
<td>108</td>
<td></td>
</tr>
<tr>
<td>23. G-Tetrads in TCF3 G4 Interfere with hUNG2 Activity</td>
<td>109</td>
<td></td>
</tr>
<tr>
<td>24. G-Tetrads in TCF3 G4 Interfere with the Cleavage Step of Base Excision Repair</td>
<td>110</td>
<td></td>
</tr>
<tr>
<td>25. LiCl Does Not Inhibit UDG Activity</td>
<td>111</td>
<td></td>
</tr>
<tr>
<td>26. LiCl Does Not Inhibit hUNG2 Activity</td>
<td>112</td>
<td></td>
</tr>
<tr>
<td>27. Native DNA Structures</td>
<td>124</td>
<td></td>
</tr>
<tr>
<td>28. Bacterial Uracil DNA Glycosylase Activity on Single-Strand and G4 Folded Substrates</td>
<td>125</td>
<td></td>
</tr>
<tr>
<td>29. SMUG1 Activity on Single-Strand and G4 Folded Substrates</td>
<td>126</td>
<td></td>
</tr>
</tbody>
</table>
CHAPTER I

URACIL IN THE GENOME: MUTAGENIC THREAT AND IMMUNOLOGICAL NECESSITY
Genomic DNA is constantly bombarded by a host of agents that modify the nucleic acid chemical structure, resulting in non-canonical base pairing and DNA damage. Threats to genomic DNA come from both external sources and also from naturally occurring cellular metabolic processes. Some examples of exogenous sources of DNA damage are UV, ionizing radiation, and many DNA modifying chemicals. Endogenous sources of DNA modifications are products of cellular metabolism such as oxidative phosphorylation, lipid peroxidation, DNA replication errors, and enzymatic processing (McCullough et al., 1999). These DNA lesions are potentially mutagenic and must be repaired prior to replication in order to maintain the integrity of the genome and to avoid genetic diseases.

One common DNA lesion that must be recognized and repaired in the genome is deoxyuracil. Spontaneous cytosine deamination (Figure 1A) to uracil is estimated to produce 100-500 uracil residues per cell, per day, with estimates orders of magnitude higher in transcriptionally active regions of the genome and in rapidly dividing cells (Lindahl, 1993). This is due to a high proportion of the genome adapting a single strand nature and are thus more susceptible to base modifications (Lindahl, 1993). Cytosine deamination leads to a U:G pairing and if left unrepaired result directly in a cytosine to thymine (C\(\rightarrow\)T) transition mutations (Olinski et al., 2010). Beyond spontaneous deamination, a cellular pathway for antibody diversification utilizes activation induced cytidine deaminase (AID) to enzymatically modify cytosine to uridine at the immunoglobulin (Ig) variable (V) locus within somatic B cells (Ganesh and Neuberger, 2011; Kato et al., 2012; Muramatsu et al., 2000; Wang, 2013). The manipulation of DNA repair in these B
cells leads to the uracil lesion not being faithfully repaired, resulting in novel antigen binding affinity (Figure 1B). This process is known as somatic hypermutation (SHM) (Chahwan et al., 2012). In humans AID is also required for the modification of the Ig heavy chain locus, which differentiates antibodies into their five different classes (IgM, IgD, IgG, IgA, and IgE). This process, known as class switch recombination (CSR), permits the cellular clearance of antibodies into restricted tissues and spaces within the body by performing programmed genomic translocations (Figure 1B) within the Ig locus (Chaudhuri et al., 2007). AID induced SHM and CSR are powerful adaptive immune tools however their processes must be tightly controlled because off target gene modifications, outside of the Ig locus, can lead to genomic instability and genetic disease (Nussenzweig and Nussenzweig, 2010).

Targeting of AID to the Ig locus is not well understood, however there is evidence that monoubiquitination of multiple proteins is an important regulator of antibody diversification in model systems (Arakawa et al., 2006; Langerak et al., 2007; Roa et al., 2008). Given that DNA repair enzymes are involved in SHM and CSR at the Ig loci, and genome wide maintenance, we speculate that histone modifications, in particular monoubiquitination, accompany hypermutation at the Ig locus. Chapter II: Histones H2A and H2B are Monoubiquitinated at AID-Targeted Loci, provides evidence that, in humans (RAMOS) and mice, AID targets the Ig loci for SHM at least in part because of monoubiquitinated histones H2A and H2B. These results demonstrate for the first time a novel chromatin architecture being associated with SHM and genomic instability and show that
monoubiquitination of histones H2A and H2B act as molecular beacons for AID targeting.

The essential role of AID on Ig gene diversification has led to the postulation that AID has high mutagenic potential at off target sequences. AID was once thought to be expressed only in B cells. However, it appears to be misexpressed in several non-lymphoid epithelial cancers and is a primary contributor to the observed genomic instability (Babbage et al., 2006; Endo et al., 2011; Matsumoto et al., 2010; Shinmura et al., 2011). This high mutagenic threat argues for the existence of several molecular mechanisms to regulate AID expression outside of activated B cells. One such level of posttranslational AID control is accomplished through the expression of microRNA’s that target and prevent translation of the AID mRNA transcripts. In Chapter III: Repression of Human Activation Induced Cytidine Deaminase by miR-93 and miR-155, the 3’ UTR of AID is characterized for the ability of miRNAs to regulate AID expression. Here we find that miR-155 and miR-93 regulate AID misexpression in the MCF-7 breast cancer cell line and loss of either miRNA induces increased AID protein levels. The results here also provide evidence that the 3’ untranslated regions of mRNA contain sequence elements that make identifying miRNA target sites difficult and more care must be taken when publishing and annotating full length mRNA transcripts.

Deaminated cytosines created enzymatically by the action of AID or by spontaneous deamination occur throughout the genome and lead to DNA transition (C→T) mutations, unless the uracil intermediate is repaired (Maul and
The uracil DNA glycosylase family of enzymes is responsible for recognizing genomic uracil lesions and initiating repair (Krokan et al., 2014). There are four uracil DNA glycosylase enzymes in mammalian cells: 1) uracil-DNA glycosylase (UNG) is expressed in two isoforms with UNG1 in the mitochondria and UNG2 in the nuclear fraction, 2) single strand selective monofunctional uracil DNA glycosylase (SMUG1), 3) thymine DNA glycosylase (TDG), and 4) methyl-binding domain 4 protein (MBD4). It remains unclear how the uracil glycosylase family of enzymes share in the uracil processing responsibilities however UNG and SMUG1 are primarily responsible for the repair of deaminated cytosines and they are not redundant because both enzymes are required to limit the accumulation of DNA sequence mutations (Hagen et al., 2006).

Much is understood about uracil DNA glycosylase processing on single strand and duplex DNA. Even so, the human genome is composed of many structural DNA conformations that deviate from the traditional Watson-Crick B-form double helix. Of particular interest are G quadruplex (G4) DNA structures as they are associated with programmed recombination of the Ig locus. In addition, some G4 capable sequences are associated with mutagenesis and genetic disease however the mechanism leading to this instability is not understood (Maizels, 2015). These observations led us to explore DNA repair abilities of uracil glycosylase enzymes near G4 capable DNA sequences. G4 DNA structures differ from Watson-Crick DNA in that G4 is a four stranded stable conformation forming from tandem guanine repeats under physiological salt and
pH conditions, resulting in a tetrad of four guanines participating in alternative DNA hydrogen bonds centered around a monovalent cation (Sen and Gilbert, 1988, 1990). Sequences capable of adopting putative G4 DNA conformations are found throughout the human genome and many show up in untranslated regions of oncogenes where they appear to effect transcriptional efficiency (Eddy and Maizels, 2006; Huppert and Balasubramanian, 2005; Todd et al., 2005). In Chapter IV: G-Quadruplex DNA Structures Can Interfere With Uracil Glycosylase Activity In Vitro, the major uracil glycosylase enzymes from E. coli and humans are characterized for their ability to process uracil lesions near guanine repeats involved in G4 DNA structure formation. Here we show that uracils positioned directly 5’ or 3’ of guanine tetrads involved in G4 structural bonds have inhibited uracil glycosylase processing activity. These results provided a potential mechanistic explanation as to why guanine rich sequences are prone to instability and demonstrate that uracil DNA glycosylase enzymes are poor initiators of repair at these loci.

In short, the work presented herein is focused on the targeting and regulation of the uracil inducing enzyme AID and the consequences of aberrant genomic uracil repair near guanine rich sequences. Chapter II identifies monoubiquitinated histones H2A and H2B as key markers for AID targeting and Ig gene diversification. Chapter III establishes both miR-93 and miR-155 as capable of regulating post-transcriptional AID expression and further characterizes the known AID 3’ untranslated region. Chapter IV explores the repair of deaminated cytosines near G4 capable DNA and demonstrates that G4
provides a unique molecular architecture and uracil located directly adjacent to guanines in G4 DNA are poorly processed by uracil glycosylase enzymes. Finally, Chapter V unifies the above findings and clarifies the future work that is needed to clarify the dynamic role that genomic uracil plays. All together this work is important because it uncovers fundamental mechanisms of mutagenesis, which are required to define the molecular sources of genomic instability.
REFERENCES


Figure 1. Cytosine Deamination and the Outcomes of Uracil DNA Repair. (A) Chemical structure of cytosine (left) and the deamination product uracil (right). (B) Outcomes of uracil repair on the genome (left) and at the Ig locus (right). Spontaneous and AID induced deaminations convert cytosine to uracil. Genomic uracil can be fully repaired (bold) or the mutations cause C→T transition mutations (grey). The Ig locus undergoes SHM and aberrant DNA repair of the deoxyuracils, which alters the nucleotide sequence for novel antigen binding (grey). During CSR the Ig locus performs DNA translocations between switch regions to generate mature antibodies of different classes (outlined).
CHAPTER II

HISTONES H2A AND H2B ARE MONOUBIQUITINATED AT AID-TARGETED LOCI

Contributions:

The contents of this chapter were previously published and cited below. My contribution involves experimentation excluding the microscopy with a minor contribution to manuscript preparation and editing.

ABSTRACT

Somatic hypermutation introduces base substitutions into the rearranged and expressed immunoglobulin (Ig) variable regions to promote immunity. This pathway requires and is initiated by the Activation Induced Deaminase (AID) protein, which deaminates cytidine to produce uracils and UG mismatches at the Ig genes. Subsequent processing of uracil by mismatch repair and base excision repair factors contributes to mutagenesis. While selective for certain genomic targets, the chromatin modifications which distinguish hypermutating from non-hypermutating loci are not defined.

Here, we show that AID-targeted loci in mammalian B cells contain ubiquitinated chromatin. Chromatin immunoprecipitation (ChIP) analysis of a constitutively hypermutating Burkitt's B cell line, Ramos, revealed the presence of monoubiquitinated forms of both histone H2A and H2B at two AID-associated loci, but not at control loci which are expressed but not hypermutated. Similar analysis using LPS activated primary murine splenocytes showed enrichment of the expressed \( V_H \) and \( S\gamma3 \) switch regions upon ChIP with antibody specific to AID and to monoubiquitinated H2A and H2B. In the mechanism of mammalian hypermutation, AID may interact with ubiquitinated chromatin because confocal immunofluorescence microscopy visualized AID colocalized with monoubiquitinated H2B within discrete nuclear foci.

Our results indicate that monoubiquitinated histones accompany active somatic hypermutation, revealing part of the histone code marking AID-targeted loci. This expands the current view of the chromatin state during hypermutation.
by identifying a specific nucleosome architecture associated with somatic hypermutation.

**INTRODUCTION**

The immunoglobulin (Ig) genes in activated B cells are diversified by somatic hypermutation and class switch recombination to promote immunity. Somatic hypermutation introduces point mutations into the rearranged and expressed Ig variable regions while class switch recombination coordinates the exchange of one Ig constant region for a downstream region, deleting the intervening DNA. Both pathways require the Activation Induced cytidine Deaminase (AID) protein (Muramatsu et al., 2000; Revy et al., 2000). AID is active on transcribed DNA and functions to convert single-stranded cytidines into uracil and uracil-guanine mismatches (reviewed in (Chelico et al., 2009; Conticello, 2008). Canonical mismatch repair and base excision repair are able correct genomic uracil (Larson et al., 2008; Pham et al., 2008; Schanz et al., 2009), but in somatic hypermutation these normally faithful repair pathways become mutagenic (Rada et al., 2004). The diversion to mutagenesis is promoted by the synthesis activities of low-fidelity DNA polymerases, such as polymerase eta (reviewed by (Di Noia and Neuberger, 2007; Peled et al., 2008), and their participation in hypermutation may be regulated by PCNA monoubiquitination (reviewed by (Langerak et al., 2009). While hypermutation is largely confined to the rearranged and expressed Ig genes, other B cell loci are prone to aberrant hypermutation, leading to lymphoma (Liu et al., 2008; Liu and Schatz, 2009). The molecular mechanisms responsible for targeting AID to
certain sites in the genome and the regulation of subsequent mutagenesis have not been established.

Monoubiquitination pathways may be important for regulating Ig gene diversification. DT40 B cells disrupted for PCNA monoubiquitination by a K164R substitution showed decreased AID-initiated Ig gene diversification (Arakawa et al., 2006). Similarly, PCNA K164R knock-in mice had an altered spectrum of Ig gene mutagenesis, paralleling mismatch repair defective mice and suggesting that monoubiquitination of PCNA influences uracil repair outcomes in mammals (Langerak et al., 2007; Roa et al., 2008). PCNA modification may be facilitated in part by the RAD6 pathway because inactivation of a RAD6 associating E3 ligase, RAD18, in DT40 B cells resulted in reduced levels of PCNA monoubiquitination and decreases in hypermutation (Arakawa et al., 2006; Bachl et al., 2006), however residual PCNA monoubiquitination in rad18 mutant DT40 suggests the involvement of more than one E3 ligase (Simpson et al., 2006). Further evidence for protein ubiquitinations in Ig gene diversification pathways comes from recent studies on the RNF8 and RNF168 E3 ligases. Silencing of these two proteins decreased class switch recombination efficiency in a murine model B cell line (Ramachandran et al., 2010), and RNF168 was identified as the RIDDLE syndrome protein, a disease characterized by immunodeficiencies and DNA repair defects (Stewart et al., 2009). Recently, RNF8 knock-out mice were shown to have impaired class switch recombination and defects in DNA damage responses (Li et al., 2010; Santos et al., 2010) further connecting histone monoubiquitination with Ig gene diversification. Together, it is likely that multiple
proteins are monoubiquitinated to promote Ig gene diversification and hypermutation specific histone E3 ligases could serve as signals for locus-specific mutagenesis.

In light of the DNA repair machinery-driven nature of hypermutation at the Ig loci, characterized roles for chromatin monoubiquitination in facilitating normal DNA repair throughout the genome (reviewed by (Game and Chernikova, 2009), and the involvement of RNF8 in class switching (Li et al., 2010; Ramachandran et al., 2010; Santos et al., 2010), we speculated that histone monoubiquitination could accompany hypermutation. We therefore assayed the nucleosomes at AID-targeted and actively hypermutating loci for ubiquitin modification. Here, we show by Chromatin Immunoprecipitations (ChIPs) that both histone H2A and H2B are monoubiquitinated at two different AID-targeted loci in the constitutively hypermutating B cell line Ramos. ChIP analysis of primary LPS activated murine splenocytes showed similar levels of enrichments for hypermutating template upon immunoprecipitation with antibody specific to AID, pol II, and monoubiquitinated forms of H2A and H2B. Furthermore, we find by immunofluorescence microscopy that AID colocalizes with monoubiquitinated H2B in discrete nuclear foci, suggesting that both are concomitant with hypermutation. Our results identify a considerable and previously uncharacterized chromatin modification associated with AID-induced somatic hypermutation.
RESULTS

AID and Hypermutation Target Two Discrete Loci in Ramos

Ramos is a human Burkitt’s lymphoma cell line with both functional and non-functionally rearranged Ig heavy chain (IgH) loci. One IgH allele contains a productive VDJ rearrangement (re-V\textsubscript{H}), which is a gene that is expressed and constitutively hypermutated (Sale and Neuberger, 1998). The other IgH allele has participated in a reciprocal translocation between the ends of chromosomes 8 and 14, replacing the antibody variable region on chromosome 14 with an intact c-MYC allele from chromosome 8 (MYC\textsubscript{14}), and this locus is also hypermutated (Bemark and Neuberger, 2000). The partially rearranged V region on chromosome 8 and the mu constant region downstream of re-V\textsubscript{H} are not diversified by hypermutation. Using EST data from the NCBI database, and previously reported c-MYC sequences (Bemark and Neuberger, 2000), we constructed a map of the rearranged and expressed heavy chain loci in Ramos (Figure 2). The documented constitutive hypermutation of Ramos re-V\textsubscript{H} and MYC\textsubscript{14} predict that AID selectively targets these two distinct loci for deamination.

