The Dissertation Committee for Hong Luo Certifies that this is the approved version of the following dissertation:

The regulation of AID function by transcription factors PU.1 and IRF4 in chicken B cells

Committee:

Ming Tian, Supervisor

Philip W Tucker

Jaquelin P Dudley

Christopher S Sullivan

Martin Poenie

The regulation of AID function by transcription factors PU.1 and IRF4 in chicken B cells

by

Hong Luo, B.S. , M.A.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December 2009

Dedication

To my parents

Acknowledgements

The journey of my dissertation has experienced both good and hard moments, and both are shared with many people. I would like to take this opportunity to express gratitude to my advisor Dr. Ming Tian. I thank him for his trust and support throughout these years. Without his teaching and helping, I would not be able to finish my dissertation. I highly appreciate the rest of my dissertation committee members: Dr. Philip W Tucker, Dr. Jaquelin P Dudley, Dr. Christopher S Sullivan, and Dr. Martin Poenie for their helpful inputs and constructive criticisms. I thank my labmate Yonghwan Kim for his help in the lab. I thank all the other people I have learned from in the microbiology program. I also appreciate microbiology graduate program coordinators: Wendy Smith and Helen M Wormington for their assistance in the past years. My husband, Dr. Yong Miao, and my parents have been a constant source of inspiration and encouragement. I thank them for always being there for me. I am extremely fortunate to meet many good friends in Austin. I thank them for all the happy moments they brought to me.

The regulation of AID function by transcription factors PU.1 and IRF4 in chicken B cells

Publication No._____

Hong Luo, Ph.D.

The University of Texas at Austin, 2009

Supervisor: Ming Tian

B cells are capable of producing antibodies of diverse antigen specificities and effector functions to counter infection by a wide range of pathogens. The diversification of immunoglobulin (Ig) is achieved through a series of programmed DNA recombination and mutagenic events during B cell maturation. A key factor involved in the Ig diversification process is Activation Induced Cytidine Deaminase (AID). AID is a B cell specific enzyme that is critical for three distinct pathways of Ig diversification: class switch recombination, somatic hypermutation and Ig gene conversion. AID functions by deaminating cytosine to uracil in target DNA at the Ig loci. Although essential for effective immunity, the mutagenic activity of AID needs to be confined to the Ig loci in order to protect genomic integrity, but the underlying mechanism is not fully understood. In this study, I show that two lymphoid specific transcription factors, PU.1 and IRF4, play important roles in regulating AID function in chicken B cells. PU.1 and IRF4 have been implicated in many aspects of B cell development and function. The two factors could form a heterodimer and regulate target gene expression cooperatively. However, we found that PU.1 and IRF4 appear to have different impacts on AID function. We

show that PU.1 is important for the expression of AID gene in chicken B cells, and the regulation appears to involve direct interaction of PU.1 with the AID gene. By comparison, IRF4 plays a minor role in AID expression. On the other hand, both PU.1 and IRF4 are required for efficient gene conversion that is mediated by AID at the Igλ locus. Moreover, the gene dosage of PU.1 is critical for AID function, since a severe gene conversion defect is observed in PU.1$^{+/-}$ cells. The function of PU.1 and IRF4 in AID-mediated gene conversion involves binding sites for the PU.1/IRF4 complex within a regulatory element at the Igλ locus. Future studies will be directed at understanding how PU.1 and IRF4 regulate AID-mediated gene conversion.

Table of Contents

List of Tables x

List of Figures xi

Chapter 1 Introduction 1

1.1 The molecular mechanisms of Immunoglobulin (Ig) diversification. 1

1.1.1 V(D)J recombination 1

1.1.2 Gene conversion 4

1.1.3 Somatic hypermutation 6

1.1.4 Class switch recombination 8

1.2 Activation-induced cytidine deaminase mediates class switch recombination, somatic hypermutation and Ig gene conversion. 10

1.3 AID function is regulated at multiple levels. 15

1.3.1 Transcriptional regulation of AID 16

1.3.2 Post-transcriptional regulation of AID 16

1.3.3 Recruitment of AID to the physiological targets at the Ig locus. 17

1.3.4 Differential repair of AID mediated mutations in Ig versus non-Ig loci 21

Chapter 2 Materials and Methods 24

2.1 Cell culture and transfection 24

2.2 Targeted mutagenesis 25

2.3 Gene conversion assay 25

2.4 Dam methylastion assay 26

2.5 I-SceI mediated gene conversion assay 27

Chapter 3 PU.1 and IRF4 regulate AID gene expression. 28

3.1 Background 28

3.2 Results 31

3.2.1 PU.1 is important for AID expression in chicken B cells. 31

3.2.2 IRF4 plays a minor role in AID expression in chicken B cells. 33

3.2.3 Mutation of a PU.1-IRF-binding site in AID gene reduces AID expression. 35
3.2.4 PU.1 and IRF4 interact with the AID gene. 36
3.3 Discussion 41
Chapter 4 IRF4 and PU.1 are important for AID-mediated gene conversion. ...44
4.1 Results 44
4.1.1 Gene conversion defect in IRF4-deficient DT40 cells. 44
4.1.2 A gene dosage effect of PU.1 on AID-mediated gene conversion.50
4.1.3 Gene conversion defect in PU.1-IRF binding site mutants at the Igλ locus. 54
4.1.4 Interaction of PU.1 and IRF4 with the Igλ locus. 57
4.2 Discussion 58
Appendix The homologous recombination defects in PU.1-deficient cells 62
A.1 Background 62
A.2 Results and discussion 63
References 67
Vita 77

List of Tables

Table 1. Gene conversion analysis of the IRF4 mutant cells. 47

Table 2. I-SceI mediated gene conversion in IRF4 deficient cells.....50

Table 3. Gene conversion analysis of PU.1 heterozygous cells. 52

Table 4. I-SceI mediated gene conversion in PU.1 heterozygous cells. 53

Table 5. Gene conversion analysis of PU.1-IRF binding site (BS) mutants. 56

Table 6. Homologous targeting efficiency in PU.1 mutant cells. 65

List of Figures

Figure 1.1. V(D)J recombination at the murine IgH locus (adapted from reference 2). 3

Figure 1.2. Gene conversion at the rearranged Ig V_λ region in chicken B cells. .6

Figure 1.3. The somatic hypermutation at the human rearranged IgH locus. 7

Figure 1.4. Class switch recombination at the human IgH locus. 9

Figure 1.5. The RNA-editing model of AID. 12

Figure 1.6. The DNA-editing model of AID. 14

Figure 3.1. AID expression is strongly reduced in PU.1-deficient cells. 32

Figure 3.2. AID expression is reduced in IRF4-deficient cells. 35

Figure 3.3. Mutation of a PU.1-IRF composite binding site near AID promoter.36

Figure 3.4. Methylation assay for the interaction of PU.1 and IRF4 with the AID gene. 41

Figure 4.1. IRF4 deficiency impairs AID mediated gene conversion. 46

Figure 4.2. IRF4 is dispensable for I-SceI-mediated gene conversion. 49

Figure 4.3. PU.1 deficiency impairs AID-mediated gene conversion. 51

Figure 4.4. PU.1 is dispensable for I-SceI mediated gene conversion. 53

Figure 4.5. PU.1-IRF composite binding sites in the Igλ locus are involved in AID-mediated gene conversion. 56

Figure 4.6. Methylation analysis of the interaction of PU.1 and IRF4 with Region A in the Igλ locus. 57

Chapter 1

Introduction

1.1 THE MOLECULAR MECHANISMS OF IMMUNOGLOBULIN (IG) DIVERSIFICATION.

In order to defend the host against infections, the immune system has evolved strategies to recognize and destroy a wide range of pathogens. A major part of the immune response is mediated by B cells, which are capable of producing a large repertoire of antibodies with distinct antigen-binding sites and effector functions. Antibody diversity is achieved through a series of programmed DNA recombination and mutagenic events that take place during various stages of B-cell maturation, including the V(D)J recombination, Ig gene conversion, somatic hypermutation, and class switch recombination. Following is a description of these Ig diversification mechanisms.

1.1.1 V(D)J recombination

Each antibody molecule consists of two Ig heavy chains (IgH) and two Ig light chains (IgL) [1], which can be divided into the variable (V_H and V_L) and constant regions (C_H and C_L) [1]. The variable regions associate with antigens [1], whereas the C_H region mediates the effector functions [1]. The coding sequences for the Ig variable regions are assembled by joining a set of variable (V), diversity (D), and joining (J) gene segments [1] through V(D)J recombination during the early stage of B-cell maturation in the bone marrow (Fig. 1.1*A*). The recombination reaction is catalyzed by a recombinase that is encoded by the recombination-activating genes 1 and 2 (RAG1 and RAG2) [2]. The two proteins constitute an endonuclease that cleaves the recombination recognition

sequences (RSSs) at the border of the V, D, and J gene segments [2], and the resultant double-stranded DNA breaks (DSBs) are joined by the nonhomologous end joining (NHEJ) system to form the recombination product [2]. V(D)J recombination introduces diversity in antibody variable region at two levels. First, there are many different V, D, and J gene segments at the IgH and IgL loci, and the random joining of these gene segments gives rise to a tremendous amount of combinatorial diversity of the assembled V region coding sequences. Moreover, the combination of different pairs of heavy and light chains further increases the range of antigen-binding sites. Second, during the repair of the RAG-mediated double stranded DNA breaks, nucleotides can be added or lost at the junctions between the gene segments [1], and the imprecise junction is another major contributor to the diversity of antibody binding site. For these reasons, a diverse population of B cells with distinct antigen binding specificities is generated during B-cell maturation.

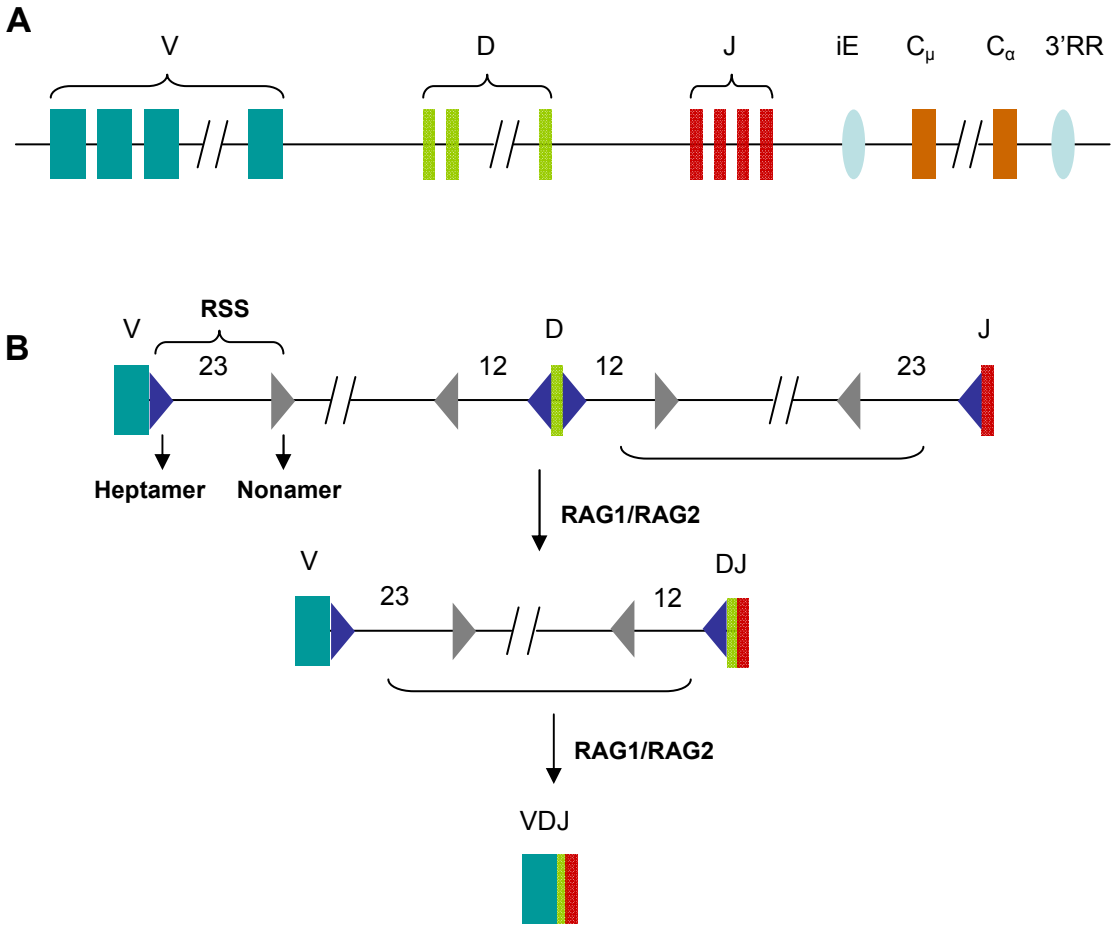


Figure 1.1. V(D)J recombination at the murine IgH locus (adapted from reference 2). *A*. The germline configuration of murine IgH locus (not to scale). iE and 3'RR refer to intronic enhancer and 3' regulatory regions, respectively. C_μ and C_α refer to the constant region of μ and α heavy chains. *B*. Scheme of V(D)J recombination at the IgH locus. RSS is composed of a conserved heptamer and a nonamer motif with a spacer of either 23 or 12-bp. The spacer length (23 or 12-bp) is indicated above the RSSs in the figure. Recombination takes place between an RSS of 12-bp spacer and an RSS of 23-bp spacer. This phenomenon is called the 12-23 rule.

In spite of the apparent diversity of V(D)J recombination, the reaction proceeds in a temporally ordered fashion during B-cell maturation in the bone marrow [1]. The IgH locus rearranges prior to the IgL loci, and takes place at the pro-B cell stage. After the successful completion of the IgH rearrangement, the expressed IgH protein associates

with the surrogate light chain to form the pre-B cell receptor, which signals the B-cell precursor to progress to the pre-B cell stage and carry out the rearrangement of the IgL loci.

There are two isotypes of IgL: κ and λ, and the κ locus generally rearranges prior to the λ locus. Besides the differential timing of the heavy and light chain rearrangements, the different gene segments also undergo V(D)J recombination in a defined order. Within the IgH locus, the D and J gene segments are recombined first, and V is subsequently joined to the combined DJ gene segments (Fig. 1.1*B*). A third level of complexity is the differential rearrangements of the two alleles of IgH and IgL. One allele of IgH or IgL is randomly chosen to undergo V(D)J recombination first. If the first rearrangement results in the formation of a functional coding sequence, the second allele is precluded from further rearrangement. This phenomenon is called allelic exclusion, which ensures that each B cell expresses a unique kind of antibody. The mechanisms that govern the orderly progression of the V(D)J recombination events, including allelic exclusion, remain poorly understood.

After the successful completion of V(D)J recombination in both heavy and light loci, a B-cell receptor is expressed at the cell surface, and the B-cell precursor becomes immature B cell [1]. After selection against self-reactivity, the surviving immature B cell leaves the bone marrow and migrates to the spleen to become mature B cells, which circulate in the peripheral lymphoid tissues to monitor against infectious agents [1].

1.1.2 Gene conversion

V(D)J recombination is a major contributor of antibody diversity in murine and human B cells. By contrast, a different mechanism is utilized to generate a diverse

repertoire of antigen receptors in avian species. In chicken, the IgH locus contains only a single functional V and J gene segment and 16 D gene segments [3], and there is only one light chain locus, Igλ, which contains a single functional V and J gene segment (Fig. 1.2) [3]. Obviously, such a simple Ig gene structure is not sufficient to generate a diverse repertoire of antibodies via V(D)J recombination alone. To compensate, chickens have evolved a gene conversion strategy to diversify the antigen-binding site. The process involves tandem arrays of pseudo-genes that are positioned upstream of the rearranged variable regions in both the IgH and Igλ loci [3] (Fig. 1.2). During gene conversion, random stretches of the pseudo-gene sequences are transferred into the rearranged variable regions [4] (Fig. 1.2), and the range of antigen-binding sites are substantially expanded as a consequence. The gene conversion reaction is accompanied by non-templated mutations in the variable region, akin to the somatic hypermutation that usually takes place during the activation of mature B cells, and such random mutations also contribute to the diversification of antigen-binding sites. Ig gene conversion takes place in an organ unique to birds, the bursa of Fabricius. In chickens, V(D)J recombination starts in the B-cell precursors in the yolk sac on embryonic day 5 to 6 and ends by embryonic day 15 [5]. These B-cell precursors seed into the bursa of Fabricius between embryonic day 10 to 15, where they undergo extensive proliferation and gene conversion as well as low levels of somatic hypermutation [5]. The bursa involutes at about 4 to 6 months after hatching, and mature B cells migrate to peripheral lymphoid organs to mediate immune responses [5].

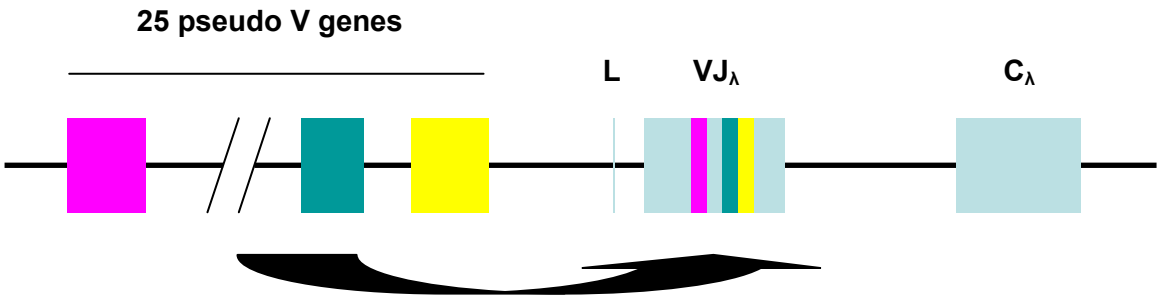


Figure 1.2. Gene conversion at the rearranged Ig V_λ region in chicken B cells. Structure of the rearranged chicken Igλ locus is illustrated. Twenty five pseudo V gene segments are situated upstream of the rearranged variable region. The arrow indicates the transfer of genetic information from the pseudo genes into the rearranged variable region. Abbreviations: L, leader exon; VJ, rearranged variable Ig region composed of joined V and J gene segments.

1.1.3 Somatic hypermutation

Even though V(D)J recombination and gene conversion are effective mechanisms of antibody diversification, the resultant antibodies are not specifically directed against any particular antigen before the naïve B cells have experienced an immune response. For this reason, the antibodies produced by naïve B cells in a primary immune response are generally of low affinity toward the offending infectious agents [6]. To remedy this deficiency, upon antigenic challenge, B cells undergo a process called somatic hypermutation to improve the specificity of the antigen-binding site [1]. This reaction takes place within the germinal centers of lymphoid tissues, such as the spleen and lymph nodes [1]. The process is characterized with high levels of primarily point mutations within the variable regions of both the IgH and L chains. The mutation rate was estimated to be in the range of 10^{-3} per base pair (bp) per cell generation, which is approximately a million fold higher relative to the spontaneous mutation rate in the

genome in general [7]. The mutations are distributed in a region that is approximately 150-bp to 2-kb downstream of the V region transcription initiation site (Fig. 1.3) [7]. The peak of mutation activity centers on the V region, and the C region is not affected by the mutation process (Fig. 1.3) [7]. Within the variable region, these points mutations are distributed in a largely random fashion, but those mutations that improve the antigen binding affinity are selected, and these B cells undergo further expansion and eventual maturation into plasma cells and memory B cells, which are capable of producing antibodies of higher affinity relative to those produced by the original naïve B cells [1]. This process is called affinity maturation, and underlies the increased potency of humoral immune response after repeated exposure to a particular antigen.

