Epigenetic Regulation of \textit{Aicda} Transcription in B Cells

by

Srikanth Kodali

Department of Biomedical Engineering
Duke University

Date:_______________________

Approved:

___________________________
William Reichert, Supervisor

___________________________
Shyam Unniraman

___________________________
Charles Gersbach

Thesis submitted in partial fulfillment of
the requirements for the degree of Master of Science in the Department of
Biomedical Engineering in the Graduate School
of Duke University

2012
ABSTRACT

Epigenetic Regulation of Aicda Transcription in B Cells

by

Srikanth Kodali

Department of Biomedical Engineering
Duke University

Date:_______________________
Approved:

___________________________
William Reichert, Supervisor

___________________________
Shyam Unniraman

___________________________
Charles Gersbach

An abstract of a thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in the Department of Biomedical Engineering in the Graduate School of Duke University

2012
Abstract

Activation-induced cytidine deaminase (AID), encoded by the *Aicda* gene, is indispensable for secondary antibody diversification through somatic hypermutation (SHM) and class switch recombination (CSR). It is expressed predominantly in germinal center (GC) B cells, where it deaminates cytosine to uracil in the DNA of immunoglobulin (Ig) genes, triggering mutagenesis (SHM) or deletional recombination events (CSR). However, when misregulated, AID can also mutate non-Ig genes, including proto-oncogenes, thereby contributing to genomic instability and cancer. Due to its potentially deleterious effects, AID expression must be tightly controlled. At the transcriptional level, *Aicda* is regulated by four highly conserved *cis*-regulatory regions (Regions 1-4). Region 1 contains the promoter, and is responsible for basal transcription of *Aicda*. Region 2 contains both enhancer and silencer elements, but functions as a negative regulatory region, restricting AID expression to antigen-activated B cells. Region 3 is reported to be essential for normal AID expression, but its role is otherwise unclear. Region 4, which contains two STAT6 sites essential for its function, provides enhancer activity in response to cytokine stimulation and is responsible for the high levels of AID expression found in GC B cells. Epigenetic mechanisms could add another layer of control to the regulation of *Aicda*; however, little research has been done in this regard. In this study, I investigated the role of DNA methylation in the *cis*-regulation of *Aicda* transcription. Through bisulfite sequencing of mouse splenic B cell DNA, I
demonstrate that Region 4 is highly methylated in naïve B cells, but becomes demethylated in activated B cells. This is in contrast to Region 2, which I found to be constitutively unmethylated, and Region 1, which only becomes partially demethylated. Using quantitative methylation-specific PCR (qMSP), I show that loss of methylation in Region 4 correlates positively with cell division number, consistent with a passive mechanism of demethylation. I also show that demethylation in Region 4 cannot be achieved by cell proliferation alone, and that it depends upon induction of AID expression (via IL-4 stimulation). However, bisulfite sequencing of CH12F3-2 cell DNA shows that demethylation by itself is insufficient to activate AID expression. Taken together, the data suggests that IL-4 induces STAT6 to bind to its cognate sites in Region 4 and inhibit DNMT1, likely through chromatin reorganization, triggering passive demethylation during the induction of Aicda transcription.
## Contents

Abstract ........................................................................................................................................ iv
List of Tables ................................................................................................................................... viii
List of Figures .............................................................................................................................. ix
Acknowledgements .................................................................................................................... x

1. Introduction ................................................................................................................................ 1
   1.1 The Humoral Immune Response ................................................................................................. 1
   1.2 Mechanisms of Antibody Diversification .................................................................................... 2
      1.2.1 V(D)J Recombination ........................................................................................................... 3
      1.2.2 Somatic Hypermutation ...................................................................................................... 4
      1.2.3 Class Switch Recombination .............................................................................................. 5
      1.2.4 Activation-induced Cytidine Deaminase .............................................................................. 7
   1.3 Cis-regulation of Aicda Expression ............................................................................................ 9
      1.3.1 Region 1 ............................................................................................................................. 9
      1.3.2 Region 2 ........................................................................................................................... 10
      1.3.3 Region 3 ........................................................................................................................... 11
      1.3.4 Region 4 ........................................................................................................................... 11
   1.4 DNA Methylation ..................................................................................................................... 15
      1.4.1 DNA Methyltransferases .................................................................................................... 15
      1.4.2 DNA Demethylation .......................................................................................................... 16
      1.4.3 Bisulfite-based DNA Methylation Analysis ......................................................................... 17
      1.4.4 Non-CpG Island Methylation and AID Expression .............................................................. 19
2. Materials and Methods.................................................................21
  2.1 Mice and Cell Lines.................................................................21
    2.1.1 Primary B Cells...............................................................21
    2.1.2 CH12F3-2 Cells...............................................................21
  2.2 Flow Cytometry.................................................................22
  2.3 DNA Preparation............................................................22
  2.4 Bisulfite Genomic Sequencing..............................................22
    2.4.1 Primers and PCR Conditions........................................22
    2.4.2 Cloning and Sequencing...............................................25
  2.5 Quantitative Methylation-specific PCR...................................26
    2.5.1 Methylation-specific PCR..............................................26
3. Results..................................................................................28
  3.1 Regions 1 and 4 Become Demethylated in Activated B cells..........28
  3.1 Demethylation in Region 4 Correlates with Cell Division..............30
4. Discussion............................................................................35
References..............................................................................41
List of Tables

Table 2.1: PCR primers used for bisulfite sequencing ................................................................. 24
Table 2.2: Bisulfite PCR setup ..................................................................................................... 25
Table 2.3: Thermal cycling conditions ......................................................................................... 25
Table 2.4: Primers used for qMSP ............................................................................................... 27
Table 2.5: qMSP setup .................................................................................................................. 27
List of Figures

Figure 1.1 Structure of a typical antibody molecule..........................................................3
Figure 1.2 Mechanisms of primary and secondary antibody diversification....................7
Figure 1.3 Organization of the mouse Aicda locus............................................................14
Figure 2.1 Aicda region 1, 2, and 4 fragments covered by bisulfite sequencing..............24
Figure 3.1 Methylation status of Aicda...........................................................................29
Figure 3.2 Division-linked demethylation of Region 4......................................................32
Figure 3.3 Average methylation at each CpG site probed in Aicda regions 1, 2, and 4 in unstimulated CH12F3-2 cells......................................................................................34
Figure 4.1 IL-4-induced AID expression via STAT6-mediated demethylation............39
Acknowledgements

My great gratitude and respect go to my advisor, Shyam Unniraman, for taking me on as a graduate student and providing outstanding advice and support throughout my time in the lab. I would like to express my appreciation and thanks to my committee members, William Reichert and Charles Gersbach, for serving on my committee and providing constructive criticism and feedback. I would like to thank the members of the Unniraman lab for all their help and advice: Rahul Arya for mentoring me when I first joined the lab, Anurodh Agrawal for helping me set up the qMSP assay, and Zach Carico for isolating, culturing, and FACS-sorting all of the mouse splenic B cells used in this study. Finally, I would like to express my deepest thanks to my family: my father Rao, my mother Veni, and my sister Santhi, for their love and encouragement throughout the years.
1. Introduction

1.1 The Humoral Immune Response

The hallmark characteristics of vertebrate adaptive immunity are the ability to discriminate between self and non-self antigens, the ability to mount an exquisitely antigen-specific immune response, and the ability to retain immunological memory of an antigenic encounter (Bartl et al., 2003; Willermain et al., 2012). In the jawed vertebrates (gnathostomes), adaptive immunity is the responsibility of the subset of leukocytes known as lymphocytes. There are two types of lymphocytes, T cells and B cells, both of which recognize and bind antigens through surface antigen recognition receptors (Bartl et al., 2003). B cells are the cellular mediators of the humoral arm of adaptive immunity. The humoral response to antigenic challenge relies upon the recruitment of antigen-specific B cell clones, each defined by the expression of a distinct antigen receptor (BCR), from a pool of mature B cells (Bartl et al., 2003).