We first established active somatic hypermutation of the re-V\textsubscript{H} and MYC\textsubscript{14} in our Ramos subclone using standard surface IgM loss assays and direct sequencing of MYC\textsubscript{14}. Constitutive sequence diversification of the expressed IgM locus results in mutations that inactivate surface display, providing a marker for hypermutation rate in Ramos. Starting from a single Ramos sIgM positive isolate, 90 days of continuous culture resulted in 6.83%
slgM loss, as measured by FACS with FITC-conjugated anti-IgM antibody (Figure 3). This compares well with the slgM loss rate reported for Ramos, showing 4% slgM loss for after 4 weeks of in vitro culture (Yabuki et al., 2005), and ~17% loss after 6 months (Sale and Neuberger, 1998). We conclude that the re-V_H is actively hypermutated. To ensure ongoing hypermutation of the other IgH locus, individual expressed MYC_{14} clones of from our Ramos population were sequenced, and 20% of our sequence reads contained one or more novel mutations (6/30, mean length=748 nt) (Figure 3B). Only the translocated c-MYC is expressed in Ramos, and MYC_{14} contains a single nucleotide polymorphism A523T (Bemark and Neuberger, 2000), distinguishing it from the non-translocated allele. This base substitution was present in all sequence reads we obtained. The MYC_{14} point mutations we identified are novel, and do not correspond to previously identified c-MYCHypermutations (Bemark and Neuberger, 2000). We conclude that both the re-V_H and MYC_{14} in our Ramos subclone are expressed and subject to hypermutation, as expected.

AID activity is required for initiating somatic hypermutation, predicting that the protein should be in physical contact with both the Ramos re-V_H and c-MYC genes. To test this, we used chromatin immunoprecipitations (ChIPs) to test for AID association with hypermutating loci in Ramos. ChIPs rely on specific antibodies to precipitate crosslinked protein-DNA complexes, followed by quantitative PCR (qPCR) analysis of the enriched genomic templates. Primer sets were designed to amplify templates immunoprecipitated from Ramos cells corresponding to the re-V_H locus (V_H 1–3), thec-MYC gene (exons 1–3), and
several non-hypermutating control loci (Figure 2). In chromatin prepared from Ramos cells, the normalized enrichment of the re-V<sub>H</sub> using anti-AID antibody was 13.8-fold (Figure 4A). Likewise, AID ChIP enriched for exons 1–3 of c-MYC (11.9 fold, 14.0 fold and 10.8-fold, respectively) (Figure 4A). AID ChIP did not enrich for any of the other loci tested, including a distal downstream (50 kb) region of MYC<sub>14</sub> (MYC4), the Cm<sub>µ</sub> constant region (V<sub>H</sub>-C<sub>µ</sub>), the non-productively rearranged heavy chain V region on chromosome 8 (V<sub>β</sub>), and the β actin gene (Figure 4A). To further test hypermutation specificity ChIPs with PCNA and pol eta antibodies enriched for the re-V<sub>H</sub> (Figure 7), and AID antibody failed to enrich for the c-MYC locus in AID negative HEK293 and A549 cells (Figure 8), and westerns with this same antibody did not detect AID in HEK293 extracts, but identified a single 24 kDa band in Ramos, as expected (Figure 9). In addition, IPs using this AID antibody and Ramos extract precipitated native AID, as detected by western with an independent AID antibody (Figure 10).

Primers designed to amplify MYC<sub>14</sub> could potentially amplify the other c-MYC allele if precipitated during ChIPs; however, only the translocated c-MYC contains the single nucleotide polymorphism A523T (Bemark and Neuberger, 2000). Further, ChIP results are suggestive of AID association with MYC<sub>14</sub> because AID tracks with pol II to deaminate transcribed DNA (reviewed by (Larson and Maizels, 2004) and the non-translocated allele of c-MYC is not expressed ((Bemark and Neuberger, 2000) yet pol II ChIPs enriched for exons 1–3 c-MYC. Most importantly, 100% (8/8) of different sequence reads from PCR product amplified from anti-AID ChIPed template corresponded to the
translocated MYC14 allele (not shown). Together, our results indicate that AID is physically associated with the Ramos re-VH and with exons 1–3 of MYC14, supporting the model that both loci are subject to AID-initiated somatic hypermutation.

**Chromatin at the Re-VH and c-MYC Gene in Ramos B Cells is Monoubiquitinated**

Having confirmed AID associations at both re-VH and c-MYC in our Ramos subculture, we next examined the chromatin state at these loci to identify histone markers of hypermutation. Based on roles for histone ubiquitination in normal DNA repair (Game and Chernikova, 2009) and apparent involvement of RNF8/168 ubiquitin E3 ligases in class switch recombination (Li et al., 2010; Ramachandran et al., 2010; Santos et al., 2010; Stewart et al., 2009), we focused our analysis on ubiquitin-modified chromatin. We used ChIPs and antibodies specific to monoubiquitinated histones to ask if the re-VH and MYC14 loci are modified by ubiquitination. Analysis of qPCR following mUb-H2A ChIP of Ramos chromatin showed a 48.3 fold normalized enrichment of the re-VH. Exons 1–3 of c-MYC were also enriched, showing 18.4 fold, 23.0 fold and 16.6 fold enrichments respectively (Figure 4B). Similar to the mUb-H2A ChIP, qPCR analysis following mUb-H2B ChIP of Ramos chromatin showed 37.6-fold normalized enrichment of the re-VH and exons 1, 2, and 3 of c-MYC showed 29.3 fold, 32.3-fold and 12.9-fold enrichments, respectively (Figure 4C). In contrast, neither the VH-Cµ, V8, β-actin gene, nor MYC4 were enriched by mUb-H2A or mUb-H2B ChIP (Figure 4B, C). Also, mUb-H2A and mUb-H2B
ChIPs with AID negative HEK293 and A549 cells did not enrich for the c-MYC loci (Figure 8). We conclude that the re-\(V_H\) and c-MYC loci in Ramos are occupied by nucleosomes that contain monoubiquitinated forms of H2A and H2B.

Previous studies connect histone monoubiquitination with transcriptional silencing and/or polymerase elongation (reviewed by (Weake and Workman, 2008)) so we next asked if histone ubiquitination is a broad consequence of high RNA polymerase II (pol II) driven expression of the heavy chain loci in Ramos, or associated with mutagenesis. ChIPs using RNA pol II antibody and Ramos chromatin showed comparable enrichment via normalized qPCR for the re-\(V_H\), exons 1–3 of c-MYC, \(\beta\)-actin, and \(V_H\)-C\(\mu\). We did not observe enrichment of the unrearranged \(V_H\) locus (\(V_8\)) or the region downstream ofMYC14, MYC4, which was anticipated because these regions are not expressed. In contrast, \(\beta\)-actin, and \(V_H\)-C\(\mu\) were all enriched by pol II ChIPs (Figure 4D), but not after ChIP using monoubiquitinated histone antibody (Figure 4B, C). These results mirror our AID analyses, and suggest that pol II-dependent gene expression alone is insufficient for establishing sustained chromatin monoubiquitination in Ramos.

If monoubiquitinated chromatin is associated with Ig gene mutagenesis, we also anticipate other hypermutation factors will be associated with these same loci. AID-induced uracil is processed by mismatch repair factors at the Ig loci, and substitutions are introduced opposite A and T bases through PCNA-dependent recruitment of pol eta synthesis in place of pol delta (reviewed by (Langerak et al., 2009). ChIP analysis suggests that the re-\(V_H\), but not the unrearranged allele (\(V_8\)), is in physical contact with PCNA and pol eta (Figure 7),
supporting the model that the monoubiquitinated loci we identify here also harbor proteins known to support mutagenesis in response to uracil in DNA. The qPCR enrichments obtained for Ramos ChIPs were independently verified by qualitative, traditional PCR amplification and ethidium bromide visualization (Figure 4E) and westerns verified specificity for AID and monoubiquitinated histone antibodies (Figure 4F).

**The Chromatin of Ig Genes in Activated Primary Murine B Cells is Ubiquitinated**

It is possible that chromatin monoubiquitination is a unique feature of Ramos hypermutation. To address this, we tested histone monoubiquitination of an AID-targeted locus in LPS activated murine B cells. Typically, hypermutation is difficult to quantify by PCR in primary cells because of the sequence variability inherent to the rearranged Ig genes. To overcome this technical limitation, we used a quasi-monoclonal mouse, which was engineered to contain a functional IgH chain allele generated by targeted replacement of heavy chain locus variable region with a defined VDJ rearrangement (17.2.25)(Cascalho et al., 1996) (Figure 5A). The 17.2.25 VDJ region in these mice is subject to diversification by somatic hypermutation(Cascalho et al., 1998).

Primary splenocytes were isolated from six healthy quasi-monoclonal mice and cultured with lipopolysaccharide (LPS) to activate primary B cells. ELISAs for secreted IgG confirmed activation of Ig gene diversification (data not shown). In chromatin prepared from these activated splenocytes, normalized enrichment of 17.2.25 after ChIP with mUb-H2A and mUb-H2B antibody showed robust
enrichment, 16.2-fold and 21.8-fold for each antibody, respectively (Figure 5). Similarly, pulldowns using these same antibodies were also enriched for the Sy3 switch region (Figure 5). AID appears to target the 17.2.25 and Sy3 loci after activation because anti-AID ChIPs resulted in 10.4 and 7.1 fold enrichments of these regions respectively (Figure 5B, C) The murine triose phosphate isomerase gene (TPI) and unexpressed neural specific gene (STEP) were used as controls, and only the TPI locus showed modest enrichment (6.9-fold) after pol II ChIP, and neither locus was precipitated upon IP with anti-AID, anti-mUb-H2A, or anti-mUb-H2B. Due to the limited antibody repertoire encoded by the quasi-monoclonal mice, selective pressure for Ig gene diversity is high, causing both hypermutation and frequent V gene replacements(Cascalho et al., 1998). Because it is unlikely that 17.2.25 gene rearrangements would be amplified by our primer sets, enrichments we obtained for the 17.2.25 locus after AID ChIP (Figure 5B) are likely to be an under-representation of the actual level of AID association. Regardless, we conclude that the 17.2.25 and Sy3 loci from primary LPS activated quasi-monoclonal murine splenocytes are expressed by RNA pol II, targeted by AID, and occupied by monoubiquitinated histone H2A and histone H2B.

**AID and MUb-H2B Colocalize in Discrete Nuclear Foci**

ChIPs examine total cellular populations, and hypermutation events may occur transiently and at unsynchronized time points. Therefore, we used immunofluorescence confocal microscopy to ask if AID and ubiquitinated chromatin co-occupy the DNA. Both mUb-H2B and AID antibody produced clear
robust signals in Ramos cell preparations. Figure 6 shows a representative image of AID and mUb-H2B foci in Ramos, and overlay of these images showed clear colocalization of AID and mUb-H2B signal in discrete nuclear foci. We also found multiple examples where more than one AID and monoubiquitinated H2B are colocalized within a single cell nucleus (Figure 11), supporting the notion that at least two and possibly more loci are diversified by somatic hypermutation in Ramos cells. We observe multiple AID foci in individual Ramos cells, highly consistent with the behavior of endogenous AID (Cattoretti et al., 2006; Patenaude et al., 2009; Yang et al., 2005). It is possible that non-Ig genes are bound and deaminated by AID in Ramos. Antibodies to mUb-H2A and mUb-H2B antibody are specific for the monoubiquitinated forms of histones, and our westerns probing Ramos whole cell extract clearly show single bands at \( \sim 25-30 \) kDa (Figure 4F) corresponding to the size of a single histone ubiquitination (\( \sim 28 \) kDa). Even though mUb-H2A, and mUb-H2B antibodies were competent for western analyses (Figure 4F) and ChIPs (Figure 4 and 5), we were unable to clearly detect mUb-H2A by immunofluorescence microscopy. Nevertheless, our imaging results suggest that AID is spatially and temporally associated with ubiquitinated nucleosomes.

**DISCUSSION**

Our results show that genomic loci that are targeted for somatic hypermutation are also occupied by monoubiquitinated chromatin. Histone H2A and H2B monoubiquitination was coincident with transcription, but not due to pol II catalyzed gene expression alone because anti-mUb-H2A and mUb-H2B ChIPs
failed to enrich for the highly expressed Ramos β actin locus or the TPI gene in primary mouse splenocytes, even though pol II ChIP enriched for these regions at levels comparable to the re-V_H, c-MYC Exons 1–3, and 17.2.25 (Figure 4 and 5). Furthermore, the only genomic regions enriched by mUb-H2A and mUb-H2B ChIP were those that were also enriched upon anti-AID ChIP (Figure 4A); and these same loci also showed significant enrichments after ChIP with antibodies to the DNA repair factors PCNA and pol eta (Figure 7), which are two proteins involved in somatic hypermutation opposite A and T bases (recently reviewed by Langerak et al., 2009). Histone monoubiquitination may occur in response to AID activity, but could also be a signal for AID recruitment because AID colocalized with mUb-H2B as visualized by immunofluorescence microscopy (Figure 6). Either way, our results identify the presence of both histone H2A and H2B tail monoubiquitinations at genes that are targeted selectively for mutagenesis, identifying this specific set of chromatin modifications with genomic loci that are deaminated by AID and mutated by responding DNA repair pathways. As such, co-occupancy of monoubiquitinated H2A and H2B within nucleosomes at the Ig genes may in part characterize the chromatin architecture associated with somatic hypermutation.

Histone termini extend away from the core histone octamer as tails that serve as acceptors for a wide range of post-transcriptional modifications. The combination of these modifications or “histone code” regulates a wide range of functions, although the specific DNA level activities triggered by various combinatorial signals have not been completely defined. In hypermutation, some
small molecule histone modifications have been described. Acetylation of histone H3 and H4 at the IgH variable region, but not the constant region, results from activation of hypermutation in BL2 cells, suggesting a role in targeting somatic hypermutation machinery (Woo et al., 2003). However, in primary mouse B cells changes in histone acetylation at hypermutating loci did not follow with B cell activation (Odegard et al., 2005). Nevertheless, H4 acetylation may be important for some types of AID-initiated diversification because disruption of the E2A transcription factor in DT40 leads to decreases in H4 acetylation and gene conversion (Kitao et al., 2008). One unique signal for V region hypermutation may be phosphorylation of histone H2A at serine 14 (Odegard et al., 2005), and it is possible that this signal may operate in concert or in response to histone monoubiquitination to promote hypermutation. More than likely, locus-specific mutagenesis in B cells depends upon the combinatorial signal generated through multiple histone tail modifications, and further studies are required to decipher the specific histone code assigned to somatic hypermutation.

H2A and H2B monoubiquitination may have active roles in the hypermutation pathway, and immunofluorescence results confirm proximity of AID and monoubiquitinated H2B (Figure 6 and 11). Chromatin modification may occur as a consequence of AID activity to promote mutagenesis at some loci, or histone monoubiquitinations could help recruit AID. These models are not necessarily mutually exclusive. If chromatin ubiquitination is important for regulating the balance between mutagenic and faithful DNA repair in Ig gene hypermutation, one might predict that AID could deaminate multiple genes but
the resulting uracils are faithfully repaired at the non-Ig loci. Indeed, we and others have observed multiple AID foci by immunofluorescence microscopy (Figure 6 and 11) (Cattoretti et al., 2006; Patenaude et al., 2009; Yang et al., 2005), and genome sequencing of germinal center B cells reveals that AID acts widely at transcribed genes, but genomic instability is largely suppressed by faithful DNA repair (Liu et al., 2008). There are likely to be signals specific to only some genes (like the Ig loci) that regulate the switch from faithful to mutagenic uracil repair, and because mUb-H2B appears to occupy only a small subset of the AID-targeted loci (Figure 6 and 11), chromatin ubiquitination is not an automatic result of AID binding but rather a locus-specific event. It is conceivable that the conspicuous presence of ubiquitinated H2A and H2B at the Ig genes represents a signal that marks these loci for mutagenesis.

Consistent with a model whereby histone monoubiquitination facilitates the repair responses to DNA deamination, recent studies have identified histone-specific E3 ubiquitin ligases in class switch recombination. Starting with histone H2A modification, ubiquitination cascades appear critical for establishing assembly of recombination factors at sites of AID activity. The loss of the RNF168 H2A E3 ligase disrupts histone ubiquitination and causes DNA repair defects and immunodeficiencies associated with RIDDLE syndrome (Stewart et al., 2009); and RNF8 works in concert with RNF168 to initiate and expand histone ubiquitinations at the Ig loci that resulting in 53BP1 activity and proper class switching (Li et al., 2010; Santos et al., 2010; Stewart et al., 2009). In normal break repair, chromatin ubiquitinations promote assembly of DNA
damage response factors at sites of ionizing radiation damage (Huen and Chen, 2010). Furthermore, the replication cofactor PCNA, which we found physically associated with the re-V_{H} (Figure 7) and may be a lynchpin for regulating A/T biased hypermutation (Langerak et al., 2009), colocalizes with monoubiquitinated H2A (Vassilev et al., 1995). Upon AID activity, chromatin ubiquitination may signal assembly of hypermutation factors and/or ubiquitination cascades that regulate the shift between faithful uracil repair and mutagenesis. However, further studies are required in order to identify the ubiquitin E3 ligases functioning in normal and aberrant somatic hypermutation.