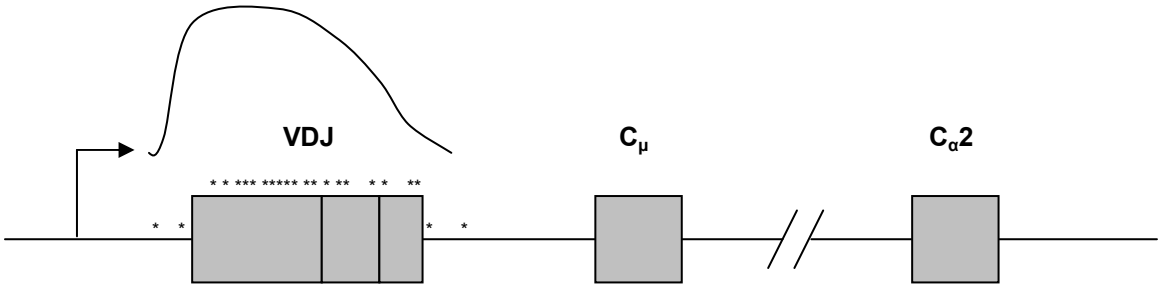


Figure 1.3. The somatic hypermutation at the human rearranged IgH locus. * represents somatic hypermutations. The curve above the rearranged IgH variable region (VDJ) depicts the mutation region and frequency. Somatic hypermutation is primarily confined to a region that is between 150-bp to 2-kb downstream of the transcription initiation site of the V region promoter, which is indicated with an arrow.

1.1.4 Class switch recombination

Antibody class and effector functions are determined by the isotypes of Ig heavy chain constant regions (C_H) [1]. There are five major isotypes of IgH: μ, δ, γ, α, and ε in human and mice, with subtypes within the γ and α families [1]. Each heavy chain isotype is encoded by a distinct C_H gene downstream of the rearranged V_H region at the IgH locus (Fig. 1.4) [1]. In naïve B cells, C_μ is the constant region gene that is most proximal to the rearranged V_H gene (Fig. 1.4). For this reason, C_μ is expressed in association with V_H through transcription and splicing to form IgM (Fig. 1.4) [1]. When naïve B cells are stimulated by antigen within peripheral lymphoid tissues, such as the spleen or lymph nodes, Ig heavy chain constant region is triggered to undergo class switch recombination (CSR) [1]. This recombination reaction takes place between switch (S) regions that precede each C_H gene except C_δ (Fig. 1.4) [1, 8]. During CSR, two S regions are cleaved and joined together, and the intervening sequence is excised as an extra-chromosomal circle (Fig. 1.4) [1, 8, 9]. As a consequence, C_μ is replaced with a downstream C_H exon, and the IgH isotype is switched to a different class (Fig. 1.4) [1, 8, 9]. Since different C_H regions elicit distinct immune responses, class switch recombination diversifies the effector functions of antibodies, and allows the production of antibody classes with the most effective effector function to eliminate the infectious agents [1].

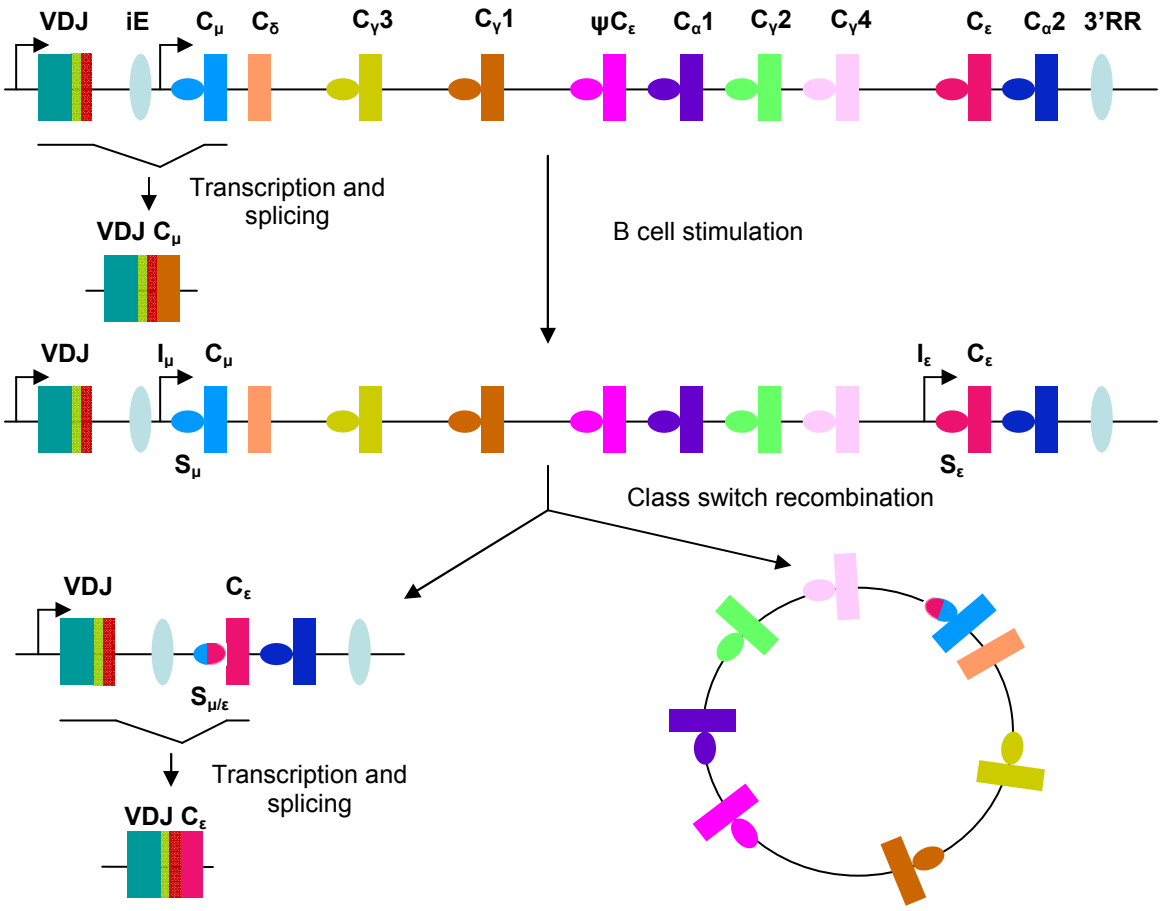


Figure 1.4. Class switch recombination at the human IgH locus. Human C_H region structure is illustrated (not to scale). In this diagram, B cells are stimulated to undergo class switching to C_ε.

In summary, four pathways contribute to the diversity of antibodies. In mice and humans, V(D)J recombination is responsible for generating the primary repertoire of antigen-binding sites during the early stages of B cell maturation in the bone marrow. By contrast, a different strategy has evolved in avian species to develop the primary pool of antigen receptors. In chickens, Ig gene conversion, rather than V(D)J recombination,

is the major contributor to Ig diversity during B-cell development in a unique avian organ, the bursa of Fabricius. After B cells have matured and migrated to the peripheral lymphoid tissues, their antigen specificity could undergo further modification during an immune response. Upon antigenic challenge, the variable regions of activated B cells accumulate somatic hypermutations, which improve the antigen-binding affinity of the antibodies. In addition, activated B cells undergo class switch recombination to expand the range of effector functions elicited by the antibodies, which facilitates the elimination of the infectious agents. Mechanistically, V(D)J recombination is mediated by the RAG-1 and RAG-2 recombinase, whereas the remaining three pathways, gene conversion, somatic hypermutation and class switch recombination, involve a distinct factor, Activation Induced Cytidine Deaminase (AID), which is introduced in the following section.

1.2 ACTIVATION-INDUCED CYTIDINE DEAMINASE MEDIATES CLASS SWITCH RECOMBINATION, SOMATIC HYPERMUTATION AND IG GENE CONVERSION.

AID was originally identified in activated B cells [10]. Its role in class switch recombination, somatic hypermutation, and Ig gene conversion was established by genetic evidence in mice, chicken B-cell lines as well as in humans [11]. Ablation of the AID gene in mice completely inhibits class switch recombination and somatic hypermutation [12]. A similar phenotype was observed in human patients who have inherited defective copies of the AID gene [13]. These patients suffer from an immune deficiency called hyper IgM syndrome, which is characterized with abnormally low levels of IgG and IgA in the blood due to defective class switch recombination in B cells, and the variable regions of their immunoglobulin genes lack somatic hypermutations [13]. Because of these defects, these patients are prone to develop recurrent infections

at an early age [13], clearly demonstrating the importance of AID and the two AID-dependent pathways, class switch recombination and somatic hypermutation, in effective immunity. In chicken B cells, ablation of the AID gene abolishes Ig gene conversion [14], establishing a mechanistic link between this pathway with the apparently distinct class switch recombination and somatic hypermutation.

Based on sequence homology, AID belongs to the family of cytidine deaminases and is most closely related to the family member, apolipoprotein B mRNA-editing enzyme, catalytic polypeptide 1 (APOBEC-1). The function of APOBEC-1 has been well established: the enzyme deaminates C2153 of apoB100 mRNA to U2153 [10]. The editing creates an in-frame stop codon, which leads to the production of a truncated protein with novel functions [10]. Based on this information, the RNA editing model was proposed by Honjo's group to explain the function of AID (Fig. 1.5) [15]. According to this model, AID was hypothesized to edit a putative pre-mRNA to allow the expression of a protein that is involved in class switch recombination, somatic hypermutation and Ig gene conversion; for example, the putative editing target of AID could be an endonuclease that cleaves the switch regions during class switch recombination. In spite of the apparent analogy between AID and APOBEC-1, no RNA editing target for AID has been identified so far.

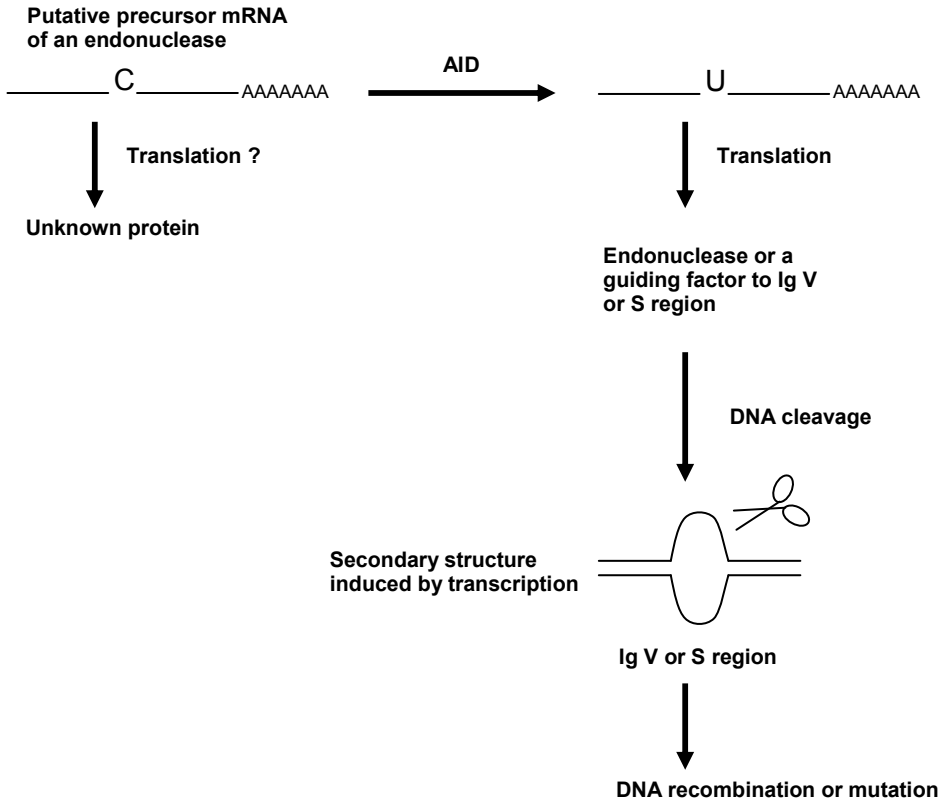


Figure 1.5. The RNA-editing model of AID.

Due to the lack of concrete evidence for the RNA editing hypothesis, an alternative model was proposed to explain the function of AID (Fig. 1.6) [11, 16]. In this model, the substrate of AID is DNA rather than RNA, and the cytidine deamination activity of AID leads to the accumulation of uridines in target DNA within the Ig loci [11, 16]. With respect to class switch recombination, the target DNA would be the switch regions, and AID is thought to deaminate cytidines within switch region DNA [8].

The resultant uridines could be removed by the uracil DNA glycosylase, and the abasic sites would be cleaved by AP endonuclease to form single-stranded nicks [8]. The deamination reaction could also activate the mismatch repair system, which recognizes U:G mismatches, and could potentially introduce DNA strand breaks in the process of removing the mismatch [17]. If the reaction occurs at high frequency in the switch region, two nicks could be formed in close proximity on both strands and result in a double stranded DNA break; ligation of the double-stranded DNA breaks in two switch regions would lead to class switch recombination [8].

The DNA deamination model could also be applied to explain somatic hypermutation and Ig gene conversion; in this case, the DNA target of AID would be the Ig variable region [7]; AID would deaminate cytidines in the Ig variable region DNA, and the downstream processing of the resultant uridines by different pathways could give rise to either somatic hypermutation or gene conversion. For example, replication of a DNA template containing the uridines would result in C:G to T:A transitions in the daughter DNA strands [16, 18, 19]. If the uridines have been removed by uracil DNA glycosylase before DNA replication, abasic sites would accumulate in the DNA and interfere with the progression of the normal replication DNA polymerase through the template [16, 18, 19]. To rescue the stalled DNA replication process, the cells would resort to error-prone DNA polymerases that are capable of translesion synthesis through abasic sites, with the price that incorrect nucleotides would be incorporated in opposition to the abasic sites in the daughter DNA strand, leading to either transition or transversion type of mutations [16, 18, 19].

Additionally, the U:G mismatch could also be the target of mismatch repair system as described earlier [17]. To correct the mismatch, the mismatch repair system normally excises a stretch of DNA containing the mismatch, and the resultant gap could

potentially be filled in by error-prone DNA polymerases to introduce mutations. Ig gene conversion could be accomplished through a similar pathway up to the stage of abasic site formation in the Ig variable region [20]. In the case of gene conversion, these abasic sites would be cleaved by the AP endonuclease to form DNA strand breaks, which could be repaired by the homologous recombination machinery, using the adjacent pseudo genes as the repair template [20, 21].

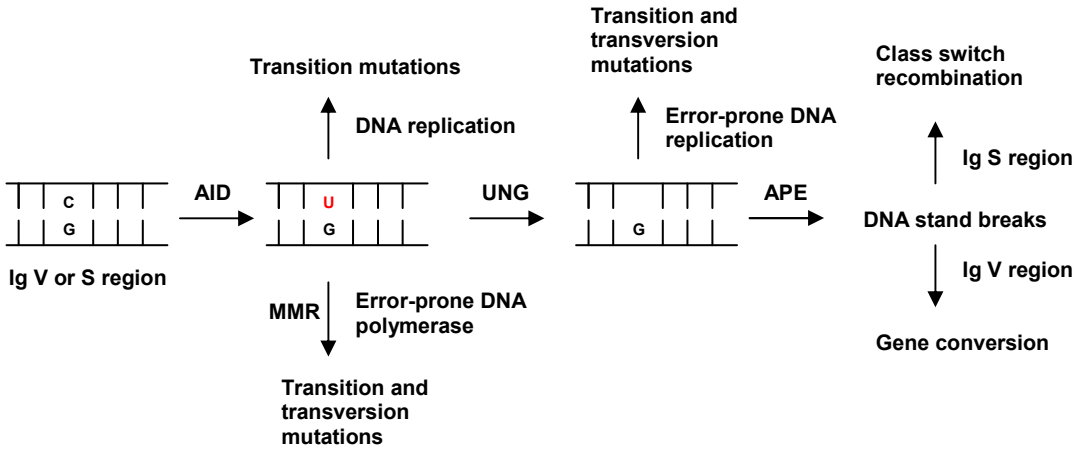


Figure 1.6. The DNA-editing model of AID. Abbreviations: UNG: uracil DNA glycosylase; APE: apurinic/apyrimidinic endonucleases; MMR: mismatch repair; V region: variable region; S region: switch region.

The DNA deamination model is consistent with several experimental observations. A key prediction of the DNA deamination model is that uracil DNA glycosylase would play an important role in AID-mediated Ig diversification events. Consistent with this prediction, genetic ablation of uracil DNA glycosylase gene in mice

leads to severe reductions in class switch recombination, and alters the pattern of somatic hypermutation [22]. Similarly, human patients with defects in the uracil DNA glycosylase gene exhibit symptoms of hyper-IgM syndrome with reduced levels of IgG and IgA in the serum, which is indicative of defects in class switch recombination [23]. Inactivation of the uracil DNA glycosylase gene in chicken B cells also compromises Ig gene conversion [24]. Besides uracil DNA glycosylase, the mismatch repair system and error-prone DNA polymerases have also been implicated in the processing of AID-induced Ig diversification events, and examination of animals deficient in these repair factors also yields results that are consistent with the prediction of the DNA deamination model [25]. Consistent with these genetic analyses, biochemical studies showed that AID is indeed capable of deaminating cytidines in the context of DNA *in vitro* [26-36].

In summary, AID has been established as the key factor for three apparently distinct pathways of Ig diversification: class switch recombination, somatic hypermutation and Ig gene conversion. Most of the current experimental evidence supports the model that AID functions by deaminating cytidines in target DNA: switch regions and variable regions at the Ig loci. Subsequent processing of the deamination product, uridine, via different repair pathways could result in class switch recombination, somatic hypermutation and gene conversion.

1.3 AID FUNCTION IS REGULATED AT MULTIPLE LEVELS.

Although the DNA deamination model could satisfactorily explain some of the key observations of AID function, it raises an important question: how is the DNA deamination activity of AID localized to the physiological targets in the Ig loci, that is, the switch regions and variable regions? The question is important since indiscriminate

deamination of the whole genome would lead to rampant mutations and genomic instability. For this reason, AID function is under stringent regulation at multiple levels. The following is a discussion of these regulatory mechanisms.