Developing B cells in the bone marrow assemble a receptor gene from a diverse pool of gene segments in order to express functional BCRs, a process known as V(D)J recombination (Xu et al., 2010). These immature B cells migrate to the periphery and complete maturation. Upon encountering cognate antigen and receiving CD40L costimulation from T helper (T<sub>H</sub>) cells, naïve B cells become activated and migrate to secondary lymphoid tissues, where they undergo rapid proliferation and differentiation in regions called germinal centers (GCs) (Xu et al., 2010). In the GC, B cells modify their receptor genes through somatic hypermutation (SHM) and class switch recombination.
(CSR) to generate highly-specific and effective BCRs (antibodies). The DNA editing enzyme activation-induced cytidine deaminase (AID), expressed predominantly in GC B cells, plays an essential role in CSR and SHM (LeBien and Tedder, 2008; Muramatsu et al., 1999; Xu et al., 2010). B cells that have successfully generated highly specific BCRs downregulate AID expression and undergo terminal differentiation into either antibody-secreting plasma cells or memory B cells (Xu et al., 2010).

### 1.2 Mechanisms of Antibody Diversification

Membrane-bound immunoglobulin (Ig) constitutes the antigen-binding subunit of the BCR (LeBien and Tedder, 2008). A typical immunoglobulin molecule, or antibody, consists of four polypeptide chains: two identical, high molecular weight heavy (IgH) chains, and two identical, low molecular weight light (IgL) chains (Barstad et al., 1974; Masat et al., 1994). Each heavy chain and each light chain has a variable region followed by a constant region (Barstad et al., 1974). The constant regions of the heavy chains form the Ig constant (C) region, which determines the class (isotype) of the antibody. The variable regions of each heavy and light chain pair form the Ig variable (V) regions, which serve as the antigen binding pockets (Masat et al., 1994). (See Figure 1.1 for an example of antibody structure.)
Figure 1.1 Structure of a typical antibody molecule. \( V_L \) and \( C_L \) correspond to the variable and constant regions of the antibody light chain, respectively. \( V_H \) and \( C_H \) indicate the variable and constant regions of the antibody heavy chain, respectively. The variable regions of each light and heavy chain pair form the antibody variable (V) region, which establishes the antigen recognition site. The constant region of the heavy chains form the antibody constant (C) region, which determines the isotype, and therefore the effector functions, of the antibody.

The vast spectrum of potential antigens dictates that there should be a similarly diverse repertoire of antibodies in order to respond promptly to antigenic challenge. Consequently, B cells undergo multiple mechanisms of primary and secondary antibody diversification, including V(D)J recombination, SHM, and CSR. Following is an overview of these processes (see Figure 1.2).

1.2.1 V(D)J Recombination

B cells undergo germline rearrangement of their Ig loci during early development in the bone marrow (pro- to pre-B cell stages), a process known as V(D)J
recombination. The IgH variable region is encoded at loci composed of numerous variable (V<sub>H</sub>), diversity (D<sub>H</sub>), and joining (J<sub>H</sub>) gene segments; the IgL variable region is encoded by V<sub>L</sub> and J<sub>L</sub> gene segments. At the IgH loci, V(D)J recombination brings together a V<sub>Hi</sub>, a D<sub>Hi</sub>, and a J<sub>Hi</sub> gene segment at random to form a contiguous VDJ exon, which encodes the IgH variable region (see Figure 1.2). Subsequent rearrangement at the IgL loci results in assembly of a VJ exon, which encodes the IgL variable region. Together, these exons code for the V region of the immunoglobulin. In addition to the combinatorial diversity resulting from the random recombination of gene segments, V(D)J recombination introduces further receptor gene variation through imprecise fusion of the V, D, and J gene segments. This results in random addition or deletion of nucleotides at the segment junctions, a phenomenon known as junctional diversity.

**1.2.2 Somatic Hypermutation**

SHM results in the rapid accumulation of point mutations in the V region of GC B cells (Rajewsky et al., 1987). The mutation rate due to SHM is a million times greater than the background mutation rate (Xu et al., 2007). Purely by chance, some of these mutations improve the affinity of the BCR to the initial antigen. Multiple rounds of SHM, positive selection of high-affinity mutants, and clonal expansion of selected cells results in the formation of B cell clones producing high affinity antibodies.

AID triggers SHM by deaminating cytosines to uracil in the Ig V region. Transcription of the V region prior to SHM is crucial, since it allows AID access to short stretches of single-stranded DNA (ssDNA), its preferred substrate, in the V region.
transcription bubbles (Roy and Lieber, 2009). If the uracils are not discovered and removed prior to DNA replication, they are mistaken for thymines by the replication machinery, resulting in transition mutations (G → A) (Casali et al., 2006). On the other hand, if the uracils are excised by uracil DNA glycosylase (UNG), the resulting abasic sites block the replication fork, and translesion synthesis (TLS) polymerases incorporate random nucleotides opposite these abasic sites (G → A, G → T, G → C) (Casali et al., 2006).

1.2.3 Class Switch Recombination

CSR of the IgH locus takes place concomitantly with SHM in GC B cells. The IgH locus contains multiple constant (C\textsubscript{H}) region gene segments, and different segments encode heavy chain constant regions of different isotypes. For each C\textsubscript{H} region (except for C\textsubscript{\delta}), there is a corresponding repetitive GC-rich switch (S) region that lies just upstream. CSR exchanges the initially expressed C\textsubscript{H} region (C\textsubscript{\mu} in naïve B cells) for one of the downstream C\textsubscript{H} regions, thus altering the isotype of the expressed immunoglobulin (see Figure 1.2).

In antigen-activated B cells, Th cell-secreted cytokines (e.g. IL-4, TGF-β) induce germline transcription from intronic promoters located immediately upstream of each S region. The extent to which a particular S region is transcribed is dependent upon the composition of the cytokine milieu. AID converts cytosines to uracil on the S region ssDNA of the transcription bubbles to initiate CSR. The resulting U:G mismatches are repaired by the BER pathway to generate single-stranded breaks (SSBs). SSBs on
opposite strands are converted into double-stranded breaks (DSBs), either spontaneously if close by, or by mismatch repair (MMR) if far apart (Shrader et al., 2007). Formation of DSBs in S_μ region and a downstream S region results in looping out and deletion of the intervening DNA, and end-joining of the two DSBs to complete CSR.

CSR has been shown to correlate with cell division number (Hasbold et al., 1998; Hodkin et al., 1996; Rush et al., 2005). CD40L+IL-4-treated mouse B cells were tracked with CFSE (a fluorescent dye used to trace cells across multiple cell divisions), and shown to undergo switching to IgG1 and IgE after three and five cell divisions, respectively, irrespective of time or CD40L dose (Hasbold et al., 1998). qRT-PCR for AID mRNA in division-sorted, class switch-stimulated mouse B cells has shown that AID transcript levels increase with successive cell divisions, coincident with the frequency of CSR, providing molecular mechanistic insight into the cell cycle dependence of CSR (Rush et al., 2005).
Figure 1.2 Mechanisms of primary and secondary antibody diversification. During the pre- and pro-B cell stages, V(D)J recombination brings together a V, a D, and a J segment from a pool of V, D, and J segments to form a contiguous variable region exon. In germinal center B cells, somatic hypermutation introduces numerous point mutations in the variable region to improve antigen binding specificity. Concomitantly, class switch recombination exchanges the current constant region for a downstream one, looping out and deleting the intervening DNA, altering the effector functions of the antibodies produced. (Adapted from Kinoshita and Honjo, 2008.)