Histone monoubiquitination is a substantial chromatin modification, and suggests the DNA at AID-targeted loci is accessible for metabolism. Unlike small molecule modifiers like phosphates or acetyl groups, ubiquitin is an 11 kDa protein. The addition of a single ubiquitin on both H2A and H2B would increase nucleosome mass by nearly 40%, which may greatly influence the association between DNA and the core histones or create an assembly site for repair factors. The significant enrichments we obtained by ChIP of the re-V_{H} and MYC_{14} with mUb-H2A and H2B antibodies (Figure 4B, C) suggest the presence of a highly ubiquitinated locus, with multiple antibody binding targets at the site of hypermutation, but apparently not extending into the IgH constant region (V_{H}1 amplicon Figure 4B, C). Thus, histone monoubiquitination is localized to the DNA regions that are also bound by AID. It follows that V_{H}1, and likely the proximal S region, were not enriched by AID ChIP (Figure 4A, C), which is expected because Ramos does not class switch. Because AID is active on
single-stranded DNA and associated with the re-V_{H} (Figure 4A), it is conceivable that histone monoubiquitination creates a chromatin architecture that is permissible to AID attack or downstream uracil repair. Indeed, nucleosomal DNA is protected from AID in the absence of transcription (Shen et al., 2009), suggesting that deamination by AID may require histone remodeling and chromatin-level signaling. Importantly, our results identify a new chromatin modification associated with hypermutation and future studies will be required to next define functional roles in mechanisms of somatic hypermutation.

**MATERIALS AND METHODS**

**Cells and Cell Lines**

Ramos cells were purchased from the ATCC and were cultured in suspension at 37°C in RPMI (10% (v/v) FBS, 1% (v/v) PS, 2 mM L-glutamine) (Life Technologies, Grand Island, NY) in 75 cm² flasks. Following limiting dilution cloning, a sIgM positive isolate was expanded for 90 days, and then analyzed for sIgM loss by flow cytometry. Ramos cells were collected by centrifugation at 300×g for 10 minutes, washed once with ice-cold PBS, and then resuspended in 100 µL PBS and incubated with 10 µL of anti-human-IgM antibody (Sigma, St Louis, MO) for 10 minutes at 4°C. Cells were again collected, washed twice with PBS, resuspended in 100 µL PBS and incubated with 3 µL of FITC-conjugated anti-goat Ig secondary for 10 minutes at 4°C in the dark. Cells were washed twice with PBS then resuspended in 400 µL PBS. Samples included 10% normal rat serum to prevent non-specific binding. Cells were analyzed immediately on a
Becton Dickinson FACS Calibur flow cytometer using CellQuest Pro software (San Jose, CA) with $1 \times 10^4$ events collected.

**Chromatin Immunoprecipitations**

Approximately $1 \times 10^7$ cells were transferred to a 15-ml Falcon tubes and incubated for 10 minutes with 1% (v/v) formaldehyde. Chromatin was sheared by sonication to generate fragments averaging 500 bps, as judged by agarose gel electrophoresis. ChIPs were performed on prepared chromatin using EZ-ChIP kit reagent buffers (Millipore, Temecula, CA, 17–371) and standard manufacturer protocol. ChIP primary antibodies were as follows: normal mouse IgG negative control (Santa Cruz Biotechnology, Santa Cruz, CA, sc-2025), anti–RNA polymerase II (Millipore, 05-623B), anti-mUb-H2A (Millipore, 05-678)(Pinato et al., 2009), anti-ub-H2B (Millipore, 05-1312) (van der Knaap et al., 2010), anti-AID (Santa Cruz, sc-25620), and polyspecific IgG (Santa Cruz, sc-2025). Crosslinks were reversed by incubating chromatin at 65°C overnight, and enriched DNA template analyzed by traditional PCR and quantified by real-time PCR (described in PCR analyses).

**Western Blotting**

Ramos cells at $\sim 1 \times 10^6$ cells/ml, were collected by centrifugation, existing media removed, and cells resuspended in SDS lysis buffer containing protease inhibitors, and transferred to 1.5 ml Eppendorf tubes. Proteins were electrophoresed through a 4–12% SDS–polyacrylamide gradient gel (Invitrogen) and transferred to immobilon-P PVDF membranes (Millipore). Membranes were blocked for 1 hour in 5% (w/v) nonfat milk in phosphate-buffered saline.
containing 0.05% Tween 20, washed, and incubated with primary antibody overnight at 4°C using the following dilution: anti-mUb-H2A (Millipore, 05-678) – 1:500, anti-ub-H2B (Millipore, 05-1312) – 1:500, and anti-AID (Santa Cruz, sc-25620) - 1:1000. Membranes were washed and incubated with secondary Abs: HRP conjugated goat anti-mouse and goat anti-rabbit (Invitrogen) at 1:10000 dilution. Immunoreactive bands were visualized with ECL Plus (Amershram, Piscataway, NJ) and signals were detected by using the Storm 840 PhosphorImager and IMAGEQUANT software (GE Healthcare Life Sciences).

**PCR Analyses**

Oligonucleotide sequences from all analyses are detailed in Table II-S1. Sequencing of Ramos c-MYC was performed as described previously (Bemark and Neuberger, 2000). Briefly, total Ramos RNA was converted to cDNA via RT-PCRs using random 20-mers and Protoscript Reverse Transcriptase (New England BioLabs (NEB), Ipswich, MA). Unless otherwise indicated, PCR amplifications were performed in 40 µl reactions at standard concentrations (1.5 mM MgCl₂, 0.2 mM dNTP, 1× NEB PCR buffer, 0.5 U Taq (NEB), 0.5 uM each primer) and then cloned into Topo TA PCR 2.1 (Invitrogen, Carlsbad, CA, K450001) for sequencing. MYC₁₄ was distinguished from c-MYC on chromosome 8 by the presence of an A at position 523 to T polymorphism found withinMYC₁₄ (Bemark and Neuberger, 2000). Traditional ChIP amplifications from recovered DNA was quantified by standard SybrGreen quantification (Invitrogen), then diluted to 0.1 ng/µl. PCR reactions were performed for 25, 30 or 35 cycles (as described above). Starting material was standardized across all reactions (0.1 ng
eluted DNA), excluding inputs, which were titrated from 0.5 ng to 10 ng per reaction. Following amplification, products were resolved by 1.5% agarose gel electrophoresis. The identities of all amplicons were verified by Topo TA PCR 2.1 cloning and sequencing. Quantitative PCR (qPCR) used a series of primer pairs generated for each genomic locus using the Integrated DNA Technology design tool, PrimerQuest (https://www.idtdna.com). Individual amplicons were each evaluated by standard PCR ethidium bromide visualization, dissociation curve determination, and direct sequencing. A final validated primer pair was then selected for each locus (Table II-S1). Recovered DNA was diluted 1:100 with ddH20 and 20 µl reactions prepared using DyNAmo SYBR Green qPCR 2× master mix (NEB, F400L) and ROX Reference Dye for Quantitative PCR (Sigma, R4526) were assembled in triplicate. Reactions were performed using an Applied Biosystems 7300 Real-Time PCR thermocycler (ABI, Foster City, CA) and DNA enrichments calculated by standard delta-delta Ct.

**Quasi-Monoclonal Mouse Primary B Cells**

Spleens from six healthy quasi-monoclonal mice (created by(Cascalho et al., 1996) were removed aseptically, placed in PBS and gently ground between frosted slides to produce a single-cell suspension. The suspension was centrifuged at 300×g for 5 minutes, the pellet was resuspended in ammonium-tris-chloride buffer to hypotonically lyse erythrocytes, and the remaining cells were washed with PBS by centrifugation at 300×g for 5 minutes. Splenocytes were cultured at 37°C/5% CO₂/95% humidity in 75 cm² flasks in RPMI 1640 (Life Technologies) supplemented with 100 U/ml of penicillin, 100 mg/ml of
streptomycin and 10% FBS. Cultures were allowed to recover for 1 hour then supplemented with 25 µg/ml LPS (Sigma, L2143) and incubated for 72 hours to induce proliferation. Active Ig gene diversification was monitored by ELISA of the RPMI media for the presence of IgG, a result of LPS induced class switch recombination, as described (Blaeser et al., 2008; Cascalho et al., 1999) and by ChIPs to show AID association with the 17.2.25 locus (Figure 5B). Cells were collected after 72 hours and immediately crosslinked with formaldehyde for ChIP analysis as described.

Confocal Immunofluorescence Microscopy

Hypermutating Ramos cells were cultured in suspension at 37°C in RPMI (10% (v/v) FBS, 1% (v/v) PS, 2 mM L-glutamine) in 75 cm² flasks. At ~1×10⁶ cells/ml, 750 µL of Ramos suspensions were transferred into individual poly-lysine/ConA-coated chambered slide wells and incubated at 37°C for 1 hour to allow adhesion. Cells were rinsed twice with PBS then fixed with 2% paraformaldehyde for 20 minutes. Cells were again rinsed twice with PBS then blocked with for 30 minutes (PBS, 1% Goat serum, 0.2% TritonX-100). Cells were next incubated with primary antibodies diluted 1:1000 in PBTG (PBS, 1% Goat serum, 0.2% TritonX-100) for 2 hours at room temperature. Cells were rinsed 3 times with PBTG then incubated in the dark for 2 hours at room temperature with secondary antibodies diluted 1:1000 in PBTG. Cells were again rinsed 3 times with PBTG then mounted with vectashield (Vector Laboratories, Burlingame, CA, H1000). Confocal images were obtained using a Leica TCS SP2 Confocal Microscope and contrast processed with Adobe Photoshop 7.0.
software. Primary antibodies included: anti-mUb-H2B (Millipore, no. 05-1312) and anti-AID (Santa Cruz, sc-25620). Anti-mUb-H2A (Millipore, 05-678) was also examined but produced no discernable immunofluorescence microscopy signal. Secondary Abs were goat anti-mouse conjugated to Alexa 555 and goat anti-rabbit conjugated to Alexa 488 (1:1000) (Invitrogen). DAPI (1:1000) Alexa 633-wheat germ agglutinin (1:1000) was used to stain nuclear envelopes (Invitrogen).
REFERENCES


Sale, J. E., and M. S. Neuberger, 1998, TdT-accessible breaks are scattered over the immunoglobulin V domain in a constitutively hypermutating B cell line: Immunity, v. 9, p. 859-69.


FIGURES

Figure 2. Diagram of the IgH Loci in Ramos Burkitt's Lymphoma. A reciprocal translocation exchanging the ends of chromosomes 8 and 14 in Ramos resulted in the formation of MYC
14 and V8. The chromosome 14 break point occurred within the heavy chain (IgH) switch region (indicated by jagged edges) just 5' of the µ constant exon (IgHM) while the chromosome 8 breakpoint occurred in a c-MYC allele promoter. (R), silent, non-coding RNA; ×1, ×2, ×3, c-MYC exons. Arrows indicate transcriptional direction. Double bar denotes amplicon location. Primers were designed to amplify the IgH constant region mu (VH1), the functional heavy chain VH4-34/D/J6 rearrangement (VH2), and an upstream non-rearranged V-region sequence (VH3). Chromosome 8 primer sets amplify three c-MYC exons (MYC1–3), and a sequence 50 kb 5' of MYC3 (MYC4). One primer set was designed for the non-expressed unrearranged VH 4–34 on chromosome 8 (V8).
Figure 3. Ramos Burkitt's Lymphoma Cells Constitutively Hypermutate at Distinct Genomic Loci. A. Flow cytometry measure of loss of surface IgM (sIgM) display over time for cultured Ramos cells. sIgM was detected using FITC conjugated antibody specific for human IgM (FL1-H). Ramos cells were continuously cultured for 90 days after which 10,000 cells were analyzed. B. Sequence analysis of the translocated and expressed MYC. Each pie wedge represents a unique sequencing read with the number of mutations identified in individual reads indicated. The total number of mutations identified in 10 unique sequence reads corresponding to individual exons is shown in the central circles.
Figure 4. ChIPs Identify MUb-H2A and MUb-H2B at Ramos Hypermutating Loci. A. Normalized template enrichment upon anti-AID ChIP. Template obtained from AID-IP was amplified and analyzed by qPCR using primer sets diagramed in Figure 1, above. Results were normalized to non-specific IgG IP and relative to input. Each bar indicates the mean of the values obtained from six amplifications (two separate experiments analyzed in triplicate) with standard deviation. B. Normalized template enrichment upon mUb-H2A ChIP. C. Normalized template enrichment upon mUb-H2B ChIP. D. Normalized template enrichment upon pol II ChIP. Enrichments were analyzed by qPCR as in Figure 3A. E. Ethidium bromide stained gel of traditional PCR amplifications from ChIP template. F. Antibodies used to ChIP mub-H2A and mub-H2B and AID are specific for monoubiquitinated H2A, H2B and AID protein. Western blot analysis of endogenous AID (24 kDa), mUb-H2A (27 kDa) and mUb-H2B (27 kDa) in Ramos cell lysates. Visible on the AID blot, the white negative stained bands of the protein marker correspond to 55, 40, 36, 25, 15 and 10 kilodaltons (top to bottom).
Figure 5. MUb-H2A and MUb-H2B Associate with Hypermutation in Activated Mouse Primary B Cells. A. Cartoon depicting the engineered rearranged heavy chain allele sequences in quasi-monoclonal mouse. The rearranged allele (17.2.25) is compared to a normal heavy chain arrangement (wt). The 17.2.25, allele formed by the targeted replacement of the joining cluster with a defined V(D)J rearrangement. J_{HΔ}, allele formed by the targeted deletion of the joining cluster. S, Sy3 switch region. IGHM, μ constant exon. V_{H} additional 5’ variable sequence segments. DQ52, diversity sequence segment immediately 5’ of the joining sequence cluster. Arrows indicate transcriptional direction. Double bar denotes amplicon location. B. ChlPs identify monoubiquitinated histones with hypermutation. Murine splenocytes were cultured for 72 hours with LPS, crosslinked, and then subjected to ChIP analysis with anti-mUb-H2A and anti-mUb-H2B, and anti-AID antibody. Normalized enrichment of the 17.2.25, Sy3, triose phosphate isomerase (TPI) and neural-specific striatum-enriched protein-tyrosine phosphatase (STEP) loci from qPCR are shown relative to input DNA. C. Ethidium bromide visualization of traditional PCRs with primer sets corresponding to the qPCRs.
Figure 6. AID and Monoubiquitinated H2B Colocalize in Discrete Nuclear Foci. Representative immunofluorescence confocal microscopy images of Ramos cells stained for AID and mUb-H2B. Ramos cells were incubated with rabbit antibody specific to AID and murine mUb-H2B antibody followed by Alexa-488 conjugated rabbit secondary and Alexa-555 conjugated murine secondary antibody. Upper left, Ramos cells imaged with 488 nm filter. Upper right, Ramos cells imaged with 555 nm filter. Lower left, merged image with white scale bar (2 microns). Lower right nuclear envelope imaged using wheat germ agglutinin (Blue). Arrows point to colocalizations of AID and mUb-H2B.
Figure 7. PCNA and Polymerase Eta are Associated with the Ramos Re-VH Region. Ramos cells were crosslinked, and sheared chromatin subjected to immunoprecipitation using antibody specific to PCNA, MSH2 and polymerase eta, and non-specific IgG antibody. Precipitated template DNA was amplified by qPCR using primers specific to a control glutamine tRNA locus (tRNA), unrearranged VH (V8), and the rearranged VH (VH2)(see Figure 1). Enrichment using poly-specific IgG for each locus was subtracted to normalize for non-specific IP. Fold enrichment relative to input DNA is shown.
Figure 8. HEK293 and A549 ChIPs Using AID, MUb-H2A and MUb-H2B Antibodies Find No Enrichment of Sequences Undergoing Hypermutation in Ramos cells. Normalized template enrichment upon anti-AID, anti-mUb-H2A and anti-mUb-H2B ChIPs. Template obtained from IPs was amplified and analyzed by qPCR using primer sets diagramed in Figure 1. Results were normalized to non-specific IgG IP and relative to input. Each bar indicates the mean of the values obtained in triplicate with standard deviation.

Figure 9. AID Antibody Used for ChIPs is Specific for AID. Western blot analysis of endogenous AID (24 kDa) in AID positive (Ramos) and in AID negative (HEK293) cell lysates. PCNA was utilized as load control.
Figure 10. AID IP Demonstrating AID Antibody Used for ChiPs is Detecting AID. Western blot analysis of endogenous AID (24 kDa) in Ramos cell lysate IP. Primary anti-AID for IP, (Santa Cruz Biotechnology, Santa Cruz, CA sc-25620). Primary Ab for western, (Santa Cruz Biotechnology, Santa Cruz, CA sc-14680). Total supernatant and total IP were loaded to ensure equivalent starting material.

Figure 11. MUb-H2b Colocalizes with AID in Multiple Discrete Foci. Representative immunofluorescence microscopy images of two distinct Ramos cells stained with AID and mUb-H2B antibodies and exhibiting colocalization of these at 1 (Panel A) and 2 (Panel B) discrete foci. In each panel: Upper left, Ramos cells imaged with Alexa 488 filter. Upper right, Ramos cells imaged with Alexa 555 filter. Lower left, merged images with white scale bar (2 microns). Lower right nuclear envelope imaged using wheat germ agglutinin (Blue). Arrows point to colocalizations of AID and mUb-H2B.
**Table 1. Oligonucleotide Table.** Oligonucleotide sequences for amplification of hypermutating genomic and control loci in this study are shown.