1.3.1 Transcriptional regulation of AID

AID expression is restricted to activated B cells [10] and bursal B cells [14], where AID function is needed for Ig diversification. The importance of this level of control is demonstrated by experiments where AID is expressed with a ubiquitous promoter in transgenic mice [37]. The ectopic expression of AID leads to the development of T-cell lymphomas as well as tumors in other tissues, such as the lung [37]. The transcriptional regulation of AID has been most extensively analyzed in mice. It has been shown that the induction of AID expression during B-cell activation involves several transcription factors, including Pax5, STAT6, NF-κB, IRF4, IRF8, and HoxC4 [38-43]. However, the mechanism responsible for AID expression in chicken bursal B cells is poorly understood. Since there is little sequence homology between the chicken and mouse AID promoters, the regulatory mechanism may differ between the two species. It has been shown that ablation of Pax5 reduces AID expression in the chicken bursal B cell line DT40 [44]. Since loss of Pax5 induces plasma cell differentiation, it is unclear whether Pax5 directly regulates the AID gene.

1.3.2 Post-transcriptional regulation of AID

Besides transcriptional regulation, AID expression is also subject to controls at various posttranscriptional levels. The stability of AID mRNA has been shown to be negatively regulated by microRNAs, miR-181b [45] and miR-155 [46, 47], which

recognize elements in the 3' untranslated region of the AID transcript. Additionally, AID protein shuttles between the nucleus and cytoplasm, and this shuttling is controlled by a nuclear localization signal in the N-terminus and a nuclear export signal in the C-terminus of the polypeptide [48]. During the steady state, most of the AID protein is localized in the cytoplasm [49, 50], which would reduce the potential for nonspecific deamination in the genome. The predominant cytoplasmic location of AID appears to be the consequence of at least two factors: the relative strength of export versus import signal [49, 50] and the preferential degradation of the nuclear form of AID [51].

1.3.3 Recruitment of AID to the physiological targets at the Ig locus

The regulation of AID levels by both the transcriptional and posttranscriptional mechanisms is unlikely to be sufficient to prevent genomic instability, since AID is still expressed in B cells at levels sufficient to induce mutation in the Ig loci. The question is how AID is targeted to the physiological targets at the Ig loci within B cells. The targets of AID during class switch recombination are the switch regions, which are composed of 1-10 kb of repetitive DNA [8]. The sequences of switch regions are characterized with two unusual features: the non-template strand is G-rich and there are abundant AGCT motifs in the repetitive DNA [8]. As described earlier, class switch recombination is closely linked with transcription from promoters situated upstream of the switch regions [8]. The transcription through the switch regions generates a non-coding RNA, which is G-rich due to the high C content in the template strand of the switch region DNA [8]. It has been shown that the G-rich RNA tends to associate with the template DNA to form an RNA-DNA hybrid and displaces the non-template strand as single-stranded DNA; this structure is called an R-loop [52]. Such R-loops have indeed been detected *in vivo* in

the switch regions engaged in class switch recombination [52-55]. Moreover, genetic experiments showed that inhibiting R-loop formation in switch regions reduces the efficiency of class switch recombination in mice [53]. The role of R-loop could be explained by the substrate requirement of AID. AID can only deaminate cytidines in the context of single-stranded DNA, since the amino group in cytidine is engaged in hydrogen bonding with the complementary base in double-stranded DNA. The displaced non-template strands in an R-loop have been shown to be good substrates for AID *in vitro* [27, 28, 30, 31], which would be consistent with a role of this structure during class switch recombination. The second feature of switch regions is the large number of AGCT motifs in the repetitive DNA [8]. It has been shown that the AGCT motif represents a hotspot for the AID deamination activity *in vitro* [56, 57]. Thus, the combination of R-loop formation and high densities of AGCT motif could render the switch regions prime targets for AID function. Besides the unusual nature of switch region sequence, other mechanisms, such as the recruitment of AID to the Ig loci via transcription factors, could also be involved, and these mechanisms are discussed in more detail below.

The targeting mechanisms for AID during class switch recombination cannot be readily applied to explain the targeting of AID to the Ig variable regions in somatic hypermutation and gene conversion. In contrast to the unusual sequence of the switch regions, the Ig variable region does not show any distinctive features relative to DNAs in non-Ig loci. In fact, non-Ig DNAs could become targets of AID when incorporated into the Ig loci [58]. Therefore, elements outside of the variable region are likely recruiting AID to the target. The prime candidate for the recruiting elements is the transcription apparatus of the Ig loci. Similar to class switch recombination, transcription is a prerequisite to both somatic hypermutation and Ig gene conversion [59-61]. Moreover,

the spatial distribution of somatic hypermutation is determined by the relative distance to the promoter: the mutations are confined to a region that is 150 bp to 2 kb downstream of the transcription initiation site [7].

In spite of the correlation, transcription *per se* cannot be sufficient to target AID function, since many highly transcribed non-Ig genes in B cells are not mutated by AID [62]. A potential explanation is that the transcription apparatus at the Ig genes is somehow different from the other genes. This reasoning has prompted extensive analyses of the Ig promoters and enhancers. It was found that non-Ig promoters could function in place of the Ig promoter in both somatic hypermutation [63, 64] and gene conversion [65]. Therefore, the Ig promoter is generally not considered to be the determinant for AID target specificity, although exceptions have been observed in one study [66].

The main focus for AID targeting is currently on the Ig enhancer elements. In mice, several Ig enhancers have been identified in the Ig loci [2]. These enhancers appear to have overlapping functions, since deletion of individual enhancers at the endogenous Ig loci did not appear to have a major impact on somatic hypermutation [67-70]. Similar analyses have been performed at the chicken Igλ locus in the chicken B-cell line DT40 [65, 71, 72]. The consensus from these studies is that multiple regulatory elements downstream of the Igλ constant region contribute to the targeting of AID to the locus. In particular, a 2.4-kb region (Region A) appears to have a more prominent role in this process, since deletion of this region results in severe reductions in AID-mediated gene conversion [65]. More importantly, the 2.4-kb region is capable of targeting AID function to ectopic sites outside of the Ig genes [65]. It remains to be established how this regulatory element facilitates AID function.

The complementary approach to the dissection of *cis*-acting elements is to identify the *trans*-acting factors that recruit AID to target DNA. So far, the only protein that has been shown to interact with AID is replication protein A (RPA) [36], a protein that binds single-stranded DNA, and is involved in both DNA replication and repair. RPA has been shown to stimulate AID mediated deamination of DNA in a transcription dependent manner *in vitro* [36]. The observation could be explained by the stabilization of transient unwound DNA strands during transcription, which would become substrates for AID. Although the finding could explain how AID gains access to cytidines in transcribed DNA, it is not sufficient to explain why AID is targeted to the Ig loci since RPA interacts with single-stranded DNA in a largely sequence non-specific manner. RPA has been shown to associate with switch regions during class switch recombination [73], but no specific interaction of RPA with the Ig variable region during somatic hypermutation or gene conversion has been reported. Thus, there must be other factors that could target AID function specifically to the Ig loci.

The transcription factor, E2A, may represent one such AID-targeting factor. E2A has been implicated in the function of enhancers at the Ig locus [74-77]. Importantly, ablation of E2A gene in chicken B cells leads to severe reductions in Ig gene conversion [78], whereas introduction of extra copies of the E2A-binding site, the E box, into the Ig variable region stimulates the levels of somatic hypermutation in mice [76]. Nonetheless, it remains to be demonstrated whether E2A indeed recruits AID to the Ig locus through direct association. Furthermore, the E2A-binding site, E box, is defined by a short consensus: CANNTG, which can be found in many places in the genome and is therefore unlikely to be sufficient to specify AID targets. If E2A indeed functions to target AID to the Ig locus, it most likely does so in collaboration with other factors, but the identities of such factors remain to be established.

In this regard, recent studies have implicated the NF-κB family of transcription factors in AID function, since inactivation of two NF-κB family members, p50 and c-Rel, impairs AID-mediated gene conversion [65]. At present, it is unclear whether the NF-κB proteins collaborate with E2A. Thus, the mechanisms that target AID to the Ig variable regions remain an open question.

1.3.4 Differential repair of AID mediated mutations in Ig versus non-Ig loci

Although AID activity is primarily targeted to the Ig loci, certain non-Ig loci do sustain mutations in activated B cells, albeit at levels substantially lower (at least a magnitude) relative to somatic hypermutation at the Ig loci [79-82]. These mutations in non-Ig genes bear the hallmark of AID-mediated somatic hypermutation: they are primarily localized to the promoter-proximal region and exhibit preference for AID mutation hotspots such as the RGYW motif (R=A,G; Y=C,T; W=A,T) [79-82]. Although these non-physiological targets are non-Ig genes, some of them do share certain characteristics with the Ig genes. For example, the genes encoding Bcl6, CD79A and CD79B genes are all highly expressed in activated B cells, and may share common transcription machinery with the Ig genes, which could explain the targeting of AID to these loci. At present, it is not clear why these genes would become mutation targets of AID in normal B cells. In fact, mutation of Bcl6 is frequently associated with the development of B-cell lymphomas in humans. It is not clear why B cells have not evolved a mechanism to prevent the accumulation of mutations in the *Bcl6* gene.

A recent study has comprehensively examined the mutation levels of a large number of non-Ig genes in activated B cells, and found that AID-mediated mutations are more widespread than previously thought [83]. In addition to the previously identified non-Ig targets, such as Bcl6, many more non-Ig genes in fact sustain a low level of AID-

mediated mutations during B-cell activation [83]. The mutation load of non-Ig genes becomes even higher in B cells that are deficient in the uracil DNA glycosylase and the mismatch repair system [83]. The observation suggests that the targeting of AID to the Ig loci is not complete, and AID does have access to other parts of the genome.

Most of the mutations outside of the Ig loci are normally eliminated by high fidelity repair. If this is true, why doesn't high fidelity repair prevent the accumulation of somatic hypermutation at the Ig loci? It is possible that, due to the preferential targeting of AID to the Ig loci, the high mutation loads at the Ig loci have overwhelmed the capacity of high fidelity repair. Alternatively, it is possible that the repair process at the Ig loci is corrupted in a way to favor the error prone pathway of mutation processing. Thus, the apparent mutation specificity in the Ig loci could also be a consequence of differential repair relative to non-Ig loci, and the above mentioned *cis*-acting elements and *trans*-acting factors could contribute to this mechanism of regulation as well. For example, a factor could recruit an error-prone repair enzyme instead of AID.

In summary, the mutagenic activity of AID needs to be under stringent control to prevent deleterious mutations in the genome in general. The control may involve multiple mechanisms, which include: the B-cell-specific expression of AID; regulation of AID mRNA stability by micro-RNAs; reduction of the nuclear concentration of AID by preferential export into the cytoplasm and degradation of the nuclear fraction of AID; recruitment of AID to the Ig loci; and error-prone repair of AID-induced lesions at the Ig loci. Many of these proposed regulatory mechanisms are not fully understood, and the study presented in this dissertation is designed to address this issue.

Specifically, I present evidence that two lymphoid specific transcription factors, PU.1 and IRF4, regulate AID function at two levels. I show that PU.1 is a major positive regulator of AID expression in chicken B cells, and this regulation appears to

involve direct interaction of PU.1 with the AID gene. Moreover, the dosage of PU.1 is critical for AID function, since a severe reduction in AID mediated gene conversion was observed in PU.1$^{+/-}$ cells, even though AID expression is normal in this situation. IRF4 is a functional partner of PU.1, and the two proteins can form a complex to interact with a combined Ets-IRF binding site. The connection led me to examine the role of IRF4 in the regulation of AID as well. I found that IRF4 plays a minor role in AID expression, but is important for AID-mediated gene conversion at the Igλ locus, and the function of IRF4 appears to involve two binding sites within a regulatory element in the Igλ locus. Thus, my study shows that PU.1 and IRF4 are important regulators of AID function in chicken B cells.

Chapter 2

Materials and Methods

2.1 CELL CULTURE AND TRANSFECTION

All the experiments in this study were performed in DT40 cells and its derivatives. The cells were maintained in RPMI media supplemented with 10% fetal bovine serum, 1% chicken serum, 27μM β-mercaptoethanol, 100 units/ml penicillin, and 100μg/ml streptomycin. For transfection, 10μg of linearized plasmid was electroporated (600V, 25μF) into $1x10^7$ cells. Drug selection was started a day after the transfection. For puromycin resistance, the drug was added to a final concentration of 0.5μg/ml. For *gpt* selection, mycophenolate was used at a concentration of 5μg/ml. The colonies from the transfection were analyzed by Southern blotting to identify the desired targeting events. In all the targeting experiments, the drug selection marker was subsequently removed through Cre-mediated recombination via flanking loxP sites. The Cre recombinase is expressed from a stably integrated transgene in the DT40 cell line, and its activity is inducible with tamoxifen [84]. For Cre-mediated excision of the drug selection marker, the cells were treated with 25nM of 4-hydroxy-tamoxifen for a day. Afterwards, single colonies were isolated through subcloning, and removal of the drug selection marker was confirmed with Southern analysis of the clones. In some experiments, the conditions for transfection and tamoxifen treatment were adjusted from the general protocol to fit the conditions of particular cell lines and reagents.

2.2 TARGETED MUTAGENESIS

This section describes the location of deletions described in this study. To specify the locations of the various deletions, the transcription initiation site is designated as position +1, and its sequence context is provided for reference. For the PU.1 gene, the sequence surrounding the transcription initiation site is CTCAG**A**$^{+1}$ACCT. The first and second targeting construct deletes the regions from -3655 to +750 and -187 to +750, respectively. For the IRF4 gene, the sequence surrounding the transcription initiation site of is TGACG**G**$^{+1}$AGAG. The first and second targeting construct deletes the regions from +620 to +3439 and +2837 to +3439, respectively. For the AID gene, the sequence around the transcription initiation site is ACAGA**G**$^{+1}$CATT. The PU.1-IRF4 composite binding site is from positions +305 to +315. In AID$^{+/\Delta}$ cells, the inactive allele contains a deletion from +2743 to +3502. For the Igλ gene, the sequence context of the transcription initiation site is CCATC**G**$^{+1}$GCGT. For the rearranged Igλ allele, the composite PU.1-IRF binding sites, the two PU.1-IRF4 composite binding sites are located from +4015 to +4025 and from +5412 to +5421, respectively. The region B deletion is from +6,450 to +14,124. Since this study involves a large number of targeting events, details of the constructs are omitted. The information is available upon request.

2.3 GENE CONVERSION ASSAY

The gene conversion reporter was generated by inserting an I-SceI recognition sequence into the SalI site of the puromycin resistance gene (*puro*): GTC Ggc **tag gga taa cag ggt aat** gcA CCC (the inserted sequences are in lower case letters and the I-SceI

recognition sequence is in bold). A fragment of *puro* serves as the donor in gene conversion; if the A of the start codon is designated as position +1, the fragment corresponds to +1 to +333 of the open reading frame. The *puro* fragment is positioned approximately 2-kb upstream of the promoter. The reporter is transcribed with the chicken Igλ promoter, which contains sequences from -1,363 to +25. The gene conversion substrate was integrated at position +4,007 of the rearranged Igλ locus.

For gene conversion experiments, the cells were subcloned by serial dilution in 96-well tissue culture plates. The day of subcloning was day 0. After 9 days, single colonies were transferred into 24-well tissue culture plates and cultured for 3 more days. At day 12, a portion of the cells was transferred into 12ml media containing 0.5μg/ml puromycin, while another aliquot of the cells was diluted into 12ml media without puromycin. The cell numbers used in each experiment are shown in the Supplementary Tables 1 through 3. Each 12ml of cells was distributed into a 96-well tissue culture plate. Colonies were counted under a microscope 6 to 7 days later.

2.4 DAM METHYLASTION ASSAY

The Dam methyltransferase cDNA was derived from *E. coli*. As explained in the text, a point mutation, R95A, was introduced into the gene to reduce the DNA-binding affinity of the Dam protein. The Dam gene or its fusion with PU.1 or IRF4 was expressed with a tamoxifen-inducible construct, which was stably integrated into the genome of DT40 cells. As described in detail in the text, Dam expression was induced in the presence of 25nM 4-hydroxy-tamoxifen. Genomic DNAs were isolated from cells after 24 or 48 hours of tamoxifen treatment, and analyzed by Southern blotting.

2.5 I-SCEI MEDIATED GENE CONVERSION ASSAY

The I-SceI expression construct was integrated into the AID locus of either wild-type or $IRF^{-/-/-}$ cells that contained the gene conversion substrate. The integration sites are between +2404 and +3502. The deletion eliminates exons 2 through 4 of the AID gene. The other allele of AID gene was inactivated by deletion from +2743 to +3502, which contains exons 3 and 4. I-SceI expression was induced with 25nM 4-hydroxy-tamoxifen for one day, and the cells were subcloned by serial dilution in 96-well tissue culture plates for gene conversion analysis as described in the previous section.

Chapter 3

PU.1 and IRF4 regulate AID gene expression.

3.1 BACKGROUND

Based on sequence homology, PU.1 belongs to the family of Ets domain-containing transcription factors, and is mainly expressed in hematopoietic lineage cells [85]. The Ets domain of PU.1 is responsible for its DNA-binding ability and recognizes a DNA sequence with the core GGAA motif [86-88]; in fact, it was originally isolated as binding factor for a purine-rich motif in the murine MHC class II 1-Ab gene promoter [89]. Based on genetic studies in mice, PU.1 regulates the expression of genes that are important for the development of both myeloid and lymphoid lineage cells [85, 90]. For example, PU.1 has been shown to interact with the *Il-7rα* gene promoter and to up-regulate the expression of IL-7 receptor, which is critical for the survival and proliferation of both B and T-cell precursors [91].

Because of the critical role of PU.1 in the early stage of lymphoid developments, mice deficient in PU.1 completely lack both B and T cells [92-94], and the role of PU.1 in mature B cells was addressed by the approach of conditional ablation of PU.1 in B cells [95]. The inactivation of PU.1 in the mature B cell stage had only minor effects on B-cell number in the peripheral lymphoid tissues [95]. In addition, B cells deficient in PU.1 undergo class switch recombination normally in response to antigenic challenge, and the affinity maturation of antibodies, a reflection of somatic hypermutation, is also unaffected [95]. These observations suggest that PU.1 is largely dispensable at the mature stage of B cell in mice. Similarly, no obvious phenotype was reported when the PU.1 gene was disrupted in a chicken B-cell line DT40 [96], although, in this case, a

truncated PU.1 protein was detected in the cell lysate and residual DNA-binding activity for a PU.1 consensus binding site was still detectable in electrophoresis mobility shift assays (EMSAs); the impact of the mutation on Ig gene conversion was also not examined. In spite of the apparent dispensability, PU.1 remains highly expressed in mature B cells [97, 98], and binding sites for PU.1 have been found in the enhancers of both Ig heavy and light chain loci in mice [90]. Therefore, it remains possible that PU.1 has a role in later stages of B-cell development, but its function is masked by redundant factors. Indeed, among the Ets family members, PU.1 is closely related to Spi-B, which is also expressed in B cells, especially in the more mature stages of B-cell development [99]. It is therefore possible that Spi-B is capable of compensating for the loss of PU.1 in the mature B-cell stage.