1.2.4 Activation-Induced Cytidine Deaminase

Activation-induced cytidine deaminase (AID), encoded by the Aicda gene, is a 24 kDa, 198 amino acid member of the APOBEC family of RNA/DNA editing enzymes (Muramatsu et al., 1999). It was first identified through subtractive hybridization of
cDNAs between class switch-stimulated and unstimulated CH12F3-2 (a murine B cell lymphoma cell line) cells as an essential factor for CSR (Muramatsu et al., 1999).

AID was initially postulated to be an RNA editing enzyme, based on its sequence homology with APOBEC1, the catalytic subunit of the multiprotein holoenzyme that edits apolipoprotein B (apoB) mRNA. APOBEC1 deaminates a cytidine to uridine at nucleotide position 6666 of the apoB mRNA, creating a premature stop codon that results in the synthesis of a truncated form of apoB (Dance et al., 2000). According to the RNA editing hypothesis, AID modifies a pre-mRNA to generate a new mRNA encoding an endonuclease that nicks the V region to initiate SHM and nicks the S region to initiate CSR. However, there is no direct evidence for the RNA editing hypothesis, because no pre-mRNA substrates for AID have been identified yet.

The DNA editing hypothesis proposes that AID directly targets DNA for mutagenesis. In support of this, ectopically expressed AID has been shown to induce mutations in the genomes of humans, mice, Chinese hamsters, yeast, and *Escherichia coli*; it is improbable that AID would edit the same pre-mRNA across these diverse species to generate a DNA mutator (Martin and Scharff, 2002; Mayorov et al., 2005; Petersen-Mahrt et al., 2002; Yoshikawa et al., 2002). Current evidence suggests that AID preferentially deaminates cytidine residues to uridine on the single-stranded DNA (ssDNA) of transcription bubbles (Larijani and Martin, 2007; Odegard and Schatz, 2006).

AID is expressed primarily in GC B cells, where it targets the Ig genes for mutagenesis and recombination, triggering SHM and CSR, respectively. Although it is
indispensable for secondary antibody diversification, AID is promiscuous, and can mutate other genes, including oncogenes, if misregulated (Komeno et al., 2010; Kotani et al., 2004; Pasqualucci et al., 2004; Robbiani et al., 2009). Indeed, a GC phenotype and dysregulated expression of AID is a feature of many B cell lymphomas (Hardianti et al., 2004; Komeno et al., 2010; Kotani et al., 2004; Pasqualucci et al., 2004). Thus AID expression must be subject to multiple layers of regulation, including at the transcriptional level.

1.3 Cis-regulation of AID Expression

The mechanism of transcriptional regulation of AID has been the subject of several previous studies. Yadav et al. (2005) have identified four cis-regulatory non-coding regions (Regions 1-4) of the mouse Aicda locus that are evolutionarily conserved between humans and mice (see Figure 1.3).

1.3.1 Region 1

Region 1, which is located just upstream of the transcription start site (TSS), functions as a promoter, but is not lymphoid-specific. Region 1 contains two Sp binding sites, which the general transcription factors Sp1 and Sp3 both compete for; Sp1 binding activates RNA polymerase II (pol II)-dependent transcription from the promoter, while Sp3 binding serves to repress it. Region 1 also contains one binding site each for the transcription factors STAT6, NF-κB, and HoxC4, which are all positive regulators of Aicda transcription. Tran et al. (2010) have performed experiments in which they transiently transfected CH12F3-2 cells with luciferase expression plasmids containing
various fragments of Regions 1-4, in order to dissect the cis-regulation of *Aicda* transcription. Through such transient transfection assays, they have found that CIT (anti-CD40L+TGF-β+IL-4) stimulation of Region 1 does not elevate transcription beyond the basal level (four- to eightfold compared to the promoter-less plasmid). They have also identified a 100 base pair (bp) fragment just upstream of the TSS, containing only the Sp and HoxC4-Oct sites, as the minimal promoter required for basal activity.

### 1.3.2 Region 2

Region 2 is an intronic regulatory region, located in the first intron of the *Aicda* locus. It contains binding sites for positive regulators, including Pax5, NF-κB, and E47, as well as negative modulators, including c-Myb and E2f.

Pax5, which is indispensable for maintaining B cell identity, is upregulated in GC B cells, which undergo SHM and CSR, but is repressed during plasma cell differentiation; thus Pax5 expression is coincident with AID expression during the GC reaction (Xu *et al.*, 2007). Moreover, enforced overexpression of Pax5 in Ba/F3 cells, a mouse pro-B cell line, has been shown to induce endogenous AID expression, implying that Pax5 directly regulates *Aicda* transcription (Nagaoka *et al.*, 2010; Xu *et al.*, 2007).

E47, a member of the E-protein family of transcription factors, is upregulated concomitantly with AID soon after B cell activation (Sayegh *et al.*, 2003). Ectopic expression of the E-protein antagonist Id3 in activated B cells hinders E47 binding to Region 2 and impairs *Aicda* transcription; conversely, overexpression of E47 in activated B cells elevates expression of AID (Sayegh *et al.*, 2003).
In contrast, E2f and c-Myb are ubiquitous repressor proteins which inhibit transcriptional activation in many cell types (Nagaoka et al., 2010). Indeed, the addition of Region 2 to a luciferase reporter plasmid carrying the minimal Aicda promoter resulted in a twofold reduction in transcriptional activity in transfected CH12F3-2 cells (Tran et al., 2010). Like Region 1, Region 2 is also unresponsive to CIT stimulation.

According to the proposed model, the silencers E2f and c-Myb act independently to repress Aicda transcription and prevent inappropriate AID expression, both in non-lymphoid cells and in unactivated B cells. Binding of the B cell-specific enhancers Pax5 and E47 can only partially derepress Aicda transcription, and so Region 2 functions as a negative regulatory region overall, in contrast to Region 1 (Tran et al., 2010).

### 1.3.3 Region 3

Region 3 is located far downstream of the fifth (last) Aicda exon, and is reportedly required for physiological expression of AID, based on experiments in transgenic mice carrying a bacterial artificial chromosome (BAC) bearing mouse Aicda (Crouch et al., 2007). Beyond that, its role remains unclear at present.

### 1.3.4 Region 4

Region 4, which is located far upstream of Region 1, enhances transcription in response to cytokine stimulation, unlike Regions 1 and 2; thus it is the region responsible for overcoming the repression of the Region 2 silencer elements and upregulating the expression of AID in activated B cells, beyond the basal level induced by Region 1 (Tran et al., 2010). This implies that Region 4 has binding sites for transcriptional activators
that are responsive to IL-4, CD40L, or TGF-β (Tran et al., 2010). Indeed it contains a number of positive regulatory cis-elements, namely three C/EBP sites, two NF-κB sites, two STAT6 sites, and a Smad3/4 site (Tran et al., 2010). NF-κB is an ubiquitous transcriptional factor that induces AID expression in B cells in response to CD40L costimulation.

The Smad transcription factor family members Smad3 and Smad4 are central to the intracellular TGF-β signaling pathway. Upon TGF-β-induced activation, TGF-β receptors, which possess serine/threonine kinase activity, phosphorylate Smad3, which then forms complexes with Smad4 (ten Dijke and Hill, 2004). These Smad3/Smad4 complexes are then shuttled to the nucleus, where they regulate transcription of TGF-β-responsive genes (ten Dijke and Hill, 2004). Mutation of the Smad3/4 binding site impairs the basal activity of Region 4, as well as its enhancer activity in response to TGF-β (Tran et al., 2010).