**Human QPCR Primers**

<table>
<thead>
<tr>
<th>Oligonucleotide</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_H^1$</td>
<td>CCCAAGGACACCCTCATGATCTCCC</td>
</tr>
<tr>
<td>$V_H^1R$</td>
<td>CATTATGCACCTCCACGCGGTCCAC</td>
</tr>
<tr>
<td>$V_H^2$</td>
<td>GCAGGACTGTTGAAGCCTTCGAG</td>
</tr>
<tr>
<td>$V_H^2R$</td>
<td>CTTACCTGAGGAGACGGTGACC</td>
</tr>
<tr>
<td>$V_H^3$</td>
<td>GACTAACCAGAGTTTCATGGCTGTGTTG</td>
</tr>
<tr>
<td>$V_H^3R$</td>
<td>GAGGTCTCAAATGATGTTTGGCGAGATG</td>
</tr>
<tr>
<td>$V_H^8$</td>
<td>GCAGGACTGTTGAAGCCTTCGAG</td>
</tr>
<tr>
<td>$V_H^8R$</td>
<td>TGCAGGGAGGTGTTTGTCTGAGG</td>
</tr>
<tr>
<td><strong>MYC1</strong></td>
<td>CCTCTCTCGCTAATCTCCGCCC</td>
</tr>
<tr>
<td><strong>MYC2</strong></td>
<td>CTGTATGTGGAGCGGCTTCTCG</td>
</tr>
<tr>
<td><strong>MYC3</strong></td>
<td>CTGTATGTGGAGCGGCTTCTCG</td>
</tr>
<tr>
<td><strong>MYC4</strong></td>
<td>CTGTATGTGGAGCGGCTTCTCG</td>
</tr>
<tr>
<td><strong>β Actin</strong></td>
<td>GGCTCCGGCATGTGCAAGG</td>
</tr>
<tr>
<td><strong>tRNA$_{glu}$</strong></td>
<td>AAGGCACATCTACTAGTCCATGAG</td>
</tr>
<tr>
<td>trna$_{glu}_F$</td>
<td>CTCTTTGCTCTTGAGTCTCAG</td>
</tr>
<tr>
<td>trna$_{glu}_R2$</td>
<td>CTCTTTGCTCTTGAGTCTCAG</td>
</tr>
</tbody>
</table>

Table Continues
Mouse QPCR Primers

STEP

F2_neural_59.7  CCAGAGGACAGGGCGTCAATCAG
R1_neural_59.9  GTCAATGAGGCAGGATCAAGTCTCAGGT
mus_tpi_2F61   CCTCTAGAAACATTGATTTGGCTCTAGAAACGCC
mus_tpi_R61    TCAGTACAAACGACGGCTCTCTC

17.2.25

F4_17.2.25_59   CTACAGGGGTCAATTCAAGGTTTCAG
R1_17.2.25_63   GGCAGTGTCCTCAGATGTCAGGC

Sγ3

Sγ3F  AGTCACCATGTAGATACGTACGCTG
Sγ3R  CAGGTTATGCAGCTCTGGTGTCAG

MYC MUTANT ANALYSIS

Exon1

MYC1F  CCTCTCTCGATAATCTCCGCCC
X1mutR CACCATCTCCAGCTGTCGCGCC

Exon2

MYC2F  CTGTATGTGAGCGGGGCTTCTC
MYC3R  CTCTGACACTGGACTTGACCCTC

Exon3

X3mutF  CCACAGCATACATCTCGTCCTCC
X3mutR  CCTACCTCTCACACTTCTACACTGCC
CHAPTER III

REPRESSION OF HUMAN ACTIVATION INDUCED CYTIDINE DEAMINASE

BY MIR-93 AND MIR-155

Contributions:

The contents of this chapter were previously published and cited below. I contributed to all levels of experimentation with minor contributions to manuscript preparation and editing.

ABSTRACT

Activation Induced cytidine Deaminase (AID) targets the immunoglobulin genes of activated B cells, where it converts cytidine to uracil to induce mutagenesis and recombination. While essential for immunoglobulin gene diversification, AID misregulation can result in genomic instability and oncogenic transformation. This is classically illustrated in Burkitt's lymphoma, which is characterized by AID-induced mutation and reciprocal translocation of the c-MYC oncogene with the IgH loci. Originally thought to be B cell-specific, AID now appears to be misexpressed in several epithelial cancers, raising the specter that AID may also participate in non-B cell carcinogenesis.

The mutagenic potential of AID argues for the existence of cellular regulators capable of repressing inappropriate AID expression. MicroRNAs (miRs) have this capacity, and we have examined the publically available human AID EST dataset for miR complementarities to the human AID 3'UTR. In this work, we have evaluated the capacity of two candidate miRs to repress human AID expression in MCF-7 breast carcinoma cells.

We have discovered moderate miR-155 and pronounced miR-93 complementary target sites encoded within the human AID mRNA. Luciferase reporter assays indicate that both miR-93 and miR-155 can interact with the 3'UTR of AID to block expression. In addition, over-expression of either miR in MCF-7 cells reduces endogenous AID protein, but not mRNA, levels. Similarly indicative of AID translational regulation, depletion of either miR in MCF-7 cells increases AID protein levels without concurrent increases in AID mRNA.
Together, our findings demonstrate that miR-93 and miR-155 constitutively suppress \textit{AID} translation in MCF-7 cells, suggesting widespread roles for these miRs in preventing genome cytidine deaminations, mutagenesis, and oncogenic transformation. In addition, our characterization of an obscured miR-93 target site located within the \textit{AID} 3'UTR supports the recent suggestion that many miR regulations have been overlooked due to the prevalence of truncated 3'UTR annotations.

**INTRODUCTION**

In antigen activated B cells the Activation Induced cytidine Deaminase (AID) protein is required for initiating mutagenesis and recombination of the immunoglobulin (Ig) genes to promote immunity. AID is a cytidine deaminase that converts single-stranded genomic cytidine into uracil, and its activity is pronounced at the Ig variable (V) and switch (S) regions (de Yébenes and Ramiro, 2006; Goodman et al., 2007; Longerich et al., 2006; Peled et al., 2008; Vallur et al., 2007). While spontaneously generated uracil in the genome is faithfully corrected as a part of normal DNA repair (Larson et al., 2008), AID-induced uracils at the Ig loci are mutagenically resolved by multiple DNA repair factors (Peled et al., 2008; Stavnezer et al., 2008). Although AID targets the Ig loci in activated B cells, activity at other genomic sites has the potential to create oncogenic DNA damage. Indeed, transcribed genes across the genome have now been found to undergo AID deamination (reviewed by (Liu and Schatz, 2009; Pérez-Durán et al., 2007), and recently, deep sequencing of AID-ChIPed template revealed broad AID-associations with expressed loci (Yamane et al., 2008).
2011). In addition, AID-dependent DNA breaks were recently identified within multiple types of repetitive elements (Staszewski et al., 2011) suggesting a broad ability for AID to target genomic regions harboring single-stranded character. AID misexpression results in DNA damage that promotes cancer (Liu et al., 2008; Nagaoka et al., 2010; Okazaki et al., 2003; Parsa et al., 2007; Staszewski et al., 2011; Zhang et al., 2001). In lymphoid cells improper AID expression has been connected with mutations at c-MYC, PIM1, RhoH and PAX5 oncogenes, promoting diffuse large B cell lymphoma (Rossi et al., 2006). Similarly, Burkitt's lymphoma is characterized by AID-induced mutation and a reciprocal translocation between the c-MYC proto-oncogene and the IgH loci (Casellas et al., 2009; Pérez-Durán et al., 2007; Ramiro et al., 2004; Robbiani et al., 2008). While AID expression was initially thought to be B cell-specific, recent evidence indicates AID may promote the development of various non-lymphoid oncologies. In gastric cancer, the upregulation of AID leads to point mutations and copy number alterations of CDKN2A and CDKN2B tumor suppressor genes (Endo et al., 2011). In the same manner, AID misexpression in human colonic cells increases the mutation rate of TP53 by ~10-fold (Matsumoto et al., 2010). In addition, AID is expressed at various levels in ~1/3 of primary lung cancers (Shinmura et al., 2011) and in numerous breast cancer cell lines (Babbage et al., 2006) where AID misexpression may be due in part to observations that estrogen is capable of inducing AID expression > 20 fold (Pauklin et al., 2009). Since errant AID activity can introduce significant genomic damage, the maintenance of
genome stability outside of the activated B cell environment likely depends upon multiple molecular AID restraints.

Recent evidence suggests that one level of AID regulation comes from microRNAs (miRs). These short, ~20 nucleotide, noncoding RNAs can regulate networks of genes and function by repressing translation or directing mRNA destruction through partial sequence complementarity to 3' untranslated regions (3'UTRs) of mRNAs (Lau et al., 2001; Lee and Ambros, 2001). MiRs are key regulators of cellular differentiation, proliferation and apoptosis, and aberrant miR expression has been associated with a myriad of human diseases, including cancer (reviewed by (Farazi et al., 2011). A subset of miRs typically misexpressed in malignancy (oncomiRs) can function as oncogenes or tumor suppressors with impacts on cellular transformation and metastasis. One such oncomiR, miR-155, represses murine AID, and the disruption of the miR-155 recognition site or miR-155 itself results in increased AID-induced c-MYC translocation (Dorsett et al., 2008) and BCL6 mutagenesis (Teng et al., 2008). This may provide one explanation for why disruption of miR-155 is associated with Burkitt's lymphoma (Kluiver et al., 2006) and suggests the existence of parallel miR repressions in humans.

In order to better define the regulators of human AID in cells other than antigen-activated B cells, we examined the 3'UTR of human AID for miR complementarities. Consistent with the role of miR-155 in regulating AID in mice (Dorsett et al., 2008; Teng et al., 2008), we find a moderate miR-155 site in human AID. Surprisingly, we also find a previously uncharacterized, yet
exceptional, complementarity to miR-93 in the AID 3'UTR. We show both of these miRs interact with the AID 3'UTR in the cell to regulate its translation, and that loss of either miR results in increased AID protein levels. Based on the involvement of AID in generating oncogenic genome mutations, our results suggest that miR-93 and miR-155 act as dual genome sentries to prevent errant translation of AID mRNA.

RESULTS

The Human AID 3'UTR Contains Sequences Complementary to MiR-93 and MiR-155

In light of the evidence supporting AID regulation by miR-155 in mouse showing miR-155 regulates AID translation and c-myc translocation (Dorsett et al., 2008; Teng et al., 2008), we asked if human AID is regulated by similar microRNA interactions. Computational analysis of the human AID 3'UTR revealed a human miR-155 target site, consistent with studies in mice (Dorsett et al., 2008; Teng et al., 2008). To our surprise, we also found pronounced complementarity between the AID 3'UTR and miR-93 (Figure 12). This miR-93 target site is markedly more extensive than the majority of characterized miR::mRNA interactions containing only a single, centrally located 4 base pair mismatch. Strikingly, we find the predicted binding affinity between miR-93 and the AID 3'UTR (-30.3 kcal/mole) to be nearly double that of miR-155 (-15.3 kcal/mole) (Figure 12A) suggesting a strong post-transcriptional regulatory relationship. Further analysis of the human AID 3'UTR revealed an internal 20 nt stretch of adenosines (sufficient for cDNA poly-T priming) separating the miR-
155 and miR-93 target sites (Figure 12A). This false polyA is located ~1,000 nt 3' of the AID stop codon and does not correspond to the full length cDNA polyadenylation as: (1) it is not preceded by the canonical 5' AAUAAA polyadenylation signal or other recognizable alternative poly adenylatation signal, and (2) the miR-93 target site is expressed in both Ramos and MCF-7 cells (Figure 12B). Importantly, examination of publically available EST datasets confirms this internal polyA stretch has repeatedly served as an oligo-dT priming site for EST reverse transcriptase leading to the improper annotation of a truncated AID cDNA (for example see NCBI accessions GI:33871601 and GI:50496022). We conclude that the human AID 3'UTR contains miR-155 and miR-93 recognition sites, which are separated by a 20 nucleotide adenosine repeat upstream from the true polyA sequence.

**MiR-93 and MiR-155 Interact with the AID 3'UTR to Inhibit its Expression**

The identification of miR-93 and miR-155 binding sites in the 3'UTR of human AID strongly suggests post-transcriptional regulatory interactions. Therefore, we next asked if miR-93 and miR-155 were capable of regulating the AID 3'UTR using standard luciferase reporter assays. Three expression constructs incorporating an endogenous human microRNA Alu promoter were created by cloning the endogenous miR-517a genomic locus and then replacing the miR-517a hairpin with the miR-93, miR-155 and shLacZ hairpins (pAL-93, pAL-155, and pAL-1 respectively) (Figure 13A), as described (Borchert et al., 2006). We also generated 3'UTR reporter constructs to testAID::miR regulatory interactions. These reporters (AID 3'LR and Ctl 3'LR) produce Renilla luciferase
transcripts tailed with the AID 3'UTR and shLacZ target sequence respectively (Figure 13). Co-transfection of human embryonic kidney (HEK293) cells with pAL-93 and AID 3'LR constructs resulted in ~80% knockdown of Renilla luciferase intensity, whereas co-transfection of pAL-155 and AID 3'LR constructs resulted in ~60% knockdown of Renilla luciferase intensity at a 4:1 transfection ratio (Figure 13C). Similarly and as expected, co-transfection of pAL-1 and Ctl 3'LR resulted in a reduction of Renilla luciferase intensity with ~95% knockdown (Figure 13D). In contrast and as expected, neither the co-transfection of pAL-1 with AID 3'LR nor co-transfection of either miR construct with Ctl 3'LR resulted in silencing (Figure 13C, D). To further ensure the repressions of AID 3'LR by pAL-93 and pAL-155 were due to specific interactions between miRs -93 and -155 and their proposed target sites in the AID 3'UTR, we constructed two additional luciferase reporters whose 3'UTRs consisted of: (1) the two target sites with intervening sequence removed and (2) scrambled versions of the two target sites with intervening sequence removed. Again, we found co-transfection of human embryonic kidney (HEK293) cells with pAL expression constructs and a luciferase reporter containing intact miR target sites resulted in significant repression of Renilla luciferase in contrast to control transfections of scrambled miR target site reporters, which were not repressed (Figure 17). Therefore, miR-93 and miR-155 facilitate molecular interactions with the AID 3'UTR sufficient to repress translation.
MiR-93 and MiR-155 can Repress Endogenous AID

Interactions between miR-93, miR-155 and the AID 3'UTR within the context of our reporter assays (Figure 13C, D) strongly support the involvement of both miRs in AID translational regulation. If so, we anticipated that if miR-93 and/or miR-155 were present in AID expressing cells they could suppress endogenous AID protein levels. We therefore asked if miR-93 and miR-155 are present in two cell lines known to express AID, Ramos and MCF-7. Ramos is a B cell line that undergoes constitutive AID-induced somatic hypermutation (Sale and Neuberger, 1998), and MCF-7 is a breast carcinoma that misexpresses AID (Babbage et al., 2006). Using standard RT-PCR analysis, we found miR -93 and -155 transcripts are expressed in both cell lines (Figure 18), consistent with the ubiquitous expression documented for both of these miRs (Landgraf et al., 2007). In order to determine if endogenous AID can be repressed by increased miR-93 and/or miR-155 expression, we introduced pAL miR over-expression plasmids (Figure 13A) into MCF-7 cells (Borchert et al., 2006). Importantly, overexpression of miR-93, miR-155, or an AID-targeting siRNA (shAID) in MCF-7 cells each reduced AID protein levels between 70 and 90%, whereas overexpression of the pAL-1 control did not significantly alter AID expression (Figure 14A). To ensure results were the consequence of miR regulation, we confirmed the miR-93 repression of AID was dose-dependent (Figure 19) and isolated total RNA from each of our transfections to monitor AID mRNA levels by quantitative PCR. Strikingly, while shAID transfection averaged a 92% reduction of AID mRNA levels, neither pAL-93 nor pAL-155 transfection resulted in significant reduction
of AID mRNA (Figure 14B) confirming that the observed reductions in AID protein levels following their transfections are mediated through translational repression. To summarize, our results demonstrate that the decreases in AID protein upon shAID transfection were due to message degradation, whereas pAL-93 and pAL-155 overexpression resulted in miR-mediated translational repression.