The transcription factor, Interferon Regulatory Factor 4 (IRF4), is a functional partner of PU.1 [100-103]. IRF4 is a lymphoid-specific member of the interferon regulatory factor (IRF) family, which is characterized by a helix-turn-helix type of DNA-binding motif at the N-terminus of the polypeptide [104]. PU.1 and IRF4 can form a heterodimeric complex and interact with a combined Ets-IRF binding site, which is composed of the recognition motifs for both PU.1 (GGAA) and IRF4 (GAAA) with a spacing of 2 to 3 bp [105]. Although PU.1 is capable of binding independently to the Ets-IRF motif, its association is enhanced in the presence of IRF4 [102]. On the other hand, IRF4 shows very weak DNA-binding activity by itself, and generally requires collaboration with PU.1 [100] or other factors, such as E2A [106], for effective interaction with DNA. The synergy between PU.1 and IRF4 is beyond DNA binding, and is exhibited at the level of transcriptional activation as well, as demonstrated in transfection studies [100].

IRF4 is expressed throughout B-cell development as well as in mature T cells [107]. Unlike other members of the IRF family, IRF4 is induced by B-cell receptor stimulation, but not by viral infection, suggesting a function independent of interferon induction [107]. Genetic studies indicate that IRF4 is crucial for the maturation and function of B cells [108]. In IRF4-deficient mice, the early B-cell development is normal, but B-cell maturation and function in the peripheral lymphoid tissues are impaired [108]. The IRF4-deficient mice have severely reduced levels of immunoglobulin in the serum; IRF4-deficient B cells do not respond properly to antigen stimulation and fail to proliferate and secrete antibodies [108]. Due to the maturation defect of B cells, it is not feasible to determine the potential role of IRF4 in class switching and somatic hypermutation in these mice.

To circumvent this problem, Klein and colleagues used a conditional knockout strategy to ablate the IRF4 gene during B-cell activation, and found a defect in class switch recombination in IRF4-deficient B cells [41]. Furthermore, they showed that AID expression is reduced in IRF4-deficient B cells, which can explain the defect in class switching [41]. A similar conclusion was reached in another study [40]. Thus, IRF4 appears to be a positive regulator of AID expression. It remains to be determined whether IRF4 directly activates AID gene transcription.

My interest in PU.1 and IRF4 stems from the presence of a consensus Ets-IRF combined binding site in an enhancer element at the chicken Igλ locus. This Ets-IRF binding site is interspersed with putative binding sites for other transcription factors implicated in Ig enhancer function such as NF-κB, E2A, as well as Mef2, and, more importantly, the arrangement of these binding sites is conserved among chicken, mouse and human Igλ enhancers, suggesting a functional relevance. This observation, coupled with information from studies in mice implicating PU.1 and IRF4 in various stages of B-

cell development and function, prompted me to examine the roles of these two factors in AID-mediated gene conversion in chicken B cells. Toward this end, I inactivated the PU.1 and IRF4 genes in the chicken B cell line DT40 and analyzed the effects of the mutation on AID function.

3.2 RESULTS

3.2.1 PU.1 is important for AID expression in chicken B cells.

My study was performed in the chicken B-cell line DT40, which was derived from virally transformed bursal B cells [109]. The cell line maintains AID-mediated Ig gene conversion in culture [110, 111]. A key advantage of this cell line is that homologous recombination is very efficient, which enables genetic manipulation of the genome in homologous targeting experiments [112]. Taking advantage of this property, we knocked out the PU.1 gene in DT40 cells to evaluate the role of PU.1 in AID function. To inactivate the PU.1 gene, we deleted the first exon together with the upstream promoter region (Fig 3.1*A*). Two different targeting constructs were used to inactivate the two alleles of PU.1 gene (Fig 3.1*A*). The strategy is based on our experience that the second targeting event tends to take place on the first targeted allele if the same targeting construct is used. In our approach, one of the homology arms of the second targeting construct is within the deleted region of the first targeted allele. Therefore, the second targeting event is limited to the remaining wild-type allele.

Using this method, we mutated both alleles of the PU.1 gene, and no PU.1 mRNA is detectable in the homozygous knockout cell line (Fig 3.1*B*). PU.1-deficient cells grow slower than wild-type cells (data not shown), presumably because certain PU.1 target genes are involved in cell proliferation. AID expression was examined in PU.1

deficient cells, and found an approximately 10-fold reduction in AID mRNA levels (Fig 3.1*C*,*D*). This result shows that PU.1 is important for AID expression in DT40 cells.

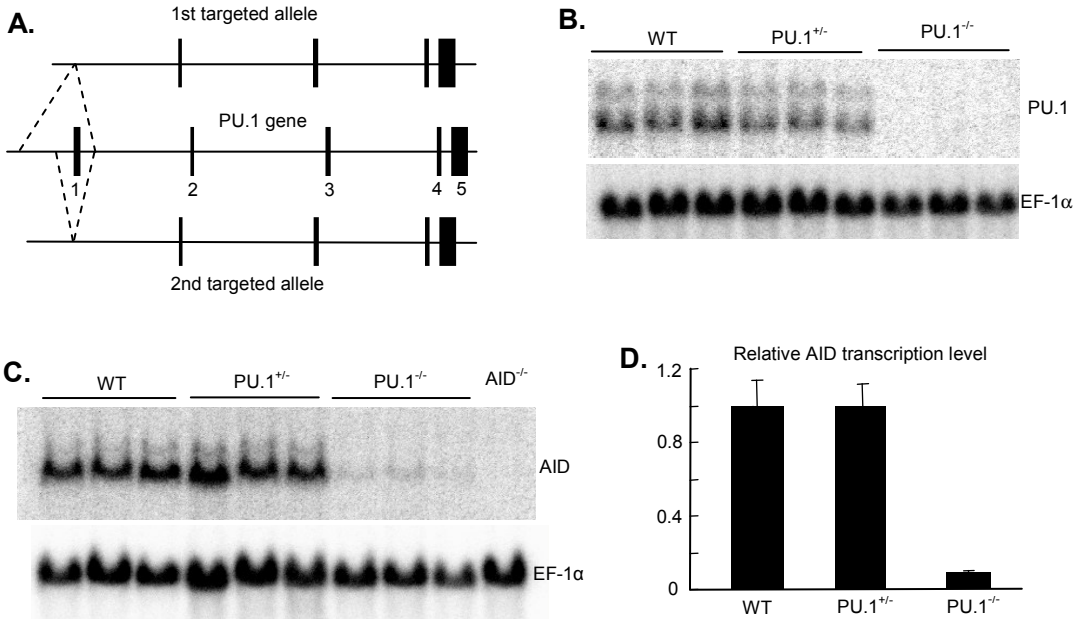


Figure 3.1. AID expression is strongly reduced in PU.1-deficient cells. *A*. This diagram illustrates the strategy for inactivating the PU.1 gene. The exon-intron organization of the PU.1 gene is drawn to scale, and the exon numbers are shown. The deleted region is shown for the targeted alleles, but details of the targeting constructs are omitted. The drug-selection marker was removed after each targeting event, and a loxP site was left in place of the deleted region. *B*. PU.1 mRNA levels were analyzed with Northern blotting. Each genotype is represented with three samples. The top autoradiograph was hybridized with a PU.1 cDNA probe. In the bottom autoradiograph, the membrane was rehybridized with an EF-1α cDNA probe to normalize RNA loading. *C*. AID mRNA levels were analyzed by Northern blotting. Except for $AID^{-/-}$ cells, each genotype is represented with three samples. WT is wild-type cells. The top autoradiography was hybridized with an AID cDNA probe. In the bottom autoradiograph, the membrane was rehybridized with an EF-1α cDNA probe to normalize for RNA loading. *D*. The histogram is the quantification of AID mRNA levels as analyzed by Northern blotting in *C*. The value for each genotype is derived from three samples, and the error bar represents standard deviation. The value for wild-type cells is set as 1, whereas $AID^{-/-}$ cells represent the base line.

3.2.2 IRF4 plays a minor role in AID expression in chicken B cells.

Since IRF4 is a functional partner of PU.1, it is conceivable that loss of IRF4 would have a similar impact on AID expression. To test this hypothesis, we also inactivated the IRF4 gene in DT40 cells. This experiment is complicated by the fact that the gene resides on chromosome 2, which is present in three copies in DT40 cells. For this reason, we performed three rounds of targeting with two different constructs (Fig. 3.2*A*). The first targeting construct deletes exons 2 through 5, which include the translation start codon and encode the DNA-binding domain. Moreover, splicing of exon 1 to exon 6 leads to a frame-shift mutation. For the second and third alleles, the construct deletes part of exon 5 together with the splice donor at the 3' end of the exon. Under this situation, exon 4 could potentially be spliced to exon 6, which generates a frame-shift mutation in exon 6.

Using this strategy, we have targeted all three alleles of the IRF4 gene in DT40 cells. In IRF4$^{-/-/-}$ cells, residual IRF4 mRNA is detectable with Northern analysis, and may correspond to the truncated transcripts from the targeted alleles (Fig. 3.2*B*). Since exon 5 is 143-bp, the truncated transcript is practically indistinguishable from the wild-type mRNA on Northern blots. To ascertain the nature of this transcript, we performed RT-PCR with primers that amplify the complete open reading frame of IRF4. Sequencing of the RT-PCR products confirmed the expected loss of exon 5 and the splicing of exon 4 to exon 6 (data not shown). This transcript could potentially encode a truncated protein containing residues 1-147 of IRF4, which includes the DNA-binding domain, but lacks the transactivation function and could potentially act as a dominant negative. However, the truncated message is present at less than 1/10th the level of normal transcripts, presumably due to nonsense-mediated RNA decay. The small amount of truncated protein, if indeed produced, should have no significant impact on the

function of partner proteins such as PU.1, and the IRF$^{-/-/-}$ cells appear to be devoid of functional IRF4 proteins. The proliferation of IRF4-deficient cells is slower relative to wild-type cells, although the defect is less severe compared to PU.1-deficient cells

The expression of AID was examined in IRF4-deficient cells, and AID mRNA levels were reduced by approximately 40% on average (Fig 3.2*C*). AID expression is variable among different subclones of IRF4-deficient cells, ranging from 30% to approximately 70% of wild-type levels. The reason for this phenomenon is unclear; one potential explanation is that loss of IRF4 renders the chromatin structure of the AID promoter unstable. Thus, relative to the more severe defect in AID expression observed in PU.1 deficient cells, IRF4 appears to be less important for AID expression.

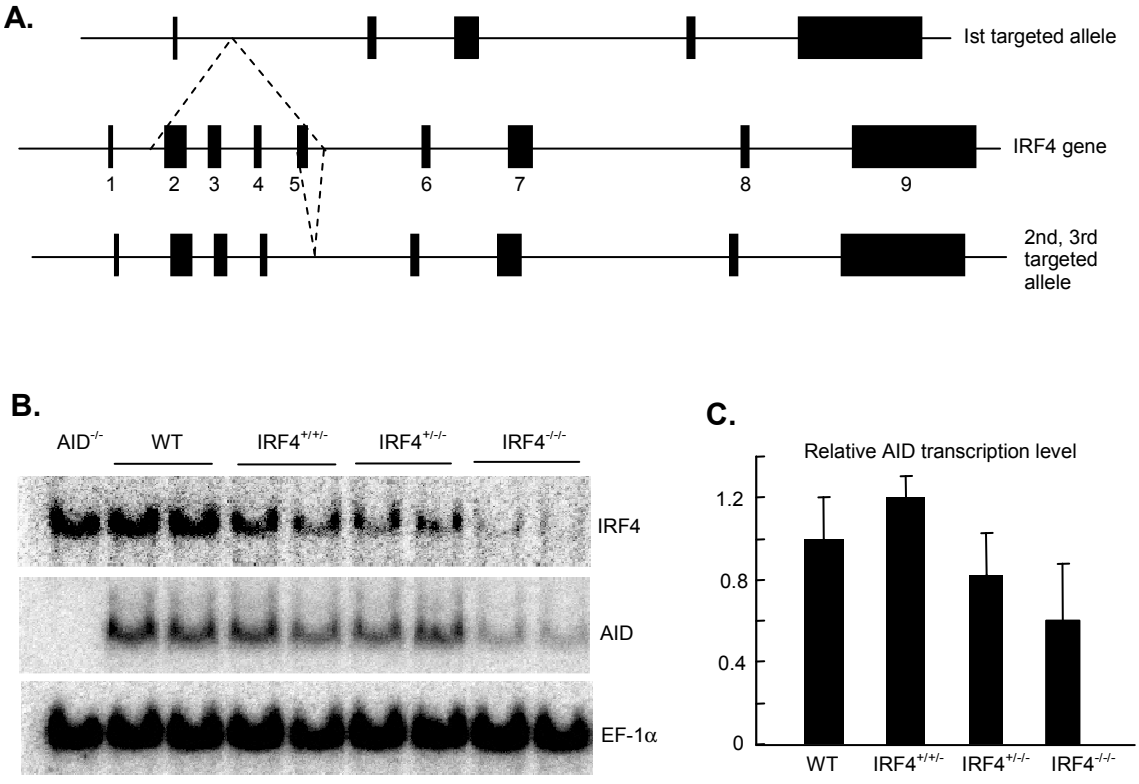


Figure 3.2. AID expression is reduced in IRF4-deficient cells. *A*. The strategy to ablate the IRF4 gene is illustrated in the same way as for PU.1 (Fig. 3.1*A*). *B*. The mRNA levels of IRF4 and AID were analyzed with Northern blotting. The membrane is sequentially hybridized with IRF4, AID and EF-1α cDNA probes, respectively. Except for $AID^{-/-}$ cells, each genotype is represented with two samples. *C*. The histogram is the quantification of AID mRNA levels as analyzed by Northern blotting in *B*. The value for each genotype is derived from at least three samples, and the error bar represents standard deviation. The value for wild-type cells is set as 1, whereas the $AID^{-/-}$ cells provide the baseline.

3.2.3 Mutation of a PU.1-IRF-binding site in AID gene reduces AID expression.

PU.1 and IRF4 are capable of interacting cooperatively with the Ets-IRF composite element (EICE). Although the distinct phenotypes of PU.1 and IRF4 knockouts suggests separate roles of the two factors in regulating AID expression, it remains possible that the collaboration of these two factors contribute, at least in part, to the regulation of AID expression. Indeed, a consensus EICE is found in the AID gene (Fig. 3.3*A*), and the element is 350-bp downstream of the transcription initiation site, a location consistent with a role in transcriptional regulation.

To test its function, the EICE was deleted through homologous recombination (Fig. 3.3*A*). The mutation was introduced into a cell line in which one allele of AID gene has been inactivated by the deletion of exon 3 and 4 (AID-Δ). The inactive allele produces low levels of a truncated transcript (AID-Δ), which is distinguishable from the wild-type mRNA (AID-WT) on Northern blots (Fig. 3.3*B*). The EICE was mutated on the wild-type allele of AID gene ($AID^{m/\Delta}$), and, relative to $AID^{+/\Delta}$cells, the mutation reduces the level of wild-type AID transcript by approximately 50% (Fig. 3.3*B*), similar to the phenotype of IRF4 knockout. Variability in AID expression was not observed as in IRF4-deficient cells, and the reason for this difference is not clear. This result

suggests that PU.1 and IRF4 complex may function through this binding site, but the major part of PU.1 function in AID expression is independent of this element and IRF4.

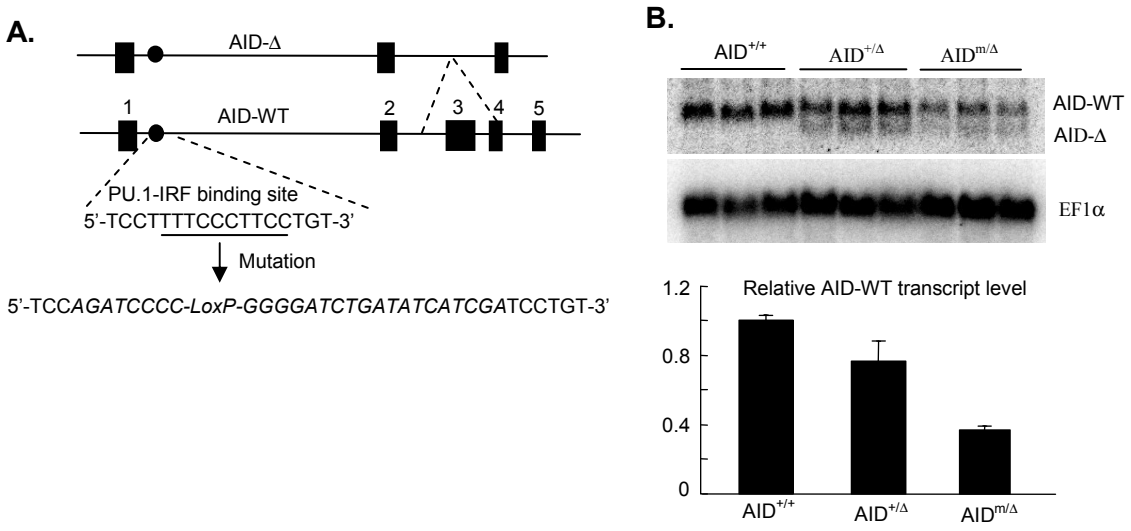


Figure 3.3. Mutation of a PU.1-IRF composite binding site near AID promoter. *A*. This diagram shows the position and sequence of the PU.1-IRF composite binding site within the AID gene. The sequence after mutation is shown below. The foreign sequences are italicized. The exon-intron organization of AID gene (AID-WT) is drawn to scale, and exon numbers are shown. The diagram for AID-Δ illustrates the inactive allele of AID, in which exons 3 and 4 are eliminated. *B*. The PU.1-IRF binding site mutant is analyzed with Northern blotting. The genotypes are indicated at the top, and "m" in the superscript represents the binding site mutation in the wild-type ($AID^{m/\Delta}$) allele of the AID gene. The top autoradiograph was hybridized with an AID cDNA probe, which revealed two bands. The upper band corresponds to the wild-type AID transcript (AID-WT), while the lower band corresponds to the transcript from the exon 3-4 deleted allele (AID-Δ). The same membrane was rehybridized with an EF-1α cDNA probe to normalize RNA loading. The relative AID-WT level is plotted in the histogram, and the value for $AID^{+/+}$ cells is set as 1. Each genotype is represented by three samples and the error bar shows standard deviation.

3.2.4 PU.1 and IRF4 interact with the AID gene.

The Dam methylation assay [113] was adapted to examine direct interaction of PU.1 and IRF4 with the AID gene in DT40 cells. The assay is based on the Dam

methyltransferase from *E. coli*, which methylates adenine in the context of GATC. There is no equivalent enzyme in eukaryotic cells. Introduction of Dam into eukaryotic cells leads to methylation of GATC motif throughout the genome. When Dam is fused to a DNA-binding protein, the fusion protein preferentially methylates GATC in the region associated with the DNA-binding protein, and this property has been employed successfully to study protein-DNA interactions *in vivo* [114]. In comparison with Chromatin Immuno-Precipitation (ChIP), the advantage of the Dam methylation assay is that it does not require an antibody against the protein of interest. Additionally, methylation within GATC motif can be conveniently detected and quantified with methylation-sensitive restriction enzymes. The drawback of this method is that Dam could potentially affect the DNA-binding property of its fusion partner.