STAT6 plays a crucial role in mediating IL-4-dependent immune responses. IL-4 is a critical Th cell-secreted cytokine that stimulates activated B cells to proliferate, and induces class switching to the IgG1 and IgE isotypes (Takeda et al., 1996). Upon binding to its receptor, IL-4 induces activation of Janus kinase (JAK) family members, which in turn activate STAT6 through tyrosine phosphorylation (Takeda et al., 1996). Using the human B cell line BL2, Dedeoglu et al. (2004) have demonstrated that STAT6 binds to a cognate sequence in Region 4 (corresponding to the second STAT6 site in mouse Region 4) following IL-4 stimulation. Moreover, they have shown that IL-4 fails to induce Aicda
mRNA synthesis in Stat6−/− mouse B cells. In vitro mutagenesis experiments demonstrate that mutation of one or both of the STAT6 sites in Region 4 severely curtails transcriptional activation by Region 4, whether induced by IL-4 alone or by CIT (Tran et al., 2010). Taken together, these observations suggest that STAT6 is required for IL-4-dependent Aicda transcription, and that the STAT6 sites in Region 4 serve as vital enhancer elements for AID expression.

The C/EBP proteins are widely-expressed transcription factors that have been reported to function together with other transcription factors, including STAT6, to promote gene expression (Nagaoka et al., 2010; Wurster et al., 2000). Indeed, mutations introduced into one or both of the two C/EBP sites upstream of the Smad3/4 site abolish Region 4 responsiveness to both TGF-β alone or CIT (Tran et al., 2010), demonstrating that the C/EBP sites are required for the enhancer activity of Region 4.

![Figure 1.3. Organization of the mouse Aicda locus. Arrow marks transcriptional start site (TSS). (Adapted from Tran et al., 2010.)](image)

**1.4 DNA Methylation**

Epigenetic marks are heritable modifications of DNA that influence gene
expression without altering the nucleotide sequence. Many higher organisms possess the ability to epigenetically remodel their genomes through DNA methylation -- the covalent addition of a methyl group to the 5-carbon position of the cytosine ring to yield 5-methylcytosine. In mammals, methylation of cytosine residues occurs predominantly in a CpG (cytosine followed by a guanine) dinucleotide context, and is typically associated with a repressed chromatin structure and transcriptional silencing. Conversely, removal of the methyl marks induces an open chromatin state that promotes transcriptional activation.

Cytosines in a CpG context exhibit significantly higher mutability than cytosines in any other dinucleotide context. This is due to the fact that 5-methylcytosine spontaneously deaminates to thymine, which, if not corrected, results in transition mutations (Fritz and Papavasiliou, 2010). Consequently, the CpG dinucleotide is severely underrepresented in the mammalian genome. However, there are small fractions of the genome, called CpG islands, where CpGs are found in clusters (Illingworth and Bird, 2009). Although it is estimated that 70-80% of all CpGs are methylated in adult mammalian cells, most CpG islands are largely unmethylated (Feltus et al., 2003; Fritz and Papavasiliou, 2010). CpG islands are frequently located in gene regulatory regions, and methylation of such islands can generate differential gene expression patterns, for example between different tissue types or stages of development (Fritz and Papavasiliou, 2010).

1.4.1 DNA Methyltransferases
A class of enzymes called DNA methyltransferases (DNMTs) is responsible for the establishment and maintenance of DNA methylation patterns by catalyzing the transfer of a methyl group from S-adenosylmethionine (SAM) to the fifth carbon of cytosine (Dhe-Paganon et al., 2011). In mice and humans, the three active methyltransferases that have been identified are DNMT1, DNMT3A, and DNMT3B.

DNMT3A and DNMT3B are present at abundant levels in embryonic stem cells (ESCs), but are significantly repressed in postnatal somatic tissues (Okano et al., 1999). Following the global erasure of DNA methylation in primordial germ cells and the pre-implantation embryo, DNMT3A and DNMT3B act synergistically to re-establish the genome-wide pattern of methylation during gametogenesis and embryogenesis (Okano et al., 1999). Previous studies have shown that deficiency in both Dnmt3a and Dnmt3b perturbs this de novo methylation, and results in embryonic lethality (Okano et al., 1999; Dhe-Paganon et al., 2011).

DNMT1 is the most ubiquitous and abundant of the three mammalian methyltransferases (Dhe-Paganon et al., 2011). During the G1 and G2 phases of the cell cycle, it is spread throughout the nucleus at low levels. However, during S phase, it is upregulated and accumulates at DNA replication foci. In contrast to DNMT3A and DNMT3B, DNMT1 has a five to thirtyfold preference for completing the methylation of hemimethylated DNA (only one strand is methylated) over catalyzing the methylation of unmethylated DNA (Okano et al., 1999; Yoder et al., 1997). Furthermore, inactivation of Dnmt1 results in gradual and extensive loss of methylation, but does not disrupt de
novo methylation activity (Dhe-Paganon et al., 2011; Lei et al., 1996). Taken together, this evidence shows that DNMT1 is primarily a maintenance methyltransferase; it copies methylation marks from the template strand onto the newly synthesized strand during DNA replication, preserving existing methylation patterns across multiple rounds of the cell cycle.

1.4.2 DNA Demethylation

The removal of DNA methylation can take place via two distinct mechanisms, passive or active demethylation. Passive demethylation occurs as a result of cell division. During DNA replication, the newly synthesized strand is formed using unmodified bases. Thus, if DNMT1 is absent or blocked from accessing replication foci, over the course of multiple cell divisions there will be gradual dilution of DNA methylation marks. In contrast, active demethylation is replication-independent, and would require a demethylase to either remove 5-methylcytosine directly (generating an abasic site that is repaired through BER), remove the 5-methyl group from 5-methylcytosine, or initiate the conversion of 5-methylcytosine into another base that is subsequently converted to cytosine (Chen and Riggs, 2011). In plants, 5-methylcytosine can be removed directly by DME/ROS1 family member DNA glycosylases, but no corresponding 5-methylcytosine-specific DNA glycosylase has been identified in mammals (Chen and Riggs, 2011). However, there is evidence to suggest that active demethylation may proceed through an oxidative mechanism involving 5-hydroxymethylcytosine as an intermediate (Zhu, 2009). In this process, 5-methylcytosine
would be enzymatically oxidized to 5-hydroxymethylcytosine, which would in turn be converted to cytosine via BER (Zhu, 2009). Furthermore, since DNMT1 does not recognize 5-hydroxymethylcytosine well, conversion of 5-methylcytosine to 5-hydroxymethylcytosine may also trigger passive demethylation (Zhu, 2009). In support of this hypothesis, the ten-eleven translocation (TET) enzymes have been shown to be able to mediate oxidation of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian cells (Booth et al., 2012; Zhu, 2009).

### 1.4.3 Bisulfite-based DNA Methylation Analysis

Many current methods for ascertaining the presence of 5-methylcytosine in target DNA sequences rely upon bisulfite-induced modification of genomic DNA. Such methods exploit the fact that sodium bisulfite efficiently deaminates unmethylated cytosine into uracil, but reacts poorly with 5-methylcytosine, leaving it unchanged (Hashimoto et al., 2007; Hayatsu, 2008; Herman et al., 1996). Upon PCR amplification, uracil and 5-methylcytosine are replaced in the resultant amplicon by thymine and cytosine, respectively (Hashimoto et al., 2007; Hayatsu, 2008; Herman et al., 1996). Bisulfite sequencing and methylation-specific PCR (MSP) are two principal techniques that have been developed for analysis of bisulfite-converted DNA. In bisulfite sequencing, primers are directed at non-CpG sequences book-ending the region of interest, in order to amplify DNA independently of methylation status (Hashimoto et al., 2007). The resulting PCR product can then be cloned into a vector and sequenced to yield a target DNA methylation profile at single-base resolution (Hashimoto et al., 2007;
Hayatsu, 2008). In contrast, primers for MSP are designed to overlap CpG sites in the target sequence, and thus selectively amplify methylated DNA (Herman et al., 1996). MSP can be used to provide a semi-quantitative assessment of the overall methylation level of the region of interest when the ratio of product generated by methylation-specific primers versus methylation-insensitive primers is determined (Hashimoto et al., 2007). The use of quantitative PCR (qPCR) rather than endpoint PCR can greatly enhance the sensitivity of this assay (Hashimoto et al., 2007).