Endogenous MiR-93 and MiR-155 Restrain AID Protein Translation

Over-expression of miR-93 and miR-155 repressed the expression of endogenous AID, supporting the model that both miRs contribute to post-transcriptional regulation of AID. To be consistent with that hypothesis, the loss of cellular miR-93 or miR-155 should result in higher AID expression. Therefore, we depleted miR-93 and miR-155 individually and then monitored changes to AID protein levels by western blotting. To achieve this, we constructed miR sponge transgenes for miR-93 and miR-155 that function by competitively inhibiting miR activity, as previously described (Ebert et al., 2007; Ebert and Sharp, 2010). Our miR sponge transgenes produce RNAs containing ~10 copies of the miR-93 or miR-155 AID 3'UTR complementary binding sites (Figure 15A). Individual miR sponges or a negative control sponge transgene were transiently transfected in MCF-7, and relative AID protein levels measured by western. MCF-7 cells expressing either miR-93 or miR-155 sponges showed clear dose-responsive increases in AID protein levels, whereas control transfections did not significantly alter AID protein expression (Figure 15B). To ensure these findings were the consequence of restricted miR activity, we isolated total RNA from each of our transfections and examined AID mRNA levels by quantitative PCR. We
found AID mRNA levels largely unaffected by transfection of sponge constructs further supporting endogenous roles for these miRs in the restraint of AID translation (Figure 15C). We conclude that one function of endogenous miR-93 and miR-155 expression is to restrict the translation of AID mRNA.

**DISCUSSION**

**Model for AID Repression**

We have shown that miR-93 and miR-155 have the capacity to repress AID translation and do so in the cellular context. Overexpression of miR-93 or miR-155 reduced AID protein levels in MCF-7 breast cancer cells (Figure 14A), whereas depletion of either endogenous miR-93 or miR-155 resulted in increased AID translation (Figure 15) directly connecting miR expression with changes in AID protein concentration. Based on our findings, we propose a model for the relationship between miRs -93 and -155 and AID regulation in maintaining genomic integrity (Figure 16). In cellular circumstances where high levels of AID transcription exist (such as antigen activated B cells or some established cancers) AID mRNA levels overwhelm miR controls. In these situations, endogenous miR-155 and miR-93 have negligible influence on AID protein levels. In contrast, miR-93 and miR-155 can greatly reduce the potential for errant AID expression which protects the genome from unwanted cytidine deamination and mutation by restricting AID protein production (Figure 16). This may be particularly important during S-phase when replication transiently denatures the DNA. Conspicuously, miR-93 is intronically encoded within the minichromosome maintenance 7 (MCM7) gene, an essential replication licensing
factor, and precursor miR-93 abundance tracks with MCM7 expression levels (Long et al., 2010; Petrocca et al., 2008). It is therefore tempting to speculate that domestication of the miR-93 microRNA along with the corresponding AID 3'UTR site occurred as a response to the selective pressure for maintaining genome fidelity during replication.

Disruptions of MiR-155 and MiR-93 are Associated with AID-Induced Oncogenesis

Highlighting the importance of tightly controlled AID expression, novel AID regulations continue to be described. In 2008, miR-155 was identified as an important repressor of AID translation in mice, and loss of this regulation is lymphomagenic (Dorsett et al., 2008; Teng et al., 2008). Our data now demonstrate directly that expression of endogenous miRs -93 and -155 represses AID translation in MCF-7 breast cancer cells. While a direct correlation between miR-93 activity and AID-induced oncogenesis remains to be described, miR-93 perturbations have been found to enhance cell survival, possess oncogenic activities, and augment tumor growth through regulating integrin-β8, the tumor suppressor gene FUS1, the Cdk inhibitor p21, and tumor protein 53-induced nuclear protein 1 (TP53INP1) (Du et al., 2009; Fang et al., 2011; Kim et al., 2009; Yeung et al., 2008). Our results suggest that in addition to these known oncogene regulations, loss of miR-93 (or miR-155) may permit AID upregulation and mutator phenotypes in non-B cell oncologies (e.g. breast, colon, stomach and lung (Babbage et al., 2006; Endo et al., 2011; Matsumoto et al., 2010; Shinmura et al., 2011). However, it is possible that additional miR AID
regulations may still be described, and our research does not seek to eliminate other potential regulators.

**MicroRNA Target Sites Obscured by Repetitive Element Insertions**

Our identification of an extensive and uncharacterized 3'UTR for human AID reveals the presence of sequence elements that may confound miR regulatory relationships for other mRNAs. We find that the AID 3'UTR is ~1000 nt longer than previously annotated likely due to mispriming from an internal stretch of 20 adenosines. The full length AID 3'UTR contains a previously unidentified regulatory sequence, an unusually pronounced miR-93 complementarity (Figure 12), conferring marked repression to targeted transcripts (Figure 13 and 17). Clearly, this finding agrees with recent suggestions that cryptic miR complementarities may actually be quite common with nearly 50% of human and mouse 3'UTRs likely extending well beyond their annotated termini (Djebali et al., 2008; Thorrez et al., 2010). Since 3'UTRs contain miR binding sites, our analysis of AID is consistent with the notion that other miR posttranscriptional regulations may have been overlooked because of truncated 3'UTR annotations. This may be particularly pronounced in humans as 5-10% of all human 3'UTRs contain at least one Alu repeat, a common miR-associated retro-element characterized by a central 10-40 nt adenosine linker (Borchert et al., 2009; Borchert et al., 2006; Häslar et al., 2007; Jurka and Gentles, 2006; Lehnert et al., 2009; Mighell et al., 1997; Smalheiser and Torvik, 2006; Yulug et al., 1995).
CONCLUSION

In the present study, two human microRNAs, miR-155 and miR-93, were each shown to repress the translation of human AID through interactions with the AID 3'UTR in the MCF-7 breast carcinoma cell line which aberrantly expresses AID. Together, our data suggest that low-level errant AID expression and subsequent genome damage may be prevented through protective miR-93 and/or miR-155 regulation. In addition, our identification of a miR-93 target site for AID located downstream of an internal adenosine repeat highlights the possibility that other miR regulations may have been overlooked because of truncated 3'UTR annotations.

MATERIALS AND METHODS

Reagents and Cell Lines

Oligonucleotide sequences are detailed in Table 1. Human embryonic kidney (HEK293) cell line was obtained from GenLantis (San Diego, CA), Burkitt's lymphoma (Ramos) and breast cancer (MCF-7) cells were both purchased from the American Type Culture Collection, (ATCC, Manassas, VA). HEK293 and MCF-7 cells were cultured in MEM (Mediatech, Herndon, VA) supplemented with 10% fetal bovine serum (Hyclone, Logan, UT). Ramos cells were cultured in RPMI (Mediatech) supplemented with 10% fetal bovine serum (Hyclone, Logan, UT). All tissue culture media were supplemented with 25 mg/ml streptomycin and 25 I.U. penicillin (Mediatech). Cells were cultured in a humidified atmosphere with 5% CO2 at 37°C. For luciferase assays, HEK293 s were cultured in MEM (10% FBS and 1% PS) in 12-well plates. At 90%
confluency, cells were transfected following the Lipofectamine 2000 (Invitrogen, Carlsbad, CA) protocol. At 35 h, existing media was replaced with 1 ml fresh media. For miR expression assays, MCF-7 s were cultured in MEM (10% FBS and 1% PS) in 6-well plates. At 90% confluency, cells were transfected following the Lipofectamine 2000 (Invitrogen) protocol and harvested after 48 h. For miR sponge assays, MCF-7 s were cultured in MEM (10% FBS and 1% PS) in 6-well plates. At 70% confluency, cells were transfected following the Lipofectamine 2000 (Invitrogen) protocol. At 48 h, existing media was replaced with 2 ml fresh media and cell transfections were repeated as initially performed. Cells were then harvested after an additional 48 h culture.

**Luciferase Assays**

At 36 h post transfection, cells were scraped from well bottoms and transferred to 1.5 ml Eppendorf tubes. Eppendorfs were centrifuged at 2000 RCF for 3 min, followed by supernatant aspiration and cell resuspension in 300 μl of PBS. Cells were lysed by freeze thaws and debris removed by centrifuging at 3000 RCF for 3 min. 50 μl of supernatant was transferred to a 96-well MicroLite plate (MTX Lab Systems, Vienna, VA) then firefly and Renilla luciferase activities measured using the Dual-glo Luciferase® Reporter System (Promega, Madison, WI) and a 96-well plate luminometer (Dynex, Worthing, West Sussex, UK). RLU’s were calculated as the quotient of Renilla/firefly RLU and normalized to mock.

**Western Blotting**

MCF-7 cells (at ~1 × 10^6 cells/ml) were pelleted by centrifugation, existing media removed, and cells resuspended in SDS lysis buffer containing protease
inhibitors and transferred to 1.5 ml Eppendorf tubes. Proteins were electrophoresed through a 4-12% SDS-polyacrylamide gradient gel (Invitrogen) and transferred to immobilon-P PVDF membranes (Millipore, Temecula, CA). Membranes were blocked for 1 hour in 5% (w/v) nonfat milk in phosphate-buffered saline containing 0.05% Tween 20, washed, and incubated with primary antibody overnight at 4°C using the following dilution: anti-AID (Santa Cruz Biotechnology, Santa Cruz, CA, sc-25620) - 1:500 and anti-PCNA (Santa Cruz) - 1:10000. Membranes were washed and incubated with secondary Abs: HRP conjugated goat anti-mouse and goat anti-rabbit (Invitrogen) at 1:10000 dilution. Immunoreactive bands were visualized with ECL Plus (GE, Piscataway, NJ) and signals were detected by using the Storm 840 PhosphorImager and IMAGEQUANT software (GE).

**Vector Construction**

Unless otherwise indicated, PCR amplifications were performed in 40 μl reactions at standard concentrations (1.5 mM MgCl₂, 0.2 mM dNTP, 1× Biolase PCR buffer, 0.5 U Taq (Bioline USA, Inc., Randolph, MA), 0.5 uM each primer) and using standard cycling parameters (94°C - 3 min, (94°C - 30 s, 55°C - 30 s, 72°C - 60 s) × 30 cycles, 72°C - 3 min) then cloned into Topo PCR 2.1 (Invitrogen). RT-PCRs were performed at 65°C using MonsterScript Reverse Transcriptase (#MSTA5110, Epicentre, Madison, WI) and gene specific or random nonamer primers. Resultant amplicons were cloned into Topo PCR 2.1 and sequenced. Antisense reporter, Ctl 3'LR, was constructed by oligonucleotide primer extension (25 cycles with 10 s extensions) with primers containing 5’ Xho-I
and 3' Spe-1 restriction enzyme sites immediately flanking sequences perfectly complementary to mature siLacZ. Antisense reporter, AID 3'LR, was constructed by standard PCR with primers containing 5' Xho-I and 3' Spe-1 restriction enzyme sites. Following digestion, amplicons were ligated into the Renilla luciferase 3'UTR of psiCheck2 (Promega) vector linearized with Xho-I and Spe-I. The presence of an independently transcribed firefly luciferase in these reporters allowed normalization for transfection efficiency. Sponge expression constructs were generated by concatamerizing PCR using the primers indicated in Table 1. Resulting amplicons were separated on a 1% agarose gel and a band excised from the appropriate lane at ~400 bp. Gel extractions were cloned into Topo PCR 2.1 and sequenced. Concatamers were next excised from Topo PCR 2.1 and cloned into the pEGFP expression vector (Clontech, Mountain View, CA) using BamHI and Not I restriction sites.
REFERENCES


Sale, J. E., and M. S. Neuberger, 1998, TdT-accessible breaks are scattered over the immunoglobulin V domain in a constitutively hypermutating B cell line: Immunity, v. 9, p. 859-69.


FIGURES

**Figure 12. The AID 3'UTR.** (A) Cartoon of AID 3'UTR miR target sites (TS) and internal (False) polyA (not to scale) (DNA sequence shown). MicroRNA alignments are shown with corresponding binding affinities (kcal/mol). Vertical lines indicate complementary base pairing. Dotted vertical lines indicate G:U base pairing. PolyA, AATAAA eukaryotic polyadenylation signal. Arrows indicate RT-PCR primer sets for confirmation of expression. (B) RT-PCR to test expression of the AID 3'UTR miR-155 and miR-93 target site (TS) sequences in Ramos and MCF-7 cells.
Figure 13. The AID 3'UTR is Recognized by MiRs-93 and -155. (A) Diagrams of miR and siRNA expression vectors. 5'Alu corresponds to a genomic fragment containing the miR-517a upstream Alu promoter. In pAL-1, shLacZ replaces the miR-517a hairpin. In pALs-155 and -93, the pre-miR-155 and -93 hairpins replace the miR-517a hairpin respectively. TTTT, RNA polymerase III termination signal. (B) Diagrams of luciferase reporters. The AID 3'UTR luciferase reporter (AID 3'LR) and control 3'UTR luciferase reporter (Ctl 3'LR) were generated respectively by placing an AID 3'UTR sequence containing both the miR-155 and miR-93 target sites and a specific region of LacZ previously shown to be targeted by the pAL-1 shLacZ hairpin in a multiple cloning site in the Renilla luciferase 3'UTR. HSV-TKp, herpes simplex virus thymidine kinase promoter; SV40p, simian virus 40 promoter; FF Luc, firefly luciferase; Ren Luc, Renilla luciferase; pA, poly(A). (C) AID 3'LR is specifically repressed by both miR-93 and miR-155 in 293 transient transfections. Luciferase assays (n = 4) of HEK293 lysates after transfection of AID 3'LR and either pAL-1, pAL-155 or pAL-93. Transfections were normalized to AID 3'LR alone. RLU, relative light units. (D) Ctl 3'LR is specifically repressed by the ctl (pAL-1) in 293 transient transfections. Cotransfections done as in (C), except Ctl 3'LR replaced AID 3'LR.
Figure 14. MiR-93 and MiR-155 Can Repress Endogenous AID. (A) Representative western demonstrating that transient transfection of pAL-93, pAL-155 and shAID expression constructs repress endogenous AID protein levels in MCF-7 cells. ShAID produces an AID-targeting siRNA from a pAL-driven hairpin expression vector as in 2A. Ctl, pAL-1 transfected; PCNA, Proliferating Cell Nuclear Antigen load control. (B) Relative AID mRNA levels corresponding to replicate transfections performed as in A (n = 3) as determined by quantitative PCR.
Figure 15. Endogenous MiR-93 and MiR-155 Restrain AID Protein Expression. (A) Diagram of miR sponge expression vector. TS indicates an AID 3’UTR miR target site detailed in 1A. The miR-93 sponge produces a polyadenylated concatamer of 11 individual 93 TSs. Analogously designed, the miR-155 sponge contains 10 individual miR-155 TSs, and the ctl sponge contains 14 individual shLacZ targets. CMVp, immediate early cytomegalovirus promoter; SV40pA, simian virus 40 polyadenylation signal. (B) Representative western demonstrating that transient transfection of MCF-7 cells with the miR-93 or miR-155 sponges results in dose responsive increases in AID protein levels as compared to control. Ctl, shLacZ target site sponge transfected; PCNA, Proliferating Cell Nuclear Antigen. (C) Relative AID mRNA levels corresponding to replicate transfections performed as in B (n = 3) as determined by quantitative PCR.
Figure 16. Model of MiR-93 and MiR-155 Regulation of AID Mediated Genomic Instability. In normal cells, top, genome stability is maintained by inhibition of AID translation by miR-93 and/or miR-155 when AID mRNA levels are low. Bottom, in antigen activated B cells or when misregulated, levels of AID mRNA exceed the capacity for miR translational repression. AID translation results in AID protein levels that support Ig gene diversification, mutagenesis and genome instability.
Figure 17. The AID 3’UTR is Recognized by MiRs -93 and -155. (A) The AID 3’UTR recognition by miRs -93 and -155 is specific. (A) Diagrams of luciferase reporters. The AID 3’UTR luciferase reporter (TS Intact) and control 3’UTR luciferase reporter (TS Scram) were generated respectively by placing the AID 3’UTR miR -93 and -155 target sites (TS Intact) (with intervening sequences deleted) or an identical sequence in which the target site nucleotides had been scrambled (TS Scram) into a multiple cloning site in the Renilla luciferase 3’UTR [32]. HSV-TKp, herpes simplex virus thymidine kinase promoter; SV40p, simian virus 40 promoter; FF Luc, firefly luciferase; Ren Luc, Renilla luciferase; pA, poly(A); 155TS, miR-155 target site; 93TS, miR-93 target site; 155SC, miR-155 target site scrambled; 93SC, miR-93 target site scrambled. (B) TS Intact is specifically repressed by both miR-93 and miR-155 in 293 transient transfections. Luciferase assays (n = 9) of HEK293 lysates after transfection of TS Intact and either pAL-1, pAL-155 or pAL-93. Transfections were normalized to TS Intact alone. RLU, relative light units. (C) TS Scram is unaffected by co-transfection of pAL-1, pAL-155 or pAL-93 in 293 transient transfections. Cotransfections done as in B, except TS Scram replaced TS Intact.