In initial trial of the methylation assay, the Dam-PU.1 fusion protein methylated the AID gene with similar efficiency as control loci, such as EF-1α (data not shown). The outcome is not particularly surprising, since Dam is expected to methylate GATC independent of the fusion partner, and the non-targeted methylation activity may be so high as to obscure the effect of recruitment by PU.1. To solve this problem, we introduced a point mutation, R95A, into the Dam gene. According to the crystal structure of Dam/DNA complex, R95 contacts the phosphates of target DNA, and stabilizes the interaction of Dam with cognate DNA in a base non-specific manner [115]. It is possible that changing R95 to a non-basic residue would reduce the DNA binding affinity of Dam without altering its target specificity. Such attenuated Dam would be more dependent on a fused DNA-binding protein to interact with DNA, thus improving the specificity of the assay.

To test this idea, we generated the Dam(R95A) mutant and evaluated its activity by complementing a *dam*$^-$ *E. coli* strain. Consistent with expectation, the mutation

substantially reduced the methylation activity of the enzyme, and the mutant Dam remains specific for GATC motif (data not shown). This result encouraged assessment of the performance of this attenuated Dam in methylation assays in DT40 cells in the hope of improving the specificity of the method. The attenuated Dam(R95A) was used in all the following experiments, and the mutation (R95A) is omitted for simplicity.

The DNA-binding domain of PU.1 is at the C-terminal part of the polypeptide. In the fusion, the Dam enzyme was fused to the N-terminus of PU.1 (Dam-PU.1) to minimize potential interference with the DNA-binding activity of PU.1. For IRF4, since the DNA-binding domain is at the N-terminal region of the polypeptide, Dam was linked to the C-terminus of IRF4 for the same reason. Free Dam serves as the control for untargeted methylation activity in the experiments. The cDNAs were expressed in DT40 cells with a tamoxifen-inducible system (Fig. 3.4*A*). In the expression construct, the Dam or Dam fusion cDNAs are positioned downstream of the *gpt* drug selection marker, which was used for the selection of stable clones. The *gpt* cDNA is positioned between the promoter and the Dam gene, and precludes the expression of Dam. To initiate methylation, the cells were treated with tamoxifen, which activates the Cre recombinase produced from a transgene in the DT40 cell line. The Cre recombinase catalyzes the excision of the *gpt* cDNA via flanking loxP sites, and the Dam cDNA is consequently positioned immediately downstream of the promoter for expression. After 24 or 48 hours of tamoxifen treatment, genomic DNAs were isolated, and the methylation pattern was analyzed by restriction digests with MboI, which only cleaves unmethylated GATC (Fig. 3.4*B-D*).

The methylation status of the AID gene was compared with two housekeeping genes, EF-1α and hsc70, which are not regulated by PU.1 or IRF4. For the AID gene, a 1-kb restriction fragment that contains the promoter as well as the PU.1-IRF4 binding site

was examined, and the comparable promoter regions of EF-1α and hsc70 gene were analyzed as controls. In cells containing free Dam, all three genes are methylated and the extent of methylation correlates with the duration of Dam expression. The hsc70 probe cross-hybridizes to an unknown fragment (* in Fig. 3.4*D*), which could be derived from a related heat shock gene family member. This unknown fragment also appears accessible to Dam methylation as indicated with the generation of a longer fragment in response to Dam expression (the upper band of *, Fig. 3.4*D*). Thus, free Dam methylates these loci in a largely indiscriminate fashion as expected.

A different pattern was observed for Dam-PU.1 fusion, which preferentially methylates the AID 5' region (Fig. 3.4*B*) relative to EF-1α (Fig. 3.4*C*), hsc70 and the cross-hybridizing fragment (Fig. 3.4*D*). We believe methylation pattern reflects specific association of PU.1 with the AID gene, in consistent with its role in AID expression. The IRF4-Dam fusion methylates the AID locus (Fig. 3.4*B*), but not the hsc70 locus (Fig. 3.4*D*). However, the fusion also methylates the EF-1α locus (Fig. 3.4*C*) and the hsc cross-hybridizing fragment (Fig. 3.4*C*) to a substantial degree. The high background makes it difficult to draw a firm conclusion regarding the specificity of IRF4 toward the AID locus.

If Dam-PU.1 is recruited to the AID locus, why does the fusion protein methylate AID gene less efficiently than the free Dam (Fig. 3.4*B*)? The apparent paradox could be explained if the intrinsic methylation activity of Dam is reduced in the context of the fusion. The point can be illustrated with the methylation pattern at the EF-1α and hsc70 loci. Methylation at these loci is expected to reflect the intrinsic activity of Dam or its fusions, and the free Dam is clearly more active than the fusion proteins. Given this consideration, the relevant comparison in this assay is among different loci for one Dam derivative.

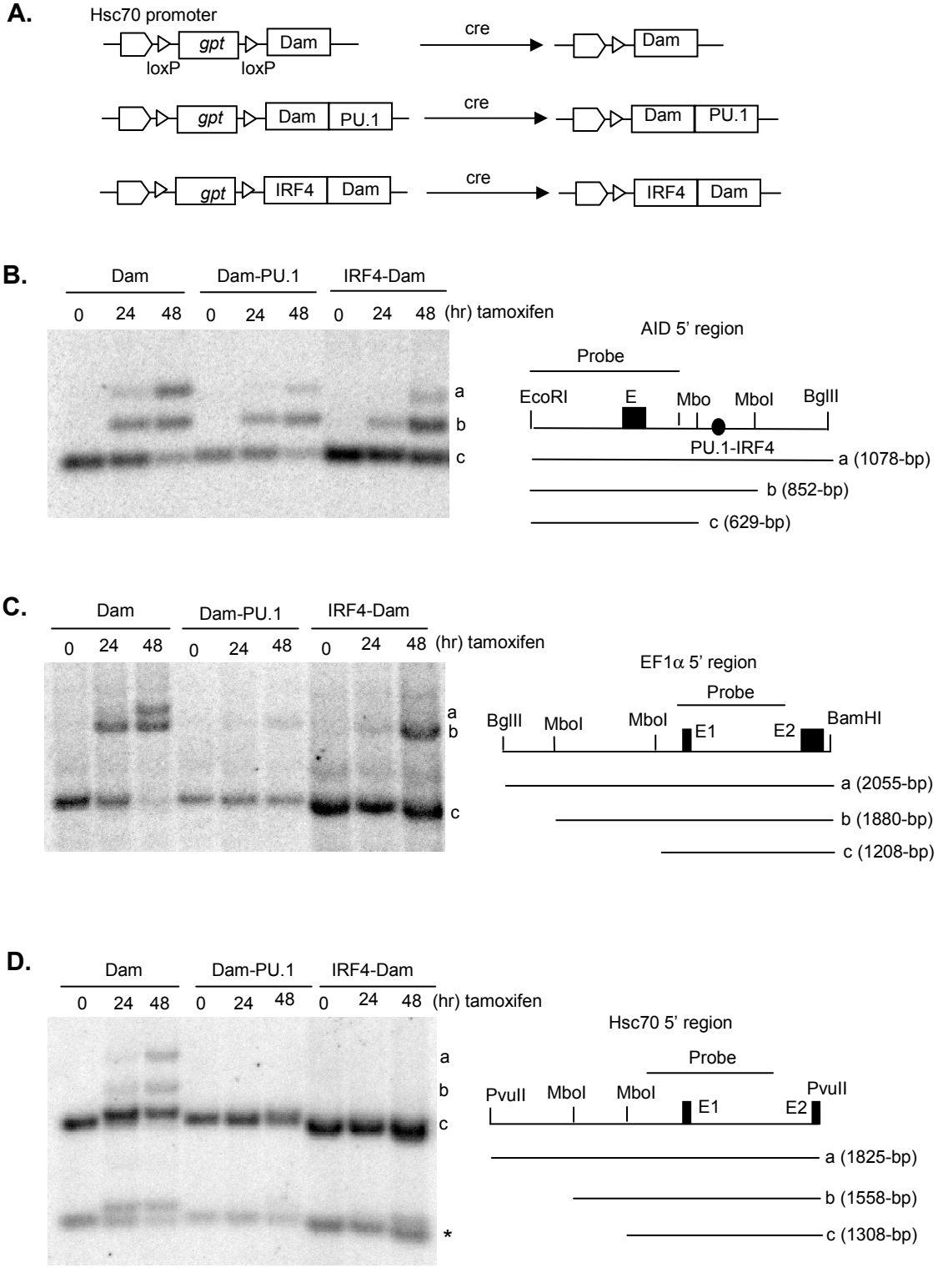

A.
Hsc70 promoter
gpt
Dam
loxP
loxP
cre
Dam
gpt
Dam
PU.1
cre
Dam
PU.1
gpt
IRF4
Dam
cre
IRF4
Dam
B.
Dam
Dam-PU.1
IRF4-Dam
0 24 48 0 24 48 0 24 48 (hr) tamoxifen
a
b
c
AID 5' region
Probe
EcoRI
E
Mbo
MboI
BglII
PU.1-IRF4
a (1078-bp)
b (852-bp)
c (629-bp)
C.
Dam
Dam-PU.1
IRF4-Dam
0 24 48 0 24 48 0 24 48 (hr) tamoxifen
a
b
c
EF1α 5' region
Probe
BglII
MboI
MboI
E1
E2
BamHI
a (2055-bp)
b (1880-bp)
c (1208-bp)
D.
Dam
Dam-PU.1
IRF4-Dam
0 24 48 0 24 48 0 24 48 (hr) tamoxifen
a
b
c
*
Hsc70 5' region
Probe
PvuII
MboI
MboI
E1
E2
PvuII
a (1825-bp)
b (1558-bp)
c (1308-bp)

Figure 3.4. Methylation assay for the interaction of PU.1 and IRF4 with the AID gene. *A*. This diagram illustrates the tamoxifen-inducible expression constructs for Dam and Dam fusions. *B*. The autoradiograph is Southern analysis of the methylation pattern at the AID 5' region. The source of the DNA and tamoxifen induction time is indicated at the top of the autoradiograph. The identity and the expected size of the restriction fragments are shown to the right of the blot. E1 represents the first exon of AID gene. The PU.1-IRF composite binding site is represented with a black circle. The location of the probe for the Southern analysis is indicated above the restriction map. *C*. The autoradiograph shows Southern analysis of the methylation pattern at the EF-1α 5' region. All the labels of this panel are analogous to *B*. *D*. The autoradiograph is Southern analysis of the methylation pattern at the hsc70 5' region. The symbol "*" marks a cross-hybridizing restriction fragment of unknown origin. All the labels of this panel are analogous to *B*.

3.3 DISCUSSION

The experiments presented in this chapter show that both PU.1 and IRF4 are involved in regulating AID gene expression, but PU.1 plays a more prominent role relative to IRF4. Based on the Dam methylation analysis, PU.1 likely interacts directly with the AID gene to regulate its expression. The role of a consensus PU.1-IRF binding site was examined near the promoter of AID gene, and found that the binding site mediates only a minor part of the PU.1 function. Further studies are needed to identify other binding sites for PU.1 in the AID gene.

PU.1 is capable of binding to the GGAA motif by itself, and such motifs are found in AID genes. However, the short consensus is widespread in the genome, and it seems unlikely that the GGAA motif alone would be sufficient for specifying the location of PU.1 function. Most likely, the function of PU.1 involves collaboration with factors other than IRF4, but the identity of such novel cofactors remains to be established. One candidate factor is IRF8, which is the closest relative of IRF4 and is expressed in DT40 cells (data not shown). IRF8 can also form a complex with PU.1 and binds to the EICE site in target gene [105]. Moreover, IRF8 has also been observed to bind to a novel type of composite Ets-IRF binding site, where the order of the PU.1 and IRF motifs are

reversed relative to the typical EICE, and such kind of binding site has been named IECS [105]. There is indeed such an IECS motif near the promoter of the AID gene, but its function remains to be evaluated in future experiments. Additionally, inactivation of the IRF8 gene in DT40 cells would help to test the role of this factor in regulating AID expression as well as other aspects of AID function.

PU.1 may also function with factors other than IRF4 and IRF8, and the identification of PU.1 interaction proteins via affinity purification or yeast two hybrid approaches could potentially shed light on this issue. Once we have identified any candidate factor, we could evaluate its function by genetic experiments as presented in this chapter. Our data also cannot rule out an indirect role of PU.1 in regulating AID expression, since PU.1 has been shown to regulate the expression of many genes involved in hematopoietic cell differentiation. Microarray analysis of PU.1-deficient cells relative to wild-type cells would be informative in this respect.

Our results have revealed a clear difference in the regulation of AID expression between chicken and mouse B cells. As described earlier, conditional ablation of PU.1 in mouse B cells does not affect AID-mediated class switching and somatic hypermutation; on the other hand, loss of IRF4 in mature B cells leads to an obvious reduction in AID expression. The situation is completely different in chicken B cells, where PU.1 appears to be more important than IRF4 with respect to the expression of AID. Several other transcription factors have been implicated in the regulation of AID expression in mouse B cells, including Pax5, STAT6, NF-κB, and HoxC4. However, the binding sites for these factors are not conserved between chicken and mouse AID promoters, and loss of NF-κB family members, p50 and c-Rel, has no impact on AID expression [65]. Most likely, the regulation of AID expression is different between

chicken and mouse, which is consistent with the fact that bursa is an organ unique to avian species with no mammalian counterpart.

Chapter 4

IRF4 and PU.1 are important for AID-mediated gene conversion.

4.1 RESULTS

4.1.1 Gene conversion defect in IRF4-deficient DT40 cells.

PU.1 and IRF4 may have roles beyond the regulation of AID expression. As described earlier, a consensus PU.1-IRF binding site was found in an enhancer at the Igλ locus, and this binding site is conserved in the Igλ locus in the mouse and human genomes. It is therefore possible that the two proteins also facilitate AID function at the Igλ locus, potentially by recruiting AID to the locus to mediate gene conversion. Thus, it would be of interest to examine AID-mediated gene conversion activity in PU.1 and IRF4 deficient cells. This experiment is not feasible in PU.1-deficient cells since AID expression is very low, and gene conversion is almost certainly defective. On the other hand, the reduction of AID expression in IRF4-deficient cells is relatively mild, which enables us to test the impact of IRF4 deficiency on AID function in these cells.

For this purpose, a gene conversion substrate was generated to determine gene conversion activity in IRF4-deficient cells (Fig. 4.1*A*). The substrate reporter is derived from the puromycin-resistance gene (*puro*) that is inactivated by the insertion of an oligonucleotide (see Material and Methods and Fig. 4.1*A*). The defective *puro* can be rescued through gene conversion, using an upstream fragment of *puro* as the donor template, and the percentage of puromycin-resistant cells represents gene conversion frequency [65]. In many studies, change in surface IgM expression, as measured with flow cytometry, was used as an assay for immunoglobulin gene conversion. Relative to

flow cytometry, the drug selection method is more sensitive, which enables quantification of low frequencies of gene conversion in mutant backgrounds.

The gene conversion substrate was integrated into a location downstream of the C region of the Igλ locus (for exact position of the integration site, see Materials and Methods) in both wild-type and IRF4-mutant cells. In a previous report, the substrate was shown to undergo AID-mediated gene conversion when integrated at this location [65]. An obvious decrease in gene conversion activity was observed in $IRF4^{-/-/-}$ cells relative to wild-type cells (compare f, g, h with a, Fig. 4.1*B*). Moreover, the dosage of IRF4 gene appears to influence gene conversion activity, which is noticeably reduced in cells with one copy of the IRF4 gene (d and e, Fig. 4.1*B*). Since AID function requires active transcription of the target DNA, the transcription status of the reporter was examined; no difference between wild-type and IRF4 deficient cells was observed (a, b, d, and f in Fig. 4.1*C*). Therefore, IRF4 regulates AID-mediated gene conversion rather than transcription of AID target gene in a dosage-dependent manner.

As observed earlier, AID expression is variable among different subclones of IRF4-deficient cells. Similarly, we observed substantial variations in gene conversion activity among the three $IRF4^{-/-/-}$ clones in gene conversion assays, and this variation seems to correlate with the level of AID expression among these subclones. However, the severe gene conversion defect of the IRF4-deficient cells is unlikely attributable entirely to the reduction in AID expression. In one of the subclone, AID is expressed at approximately 60% the level of wild-type cells, which is comparable to that in $AID^{+/-}$ cells. The gene conversion activity in the $AID^{+/-}$ cells is reduced by about 2-fold relative to wild-type cells (i in Fig. 4.1*B*), which is less severe compared to that observed in the IRF4-deficient clone (10-fold) with comparable AID expression levels.

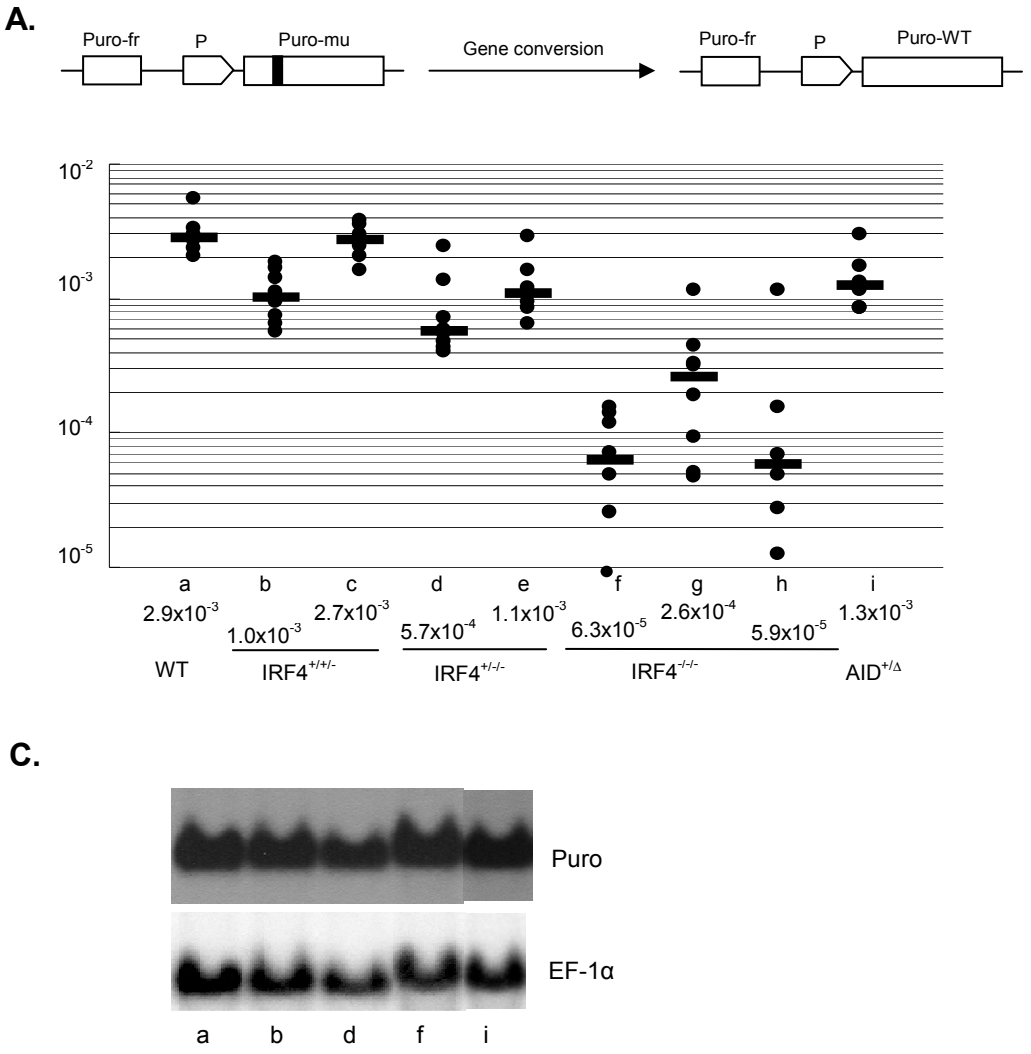


Figure 4.1. IRF4 deficiency impairs AID mediated gene conversion. *A*. This diagram illustrates the gene conversion substrate. Puro-fr represents the fragment of puromycin resistance gene that serves as the donor in gene conversion. Puro-mu represents the mutated *puro* reporter before gene conversion, whereas Puro-WT represents the wild-type *puro* reporter after gene conversion. *B*. The dot plot presents the gene conversion analysis of cells with different copy numbers of IRF4 genes. Each IRF4 mutant is represented with two or three clones ($IRF^{+/+/-}$: b and c; $IRF^{+/-/-}$: d and e; $IRF^{-/-/-}$: f, g, and h). Each dot represents result from one experiment (for colony numbers, see Table 1). In two experiments for clone f, no puromycin-resistant colonies were observed, and these data points are shown below the baseline of the plot. The bar represents the median GC frequency of a data set of at least six experiments, and the value is written below the plot. *C*. The transcript levels of the *puro* reporter in various cell lines are analyzed with Northern blotting. The same blot was rehybridized with an EF-1α cDNA probe to normalize RNA loading.