### 1.4.4 Non-CpG Island Methylation and AID Expression

*Aicda* is a gene whose expression must be tightly regulated, and the balance between enhancer elements and repressor elements in its multiple *cis*-regulatory regions serves as one layer of transcriptional control. However, given the significance of *cis*-regulatory regions, and the ability of DNA methylation to modulate transcription, methylation of *cis*-regulatory regions themselves may provide an additional layer of transcriptional control. Although there are no CpG islands located in Regions 1-4, even in CpG-poor gene regulatory elements, DNA methylation may still be an important mechanism of transcriptional regulation. Many transcription factors have been shown to bind inefficiently to DNA sequences that include, or are adjacent to, methylated CpGs. Thus CpG-containing *cis*-elements, such as the first Sp site in the *Aicda* minimal promoter, are potential candidates for epigenetic regulation. Moreover, although 45% of human gene promoters lack a CpG island, many have still been reported to have tissue-specific methylation patterns (Han et al., 2011). Methylation of the non-CpG island
human β-globin, human γ-globin, and rat α-actin gene promoters resulted in transcriptional repression, albeit only in cells where these promoters have weak activity (Boyes and Bird, 1992). On the other hand, methylation of the CpG-poor human LAMB3 and RUNX3 promoters has been shown to directly cause transcriptional silencing even in cells that express these genes efficiently (Han et al., 2011). No previous studies have investigated whether DNA methylation influences Aicda transcription. Therefore, I aim to elucidate the role of DNA methylation in regulating the expression of AID in B cells, in the context of its cis-regulatory network.
2. Materials and Methods

2.1 Mice and Cell Lines

2.1.1 Primary B Cells

Naïve splenic B cells were isolated from 13-16 wk-old female mice of the C57BL/6 (B6) inbred strain using the EasySep Mouse B Cell Enrichment Kit (Stem Cell Technologies). Cells were grown in RPMI 1640 medium (Sigma-Aldrich) supplemented with 10% (vol/vol) fetal bovine serum (FBS), 20 mM HEPES, 100 μM non-essential amino acids, 0.1% (vol/vol) β-mercaptoethanol, 1 mM sodium pyruvate, and 1 u/mL penicillin/streptomycin/L-glutamine (PGS). Cells were either harvested at day 0 with no stimulus, or cultured for three days in the presence of LPS (10 μg/mL) and IL-4 (10 ng/mL), anti-CD40 (1 μg/mL) and IL-4 (10 ng/mL), or anti-CD40 (1 μg/mL) alone. For experiments investigating the relationship between demethylation of Aicda and cell division, anti-CD40+IL-4-activated cells were labeled with carboxyfluorescein succinimidyl ester (CFSE) prior to culture.

2.1.2 CH12F3-2 Cells

CH12F3-2 cells were seeded at a density of 5 × 10⁴ cells/mL in RPMI 1640 medium supplemented with 10% (vol/vol) FBS, 1 u/mL PGS, 0.1% (vol/vol) β-mercaptoethanol, and 10mM HEPES, and were harvested after two days of culture. Cells were either grown with no stimulus, or were induced to switch to IgA with a combination of IL-4 (2.5 ng/mL), TGF-β (0.15 ng/mL), and anti-CD40 (100 ng/mL).

2.2 Flow Cytometry
CFSE-labeled anti-CD40+IL-4-activated B cells were harvested at day 3 and separated by division number for the first four divisions, using a MoFlo cell sorter (Beckman Coulter). Flow cytometric data was analyzed using FlowJo software (TreeStar).

2.3 DNA Preparation

Briefly, genomic DNA was isolated from CH12F3-2 and splenic B cells using the Wizard SV Genomic Purification System (Promega), concentrated by ethanol precipitation, and quantified using an Infinite 200 PRO NanoQuant absorbance reader (Tecan). The DNA (1 μg per sample) was subjected to bisulfite modification using the EpiTect Bisulfite Kit (QIAGEN), according to the manufacturer’s protocol. The bisulfite-treated DNA was then eluted in 40 μL of buffer EB (QIAGEN, 10 mM Tris-Cl, pH 8.5).

2.4 Bisulfite Genomic Sequencing

2.4.1 Primers and PCR Conditions

Bisulfite sequencing of Aicda regions 1, 2, and 4 was performed in order to identify differentially methylated cis-regulatory regions of the Aicda locus in B cells. Bisulfite PCR primers that amplify fragments of Regions 1, 2, and 4 were designed using the online program Bisearch (http://bisearch.enzim.hu/) and purchased from Sigma-Aldrich (St. Louis, MO). Analysis with MethPrimer software (http://www.urogene.org/methprimer/index.html) did not reveal any CpG islands in Regions 1, 2, or 4, so primers were simply selected to include as many CpG dinucleotides in the product as possible, with emphasis placed on CpGs in or adjacent to
transcription factor binding sites (see Figure 2.1). Furthermore, all primers were constrained to avoid any CpGs within their sequences in order to amplify DNA independently of methylation status. One primer set (AID_BiS(F), AID_BiS(R)) covers most of the minimal promoter element of Region 1 (-241 to -30 relative to transcription start site (TSS)), and probes five CpGs, including a CpG at the center of a Sp site. The second primer set (AID_2(2)_BiS_F, AID_2(2)BiS_R) spans part of Region 2 (+1621 to +1867 relative to TSS), and probes four CpGs, which includes one at the 3’ end of an NF-κB site and one near the 3’ end of an E-box. The third primer set (AID_4(1)_BiS_F, AID_4(1)_BiS_R) covers Region 4 (-8228 to -7996 relative to TSS), and probes seven CpGs, but none in or adjacent to transcription factor binding sites. (See Table 2.1 for primer information.)

Sp
Pax5
E-box
NF-κB
E2f
Cp2/c-Myb
Smad3/4
C/EBP

Aicda region 1 (-241 to -30)
TGGTACCTGGGCTGGCTTTCAGAGGAACAGCTCTGAAGGAAGTTGGACATTAAGCATGAGCAGAAGCTGCCTCCCATCCCACTTTAATCCTGGTGGCTCTGCCACCCACACAGGCCACGCCACCCTCTTTTACTGGACCCAACCAGGAGGCAGATGTTGGATACCTGGTGGTAGTGATGCTGTGTGCTGGGGAGGAGCCCCACAGAGACAGCTGC

Aicda region 2 (+1621 to +1867)
TGCTATTGCTGGGCTGGCTTCAGAGGAACAGCTCTGAAGGAAGTTGGACATTAAGCATGAGCAGAAGCTGCCTCCCATCCCACTTTAATCCTGGTGGCTCTGCCACCCACACAGGCCACGCCACCCTCTTTTACTGGACCCAACCAGGAGGCAGATGTTGGATACCTGGTGGTAGTGATGCTGTGTGCTGGGGAGGAGCCCCACAGAGACAGCTGC

Aicda region 4 (-8228 to -7996)
TGTGCTGGGCTGGCTTCAGAGGAACAGCTCTGAAGGAAGTTGGACATTAAGCATGAGCAGAAGCTGCCTCCCATCCCACTTTAATCCTGGTGGCTCTGCCACCCACACAGGCCACGCCACCCTCTTTTACTGGACCCAACCAGGAGGCAGATGTTGGATACCTGGTGGTAGTGATGCTGTGTGCTGGGGAGGAGCCCCACAGAGACAGCTGC
Figure 2.1. *Aicda* region 1, 2, and 4 fragments covered by bisulfite sequencing. Transcription factor binding sites are highlighted. For each region, CpG sites (underlined and in blue font) are numbered in ascending order, starting from 1, relative to the TSS in the 5’ to 3’ direction.