### Diagrams

- **TS Intact**: HSV-TKp FF Luc pA SV40p Ren Luc 155TS 93TS pA
- **TS Scram**: HSV-TKp FF Luc pA SV40p Ren Luc 155SC 93SC pA

### Graphs

**TS Intact**
- RLU vs. pAL-1
- RLU vs. pAL-155
- RLU vs. pAL-93

**TS Scram**
- RLU vs. pAL-1
- RLU vs. pAL-155
- RLU vs. pAL-93
Figure 18. MicroRNAs MiR-93 and MiR-155 are Widely Expressed. RT-PCR of pre-miR template in three different cell lines, HEK293 embryonic kidney cells, MCF-7 breast cancer line, and Ramos Burkitt's lymphoma line. (top) RT-PCR products of the miR-93 locus amplified from indicated cellular RNAs. m, miR-93; β, Beta actin positive control; -Ctl, no template. (bottom) RT-PCR products of the miR-155 locus amplified from indicated cellular RNAs. Annotations are same as above except (m) denotes pre-miR-155 amplicon. PCR of source RNAs failed, indicating above RT-PCRs templated from the generated cDNA.
Figure 19. MiR-93 Represses Endogenous AID in a Dose Responsive Manner. (A) Representative western demonstrating that transient transfection of pAL-93 expression constructs repress endogenous AID protein levels in MCF-7 cells in a dose responsive manner. (B) Relative AID mRNA levels corresponding to replicate transfections performed as in A (n = 2) determined by quantitative PCR.
### Table 2. Oligonucleotide Master List

All oligonucleotides were synthesized commercially (Fisher, Operon) at either 25 or 100 nM scale. pAL refers to the human miR-517a Alu promoter. 93 and 155 refer to human miRs. TS refers to a siRNA or miR target site. lowercase indicates imperfect antisense, AID 3'LR was generated using TS-155F and TS-93R.

#### pAL Construct Primers

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>5'Alu Forward</td>
<td>CCCAAGGACACCCTCATGATCTCCC</td>
</tr>
<tr>
<td>3'pAL-93R</td>
<td>AAAAACTCGGGAAGTGCTAGCTCAGCAGTAGGGTTGGGTAATCACA</td>
</tr>
<tr>
<td></td>
<td>CTACCTACTTTGGGAGTTGATATCTGCAGAATTTCG</td>
</tr>
<tr>
<td>3'pAL-155R</td>
<td>AAAAACTGTTAATGCTAATATGTAGGAGTCAGTTGGGAGGCAAAAAAA</td>
</tr>
<tr>
<td></td>
<td>CACCCTCTATGCAGAATTTCG</td>
</tr>
<tr>
<td>3'pAL-1R3</td>
<td>AAAAAGCTGACCTATCCCATTACGTGACAGGAAGCGTAATGGGAGG</td>
</tr>
<tr>
<td></td>
<td>ATGATATCTGCAGAATTTCG</td>
</tr>
<tr>
<td>3'pAL-shAIDr</td>
<td>AAAAAGAATGTCCGCTGGGCTAAGGTGACAGGAAGCCTTAGCC</td>
</tr>
<tr>
<td></td>
<td>AGGATATCTGCAGAATTTCG</td>
</tr>
</tbody>
</table>

#### Sponge Construct Primers

<table>
<thead>
<tr>
<th>Sponge</th>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>93 Sponge</td>
<td>Gb130SPG93TSf</td>
<td>TGAATATAACGCTGTAATCCAGCAGCATTTGGAGGTGAATAA</td>
</tr>
<tr>
<td></td>
<td>Gb131SPG93TSr</td>
<td>TATTTGATGGATATCTGCAGAATTTCG</td>
</tr>
<tr>
<td>155 Sponge</td>
<td>Gb132Spg155TSf</td>
<td>GGTGAATATATGGGGATAGTGCTAATCCAGGAATTGGTGAATAA</td>
</tr>
<tr>
<td></td>
<td>Gb133Spg155TSr</td>
<td>TATATTCAACATTACGAGTACCTCAGCATCCCATATAATTACC</td>
</tr>
</tbody>
</table>
**Sponge Control1**
Gb140spgsiLacZ5p10f
GTGAATATATTAATACGTGACCTATTaaTGGTGAAATATA
Gb141spgsiLacZ5p10r
TATATTCCACCAAttAATAGGTCACGTATTAATATATTTACCC

**AID 3’UTR Target Sites and Reporter Primers**

**MiR-93-TS**
TS-93F  GGAAGGAAGTTGCTTGAATGTTGG
TS-93R2 GAGACAGAGTCTTGCTCTTTGTCG
MiR-155-TS TS-155F2  CCTGGGAGCATCCTAAAGTGTCAAC
TS-155R  TCTTGTCACGATCCTCCTGCCC

**Ctl 3’LR3 LacZ_TS_F**
TTCTCGAGCAGTTACGATGCGCCCATCTACACCAACGTGACCTATCCCATTA
CGGTCAA
LacZ_TS_R
AAGCGGCCGCAGGAACAAACGGCGGATTGACCGTAATGGGATAGGTTCAGC
TTGGTGT

**TS Intact**
TSIntactF
AAACTCGAGTGATTGTGAACCCCAAAACCATCTCTCCTCCTAAACATTAATATCCAA
TCTATGCGCGCGTGGTGCTC
TSIntactR
TTTGCAGGTCGCTCGCCCTCCCAAAGTGCTGGATTACAGGCGTGAGCC
ACCACCGCCGCGCATGATTGG

**TS Scram**
TSScramF
AAACTCGAGTGATTGTGAACCCCAAAACCATCTCTCCTCCTCCTCCTAAGAACTCATCCAA
TCTATGCGCGCGTGGTGCTC
TSScramR
TTTGCAGGTCGCTCGCCCTCCCAAAGTGCTGGATTACAGGCGTGAGCC
ACCACCGCCGCGCATGATTGG

**RT-PCR Primers for miR Expression**

**Pre-miR-93**
miR-93F  GAGATGAGGCGAGAGGCTTGGG
miR-93R  CCTTTGACATCTCCATTAGCCTGAGG

**Pre-miR-155**
miR-155F GCCTCCAAGTGAACCTCCTACATATTAGCC
miR-155R GTTAAGGGTAAACATCCAGTACCAG

Table Continues
β Actin
gb204_bAct2F2_62 CCAACCGCGAGAAGATGACCC
gb205_bAct2R1_61 CTCCTTAATGTCACGCACGATTTCCC

Quantitative PCR Primers
AID
AID-F GTGACCCCAAACCATCTCTCCAAAGC
AID-R TCCCCAACATTCAAGCAACTTCTTCTTCC

β Actin
bActF GGCTCCGGCATGTGCAAGG
bActR GGTGAGGATGCCTCTCTTGCTCTG

tRNAglu
trna_glu_F AAGGCGCATCTCTAGTTGAGTG
trna_glu_R2 CTCTTTTGCTCTGAGTCTCTCAG
CHAPTER IV

G-QUADRUPLEX DNA STRUCTURES CAN INTERFERE WITH URACIL GLYCOSYLASE ACTIVITY IN VITRO

Contributions:
The contents of this chapter have been submitted to the journal Mutagenesis and the manuscript is currently under review for acceptance and publication (October 2015). I performed all experimentation and fully prepared the manuscript with Erik Larson.
ABSTRACT

Genome sequences that contain tandem repeats of guanine can form stable four-stranded structures known as G-quadruplex, or G4 DNA. While the molecular mechanisms are not fully defined, such guanine-rich loci are prone to mutagenesis and recombination. Various repair pathways function to reduce the potential for genome instability by correcting base damage and replication errors; however, it is not yet fully defined how well these processes function at G4 DNA. One frequent form of base damage occurs from cytidine deamination, resulting in deoxyuracil and UG mismatches. In duplex and single-stranded DNA, uracil bases are recognized and excised by uracil glycosylases. Here, we tested the efficiency of uracil glycosylase activity in vitro on deoxyuracil bases located directly adjacent to guanine repeats and G4 DNA. We show that uracil excision by bacterial UDG and hUNG2 is reduced at uracils positioned directly 5’ or 3’ of a guanine tetrad. Control reactions using oligonucleotides disrupted for G4 formation or reaction conditions that do not favor G4 formation resulted in full uracil excision activity. Based on these in vitro results, we suggest that folding of guanine-rich DNA into G4 DNA results in a DNA conformation that is resistant to uracil glycosylase-initiated repair, and this has the potential to increase the risk of instability at guanine repeats in the genome.

INTRODUCTION

Guanine-rich and repetitive genomic DNA has the unique ability to adopt four-stranded conformations called G4 DNA, and the impact of those structures on DNA metabolism is important for clarifying the molecular sources of genetic
Physiological salt and pH conditions promote G4 folding into structures built from stacks of guanine “tetrads”, each of which is composed of four guanine bases paired to each other via Hoogsteen bonds (Sen and Gilbert, 1988, 1990). The specific conformation and overall stability of a given G4 is influenced by the sequences, ions, and reaction conditions (Burge et al., 2006). In higher eukaryotes, G4 sequences are located at sites of programmed recombination and at a variety of unstable intergenic regions where G4 structure formation promotes site-specific instability and disease (Cea et al., 2015; Davis and Maizels, 2011; Maizels, 2015). Of particular note are G4-capable sequences enriched at oncogenes, where G4 structure formation may contribute to gene regulation and oncogenesis (Eddy and Maizels, 2006). G4 structures have been shown to form from guanine-repeat sequences at rearrangements sites, such as the \textit{HOX11} (Nambiar et al., 2013), \textit{BCL2} (Nambiar et al., 2011), and \textit{TCF3} (Williams et al., 2015) cancer-related genes.

When DNA becomes transiently denatured during transcription or replication guanine-rich and repetitive DNA is freed from the complement and permitted to adopt structures like G4, where they can subsequently interfere with various DNA metabolic pathways (Bochman et al., 2012; Brooks et al., 2010; Tarsounas and Tijsterman, 2013; van Kregten and Tijsterman, 2014; Wu and Brosh, 2010). This raises the possibility that some DNA repair activities may be inhibited at G4 DNA, contributing to the apparent instability that characterizes many guanine-rich loci. One major pathway, base excision repair (BER), reduces
mutagenesis by correcting bases that were damaged by deamination or oxidation (Krokan and Bjørås, 2013; Krokan et al., 2014). BER is initiated by a lesion-specific glycosylase that removes the damaged base, followed by excision of the resulting abasic site to permit subsequent resynthesis to reconstitute the original base pairing. Emerging evidence suggests that some DNA lesions in G4 DNA may not be processed efficiently by BER. In telomeric G4, activity assays have shown that 8-oxoguanine is not excised by human glycosylases, even though further oxidation products can be processed (Zhou et al., 2015; Zhou et al., 2013). The APE1 enzyme, which cleaves abasic sites, was also shown to have reduced enzymatic activity at abasic sites residing within certain G4 DNA structures (Broxson et al., 2014; Zhou et al., 2015). These findings support the model that G4 DNA can interfere with DNA correction activities, and could therefore contribute to mutagenesis at guanine-rich loci by reducing DNA repair efficacy at those sites.

Here we asked if G4 DNA interferes with excision of a common form of DNA damage, deaminated cytidine (deoxyuracil), by uracil DNA glycosylase in vitro. Using synthetic oligonucleotide substrates, we placed a single deoxyuracil either directly 3’ or 5’ of a guanine repeat sequence or in between repeats, and then measured the activities of two uracil glycosylases, hUNG2 (human) and UDG (E. coli) on those DNAs under conditions that allow G4 to fold. Activities were compared to controls that cannot adopt G4 structures. We find that the excision of uracil is reduced when deoxyuracil is positioned directly adjacent to guanine repeats participating in G4 formation, but not when deoxyuracil is
located three bases away from the tetrad. Addition of APE1 to the reactions did not increase abasic site cleavage, suggesting that the initiating steps of BER of uracil are reduced at G4 DNA. Together, our results suggest that deaminated cytidines positioned adjacent to G4 DNA structures are poor substrates for the BER pathway.

RESULTS

G4 Formation with Deoxyuracil Oligonucleotides

We selected a G4 DNA sequence from the TCF3 (E2A) gene as a model structure, which we previously showed folds into G4 DNA at physiological salt and pH conditions in vitro (Williams et al., 2015) (Figure 20). The goal of this study was to simply test the efficiency of uracil excision activities on DNA containing deoxyuracil bases located directly next to (5’ or 3’) of a guanine tetrad. Therefore, oligonucleotides were synthesized to contain a single uracil lesion at defined positions and as such the U7, U15, and U24 sequences contain a deoxyuracil at position 7, 15 and 24, respectively (Figure 20A). The deoxyuracil is placed directly 5’ (U7), 3’ (U24) or in between (U15) the guanine repeats participating of G4 formation (Figure 20A). Control DNAs were sequence paired to the uracil-containing G4 oligonucleotide, except that the guanine repeats were interrupted with thymine to eliminate stable G4 folding (Figure 20A). These oligonucleotides are distinguished from one another by the terminology “G4” (with guanine repeats) or “GT” (thymine substituted). One important experimental consideration regarding G4 DNA is that a sequence capable of intra-molecular G4 formation will also permit inter-molecular G4 conformations in solution (Sen
and Gilbert, 1990, 1992). Therefore, we expected to observe a mixture of both conformations (Figure 20B). Independent of the precise G4 structure that forms, the U7 and U24 uracil bases will be adjacent to a tetrad. Structure prediction using the Mfold server (Zuker, 2003) did not return any significant fold-back structures for the GT oligonucleotides ($\Delta G < -1.0$ kcal/mol). For the G4-capable sequences, the QGRS G4 prediction software (Kikin et al., 2006) verified that the most likely G4 conformation will involve the longest guanine repeats (using an input with a G-group of four and minimum loop of 1). That results in placement of the U7 and U24 deoxyuracils directly next to a tetrad in either intra-molecular or inter-molecular G4, and this is depicted in Figure 20B. This remains the predominant structure predicted by the program even when the input parameters are reduced to a G-group size of three. Alternate G4 conformations are theoretically possible, but if they do form they would most likely exist as a collection of various minor G4 species in addition to the two predominant inter- and intra-molecular conformations predicted by QGRS mapper (Figure 20B).

Deoxyuracil differs from deoxythymine by only a single methyl group, so we predicted that substituting thymine for uracil in the TCF3 oligonucleotide would not disrupt G4 formation. Further, the deoxyuracils used here reside within a loop, or between G-tetrads (Figure 20A and 20B), so we did not expect to see alterations in guanine-guanine pairing. In other G4 studies, telomeric G4 structures were shown to maintain stability when thymine was substituted for 5-hydroxymethyluracil, a major oxidation product (Konvalinová et al., 2015; Virgilio et al., 2015), so we infer that changes to bases not directly participating in G4
folding will not disrupt the structure. Guanine repeats are needed though, and interruption of those sequences by thymine substitution would eliminate the possibility of G4 formation. Consistent with that, QGRS mapper returned no scores for the GT oligonucleotides used here, indicating the absence of G4 folding potential. In order to experimentally verify G4 formation for the deoxyuracil oligonucleotides used here we applied two different approaches, native PAGE and Circular Dichroism (CD). In native PAGE, G4 DNA structures migrate faster (intramolecular) or slower (intermolecular) compared to unfolded controls of similar molecular weight (GT). In the presence of 100 mM KCl, the migration patterns for G4(U7), G4(U15), and G4(U24) are entirely consistent with G4 DNA (Figure 21A). This matches previous results showing G4 formation with TCF3 sequences in KCl salt, except those oligonucleotides did not contain deoxyuracil (Williams et al., 2015). The GT controls for each G4 sequence migrate predominately as a single species (Figure 21A), as expected. The G4(U24) oligonucleotide paired with complement (double stranded) migrates predominately as a single species and slower than the single-stranded counterparts, as expected (Figure 21A).

CD measures the absorbance of circularly polarized light by chiral molecules and G4 DNA structures produce characteristic CD spectra, with ellipticity maximums at either 264 or 295 nm, and minima at either 265 or 240 nm (Balagurumoorthy et al., 1992; Dapić et al., 2003; Kypr et al., 2009; Vorlíčková et al., 2012). We applied CD analysis to the deoxyuracil containing GT and G4 oligonucleotides shown in Figure 20A to measure their ability to fold into G4 DNA.
in the presence of 100 mM KCl. Only the oligonucleotides containing guanine repeats (G4) produced spectra characteristic of G4 DNA (Figure 21B), consistent with earlier results (Williams et al., 2015). Interruption of the guanine repeats (GT) resulted in a CD spectral shift for each oligonucleotide (Figure 21B). We conclude that the TCF3 oligonucleotides containing deoxyuracil fold into G4 DNA structures, and the GT control sequences do not adopt G4 DNA.