Table 1. Gene conversion analysis of the IRF4 mutant cells.

Genotype	Puro		Control		GC
	a	b	c	d	e
Wild-type	28	$2x10^4$	68	100	$2.1x10^{-3}$
	42		70		$3.0x10^{-3}$
	20		37		$2.7x10^{-3}$
	45		66		$3.4x10^{-3}$
	34		72		$2.4x10^{-3}$
	61		54		$5.7x10^{-3}$
$IRF4^{+/+/-}$ (Clone 1)	54	$1x10^5$	38	100	$1.4x10^{-3}$
	72		64		$1.1x10^{-3}$
	114		60		$1.9x10^{-3}$
	99		59		$1.7x10^{-3}$
	45		47		$9.6x10^{-4}$
	29		39		$7.4x10^{-4}$
	47		72		$6.5x10^{-4}$
	40		70		$5.7x10^{-4}$
$IRF4^{+/+/-}$ (Clone 2)	123	$1x10^5$	50	100	$2.5x10^{-3}$
	200		56		$3.6x10^{-3}$
	112	$5x10^4$	75		$3.0x10^{-3}$
	106		55		$3.9x10^{-3}$
	67		65		$2.1x10^{-3}$
	56		67		$1.7x10^{-3}$
$IRF4^{+/-/-}$ (Clone 1)	129	$1x10^5$	44	100	$2.9x10^{-3}$
	55		63		$8.7x10^{-4}$
	54		83		$6.5x10^{-4}$
	49		40		$1.2x10^{-3}$
	49		30		$1.6x10^{-3}$
	63		66		$9.6x10^{-4}$
$IRF4^{+/-/-}$ (Clone 2)	98	$1x10^5$	70	100	$1.4x10^{-3}$
	22		45		$4.9x10^{-4}$
	23		56		$4.1x10^{-4}$
	24		54		$4.4x10^{-4}$
	164		65		$2.5x10^{-3}$
	25		42		$6.0x10^{-4}$
	33		60		$5.5x10^{-4}$
	51		69		$7.4x10^{-4}$

Genotype	Puro		Control		GC
	a	b	c	d	e
$IRF4^{-/-/-}$ (Clone 1)	0	$2x10^5$	41	200	0
	12		85		$1.4x10^{-4}$
	3		84		$7.1x10^{-5}$
	4		68		$1.2x10^{-4}$
	1		77		$2.6x10^{-5}$
	0		70		0
	6		75		$1.6x10^{-4}$
	1		40		$5.0x10^{-5}$
$IRF4^{-/-/-}$ (Clone 2)	4	$2x10^5$	39	100	$5.1x10^{-5}$
	8		42		$9.5x10^{-5}$
	51		22		$1.2x10^{-3}$
	3		31		$4.8x10^{-5}$
	26		39		$3.3x10^{-4}$
	18		46		$2.0x10^{-4}$
	48		53		$4.5x10^{-4}$
	23		36		$3.2x10^{-4}$
$IRF4^{-/-/-}$ (Clone 3)	2	$5x10^4$	25	100	$1.6x10^{-4}$
	1		41		$4.9x10^{-5}$
	1		29		$6.9x10^{-5}$
	1	$2x10^5$	40		$1.3x10^{-5}$
	89		38		$1.2x10^{-3}$
	3		54		$2.8x10^{-5}$
$AID^{+/\Delta}$	58	$1x10^5$	43	100	$1.4x10^{-3}$
	51		44		$1.2x10^{-3}$
	53		61		$8.7x10^{-4}$
	133		43		$3.1x10^{-3}$
	86		48		$1.8x10^{-3}$
	56		66		$8.5x10^{-4}$

a: puromycin (puro)-resistant colony number; b: total number of cells distributed into media containing puromycin; c: colony number in media without puromycin (control); d: total number of cells distributed into media without puromycin; e: gene conversion frequency (GC) is calculated by (a/b)/(c/d).

To test whether the IRF4 function is specific for AID, the effect of IRF4 deficiency on I-SceI mediated gene conversion was examined. I-SceI is a homing endonuclease that is capable of introducing a double-stranded DNA (dsDNA) strand break within an 18-bp recognition sequence. The puro reporter in the substrate contains an I-SceI recognition sequence, and cleavage at this site by I-SceI ought to induce gene conversion that is independent of AID [116]. An I-SceI expression construct (Fig. 4.2*A*) was introduced into both wild-type and IRF4-deficient cells that contain the gene conversion substrate. Both alleles of the AID gene were inactivated in these cells (Fig. 4.2*B*) so that gene conversion relies exclusively on I-SceI. In fact, the I-SceI expression construct was integrated into the AID gene. This way, both wild-type and IRF4-deficient cells contain a single copy of the I-SceI transgene at the same genomic location, which would ensure comparable levels of I-SceI expression in the two cell types. The expression of I-SceI cDNA was controlled with the same tamoxifen-inducible system employed for the methylation experiments (Fig. 4.2*A*). The inducible system prevents gene conversion before tamoxifen induction (Table 2). Addition of tamoxifen to the cells readily induced I-SceI expression (Fig. 4.2*B*), and similar levels of gene conversion was observed in wild-type and IRF4-deficient cells (Fig. 4.2*C*). Thus, IRF4 is dispensable for I-SceI mediated gene conversion, and it is specifically required for facilitating AID function, presumably in generating strand breaks in target DNA.

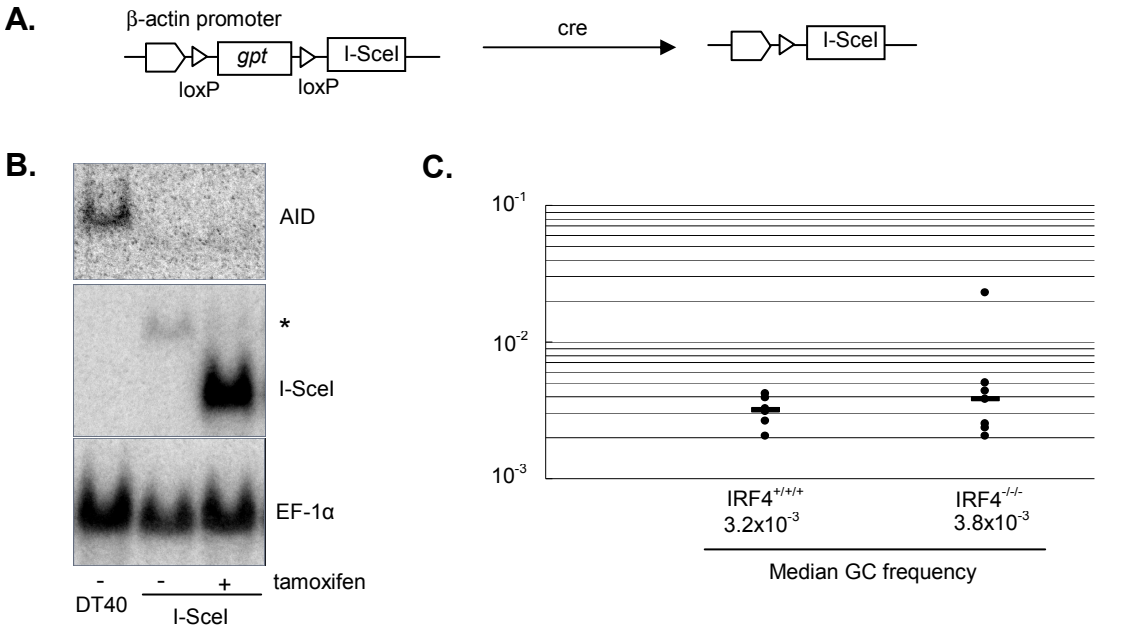


Figure 4.2. IRF4 is dispensable for I-SceI-mediated gene conversion. *A*. The diagram illustrates the tamoxifen inducible expression construct for I-SceI. *B*. The top panel of Northern blotting shows that AID is not expressed in the cells containing the I-SceI expression construct. DT40 cells without the construct serves as control. The middle panel of Northern blotting shows that I-SceI expression is induced with tamoxifen. The band marked with * is most likely a readthrough transcript due to incomplete polyadenylation downstream of the *gpt*. The bottom panel of Northern blot is hybridized with an EF-1α cDNA probe to normalize RNA loading. *C*. The dot plot compares the frequency of I-SceI-mediated gene conversion in wild-type versus IRF deficient cells. The bar indicates the median GC frequency of a data set of seven experiments, and the value is written below the plot (for colony numbers, see Table 2).

Table 2. I-SceI mediated gene conversion in IRF4 deficient cells.

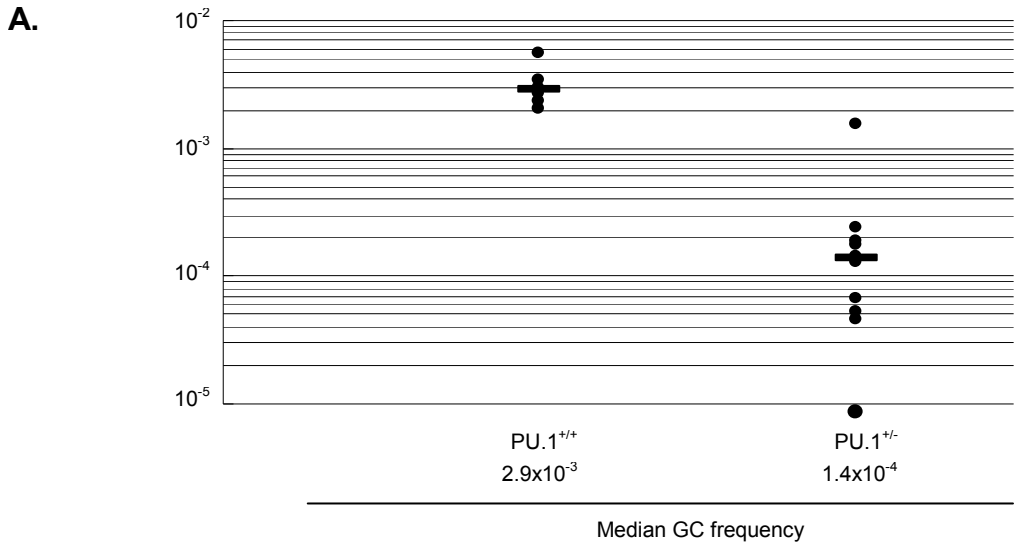
Genotype	Puro		Control		GC
	a	b	c	d	e
Wild-type (I-SceI induced)	102	$5x10^4$	52	100	$3.9x10^{-3}$
	66		50		$2.6x10^{-3}$
	59		37		$3.2x10^{-3}$
	69		45		$3.1x10^{-3}$
	81		80		$2.0x10^{-3}$
	71		34		$4.2x10^{-3}$
	94		58		$3.2x10^{-3}$
Wild-type (I-SceI uninduced)	0	$1x10^5$	26	100	0
	0		28		0
	0		36		0
	0		52		0
$IRF4^{-/-/-}$ (I-SceI induced)	141	$1x10^5$	32	100	$4.4x10^{-3}$
	76		20		$3.8x10^{-3}$
	19	$5x10^4$	30		$2.5x10^{-3}$
	56		45		$5.0x10^{-3}$
	40		78		$2.1x10^{-3}$
	37		61		$2.4x10^{-3}$
	238		42		$2.3x10^{-2}$
$IRF4^{-/-/-}$ (I-SceI uninduced)	0	$1x10^5$	30	100	0

a, b, c, d, e are defined as in Table 1.

4.1.2 A gene dosage effect of PU.1 on AID-mediated gene conversion.

In PU.1-deficient cells, AID expression is dramatically reduced, which precludes the evaluation of AID-mediated gene conversion activity in this cell line. However, AID expression is normal in $PU.1^{+/-}$ cells, and in light of the gene dosage effect of IRF4 on AID-mediated gene conversion, it is possible that a gene conversion defect could be observed in $PU.1^{+/-}$ cells as well. Using the same gene conversion substrate, the gene conversion activity in $PU.1^{+/-}$ cells was analyzed, and, unexpectedly, an approximately 20-fold reduction in gene conversion frequency was observed (Fig. 4.3*A*). Moreover,

the transcription level of the substrate is not affected in PU.1$^{+/-}$ cells. Therefore the phenotype is not an indirect consequence of defect in transcription (Fig. 4.3*B*).



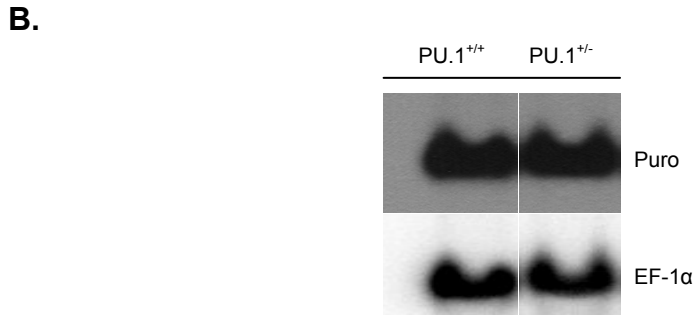


Figure 4.3. PU.1 deficiency impairs AID-mediated gene conversion. *A*. The dot plot presents the gene conversion analysis of wild-type and PU.1 heterozygous cells. Each dot represents result from one experiment (for colony numbers, see Table 3). In one experiment for clone f, no puromycin resistant colonies were observed, and this data point is shown below the baseline of the plot. The bar represents the median GC frequency of a data set of at least six experiments, and values are written below the plot. *B*. The transcript levels of the *puro* reporter in both cell lines were analyzed with Northern blotting. The same blot was rehybridized with an EF-1α cDNA probe to normalize RNA loading.

Table 3. Gene conversion analysis of PU.1 heterozygous cells.

Genotype	Puro		Control		GC
	a	b	c	d	e
Wild-type	28	$2x10^4$	68	100	$2.1x10^{-3}$
	42		70		$3.0x10^{-3}$
	20		37		$2.7x10^{-3}$
	45		66		$3.4x10^{-3}$
	34		72		$2.4x10^{-3}$
	61		54		$5.7x10^{-3}$
$PU.1^{+/-}$	9	$1x10^5$	37	100	$2.4x10^{-4}$
	2		38		$5.3x10^{-5}$
	36		23		$1.6x10^{-3}$
	0		28		0
	4		21		$1.9x10^{-4}$
	1		22		$4.6x10^{-5}$
	10		69		$1.5x10^{-4}$
	8		45		$1.8x10^{-4}$
	4		31		$1.3x10^{-4}$
	2		30		$6.7x10^{-5}$

a, b, c, d, e are defined as in Table 1.

To address whether the gene conversion defect is specific to AID, the I-SceI mediated gene conversion system was applied to $PU.1^{+/-}$ cells. In contrast to the severe reduction in AID-mediated gene conversion, I-SceI mediated gene conversion is in fact increased by about 3-fold in $PU.1^{+/-}$ cells (Fig. 4.4). The reason behind the increase in I-SceI mediated gene conversion in $PU.1^{+/-}$ cells is presently unclear, but the result clearly shows that the gene conversion defect is specific for AID. Thus, PU.1 is important for AID-mediated gene conversion at the Ig locus, and moreover, the dosage of PU.1 is critical in this respect.

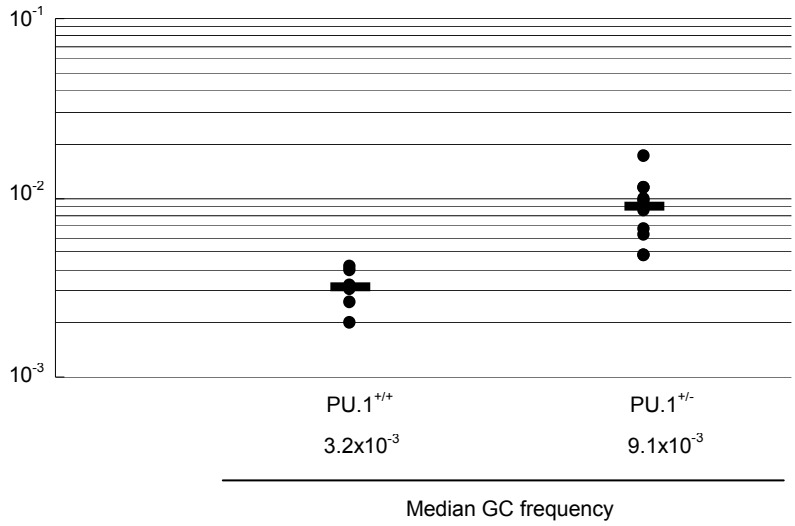


Figure 4.4. PU.1 is dispensable for I-SceI mediated gene conversion. The dot plot compares the frequency of I-SceI mediated gene conversion in wild-type versus PU.1 heterozygous cells. The bar indicates the median GC frequency of a data set of at least six experiments (for colony numbers, see Table 4), and the value is written below the plot.

Table 4. I-SceI mediated gene conversion in PU.1 heterozygous cells.