**Table 2.1: PCR primers used for bisulfite sequencing**

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Sequence (5’ to 3’)</th>
<th>T\text{Annealing}</th>
<th>Product Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>AID_BiS(F)</td>
<td>TGGTATTTGGTGGTTTTTTTTTTT</td>
<td>52.1 °C</td>
<td>212 bp</td>
</tr>
<tr>
<td>AID_BiS(R)</td>
<td>AACTTACTCTTTATAAACCTCC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AID_2(2)_BiS_F</td>
<td>TGTATTGTTGGGAAGTAGTTAAGG</td>
<td>58.5 °C</td>
<td>247 bp</td>
</tr>
<tr>
<td>AID_2(2)_BiS_R</td>
<td>AACTATTTCTATTTAATTTTTTCAACCAC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AID_4(1)_BiS_F</td>
<td>TGTGTTGGTTTTGAGATATAGGAAT</td>
<td>58.5 °C</td>
<td>233 bp</td>
</tr>
<tr>
<td>AID_4(1)_BiS_R</td>
<td>TAAAAAACCACAACAACTTAATC</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Gradient PCR was used to optimize annealing temperatures for each primer pair, in order to avoid nonspecific amplification and to achieve maximum product yield. Each PCR was set up using the GoTaq Hot Start Polymerase kit (Promega) according to manufacturer’s recommendations (see Table 2.2). All PCRs were performed in a Mastercycler ep Gradient S thermocycler (Eppendorf) under the following conditions: initial 2 min incubation at 95 °C to activate the DNA polymerase, 40 cycles of denaturation at 95 °C for 30 s, annealing at optimal annealing temperature for 30 s, and extension at 72 °C for 30 s, followed by a final 5 min extension at 72 °C (see Table 2.3). If not used immediately, all PCRs were stored at -20 °C.

**Table 2.2: Bisulfite PCR setup**

<table>
<thead>
<tr>
<th>Component</th>
<th>Final Volume</th>
<th>Final Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>5X Green GoTaq Flexi Buffer</td>
<td>5 μL</td>
<td>1X</td>
</tr>
<tr>
<td>MgCl\textsubscript{2} Solution, 25 mM</td>
<td>3 μL</td>
<td>3 mM</td>
</tr>
<tr>
<td>KAPA dNTP Mix, 10 mM each</td>
<td>0.5 μL</td>
<td>0.2 mM each dNTP</td>
</tr>
<tr>
<td>Forward primer, 10 μM</td>
<td>0.25 μL</td>
<td>0.1 μM</td>
</tr>
<tr>
<td>Reverse primer, 10 μM</td>
<td>0.25 μL</td>
<td>0.1 μM</td>
</tr>
<tr>
<td>GoTaq Hot Start Polymerase (5 u/μL)</td>
<td>0.25 μL</td>
<td>1.25 u</td>
</tr>
<tr>
<td>Template DNA, 25 ng/μL</td>
<td>1 μL</td>
<td>25 ng</td>
</tr>
</tbody>
</table>
Table 2.3: Thermal cycling conditions

<table>
<thead>
<tr>
<th>Step</th>
<th>Temperature</th>
<th>Time</th>
<th>Number of Cycles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial incubation</td>
<td>95 °C</td>
<td>2 min</td>
<td>1</td>
</tr>
<tr>
<td>Denaturation</td>
<td>95°C</td>
<td>30 s</td>
<td></td>
</tr>
<tr>
<td>Annealing</td>
<td>X °C*</td>
<td>30 s</td>
<td>40</td>
</tr>
<tr>
<td>Extension</td>
<td>72 °C</td>
<td>30 s</td>
<td></td>
</tr>
<tr>
<td>Final extension</td>
<td>72 °C</td>
<td>5 min</td>
<td>1</td>
</tr>
<tr>
<td>Soak</td>
<td>4 °C</td>
<td>Indefinite</td>
<td>1</td>
</tr>
</tbody>
</table>

*Optimum annealing temperature for the primer pair.

2.4.2 Cloning and Sequencing

PCR products were electrophoresed on 2% agarose-TAE gels containing 0.5 μg/mL ethidium bromide, for 25 min at 5 V/cm. A 100 bp ladder (New England Biolabs) was used as a marker of molecular size. The resulting DNA bands were visualized with a UV transilluminator. Desired bands were excised using clean razor blades and were gel-purified using the Wizard SV Gel and PCR Clean-Up System according to the manufacturer’s instructions (Promega). Each purified PCR product was ligated into the blue-white selection vector pCR 2.1-TOPO using the TOPO TA Cloning Kit as per the manufacturer’s instructions (Invitrogen). Briefly, ligation was performed by incubating a mixture of PCR product (4 μL), pCR 2.1-TOPO vector (1 μL), and salt solution (1 μL) at room temperature for 20 min. NEB5-α competent cells (New England Biolabs) were transformed with 3 μL of ligation reaction, grown in 100 μL of SOC medium for 1 h at 37 °C, combined with 60 μL of X-Gal and 8 μL of IPTG, and subjected to blue-white screening on Luria-Bertani (LB) agar plates containing 50 μg/mL ampicillin (selective for pCR 2.1-TOPO). After 14 h of selection, successful recombinants (white colonies) were
harvested using sterile pipette tips and sent to Eton Bioscience (Durham, NC) for direct sequencing using the M13F(-21) universal sequencing primer (5’-TGTAAAACGACGGCCAGT-3’). Sequencing traces were analyzed using FinchTV software (Geospiza). The statistical significance of the methylation change at each CpG site was established using a two-tailed Student’s T-test (significance level \( \alpha = 0.05 \)).

Three independent sequencing replicates (each from a different mouse) were obtained.

2.5 Quantitative Methylation-specific PCR

2.5.1 Methylation-specific PCR

Quantitative methylation-specific PCR (qMSP) was performed to determine the relationship between demethylation of Region 4 and cell division. A methylation-specific primer pair (AID4.1_MF, AID4.1M_R) that amplifies the most differentially methylated part of the previously-sequenced Region 4 fragment (-8156 to -8042 relative to TSS) was designed using MethPrimer and purchased from Sigma-Aldrich. Both forward and reverse primers were designed to include 2-3 CpGs within their sequence and to preferentially amplify methylated DNA (see Table 2.4). As done previously, gradient PCR was used to optimize the annealing temperature of the primer pair. All qMSPs were performed in triplicate in a 96-well plate (Eppendorf), in a 10 \( \mu \)L volume containing 1X SYBR FAST qPCR Master Mix (Kapa Biosystems), 25 ng of bisulfite-modified DNA, and forward and reverse methylation-specific primers at a concentration of 0.25 \( \mu \)M each (see Table 2.5). In order to control for the amount of input DNA in each qMSP, qPCRs using the methylation-independent primer pair for Region 4
(AID_4(1)_BiS_F, AID_4(1)_BiS_R) were run in parallel under otherwise identical conditions. All qPCRs were carried out in a Mastercycler ep realplex thermocycler (Eppendorf) according to the SYBR FAST qPCR protocol (Kapa Biosystems).