G-Tetrads Interfere with Bacterial UDG Activity

We next asked if uracil glycosylase is active at deoxyuracils residing within guanine-rich DNA and next to a G4 tetrad. We first tested E. coli UDG. Each 5′32P end-labeled and uracil-containing G4 and GT oligonucleotide was incubated with UDG, and after alkaline lysis of the abasic sites the cleavage products were resolved by denaturing PAGE. Generally, G4 is stabilized in solutions containing physiological concentrations of K⁺ and in neutral pH, and it is less stable in the absence of K⁺ salts or in the presence of Li⁺ (Sen and Gilbert, 1990; Sundquist and Klug, 1989; Williamson et al., 1989) Therefore, we performed UDG reactions in either 0 mM, 50 mM or 100 mM KCl, anticipating that 100 mM KCl will result in the most stable G4 DNA. One representative phosphorimage for each PAGE-resolved cleavage assay is shown in figure 22 (left), and to the right of each image are graphs depicting the quantitation of cleavage from at least three independent experiments. In the absence of KCl, all six oligonucleotides were cleaved with near equal efficiency, regardless of the position of the deoxyuracil (Figure 22 A-C). In contrast, UDG activity on G4(U7) and G4(U24) was sharply reduced in the presence of KCl compared to each companion GT control (Figure
KCl had only a marginal effect on the cleavage activity of UDG on G4(U15) (Figure 22B). In contrast, the presence of KCl had a strong effect on the ability for UDG to cleave G4(U7) and G4(U24), with more than 5-fold fewer uracils removed compared to reactions that did not contain KCl (Figures 22A and 22C). KCl at concentrations up to 100 mM did not measurably effect UDG activity on any of the GT oligonucleotides, so the reduction in UDG activity observed for G4(U7) and G4(U24) was not due to the ionic strength of the reaction. Furthermore, UDG showed equivalent activity on all oligonucleotides tested when LiCl was substituted for KCl, up to 100 mM (Figure 20). Together, we concluded that uracil bases located next to guanine repeats are excised by UDG. However, UDG does not cleave the uracil as efficiently when those repeat participate in G4 formation compared single-stranded DNA of identical sequence.

**G-Tetrads Interfere with hUNG2 Activity**

Even though uracil DNA glycosylase is a highly conserved enzyme (Olsen et al., 1989), we next asked if the human UNG2 enzyme shows the same reluctance to process uracil near G4 DNA structures. We performed an identical set of cleavage assays, except that hUNG2 was substituted for E. coli UDG (Figure 23). The activities observed for hUNG2 on each oligonucleotide paralleled results obtained for E. coli UDG (Figure 22). Compared to the GT oligonucleotide, hUNG2 showed reduced activity on G4(U7) and G4(U24) oligonucleotides, but only when KCl and guanine repeats are present (Figures 23A and 23C), whereas activity on G4(15) was indistinguishable from GT(U15) (Figure 23B). The reduced activity is most likely due to the formation of G4
structures because it was only observed when K\(^+\) ions are included in the reaction. Identical reactions performed in the presence of LiCl instead of KCl resulted in full hUNG2 activity (Figure 26). We conclude that hUNG2 activity is inhibited at uracils positioned adjacent to the G4-tetrads, at least for G4 DNA structures formed by the TCF3 oligonucleotide. Further, it appears that weak cleavage activity at G4 DNA is a likely a conserved property of the human (hUNG2) and E. coli UDG enzymes.

In humans, abasic DNA is cleaved by the AP-endonuclease (APE1) enzyme and it was recently demonstrated that APE1 activity is inhibited at G4 DNA (Broxson et al., 2014), suggesting that BER may be compromised at guanine-rich loci. We further tested the efficiency of uracil repair by co-incubating hUNG2 and APE1 with U7, U15 and U24 G4 and GT oligonucleotides all under G4 folding conditions. As expected, deoxyuracil was poorly cleaved from the G4(U7) and G4(U24) oligonucleotides by hUNG2 and APE1 compared to GT(U7) and GT(U24) (Figure 24A and C). The cleavage difference between G4 and GT was highly significant (two tailed t-test, \(P<0.002\)). The G4(U15) oligonucleotide was processed at essentially the same efficiency as GT(U15), although there was as modest, but not significant (\(P>0.20\), reduction in abasic site cleavage for the G4(U15) oligonucleotide (Figure 24B). This compares well to alkaline cleavage results (Figure 23), with the exception that APE1 displayed a lower overall ability to cleave (Figure 24). This suggests that not all abasic sites were cut by APE1 because treatment with NaOH resulted in ~90% of the GT oligonucleotides cleaved in every uracil glycosylase assay (Figure 22, 23, 25 and
Even so, the GT controls are used here as substrate comparisons for the G4 structures, and we conclude that deoxyuracils located next to a G-tetrad are poorly excised by hUNG2 even when APE1 is included in the reaction.

DISCUSSION

G4 DNA structures that form in the genome promote genetic instability and human disease (Cea et al., 2015; Maizels, 2015), yet the reasons why are not fully defined. In this work we asked if G4 DNA structures, or more precisely the presence of G-tetrads, are an impediment to uracil repair in vitro. Both UDG and hUNG2 have been well characterized for their activity on uracils residing in either single or double stranded DNA substrates. We found that UDG and hUNG2 efficiently cleaved uracil from the unstructured single-stranded TCF3 DNA and from the repeat disrupted (GT) oligonucleotides (Figures 22 and 23), showing equal activity for the repeat containing and control DNAs. This was in contrast to the lower activity observed on the G4 oligonucleotides U7 and U24 under G4 folding conditions (Figure 22A, C and 23A, C). The addition of salt to the reactions did not reduce excision activity on the G4(U15) oligonucleotide (Figure 22B, 23B), arguing that it is the position of the deoxyuracil relative to G4 and not the ionic strength of the reaction that affected cleavage efficiency. Furthermore, reactions that replaced KCl with a salt that does not support TCF3 G4, LiCl (Williams et al., 2015), returned full cleavage activity on the G4(U7) and G4(U24) (Figure 25 and 26). We suggest that the G4 structure formation within G4(U7) and G4(U24) place the deoxyuracil in a molecular geometry that is less accessible to either UDG or hUNG2 activity. Indeed, moving the deoxyuracil
three nucleotides from a tetrad G4(U15) results in nearly full glycosylase activity compared to single-stranded conformations (Figure 22B and 23B). While these experiments do not resolve whether G-tetrads are an impediment to substrate recognition or cleavage by UDG and hUNG2, both enzymes do appear to share the property of having relatively weak activity next to G4 DNA structures.

Our study is limited to the analysis of a single deoxyuracil placed at three different positions relative to guanine repeats in a synthetic oligonucleotide that can participate in G4 formation. Based on our results, it seems likely that a deoxyuracil located directly adjacent to a guanine tetrad will be poorly repaired by UNG-initiated BER. Since we have used a single sequence (albeit more than one structural conformation) we cannot exclude the possibility that uracil excision activity may be more efficient at G4s folded from other genomic sequences or other G4 conformations. It is possible that G-tetrads are a general impediment to uracil excision activity by UNG enzymes, but cytidine deaminations within sequence contexts that support alternate G4 conformations, such as anti-parallel or Na\(^+\) stabilized G4s, could feasibly be better substrates. Furthermore, considering the apparent abundance and complexity of unstable G4 loci in the human genome (Cea et al., 2015; Davis and Maizels, 2011; Maizels, 2015), it will be important to further characterize the limits of uracil excision activity on a wider range of G4 structural conformations.

During transcription or replication, when DNA is transiently denatured, cytosine is exposed to hydrolytic deamination at a rate orders of magnitude higher than that of duplex DNA (Barnes and Lindahl, 2004). Presumably,
unpaired cytidines present within G4 DNA structures will also be prone to
deamination, and our results suggest hUNG2 may not efficiently remove these
particular base lesions. Even so, there are three other glycosylases that are
capable of removing uracil from DNA with varying efficiencies (Krokan and
Bjørås, 2013; Krokan et al., 2002; Krokan et al., 2014), and some or all could
feasibly substitute for hUNG2 at G4 DNA. Using telomeric G4 substrates and
glycosylases specific to oxidized DNA damage, it was shown that 8-oxyguanine
was not processed, but other oxidation products were excised by the NEIL
glycosylases (Zhou et al., 2013) suggesting that G4 DNA does not inhibit every
glycosylase. Furthermore, some repair pathways appear to not be affected by G4
DNA, O(6)-alkylguanine in telomeric G4 was shown to be repaired by
alkylguanine alkyltransferase at levels comparable to duplex DNA (Hellman et
al., 2014).

One prediction from our results is that the occurrence of cytidines in the
genome would be uncommon next to guanine repeats that are frequently
participating in G4 formation, such as the telomeres (Sundquist and Klug, 1989;
Williamson et al., 1989). This is because unrepaired cytidine deaminations due to
reduced BER at those sites would cause C to T mutagenesis. There is some
evidence to support that model. In vertebrates, telomere repeats do not contain
any cytosines on the guanine-rich strand. Moreover, a compilation of known G-
rich telomere repeat sequences from animals, plants and fungi species shows
that cytosines are rare and, if they are present in the repeat, they do not reside
directly next to tandem guanines (Podlevsky et al., 2008) and would therefore not
be directly adjacent to a G4 tetrad. It is therefore interesting to speculate that the frequency of G4 formation in the cell may be a reflection of the sequence context and size of the looped regions, with the possibility that in the genome G4-capable sequences containing cytosine loops infrequently participate in stable or persistent G4 formation. Indeed, it was recently shown in yeast that both minisatellite instability and the thermal stability of G4 structures increases for sequence variants that contain shortened pyrimidine loops, which happen to be rare G4 sequences in the yeast genome (Piazza et al., 2015). If selection acts to favor genetic stability at a given G4 locus, our results provide a possible molecular mechanism for that evolutionary force and an explanation for the paucity of short pyrimidine loops between G-tetrads. Cytidine damage may be poorly repaired in the context of G4 DNA, and if damage occurs within stable G4 structures, this could result in unresolved DNA repair intermediates, or mutagenesis, at those sites.

In conclusion, we have used a model G4 forming sequence from the TCF3 gene to ask if G4 DNA structures have the potential to interfere with uracil excision activities. Both bacterial UDG and hUNG2 shared a reluctance to cleave uracils next to G-tetrads, but showed full activity on control oligonucleotides that cannot adopt DNA structure, and full activity on guanine repeat oligonucleotides in conditions that do not support G4 formation. The addition of APE1 to the reaction did not improve excision activity, supporting the notion that the base excision repair pathway is not fully activated at uracil bases in G4 DNA. Guanine-rich loci are known to be genetically unstable, and our results suggest that
defects in base excision repair of uracil at G4 DNA can contribute to mutagenesis.

MATERIALS AND METHODS

Substrate Preparation

Deoxyuracil-containing G4 sequences were based on previously described G4 sequence from the TCF3 gene (Williams et al., 2015). All oligonucleotides were synthesized by Eurofins MWG Operon (Huntsville, AL) and PAGE purified. Oligonucleotides were 5’ end labeled with γ-P$^{32}$ ATP from MP Biomedicals (Solon, OH) and with T4 PNK, New England Biolabs (NEB) (Ipswich, MA). Unincorporated label was removed by spin chromatography using an Illusttra Microspin G-50 column from GE Healthcare (Pittsburg, PA) using the manufacturer's instructions.

PAGE Analysis

G4 and GT oligonucleotides were both subjected to the same G4 folding conditions. Reactions (10 µl) contained 5 pmols of labeled oligonucleotide and 100 mM KCL in TE. The DNA was first denatured in a small > 98°C water bath, and which was allowed to slowly cool to room temperature to permit structure folding. Samples were resolved by either native 16% PAGE (29:1) for 600Vhrs in 0.5X TBE, or denaturing PAGE containing 7 M Urea 550 V for 20-35 minutes.

Circular Dichroism

CD detection of TCF3 G4 was performed as previously described (Williams et al., 2015). Briefly, we used an Aviv model 215 CD spectrometer at 37°C and a 1 cm cuvette. Oligonucleotides were prepared at 12.5 µM in 10 mM
Tris-HCl (pH 7.6), 1 mM EDTA, and 100 mM KCl. Molar ellipticity was taken from 200 nm to 300 nm in 1 nm increments. The averages of three different scans for each oligonucleotide analyzed were used.

**Uracil Glycosylase and APE1 Cleavage Assays**

Purified E. coli UDG and hAPE1 were purchased from NEB (Ipswich, MA) and hUNG2 from Enzymax (Lexington, KY). G4 DNA oligonucleotide structures were formed from 5 pmol of labeled DNA in 10 μl of 100 mM KCl and 1X UDG reaction buffer (20 mM Tris-HCl, 1 mM DTT, 1 mM EDTA, pH 8.0). Structures were folded in a ABI 2720 Thermal-cycler (Foster City, CA) set at the following temperatures held a 5 minute intervals: 98°C, 90°C, 75°C, 60°C, 45°C, 30°C, 20°C. Samples were then immediately placed on ice. Cleavage assays contained indicated enzymes and incubated at 37°C for 20 minutes (glycosylase reactions) and 60 minutes for reactions containing both hUNG2 and APE1. Reactions were halted by addition of an equal volume addition of 1.8% SDS and 100 μg/ml Proteinase K Fermentas (Waltham, MA) solution. The abasic sites of the UDG and hUNG2 reactions were cleaved with 333 mM NaOH and by heating at 55°C or 10 minutes. All enzymatic reactions were resolved using a 16% polyacrylamide (19:1) with 7 M urea and 0.5X TBE and with formamide load dye. 100 mM KCl was included in both the gel and buffer to stabilize G4 structures. Gels were resolved at 550 V for 20-35 minutes, depending on the oligonucleotide. Images were collected using a Typhoon FLA 7000 phosphorimager and bands quantified using ImageQuant TL software (GE Healthcare). Pixel density ratios were used to score enzyme activity and graphed.
as the percentage of the total cleaved. Enzymatic reactions were repeated at least three times (n=3) for each oligonucleotide and displayed with standard deviation. One representative phosphorimage was selected for each figure.
REFERENCES


Williams, J. D., S. Fleetwood, A. Berroyer, N. Kim, and E. D. Larson, 2015, Sites of instability in the human TCF3 (E2A) gene adopt G-quadruplex DNA structures in vitro: Front Genet, v. 6, p. 177.


**FIGURES**

**Figure 20. Sequences Used and Diagrams of G4 Structures.** (A) The name (left) and sequence (right) for each oligonucleotide are shown. G4 and GT indicate oligonucleotides capable or incapable of adopting G4, respectively. The nucleotide position of the deoxyuracil base, U7, U15, and U24 relative to the 5’ end is indicated within each oligonucleotide name. The positions of deoxyuracil bases within each sequence are underlined and guanine repeats are bolded. Substitution of thymine for guanine interrupts G4 folding potential, shown with an italicized “T”. (B) Diagram depicting intramolecular, left, or intermolecular, right, G4 conformations and the relative positions of each deoxyuracil (asterisks). All three uracils positions are depicted on a single G4 conformation for convenience. Diagrams are not scaled models. Actual G4 conformations in solution are predicted to be a mixture of intra and inter-molecular species, but U7 and U15 will be adjacent to a tetrad.

<table>
<thead>
<tr>
<th>Oligo Name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>G4 (U7)</td>
<td>AGGGAGUGGGACGTGAATGGGGTCGGAGGGGCGGGTG</td>
</tr>
<tr>
<td>GT (U7)</td>
<td>AGTGAGUGTTGACGTGAATGTTGTGCGAGTTTCGTTGTG</td>
</tr>
<tr>
<td>G4 (U15)</td>
<td>AGGGAGTGGGAGGUGAATTGGGGTCGGAGGGGCGGGTG</td>
</tr>
<tr>
<td>GT (U15)</td>
<td>AGTGAGTGTTGACGUGAATGTTGTGCGAGTTTCGTTGTG</td>
</tr>
<tr>
<td>G4 (U24)</td>
<td>AGGGAGTGGGAGCCTGAATGGGGUGCGAGGGGCGGGTG</td>
</tr>
<tr>
<td>GT (U24)</td>
<td>AGTGAGTGTTGACGTGAATGTTGGCGAGTTTCGTTGTG</td>
</tr>
</tbody>
</table>

Figure 21. Detection of G4 Formation. (A) Phosphorimages showing native PAGE resolution of 5' 32P end labeled G4 and GT oligonucleotides in neutral pH and 100 mM KCl. The name of each oligonucleotide is shown at the top. Oligonucleotides where guanine repeats are interrupted by thymine to prohibit G4 formation (GT) migrate predominately as a single species. Guanine repeat oligonucleotides (G4) migrate faster (arrow) and slower (bracket) than each control (GT). All oligonucleotides are equal nucleotide length and migration of G4(U24) duplex DNA (DS) is shown as a reference for double-stranded DNA. (B) CD of each G4 and GT deoxyuracil-containing oligonucleotide. Ellipticity is shown on the Y axis, wavelength (nm) on the X axis. Spectra for uracil containing oligonucleotides with guanine repeats (G4) are in black, repeat interrupted (GT) oligonucleotides in grey. All oligonucleotides were assayed in 100 mM KCl solution at pH 7.6.
Figure 22. G-Tetrads in TCF3 G4 Interfere with UDG Activity. (Left) Representative phosphorimages of E. coli UDG cleavage assays for each 5' radiolabeled oligonucleotide are shown. Reactions were performed at the indicated salt conditions (0, 50 and 100 mM KCl), followed by alkaline lysis and resolution of the cleavage products by denaturing PAGE. Oligonucleotides contained a single deoxyuracil at position 7 (A), 15 (B), or 24 (C) in sequences capable of forming G4 (G4) or controls where the guanine repeats were interrupted with thymine to prohibit G4 formation (GT). The presence (+) or absence (-) of UDG is indicated at the top. An arrow shows the position of the cleavage products. (Right, A-C) Quantitation of the percentage cleaved by UDG (Y axis) for each oligonucleotide at the indicated salt concentration (X axis). Values are means from three independent experiments, with standard deviation.
Figure 23. G-Tetrads in TCF3 G4 Interfere with hUNG2 Activity. (Left) Representative phosphorimages of hUNG2 cleavage assays for each 5' radiolabeled oligonucleotide are shown. Reactions were performed at the indicated salt conditions (0, 50 and 100 mM KCl), followed by alkaline lysis and resolution of the cleavage products by denaturing PAGE. Oligonucleotides contained a single deoxyuracil at position 7 (A), 15 (B), or 24 (C) in sequences capable of forming G4 (G4) or controls where the guanine repeats were interrupted with thymine to prohibit G4 formation (GT). The presence (+) or absence (-) of hUNG2 is indicated at the top. An arrow shows the position of the cleavage products. (Right, A-C) Quantitation of the percentage cleaved by hUNG2 (Y axis) for each oligonucleotide at the indicated salt concentration (X axis). Values are means from three independent experiments, with standard deviation.
Figure 24. G-Tetrads in TCF3 G4 Interfere with the Cleavage Step of Base Excision Repair. (Left) Representative phosphorimages showing the activity of both hUNG2 and APE1 on single deoxyuracil bases placed at positions 7 (A), 15 (B), or 24 (C) in G4 DNA folded from the TCF3 oligonucleotide (G4) or control oligonucleotides that cannot fold into G4 structures (GT). Reactions contained 100 mM KCl and the presence (+) or absence (-) of hUNG2 or APE1. Cleavage products (arrow) were resolved by denaturing PAGE. (Right) Quantitation of at least three independent experiments are displayed on the right of each representative phosphorimage. The percentage cleaved by APE1 activity (Y axis) is shown for each G4 (X axis left) or GT (X axis right), with standard deviation and $P$ values (two-tailed t test).
Figure 25. LiCl Does Not Inhibit UDG Activity

(Left) Phosphorimages of UDG activity assays resolved by denaturing PAGE. Assays were identical to those presented in Figure 3, except that LiCl was substituted for KCl. (Right) Quantitation of UDG cleavage in the presence of LiCl at the indicated concentrations. Pixel means were derived from at least three independent experiments, with standard deviation.