Genotype	Puro		Control		GC
	a	b	c	d	e
Wild-type (I-SceI induced)	102	$5x10^4$	52	100	$3.9x10^{-3}$
	66		50		$2.6x10^{-3}$
	59		37		$3.2x10^{-3}$
	69		45		$3.1x10^{-3}$
	81		80		$2.0x10^{-3}$
	71		34		$4.2x10^{-3}$
	94		58		$3.2x10^{-3}$
Wild-type (I-SceI uninduced)	0	$1x10^5$	26	100	0
	0		28		0
	0		36		0
	0		52		0
$PU.1^{+/-}$ (I-SceI induced)	67	$5x10^4$	28	100	$4.8x10^{-3}$
	79		14		$1.1x10^{-2}$
	104		12		$1.8x10^{-2}$
	78		23		$6.8x10^{-3}$
	27	$2x10^4$	63	400	$8.6x10^{-3}$
	32		102		$6.3x10^{-3}$
	31		63		$9.8x10^{-3}$
	16		67		$4.8x10^{-3}$
	35		73		$9.6x10^{-3}$
	12		21		$1.1x10^{-2}$

a, b, c, d, e are defined as in Table 1.

4.1.3 Gene conversion defect in PU.1-IRF binding site mutants at the Igλ locus.

IRF4 and PU.1 could facilitate AID function by recruiting it to the Igλ locus or by increasing local chromatin accessibility, and both mechanisms would require that IRF4 and PU.1 associate with the Igλ locus directly. The two PU.1-IRF composite binding sites within the 2.4-kb regulatory region may mediate this association. To evaluate their function, both binding sites were mutated through homologous recombination (Fig. 4.5*A*), and the gene conversion substrate was integrated into the Igλ locus (Fig. 4.5*A*) with (j, Fig. 4.5*B*) or without (a, Fig. 4.5*B*) the binding site mutations. Mutation of the PU.1-IRF binding sites reduces gene conversion frequency by approximately 3-fold (compare j with a in Fig. 4.5*B*). The relatively mild effect of the mutation could be due to the presence of potentially redundant elements in the Igλ locus, which is known to contain multiple regulatory elements. To address this possibility, a 7.4-kb region (Region B, Fig. 4.5*A* and *B*) downstream of the PU.1-IRF binding sites was deleted. Deletion of this 7.4-kb region leads to a 4-fold reduction in gene conversion (k, Fig. 4.5*B*). Introduction of the binding site mutation into this deletion background decreases gene conversion further by approximately 5-fold (compare l with k, Fig. 4.5*B*). Thus, in the absence of redundant regulatory elements, gene conversion is more dependent on the PU.1-IRF4 binding sites. The binding site mutation does not have major impacts on the transcription of the substrate (compare j with a, l with k, Fig. 4.5*C*). Therefore, the gene conversion defect is not simply a consequence of reduced transcription.

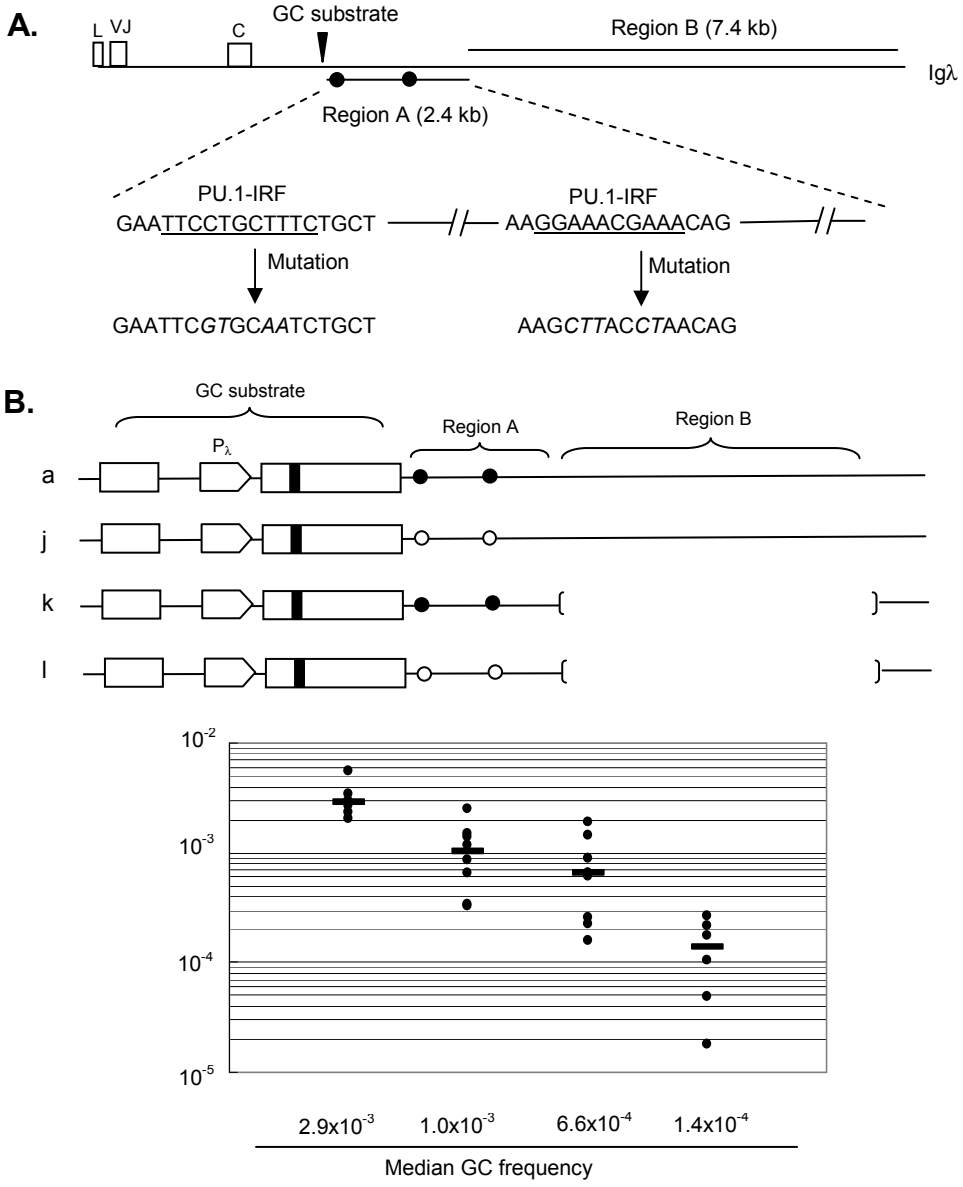

A.
L VJ
C
GC substrate
Region B (7.4 kb)
Igλ
Region A (2.4 kb)
PU.1-IRF
GAATTCCTGCTTTCTGCT
PU.1-IRF
AAGGAAACGAAACAG
Mutation
Mutation
GAATTCGTGCAATCTGCT
AAGCTTACCTAACAG
B.
GC substrate
Region A
Region B
Pλ
a
j
k
l
10-2
10-3
10-4
10-5
2.9x10-3
1.0x10-3
6.6x10-4
1.4x10-4
Median GC frequency
C.
Puro
EF-1α
a
j
k
l

Figure 4.5. PU.1-IRF composite binding sites in the Igλ locus are involved in AID-mediated gene conversion. *A*. This diagram shows the location (filled circle) and sequence (underlined) of the two PU.1-IRF composite binding sites in the rearranged allele of Igλ. The two binding sites differ in the relative order of the PU.1 and IRF binding motifs. The mutations are italicized. The diagram of Igλ is drawn to scale. L, VJ, and C represent the leader exon, rearranged VJ region, and constant region, respectively. The integration site for the gene conversion substrate is marked with an arrowhead. *B*. The four gene conversion substrates (a, j, k, and l) differ in mutation of the PU.1-IRF binding site (open circle) and deletion of Region B (bracket). The gene conversion frequencies of these four substrates are presented in the dot plot. The data for a is the same shown in Fig. 4.1*B*. Each dot represents results from one experiment (for colony numbers, see Table 5). The bar represents the median gene conversion (GC) frequency of a data set of at least six experiments, and value is written below the plot. *C*. The transcript levels of the *puro* reporter in various cell lines are analyzed with Northern blotting. The same blot was rehybridized with an EF-1α cDNA probe to normalize RNA loading.

Table 5. Gene conversion analysis of PU.1-IRF binding site (BS) mutants.

Genotype	Puro		Control		GC
	a	b	c	d	e
Wild-type	28	$2x10^4$	68	100	$2.1x10^{-3}$
	42		70		$3.0x10^{-3}$
	20		37		$2.7x10^{-3}$
	45		66		$3.4x10^{-3}$
	34		72		$2.4x10^{-3}$
	61		54		$5.7x10^{-3}$
BS mutant	21	$2x10^4$	74	100	$1.4x10^{-3}$
	11		63		$8.7x10^{-4}$
	5		75		$3.3x10^{-4}$
	28		55		$2.6x10^{-3}$
	5		38		$6.6x10^{-4}$
	3		44		$3.4x10^{-4}$
	16		54		$1.5x10^{-3}$
	10		42		$1.2x10^{-3}$

Genotype	Puro		Control		GC
	a	b	c	d	e
Deletion mutant	14	$1x10^5$	63	100	$2.2x10^{-4}$
	16		61		$2.6x10^{-4}$
	38		42		$9.1x10^{-4}$
	34		53		$6.4x10^{-4}$
	33		55		$6.0x10^{-4}$
	108		68		$1.6x10^{-4}$
	36		55		$6.6x10^{-4}$
	75		51		$1.5x10^{-3}$
	90		47		$1.9x10^{-3}$
BS mutant with deletion	9	$1x10^5$	42	100	$2.1x10^{-4}$
	1		55		$1.8x10^{-5}$
	10		57		$1.8x10^{-4}$
	3		62		$4.8x10^{-5}$
	6		57		$1.1x10^{-4}$
	10		37		$2.7x10^{-4}$

a, b, c, d, e are defined as in Table 1.

4.1.4 Interaction of PU.1 and IRF4 with the Igλ locus.

The phenotype of the PU.1-IRF binding sites mutations is consistent with the possibility that both proteins interact with the Igλ locus. To address this issue in a more direct manner, the Dam methylation assay again was used with region A that contains both of the composite PU.1-IRF binding sites (Fig. 4.5*A*). There are restriction site polymorphisms between the two alleles of Igλ in this region, which enables us to distinguish them in Southern analysis (Fig. 4.6). Both Dam-PU.1 and IRF4-Dam efficiently methylate the regulatory region on both alleles of Igλ (Fig. 4.6), and the extent of methylation appears to be higher relative to EF-1α (Fig. 3.4*C*), hsc70 and the cross-hybridizing fragment (Fig. 3.4*D*). This result suggests that both PU.1 and IRF4 interact directly with the regulatory region.

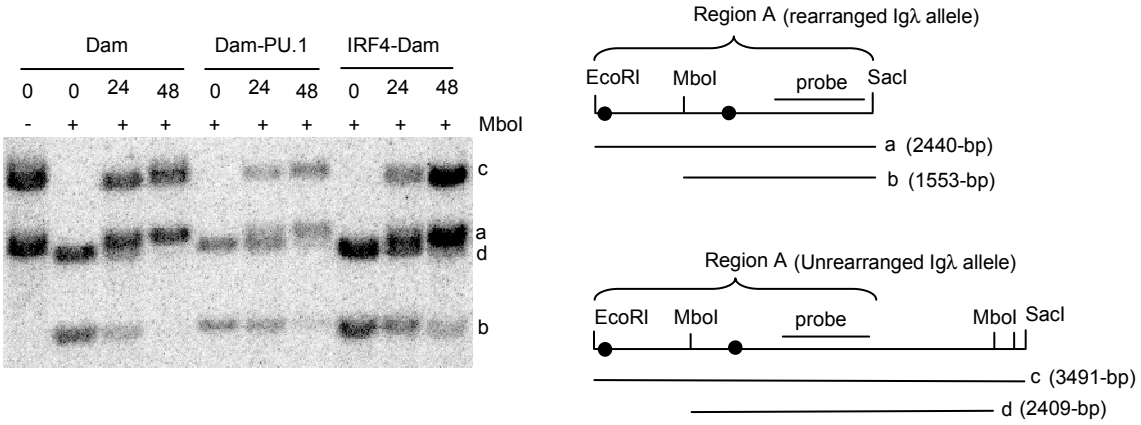


Figure 4.6. Methylation analysis of the interaction of PU.1 and IRF4 with Region A in the Igλ locus. The autoradiograph is a Southern analysis of the methylation pattern in Region A. The sample in the first lane is digested with EcoRI and SacI, but without MboI, to illustrate the restriction site polymorphism of the two Igλ alleles. This figure is labeled in an analogous manner as Fig. 3.4*B*.

4.2 DISCUSSION

The experiments presented in this chapter address the roles of PU.1 and IRF4 in AID-mediated gene conversion at the Igλ locus. Deficiencies were shown in either PU.1 or IRF4 impair AID-mediated gene conversion, and the dosage of PU.1 appears to be critical in this respect. Two composite PU.1-IRF binding sites were found in a regulatory region in the Igλ locus, and the roles of these two binding sites in AID-mediated gene conversion were addressed. Elimination of these two PU.1-IRF binding sites impairs AID function at the locus, suggesting that these binding sites may indeed mediate, at least in part, the functions of PU.1 and IRF4. Consistent with this possibility, both PU.1 and IRF4 appear to interact directly with the regulatory region based on methyltransferase recruitment assays. On the other hand, the gene conversion defect observed in the binding site mutations is less severe compared with IRF4 and PU.1 deficient cells. The discrepancy suggests that these two factors have additional functions, which may involve collaboration with other factors through additional binding sites. As described previously, E2A and NF-κB have also been implicated in AID mediated gene conversion. Moreover, the Igλ regulatory element contains binding sites for the transcription factors, Octamer binding proteins and Mef2; mutations of these binding sites have been shown to impair the function of this regulatory element in targeting AID function to ectopic sites [65]. All of these factors could potentially collaborate with PU.1 and IRF4. In fact, E2A is known to bind to the mouse Igκ 3' enhancer synergistically with IRF4 [106]. The potential collaboration of PU.1 and IRF4 with the other factors remains to be evaluated.

The next important question is how PU.1 and IRF4 facilitate AID function. Since both factors can function as transcriptional activators, they may help to open up the local chromatin structure and facilitate the access of AID to target DNA. However, the

results from the I-SceI gene conversion experiments seem to be inconsistent with the explanation. The gene conversion reaction is initiated by the endonuclease cleavage of the target site, and the cleavage efficiency is expected to correlate with the accessibility of target DNA. If deficiency in IRF4 or PU.1 leads to a reduction in chromatin accessibility at the Igλ locus, a reduction in I-SceI mediated gene conversion would be expected, which is not the case. The caveat of this argument is that gene conversion is an indirect readout of I-SceI cleavage activity. The cleavage product by Southern analysis could not be observed, presumably because of low abundance and transient nature of the double strand break. In future experiments, DNase I hypersensitivity assay, methylation assay and histone modification status may further address this issue.

PU.1 and IRF4 could also function by recruiting AID to the Igλ locus, potentially through direct interaction with AID. As discussed in chapter 1, RPA is the only factor that previously has been shown to interact with AID, but this interaction is unlikely to be sufficient to explain the target specificity of AID. Since PU.1 and IRF4 can recognize binding sites in the Igλ locus, they would fit the requirement for the specific AID-recruiting factor, and they may achieve this function through collaboration with additional factors, such as E2A. It is possible that the recruitment of AID may involve a multi-protein complex rather than single factors, which may explain the difficulty in identifying a specific AID recruitment factor.

Applying the methylation recruitment assay to AID localization, preliminary evidence has been obtained that AID is indeed preferentially targeted to the Igλ locus, and this recruitment requires the regulatory element that contains the PU.1-IRF binding sites (Kim and Tian, unpublished observations). In future experiments, this approach could be used to address whether PU.1 and IRF4 are important for recruiting AID to the Igλ locus. If these factors do play a role in targeting AID, potential direct interactions

between AID and these factors could be tested, either by themselves or together with other factors such as E2A. Moreover, affinity purification approaches could be used to isolate proteins associated with AID or PU.1 and IRF4 to see if these proteins exist in a common complex, and if so, what other proteins are within this complex. The advantage of the system is that, once a candidate factor or association is detected, its functional relevance could be evaluated by genetic experiments.

PU.1 and IRF4 could also regulate AID function by other indirect means, for example by controlling the expression of other factors involved in AID function. In mice, PU.1 and IRF4 control the expression of many factors involved in B-cell differentiation. The same could be true in chicken B cells. This might explain the severe defect in gene conversion observed in PU.1$^{+/-}$ cells. It is possible that function of PU.1 is counter balanced by competing factors, which would favor an alternative cell fate that is incompatible with AID function. In such a scenario, the dosage of PU.1 could be critical since a small change in PU.1 could tip the balance between the two competing cell fates and strongly influence AID function. This issue could be addressed by systematically comparing the transcript profiles of wild-type, PU.1$^{+/-}$, and PU.1$^{-/-}$ cells, for example by microarray analysis. As described in chapter 3, such experiments may also help to understand how PU.1 regulate the expression of AID, since it remains possible that PU.1 may stimulate AID expression, at least in part, by controlling the expression of other factors involved in AID gene transcription.

An intriguing observation made in this part of the experiments is that there seems to be a reciprocal relationship between AID and I-SceI-mediate gene conversion in PU.1$^{+/-}$ cells. Whereas AID-mediated gene conversion is reduced in PU.1$^{+/-}$ cells, I-SceI-mediated gene conversion is increased in the same cell type. The difference is consistent, and does not appear to simply represent experimental variation. Moreover,

the same reciprocal relationship was observed when the same approach was used to compare the effect of deleting the Igλ regulatory region on AID versus I-SceI-mediated gene conversion (Kim and Tian, unpublished observations). In those experiments, deletion of the Igλ regulatory element strongly impairs AID-mediated gene conversion, but I-SceI mediated gene conversion appears to be more efficient in the absence of the regulatory element, a result analogous to that observed in PU.1$^{+/-}$ cells. At present, we do not have a satisfactory explanation for this unusual phenomenon.

In broad terms, it is possible that the regulatory element and PU.1 are involved in recruiting certain factors to the Igλ locus that facilitate AID function, and these factors interfere with I-SceI cleavage of target DNA, potentially by competing for the interaction with target DNA. Another possibility is that the regulatory element and PU.1 help to unwind target DNA. As discussed in chapter 1, AID can only deaminate cytosines in the context of single-stranded DNA. By contrast, I-SceI can only cleave double-stranded DNA. So, there is mutually exclusive relationship between the two enzymes in terms of their substrate requirements. It is possible that the regulatory element and PU.1 may modify the transcription apparatus in a way that allows it more effective unwinding of template DNA. Another possibility is that the regulatory element and PU.1 are required for high levels of transcription elongation through the target DNA, which would increase the extent of DNA unwinding. Although obvious reductions in the steady state transcript levels of the gene conversion substrate were not observed in the absence of PU.1, IRF4, or the regulatory element, it remains possible that the actual numbers of RNA polymerases engaged in elongation on the template DNA is reduced under the circumstances. This issue could be resolved by performing Chromatin Immuno-Precipitation (ChIP) analysis on RNA polymerase II or nuclear run-on analysis.