Realplex software was used for data acquisition. Data from all qPCR experiments (three independent replicates) were analyzed using the comparative Ct method (Schmittgen and Livak, 2008). All levels of methylation are reported relative to a previously-sequenced naïve B cell sample, known to be hypermethylated at Region 4.

**Table 2.4: PCR primers used for qMSP**

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Sequence</th>
<th>T\text{Annealing}</th>
<th>Product Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>AID4.1M_F</td>
<td>TCGTTAGTTTTTATGTTGTTATAAC</td>
<td>58 °C</td>
<td>115 bp</td>
</tr>
<tr>
<td>AID4.1M_R</td>
<td>AAAATCGAATCGCTACGAC</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 2.5: qMSP setup**

<table>
<thead>
<tr>
<th>Component</th>
<th>Final Volume</th>
<th>Final Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>2X SYBR Green Master Mix</td>
<td>5 μL</td>
<td>1X</td>
</tr>
<tr>
<td>Forward primer, 10 μM</td>
<td>0.25 μL</td>
<td>0.25 μM</td>
</tr>
<tr>
<td>Reverse primer, 10 μM</td>
<td>0.25 μL</td>
<td>0.25 μM</td>
</tr>
<tr>
<td>Template DNA</td>
<td>X μL</td>
<td>25 ng</td>
</tr>
<tr>
<td>Nuclease-free water to 10 μL</td>
<td>Y μL</td>
<td>N/A</td>
</tr>
</tbody>
</table>
3. Results

3.1 Regions 1 and 4 Become Demethylated in Activated B Cells

AID expression is absent in naïve B cells, but is present at high levels in antigen-activated B cells during the GC reaction. Given this, I sought to identify *Aicda cis*-regulatory regions which are differentially methylated between unactivated and activated B cells. Thus, I performed bisulfite sequencing of portions of Regions 1 (promoter region), 2 (negative intronic regulatory region), and 4 (enhancer region) (see Figure 1.3) in genomic DNA isolated from naïve and LPS+IL-4-activated mouse splenic B cells. Region 2 is almost completely unmethylated in both unactivated and activated cells (see Figures 3.1b and 3.1e). The decrease in methylation at CpG site 1 (see Figure 2.1), the E-box proximal CpG, is not statistically significant (P = 0.26) (see Figure 3.1e).

Region 1 is substantially methylated in unactivated cells, but becomes somewhat demethylated in activated cells (see Figures 3.1a and 3.1d). Of particular interest is CpG site 4 (see Figure 2.1), which is located within the first Sp binding site and undergoes a significant (P = 0.011912) decrease in methylation of 40.34 percentage points following activation (79.17% to 38.83%) (see Figure 3.1d). Region 4 exhibits the most striking change in methylation status, going from hypermethylated in unactivated cells to largely unmethylated in activated cells (see Figures 3.1c and 3.1f). Given the magnitude of this methylation change, I devoted most of my subsequent efforts toward understanding the mechanism of demethylation in Region 4. As discussed previously, Region 4 is indispensable for AID induction in response to cytokine stimulation (Tran et
al., 2010).

Figure 3.1. Methylation status of *Aicda*. (a-c) Lollipop diagrams showing the CpG methylation pattern of *Aicda* regulatory regions 1, 2, and 4 in B6 mouse B cells. Rows represent individual clones from bisulfite sequencing. For a given clone, each circle indicates the methylation status at a single CpG site. Black circles and white circles represent methylated and unmethylated CpGs, respectively. Unstimulated B cell sequences are on top, day 3 LPS+IL-4-activated B cell sequences are on the bottom. Data from B cells prepared from three different mice are shown. (d-f) Average methylation at each CpG site probed in *Aicda* regions 1, 2, and 4 in unactivated (black bars) and day 3 LPS+IL-4-activated (white bars) B cells.

3.2 Demethylation in Region 4 Correlates with Cell Division

28
It has been shown that CSR is cell division-dependent, and that *Aicda* transcription increases with successive cell divisions, coincident with the frequency of class switching (Rush *et al.*, 2005). Given this, I sought to address two specific questions: (i) Is demethylation of *Aicda* due to cell division alone, or does it coincide with the induction of *Aicda* expression? (ii) If the latter is true, then what is the relationship between cell division and demethylation of *Aicda*?

For this experiment, I made use of the fact that, in mouse B cells, anti-CD40, both alone or in conjunction with IL-4, stimulates proliferation, while anti-CD40+IL-4, but not anti-CD40 alone, induces robust *Aicda* transcription (Dedeoglu *et al.*, 2003; Nomura *et al.*, 1995). Naïve mouse splenic B cells were harvested at day 0 with no stimulus, or cultured for three days in the presence of anti-CD40, anti-CD40+IL-4, or LPS+IL-4. In order to analyze the relationship between AID expression and cell division, some of the anti-CD40+IL-4-activated B cells were labeled with the fluorescent dye CFSE prior to culture. CFSE irreversibly binds the amine groups of intracellular proteins via its succinimidyl ester moiety, and thus persists in cells for extended periods of time. Cell division results in halving of CFSE fluorescence intensity, and hence flow cytometry can be used to track division progression (Jedema *et al.*, 2004). All activated cells were harvested at day 3, and the labeled cells were then FACS-sorted into populations of uniform division number across the first four divisions, as measured by CFSE dilution. However, the relatively low numbers of cells obtained for each division posed a challenge for bisulfite sequencing. To overcome this problem, I used a quantitative methylation-specific PCR
(qMSP) approach to detect the methylation status of Region 4. Specifically, I amplified a 115 bp fragment of Region 4, using a forward primer that overlaps CpG sites 2 and 3, and a reverse primer that overlaps CpG sites 5, 6, and 7, with both primers designed to preferentially amplify methylated DNA. Naïve (~85% methylated at Region 4) and day 3 LPS+IL-4-activated (~0% methylated at Region 4) B cell DNA, whose methylation status had previously been determined by bisulfite sequencing, were used as positive and negative controls, respectively, for the specificity of the qMSP primers.

As in day 3 LPS+IL-4-activated cells, I found Region 4 to be substantially demethylated in day 3 anti-CD40+IL-4-activated cells. However, Region 4 was hypermethylated in cells cultured in the presence of anti-CD40 alone, much like in cells that were prepared without stimulation (day 0) (see Figure 3.2a). This finding was semi-quantitatively corroborated by bisulfite sequencing of Region 4 in genomic DNA prepared from the day 3 anti-CD40 and anti-CD40+IL-4 samples (see Figure 3.2c). Based on the data, demethylation of Region 4 correlates with the induction of AID gene expression.

Methylation analysis of the division-sorted B cells demonstrates that Region 4 loses methylation gradually over successive cell divisions following IL-4 stimulation (see Figure 3.2b). This is consistent with a passive, rather than active, mechanism of demethylation. In principle, downregulation of DNMT1, the maintenance DNA methyltransferase, could be responsible for the demethylation of Region 4. However, DNMT1 is highly expressed in GC B cells, and so this cannot be the case (Shaknovich et
al., 2011). Instead, taken together, the data suggests that DNMT1 is blocked from binding Region 4 in activated B cells due to the effects of IL-4.

**Figure 3.2 Division-linked demethylation of Region 4.** (a, b) The data is presented as a fold change compared to the *Aicda* region 4 methylation level in a previously sequenced unactivated (day 0) B cell DNA sample. (c) Average methylation at each CpG site probed in Region 4 in day 3 anti-CD40+IL-4-activated (green bars) and day 3 anti-CD40-activated (blue bars) B cells.