A. U7

B. U15

C. U24
Figure 26. LiCl Does Not Inhibit hUNG2 Activity (Left) Phosphorimages of hUNG2 activity assays resolved by denaturing PAGE. Assays were identical to those presented in Figure 4, except that LiCl was substituted for KCl. (Right) Quantitation of hUNG2 cleavage in the presence of LiCl at the indicated concentrations. Pixel means were derived from at least three independent experiments, with standard deviation.
CHAPTER V

FURTHER CHARACTERIZATION OF URACIL GLYCOSYLASE ACTIVITY ON G4 DNA SUBSTRATES
INTRODUCTION

As demonstrated in chapter II, the monoubiquitination of histones is one of many potential key molecular beacons that may recruit the DNA modifying enzyme activation induced cytosine deaminase (AID). However, the full complement of molecular interactions responsible for AID targeting to the immunoglobulin (Ig) locus remains a critical question that must be resolved. The specificity and efficiency of AID targeting to the Ig locus is most likely regulated at multiple levels using independent cellular mechanisms and it will probably require collaborative efforts to determine how all these regulatory elements interact and direct the targeting of AID (Delker et al., 2009; Stavnezer, 2011). The Ig locus contains a “perfect storm” of molecular and sequence conditions that ensure targeted genomic rearrangements to produce novel antibodies. Although many other loci share some of the same molecular beacons, AID induced lesions at non Ig loci remain below the mutagenic threshold and DNA repair is able to overcome the mutagenic threat and maintain genomic integrity.

Recently, CHIP-seq analysis of B cells demonstrated that AID targets over 5,000 transcribed genes and some of these non-Ig loci have increased levels of mutagenesis (Yamane et al., 2011). This clearly highlights the potency of AID to act as a global genomic mutator and argues for the existence of powerful regulators to tightly control AID expression and deamination activity. Chapter III of this manuscript characterizes the 3’ untranslated region of AID mRNA and shows that miR-93 and miR-155 are able to post transcriptionally regulate the levels of functional AID within RAMOS B cells. The model suggests, miR-93 and
miR155 act as genomic sentries that protect DNA from the mutational processes of AID.

One characteristic that is shared between the Ig locus and off target genes deaminated by AID is that they are highly transcribed (Chaudhuri et al., 2003), which also provides the conditions necessary to form G4 DNA structures at these loci. AID preferentially deaminates single strand DNA and is capable of acting on both the template and non-transcribed strands (Bransteitter et al., 2003; Dickerson et al., 2003; Pham et al., 2003; Ramiro et al., 2003; Shen and Storb, 2004; Sohail et al., 2003). As stated above, off target mutations created by AID are corrected by DNA repair mechanisms. Identifying how these pathways can be overwhelmed is a critical component in understanding genomic instability and genetic disease. In Chapter IV evidence is provided showing the base excision repair (BER) pathway is poorly initiated by uracil DNA glycosylase (UDG) enzymes when the lesion is directly adjacent to guanine repeats that are involved in G4 DNA structure formation. These experiments were designed simply to determine if G4 DNA affects UDG activity on a pool of substrates containing both intra- and inter-molecular G4 DNA conformations. The rest of this manuscript outlines experiments that are underway to further elucidate how G4 DNA structures alter the efficiency of BER.

In order to further characterize UDG activity on G4 DNA another oligonucleotide substrate was designed. The sequence allows for the investigation of glycosylase activity on only two species of multi-molecular G4 DNA substrates (G4 and G2’) (Figure 27B and C), which have been well
characterized (Sen and Gilbert, 1992). The advantages to using this substrate
design are twofold. First, it controls for the number of potential G4 structures due
to the fact that only multi-molecular G4 motifs can fold. Second, the structures
are stable in lower concentrations of KCl (10 mM), which allows for the
investigation of UDG as well as single-strand selective monofunctional uracil
DNA glycosylase (SMUG1) an enzyme that is salt sensitive. The following
experiments will demonstrate that UDG inhibition extends to SMUG1 when the
uracil is located within guanine repeats under G4 stabilizing conditions.

RESULTS

Uracil Does Not Disrupt Inter-Molecular G4 DNA Formation

Native PAGE was performed comparing DNA migration of samples under G4
folding conditions to those that promote a single stranded conformation (Figure
27A). When multi-molecular G4 DNA is present it involves the interaction of two
or four oligonucleotides, making it a much larger structure forcing it to migrate
more slowly through the gel relative to single-strand DNA (Figure 27A). For each
sequence analyzed there was a noticeable shift up when the DNA was subjected
to G4 folding conditions (100 mM KCl) (Figure 27A). The folded samples show a
mix of multi-molecular G4 DNA structures with G2’ migrating faster than G4 DNA
but still much slower than single strand substrates (Figure 27A). The
oligonucleotide containing the scrambled (520scr) sequence disrupting the
guanine repeats shows no such mobility shift during native PAGE when KCl is
present (Figure 27A). The highly repetitive nature of the DNA oligonucleotide
sequence allows for hairpin formation and accounts for the doublet that is seen
when the samples are under single strand conditions (Figure 27A). The DNA sequences are outlined in Table V-1 and show a single uracil substitution within each oligonucleotide. The structures depicted in Figure 27A and 27B have an asterisk at the nucleotide position representing the location of the single uracil substitution.

**UDG is Inhibited When the Lesion is Located Within G4 DNA Structures**

The uracil glycosylase activity assays were carried out in non-denaturing reaction conditions before being subjected to DNA alkaline lyses and denaturing PAGE. The oligonucleotides were used to assay the ability of UDG to process uracil in conditions that promote G4 structure formation. The ratio of substrate to product created by UDG is used to quantify the ability of UDG to act of G4 DNA. When the guanine repeats are disrupted and the oligo is unable to form a G quadruplex structures there is no difference in UDG processing between linear and G4 DNA conditions (Figure 28A, 520scr). However when the sequence and reaction conditions allow for G4 to form, a significant reduction in uracil processing is evident by the presence of an increased amount of intact substrate (Figure 28A, EDL520, EDL521, EDL523, EDL524). Interestingly, when the uracil lesion is found outside of guanine repeats that participate in G4 folding there is no difference in UDG processing between linear and G4 folding conditions (Figure 28A, EDL522). The trend of UDG inhibition in a G4 DNA context can clearly be seen when average uracil processing of several gels (n=3) is graphed (Figure 28). Under linear conditions there is no change in UDG processing between the different oligonucleotides examined. When the reaction conditions promote G4
DNA there is a significant reduction in uracil processing of between approximately 35% and 55% when the lesion is located within guanine repeats that participate in the G4 DNA structure formation.

**SMUG1 is Inhibited When the Lesion is Located Within G4 DNA Structures**

The same uracil glycosylase activity assay was repeated using hSMUG1 instead of UDG and similar results were observed. The scrambled sequence and the oligonucleotidess with the uracil lesions located outside of the guanine repeats involved in the G4 structure again showed no difference in SMUG1 processing between the linear (SS) and G4 folded DNA structures (Figure 29A, 520scr, 522). EDL522 reaches a maximum processing of just over 50% when it is linear or under G4 folding conditions. This may be due to SMUG1 having different specificity for single-strand and duplex DNA and the presence of hairpins in the oligo sequence. Again, the DNA sequences with a uracil within guanine repeats displayed SMUG1 processing inhibition within G4 structures compared to that of the single strand DNA samples (Figure 28A., 520, 521, 523, 524). A graph of SMUG1 processing was constructed using the average pixel density ratio of multiple gels (n=3) and the averages display the inability of SMUG1 to process uracil when found within guanine repeats participating in G4 structure formation.

**DISCUSSION**

Chapter IV demonstrated that both E. coli UDG and human UNG2 have a reduced processing ability near G4 DNA that is capable of forming both intramolecular and intermolecular motifs. The sequence used here in Chapter V represents a generic G4 locus and was chosen because it forms distinct, well
characterized, intermolecular G4 DNA structures (Figure 27A) (Sen and Gilbert, 1992). Experimentation on these oligonucleotides is crucial in characterizing the limits of uracil glycosylase sensitivity on intermolecular G4 substrates. When the oligonucleotides are subjected to UNG and SMUG1 processing it is clear that fewer uracil containing lesions are being processed when the DNA oligonucleotide is under G4 stabilizing conditions and the uracil is within guanine repeats (Figure 29 and Figure 29). Experiments using these oligonucleotides are still currently being performed in order to attain publication quality gels and replicates for statistical analysis. The glycosylase sensitivity demonstrated in Figures 28 and V-3 follows with recent data (not shown), in which inhibition of the glycosylase is only present when the lesion is found within the G4 DNA structure between two sets of tetrads (EDL520, 521,523,524) but the other uracil positions are accessible by the enzyme. This again shows that G4 DNA is an impediment to the efficient initiation of the BER pathway and provides a molecular mechanism that explains the association with G4 DNA and instability.

CONCLUSION

The work compiled in this manuscript elucidate several molecular mechanisms that are involved in maintaining genomic stability and provide significant contributions to the gaps in our current understanding of both targeted and spontaneous mutagenesis. The effect of structures on genomic stability is a recently discovered phenomenon and will remain relevant for some time due to the unique characteristics of individual DNA structures which provide a substantial substrate pool to be examined.
MATERIALS AND METHODS

Substrates

As a control to ensure that the G4 folding capability is sequence specific, a scrambled sequence was generated disrupting the guanine repeats and was unable to fold into G4 DNA (520scr). The deoxyuracil was placed at positions within guanine repeats participating in G4 structure folding and outside of the guanine repeats. All deoxyoligonucleotides were purchased from Eurofins MWG Operon (Huntsville, AL) and the uracil containing strands were PAGE purified.

The G4 folding conditions were, 500µM of DNA was incubated with TE (pH 7.9) in a 5µl volume at 98°C for 2 min and placed immediately on ice. A 1:1 addition of salt mix (1.9M NaCl, 0.1M KCl) was added and incubated at 60°C for 90 min followed by 7 hours at 37°C. Reactions were performed in thin walled PCR tubes (Fischer). DNA was stored at -20°C and appropriate dilutions were 32P radio labeled and visualized by PAGE. The 32P radio label was ordered from MP Biomedicals (Santa Ana, Ca) and 5’ DNA end labeling was performed using NEB Polynucleotide Kinase following the supplier protocol to a final concentration of 5 µM. Unincorporated radio label was removed by spinning the sample through a G-50 micro columns (GE Healthcare) either equilibrated in TE or in TE with 10mM KCl (G4).

Enzymes

The purified enzymes were purchased commercially from NEB (UDG), enzymax (hUNG2) and Fermentas (SMUG1)
Polyacrylamide Gel Electrophoresis

To visualize the multi-molecular parallel DNA structures Native PAGE gels were prepared at 6% acrylamide (29:1) and run in 0.5x TBE with 50mM KCl. The samples were loaded in 50% glycerol with 10mM EDTA and bromophenol blue and run at 100 volts for 2 hours. The gel was exposed to the PhosphorImager screen and read using the Storm840 scanner (GE Healthcare).

The enzymes and appropriate DNA substrates were incubated at 37°C for 20 minutes under native conditions in buffers outlined by the enzyme manufacturer. Next, alkaline lyses of the abasic sites was performed by bringing the sample to 300 mM NaOH and incubated at 55°C for 10 minutes. An equal volume of 95% formamide with bromophenol blue was added to each sample and heated to 96°C for two minutes. The DNA was then loaded onto 15% denaturing urea gels and run at either 300 V or 700 V to resolve DNA fragments. Quantification of the image and a pixel density ratios of substrate to product formation were used to score the activity of each glycosylase on single-strand DNA and G4 DNA.
REFERENCES


FIGURES

Figure 27. Native DNA Structures. A) Native PAGE with oligonucleotides under single stranded (SS) and G4 conditions. Oligonucleotide structural conformation listed above, construct name listed below and DNA structure motif indicated on the right of the gel. B) G4 DNA oligo sequence and parallel G4 structure with asterisks representing nucleotide positions with a single uracil substitutions. C) Alternative G4 DNA structure (G2') consisting of two oligonucleotides in an anti-parallel conformation with asterisks indicating a single uracil substitutions.
Figure 28. Bacterial Uracil DNA Glycosylase Activity on Single-Strand and G4 Folded Substrates. A) Representative denaturing PAGE of individual oligonucleotides consisting of linear (SS) DNA and G4 folded (G4) structures incubated with or without uracil DNA glycosylase. B) Average UNG processing (n=3) for each individual oligonucleotide under single strand (black) and G4 conditions (white).
Figure 29. SMUG1 Activity on Single-Strand and G4 Folded Substrates. A) Representative denaturing PAGE of individual oligonucleotides consisting of linear (SS) and G4 folded (G4) structures incubated with or without SMUG1. B) Average SMUG1 processing (n=3) for each individual oligo under linear (black) and G4 conditions (white).
**Table 3. DNA Oligonucleotide Sequence.** The left column indicates the oligo name. The right column is the uracil containing DNA oligo sequence with uracil underlined and written 5’ to 3’.

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>EDLSCR520</td>
<td>ATAGAGGCGACGGCGGTGAGAATAGAGTGUGCAGGCGAGTGATAGTGCAGG</td>
</tr>
<tr>
<td>EDL520</td>
<td>TGGACCAGACCTAGCAGCTATGGGGAGCCUGCGAAGGTGGGATGGAATGTGTA</td>
</tr>
<tr>
<td>EDL521</td>
<td>TGGACCAGACCTAGCAGCTATGGGGAGUUUGGGAAGGTGGGATGGAATGTGTA</td>
</tr>
<tr>
<td>EDL522</td>
<td>TGGACCAGACCTAGCAGCUATGGGGAGCTGGGGAAGGTGGGATGGAATGTGTA</td>
</tr>
<tr>
<td>EDL523</td>
<td>TGGACCAGACCTAGCAGCTATGGGGAGCUUGGGGAAGGTGGGATGGAATGTGTA</td>
</tr>
<tr>
<td>EDL524</td>
<td>TGGACCAGACCTAGCAGCTATGGGGAGCUUGGGGAAGGTGGGATGGAATGTGTA</td>
</tr>
<tr>
<td>EDL522X</td>
<td>TGGACCAGACCTAGCAGCTATGGGGAGCUUGGGGAAGGTGGGATGGAATGTGTA</td>
</tr>
<tr>
<td>EDL522XX</td>
<td>TGGACCAGACCTAGCAGCTUUGGGGAAGGTGGGATGGAATGTGTA</td>
</tr>
<tr>
<td>EDL525X</td>
<td>TGGACCAGACCTAGCAGCTATGGGGAGCGTGGGGAAGGTGGGATGGAATGTGTA</td>
</tr>
<tr>
<td>EDL525XX</td>
<td>TGGACCAGACCTAGCAGCTATGGGGAGCGTGGGGAAGGTGGGATGGAATGTGTA</td>
</tr>
<tr>
<td>EDL525XXX</td>
<td>TGGACCAGACCTAGCAGCTATGGGGAGCGTGGGGAAGGTGGGATGGAATGTGTA</td>
</tr>
</tbody>
</table>