I next sought to determine whether the DNA methylation patterns of the *Aicda* loci interrogated in the mouse splenic B cells could be recapitulated in CH12F3-2 cells (a mouse B lymphoma cell line used as a B cell model). Similar to primary B cells, CH12F3-2 cells can be induced to undergo CSR upon stimulation with CIT (anti-CD40, IL-4, and TGF-β). Furthermore, basal AID expression in unstimulated CH12F3-2 cells is negligible
(Tran et al., 2010). Taken together, these observations suggest that regulation of Aicda in CH12F3-2 cells is a good approximation of physiological reality (Tran et al., 2010). Thus, I carried out bisulfite sequencing of Regions 1, 2, and 4 in genomic DNA isolated from unstimulated CH12F3-2 cells. As expected, Region 2 was unmethylated, but surprisingly, I found Regions 1 and 4 to be mostly unmethylated as well (see Figure 3.3), despite the absence of extrinsic cytokine signaling, in stark contrast to what I had observed in naïve splenic B cells (see Figure 3.2). Since AID expression is barely detectable in CH12F3-2 cells without CIT stimulation, it appears that demethylation of Region 4 is necessary, but not sufficient for induction of AID expression.

![Figure 3.3](image)

Figure 3.3 Average methylation at each CpG site probed in Aicda regions 1, 2, and 4 in unstimulated CH12F3-2 cells.
4. Discussion

Methylation in cis-regulatory regions may play an important role in controlling gene expression, even in the absence of CpG islands. In Region 1, a CpG (CpG site 4) was identified that exhibited a significant decrease in methylation following activation. Previous studies regarding the effect of Sp site methylation upon Sp1 binding are conflicting. Some investigators have reported that both Sp1 binding and transcriptional activation are unaffected by methylation at the Sp site (Harrington et al., 1987; Holler et al., 1988). However, other studies show that methylation of a CpCpG trinucleotide within a Sp site can hinder Sp1 binding (Clark et al., 1998; Mancini et al., 1999).

Specifically, Clark et al. (1998) show that methylation of both cytosines decreases Sp1 binding by 95%, while methylation of the CpG cytosine alone only reduces Sp1 binding by 20-25%. CpG site 4 is found in a CpCpG trinucleotide context, but the outer cytosine is constitutively unmethylated in our analyses. However, given the decrease in methylation at CpG site 4 upon activation, it is conceivable that Sp1 binding in Region 1, and hence basal transcription from the Aicda promoter, may be partly sensitive to DNA methylation.

Region 2 is almost completely unmethylated in both unactivated and activated B cells. This is unsurprising since it functions overall as a negative regulator of AID expression. Given the strength of the repressor elements in Region 2, DNA methylation is probably not required for silencing of Aicda transcription in unactivated B cells.

Demethylation of Region 4 of the Aicda locus in activated B cells appears to take
place as a passive process during cell division. It also appears that IL-4 signaling is necessary for efficient demethylation of Region 4, while CD40 ligation (which is simply a B cell mitogen) by itself is not sufficient. Furthermore, as indicated by the CH12F3-2 data, it appears that AID expression cannot be induced by demethylation of the Aicda locus alone, and that it requires extrinsic cytokine stimulation as well. The two STAT6 binding sites in Region 4 are known to be crucial for the induction of AID expression in B cells under conditions of IL-4 stimulation (Tran et al., 2010).

STAT6 plays a pivotal role in mediating IL-4-dependent gene expression, and is known to be critical for lineage commitment in several immune cell types (Ishii et al., 2009; Kaplan et al., 1996; Schroder et al., 2002). In recent years, STAT6 has been demonstrated to transduce extrinsic cytokine signals into epigenetic changes in cell type-specific genes (Ishii et al., 2009; Kim et al., 2007; Wei et al., 2010). For example, the Th2 locus control region (LCR), which regulates expression of the Th2 cytokine genes (IL-4, IL-5, and IL-13), contains a hypersensitive site that undergoes STAT6-induced demethylation in T cells stimulated by IL-4 (Kim et al., 2007).

We postulate that STAT6 binds its cognate sequences in Region 4 in response to IL-4 signaling, and triggers passive demethylation during the induction of AID gene expression (see Figure 4.1). Such passive demethylation would require exclusion of DNMT1, the maintenance DNA methyltransferase, from Region 4. Recruitment of DNMT1 could potentially be inhibited by histone modifications induced by STAT6 binding. It has been reported that methylation of histone H3 lysine 9 (H3K9me)}
establishes binding sites for heterochromatin protein 1 (HP1) family proteins, which directly interact with DNMT1 to mediate DNA methylation (Paul et al., 2010; Smallwood et al., 2007). Similarly, trimethylation of histone H3 lysine 27 (H3K27me3) is recognized by enhancer of zeste homolog 2 (EZH2), a polycomb group protein which recruits DNMT1 (Paul et al., 2010; Viré, et al., 2006). In CD4+ T cells, STAT6 has been reported to decrease H3K27 trimethylation of its target genes (Wei et al., 2010). In IL-4-stimulated macrophages, STAT6 induces upregulation of H3K27 demethylase Jumonji domain containing 3 (JMJD3), a histone demethylase which removes the repressive H3K27 methylation marks from specific inflammatory genes (Ishii et al., 2009). Thus, STAT6 could mediate IL-4-dependent histone demethylation of Region 4, resulting in a chromatin environment that fails to recruit DNMT1 for maintenance methylation.

DNA demethylation in Region 4 would likely be accompanied by the formation of an active chromatin configuration that is accessible to crucial trans-activating factors. At least in the case of NF-κB, it appears that a closed chromatin architecture inhibits binding of its cognate sequence (Rau et al., 2012). Still, whether NF-κB, C/EBP, and/or Smad3/4 require Region 4 to be demethylated in order to bind remains unclear. However, a chromatin immunoprecipitation (ChIP) sequencing approach should allow this question to be addressed directly in future.

Ultimately, developing a comprehensive understanding of how Aicda is regulated will allow us to better understand pathways of lymphomagenesis, given that many B cell lymphomas express high levels of AID (Hardianti et al., 2004; Komeno et al.,
2010; Kotani et al., 2004; Pasqualucci et al., 2004). In addition to advancing medical knowledge, understanding Aicda regulation in greater detail may help us to harness AID for directed evolution of proteins with desirable properties. AID has been used to artificially improve the affinity of antibodies through SHM, but given its promiscuity, it can be applied even more broadly (Chen et al., 2012). Indeed, AID has been used to engineer diverse functional libraries of non-antibody proteins, such as novel fluorescent proteins and more powerful anti-apoptosis proteins (Majors, et al., 2011).
Figure 4.1 IL-4-induced AID expression via STAT6-mediated demethylation.
(a) In naïve B cells, Region 4 is hypermethylated, and DNMT1 maintains this methylation pattern across multiple rounds of cell division. As a result, Region 4 adopts a closed chromatin configuration, preventing the binding of trans-activating factors. Thus, the enhancer activity of Region 4 is silenced, and Aicda transcription is repressed. (b) IL-4 induces STAT6 to bind its cognate sequences in Region 4. STAT6 inhibits DNMT1 from accessing hemimethylated CpGs during replication (possibly
by antagonizing repressive chromatin marks that recruit DNMT1), resulting in passive demethylation. Demethylation is accompanied by formation of an open chromatin architecture, allowing trans-activating factors such as NF-κB to bind Region 4 and induce Aicda transcription.
References


of activation-induced cytidine deaminase gene expression by IL-4 and CD40 ligation is dependent on STAT6 and NF-κB. International Immunology 16:395-404.


Lei, H., S. P. Oh, M. Okanao, R. Juttermann, K. A. Goss, R. Jaenisch, and E. Li,