Appendix

The homologous recombination defects in PU.1-deficient cells

A.1 BACKGROUND

One unique feature of DT40 is the high ratio of homologous targeting events versus random integration events [112], but the molecular basis for this phenomenon is presently unclear. The high levels of homologous recombination probably reflect the need of the bursa B cells to perform gene conversion, which is mediated by the homologous recombination machinery. Broadly speaking, the high ratio of homologous targeting efficiency could be the consequence of either elevated levels of homologous recombination or reduced levels of non-homologous end joining, which is responsible for random integration; the two recombination pathways occur through completely different mechanisms and involve distinct factors.

Homologous recombination is generally initiated by a double-stranded DNA break [117]. The 5' end of the double-strand break is then resected by exonucleases to generate a 3' ended single-stranded tail [117]. The homologous recombination machinery catalyzes the invasion of this 3' tail into a homologous duplex to initiate strand exchange [117]. The strand exchange intermediate could eventually be resolved by various pathways, such as Holiday junction resolvase, to generate the recombination product [117]. By comparison with the elaborate pathway of homologous recombination, non-homologous end joining is achieved by the straight ligation of DNA ends from double stranded breaks [118]. It is possible that DT40 may simply have higher levels of the factors involved in homologous recombination or lower levels of the factors responsible for non-homologous end joining. Alternatively, DT40 may have

novel factors that either up-regulate homologous recombination or repress non-homologous end joining. No concrete evidence supporting any of these scenarios has been reported so far.

My interest in this problem stems from the unexpected finding that the gene conversion substrate could not be integrated into the Igλ locus of PU.1$^{-/-}$ cells through homologous targeting, which is usually an efficient event in wild-type DT40 cells (Table 6). The question is: do PU.1$^{-/-}$ cells have a generalized defect in homologous recombination? If so, does PU.1 control the expression of the factor that is responsible for the unusual property of DT40 cells in terms of high homologous targeting ratio relative to random integration? In the following section, I present preliminary experiments to address these issues.

A.2 RESULTS AND DISCUSSION

As described above, the gene conversion substrate could not be stably integrated into the Igλ locus of PU.1$^{-/-}$ cells. The problem is not caused by inefficient transfection since stable integrants that contain the transfected construct were obtained; but all of the integration events were random. To test whether this phenomenon is specific to this targeting construct, a different targeting construct was designed to integrate into a different location in the Igλ locus (Table 6). In wild-type DT40 cells, 8 out of 24 screened colonies integrated the construct into the Igλ locus. By contrast, none of the 24 transfectants of PU.1$^{-/-}$ cells contain homologous integration with the same construct. Thus, the problem is not unique to one particular construct or transfection event.

The targeting defect was analyzed for loci other than Igλ and to determine if there was a general defect in homologous recombination. Toward this end, the homologous

targeting efficiency was tested at two other genes: CBP and p300, which encode transcription coactivators. In wild-type DT40 cells, these targeting constructs integrate into their homologous targets at 50% frequency (Table 6). In PU.1$^{-/-}$ cells, the targeting frequencies at these two loci depend on the way the targeting constructs were linearized prior to transfection (Table 6). When the construct was linearized with PvuI, which generates a 3' protruding end, the targeting frequencies into either the p300 or the CBP locus was around 30% in PU.1$^{-/-}$ cells. By contrast, if the constructs were linearized with a restriction enzyme to produce a 5' overhang, the frequency of homologous targeting at both loci was severely reduced. This observation led us to redo the targeting experiments at the Igλ locus with a construct that was linearized with PvuI. However, contrary to expectations, Igλ remains refractory to homologous targeting even under this situation.

Moreover, the position of the linearization site at the targeting construct seems to affect the targeting efficiency as well (Table 6). The linearization sites described so far are all within the cloning vector, and the construct was cleaved only once; for this reason, the heterologous vector sequence is still attached to the homology arms. In the CBP locus, if the vector sequence was removed by restriction digestion on both sides of the homology arm, the targeting frequency was more than 90%, even though the restriction enzyme generates 5' protruding ends, which previously gave poor targeting frequencies (Table 6). By contrast, the same treatment for both the p300 and Igλ targeting constructs failed to improve the targeting frequency (Table 6).

Thus, PU.1$^{-/-}$ cells clearly exhibit certain defects in homologous targeting, but the defect is variable with respect to the targeted loci and the manipulation of the targeting construct. The defect is most consistent at the Igλ locus, which failed to achieve homologous targeting with any of the constructs or linearization methods (Table 6). At

the p300 and CBP loci, the targeting frequency depends on the linearization method of the targeting construct, and the same linearization method has different effects in the two loci. At present, no satisfactory explanation for these observations has been obtained. Since homologous targeting works at least under some circumstances, the PU.1$^{-/-}$ cells do not appear to have a generalized defect in homologous recombination. Moreover, the expression levels of many homologous recombination factors in PU.1$^{-/-}$ cells have been examined, but no difference relative to wild-type cells were found (data not shown). The locus dependence of the integration defect suggests that the problem may lie in the relative accessibility of the different loci to homologous recombination machinery. It is possible that PU.1 is important for the access of homologous recombination machinery to the Igλ locus, and this function may be relevant to the regulation of Ig gene conversion.

Table 6. Homologous targeting efficiency in PU.1 mutant cells.

Targeted Locus	Targeting plasmid	Linearization (protruding ends)	Non-homologous sequence	Number of targeted colonies	Number of screened colonies	Genotype
Igλ	M173	XhoI (5')	Yes[a]	6	7	Wild-type
			Yes	7	36	PU.1$^{+/-}$
			Yes	0	54	PU.1$^{-/-}$
		BglII (5') and XhoI (5')	No[b]	0	7	
	pHL255	EcoRI (5')	Yes	8	24	Wild-type
		PvuI (3')	Yes	0	24	PU.1$^{-/-}$
p300	M497	NdeI (5')	Yes	6	11	Wild-type
		NdeI (5')	Yes	0[c]	24	PU.1$^{-/-}$
		PvuI (3')	Yes	12	36	
		EcoRI (5')	No	1	24	
CBP	M504	PvuI (3')	Yes	5	11	Wild-type
		PvuI (3')	Yes	8	24	PU.1$^{-/-}$
		NdeI (5')	No	22	24	
	M765[d]	XhoI (5')	Yes	0	6	

a: Non-homologous sequences are retained on the targeting constructs after linearization.
b: Non-homologous sequences are removed from targeting construct after linearization.
c: On one experiment, there is a faint band for targeted allele on the Southern blot. The sample for that experiment may contain more than one colony.
d: M765 is derived from M504 and differs from M504 by a slightly shorter 3' homology arm.

References

1. Charles A. Janeway, J., Paul Travers, Mark Walport, Mark J. Shlomchik, *Immunobiology - the immune system in health and disease*. 6 ed. 2005, New York: Garland Science Publishing.

2. Jung, D., et al., *Mechanism and control of V(D)J recombination at the immunoglobulin heavy chain locus.* Annu Rev Immunol, 2006. **24**: p. 541-70.

3. Ratcliffe, M.J., *Antibodies, immunoglobulin genes and the bursa of Fabricius in chicken B cell development.* Dev Comp Immunol, 2006. **30**(1-2): p. 101-18.

4. Reynaud, C.A., et al., *A hyperconversion mechanism generates the chicken light chain preimmune repertoire.* Cell, 1987. **48**(3): p. 379-88.

5. Arakawa, H. and J.M. Buerstedde, *Immunoglobulin gene conversion: insights from bursal B cells and the DT40 cell line.* Dev Dyn, 2004. **229**(3): p. 458-64.

6. Papavasiliou, F.N. and D.G. Schatz, *Somatic hypermutation of immunoglobulin genes: merging mechanisms for genetic diversity.* Cell, 2002. **109 Suppl**: p. S35-44.

7. Odegard, V.H. and D.G. Schatz, *Targeting of somatic hypermutation.* Nat Rev Immunol, 2006. **6**(8): p. 573-83.

8. Manis, J.P., M. Tian, and F.W. Alt, *Mechanism and control of class-switch recombination.* Trends Immunol, 2002. **23**(1): p. 31-9.

9. Selsing, E., *Ig class switching: targeting the recombinational mechanism.* Curr Opin Immunol, 2006. **18**(3): p. 249-54.

10. Muramatsu, M., et al., *Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells.* J Biol Chem, 1999. **274**(26): p. 18470-6.

11. Delker, R.K., S.D. Fugmann, and F.N. Papavasiliou, *A coming-of-age story: activation-induced cytidine deaminase turns 10.* Nat Immunol, 2009. **10**(11): p. 1147-53.

12. Muramatsu, M., et al., *Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme.* Cell, 2000. **102**(5): p. 553-63.

13. Revy, P., et al., *Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2).* Cell, 2000. **102**(5): p. 565-75.

14. Arakawa, H., J. Hauschild, and J.M. Buerstedde, *Requirement of the activation-induced deaminase (AID) gene for immunoglobulin gene conversion.* Science, 2002. **295**(5558): p. 1301-6.

15. Honjo, T., et al., *AID to overcome the limitations of genomic information.* Nat Immunol, 2005. **6**(7): p. 655-61.

16. Casali, P., et al., *DNA repair in antibody somatic hypermutation.* Trends Immunol, 2006. **27**(7): p. 313-21.

17. Martin, A. and M.D. Scharff, *AID and mismatch repair in antibody diversification.* Nat Rev Immunol, 2002. **2**(8): p. 605-14.

18. Teng, G. and F.N. Papavasiliou, *Immunoglobulin Somatic Hypermutation.* Annu Rev Genet, 2007.

19. Di Noia, J.M. and M.S. Neuberger, *Molecular mechanisms of antibody somatic hypermutation.* Annu Rev Biochem, 2007. **76**: p. 1-22.

20. Tang, E.S. and A. Martin, *Immunoglobulin gene conversion: Synthesizing antibody diversification and DNA repair.* DNA Repair (Amst), 2007.

21. Tang, E.S. and A. Martin, *NHEJ-deficient DT40 cells have increased levels of immunoglobulin gene conversion: evidence for a double strand break intermediate.* Nucleic Acids Res, 2006. **34**(21): p. 6345-51.

22. Rada, C., et al., *Immunoglobulin isotype switching is inhibited and somatic hypermutation perturbed in UNG-deficient mice.* Curr Biol, 2002. **12**(20): p. 1748-55.

23. Imai, K., et al., *Human uracil-DNA glycosylase deficiency associated with profoundly impaired immunoglobulin class-switch recombination.* Nat Immunol, 2003. **4**(10): p. 1023-8.

24. Saribasak, H., et al., *Uracil DNA glycosylase disruption blocks Ig gene conversion and induces transition mutations.* J Immunol, 2006. **176**(1): p. 365-71.

25. Rada, C., J.M. Di Noia, and M.S. Neuberger, *Mismatch recognition and uracil excision provide complementary paths to both Ig switching and the A/T-focused phase of somatic mutation.* Mol Cell, 2004. **16**(2): p. 163-71.

26. Petersen-Mahrt, S.K., R.S. Harris, and M.S. Neuberger, *AID mutates E. coli suggesting a DNA deamination mechanism for antibody diversification.* Nature, 2002. **418**(6893): p. 99-103.

27. Bransteitter, R., et al., *Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase.* Proc Natl Acad Sci U S A, 2003. **100**(7): p. 4102-7.

28. Dickerson, S.K., et al., *AID mediates hypermutation by deaminating single stranded DNA.* J Exp Med, 2003. **197**(10): p. 1291-6.

29. Sohail, A., et al., *Human activation-induced cytidine deaminase causes transcription-dependent, strand-biased C to U deaminations.* Nucleic Acids Res, 2003. **31**(12): p. 2990-4.

30. Pham, P., et al., *Processive AID-catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation.* Nature, 2003. **424**(6944): p. 103-7.

31. Ramiro, A.R., et al., *Transcription enhances AID-mediated cytidine deamination by exposing single-stranded DNA on the nontemplate strand.* Nat Immunol, 2003. **4**(5): p. 452-6.

32. Chaudhuri, J., et al., *Transcription-targeted DNA deamination by the AID antibody diversification enzyme.* Nature, 2003. **422**(6933): p. 726-30.

33. Morgan, H.D., et al., *Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming.* J Biol Chem, 2004. **279**(50): p. 52353-60.

34. Bransteitter, R., et al., *Biochemical analysis of hypermutational targeting by wild type and mutant activation-induced cytidine deaminase.* J Biol Chem, 2004. **279**(49): p. 51612-21.

35. Yu, K., F.T. Huang, and M.R. Lieber, *DNA substrate length and surrounding sequence affect the activation-induced deaminase activity at cytidine.* J Biol Chem, 2004. **279**(8): p. 6496-500.

36. Chaudhuri, J., C. Khuong, and F.W. Alt, *Replication protein A interacts with AID to promote deamination of somatic hypermutation targets.* Nature, 2004. **430**(7003): p. 992-8.

37. Okazaki, I.M., et al., *Constitutive expression of AID leads to tumorigenesis.* J Exp Med, 2003. **197**(9): p. 1173-81.

38. Gonda, H., et al., *The balance between Pax5 and Id2 activities is the key to AID gene expression.* J Exp Med, 2003. **198**(9): p. 1427-37.

39. Dedeoglu, F., et al., *Induction of activation-induced cytidine deaminase gene expression by IL-4 and CD40 ligation is dependent on STAT6 and NFkappaB.* Int Immunol, 2004. **16**(3): p. 395-404.

40. Sciammas, R., et al., *Graded expression of interferon regulatory factor-4 coordinates isotype switching with plasma cell differentiation.* Immunity, 2006. **25**(2): p. 225-36.

41. Klein, U., et al., *Transcription factor IRF4 controls plasma cell differentiation and class-switch recombination.* Nat Immunol, 2006. **7**(7): p. 773-82.

42. Lee, C.H., et al., *Regulation of the germinal center gene program by interferon (IFN) regulatory factor 8/IFN consensus sequence-binding protein.* J Exp Med, 2006. **203**(1): p. 63-72.

43. Park, S.R., et al., *HoxC4 binds to the promoter of the cytidine deaminase AID gene to induce AID expression, class-switch DNA recombination and somatic hypermutation.* Nat Immunol, 2009. **10**(5): p. 540-50.

44. Nera, K.P., et al., *Loss of Pax5 promotes plasma cell differentiation.* Immunity, 2006. **24**(3): p. 283-93.

45. de Yebenes, V.G., et al., *miR-181b negatively regulates activation-induced cytidine deaminase in B cells.* J Exp Med, 2008. **205**(10): p. 2199-206.

46. Teng, G., et al., *MicroRNA-155 is a negative regulator of activation-induced cytidine deaminase.* Immunity, 2008. **28**(5): p. 621-9.

47. Dorsett, Y., et al., *MicroRNA-155 suppresses activation-induced cytidine deaminase-mediated Myc-Igh translocation.* Immunity, 2008. **28**(5): p. 630-8.

48. Ito, S., et al., *Activation-induced cytidine deaminase shuttles between nucleus and cytoplasm like apolipoprotein B mRNA editing catalytic polypeptide 1.* Proc Natl Acad Sci U S A, 2004. **101**(7): p. 1975-80.

49. Brar, S.S., M. Watson, and M. Diaz, *Activation-induced cytosine deaminase (AID) is actively exported out of the nucleus but retained by the induction of DNA breaks.* J Biol Chem, 2004. **279**(25): p. 26395-401.

50. Patenaude, A.M., et al., *Active nuclear import and cytoplasmic retention of activation-induced deaminase.* Nat Struct Mol Biol, 2009. **16**(5): p. 517-27.

51. Aoufouchi, S., et al., *Proteasomal degradation restricts the nuclear lifespan of AID.* J Exp Med, 2008. **205**(6): p. 1357-68.

52. Daniels, G.A. and M.R. Lieber, *RNA:DNA complex formation upon transcription of immunoglobulin switch regions: implications for the mechanism and regulation of class switch recombination.* Nucleic Acids Res, 1995. **23**(24): p. 5006-11.

53. Huang, F.T., et al., *Sequence dependence of chromosomal R-loops at the immunoglobulin heavy-chain Smu class switch region.* Mol Cell Biol, 2007. **27**(16): p. 5921-32.

54. Huang, F.T., et al., *Downstream boundary of chromosomal R-loops at murine switch regions: implications for the mechanism of class switch recombination.* Proc Natl Acad Sci U S A, 2006. **103**(13): p. 5030-5.

55. Yu, K., et al., *R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells.* Nat Immunol, 2003. **4**(5): p. 442-51.

56. Rogozin, I.B. and M. Diaz, *Cutting edge: DGYW/WRCH is a better predictor of mutability at G:C bases in Ig hypermutation than the widely accepted RGYW/WRCY motif and probably reflects a two-step activation-induced cytidine deaminase-triggered process.* J Immunol, 2004. **172**(6): p. 3382-4.

57. Michael, N., et al., *Effects of sequence and structure on the hypermutability of immunoglobulin genes.* Immunity, 2002. **16**(1): p. 123-34.

58. Yelamos, J., et al., *Targeting of non-Ig sequences in place of the V segment by somatic hypermutation.* Nature, 1995. **376**(6537): p. 225-9.

59. Yang, S.Y., S.D. Fugmann, and D.G. Schatz, *Control of gene conversion and somatic hypermutation by immunoglobulin promoter and enhancer sequences.* J Exp Med, 2006. **203**(13): p. 2919-28.

60. Yang, S.Y., et al., *Activation-induced cytidine deaminase (AID)-mediated sequence diversification is transiently targeted to newly integrated DNA substrates.* J Biol Chem, 2007.

61. Fukita, Y., H. Jacobs, and K. Rajewsky, *Somatic hypermutation in the heavy chain locus correlates with transcription.* Immunity, 1998. **9**(1): p. 105-14.

62. Shen, H.M., et al., *The TATA binding protein, c-Myc and survivin genes are not somatically hypermutated, while Ig and BCL6 genes are hypermutated in human memory B cells.* Int Immunol, 2000. **12**(7): p. 1085-93.

63. Betz, A.G., et al., *Elements regulating somatic hypermutation of an immunoglobulin kappa gene: critical role for the intron enhancer/matrix attachment region.* Cell, 1994. **77**(2): p. 239-48.

64. Tumas-Brundage, K. and T. Manser, *The transcriptional promoter regulates hypermutation of the antibody heavy chain locus.* J Exp Med, 1997. **185**(2): p. 239-50.

65. Kim, Y. and M. Tian, *NF-kappaB family of transcription factor facilitates gene conversion in chicken B cells.* Mol Immunol, 2009. **46**(16): p. 3283-91.

66. Peters, A. and U. Storb, *Somatic hypermutation of immunoglobulin genes is linked to transcription initiation.* Immunity, 1996. **4**(1): p. 57-65.

67. Gorman, J.R., et al., *The Ig(kappa) enhancer influences the ratio of Ig(kappa) versus Ig(lambda) B lymphocytes.* Immunity, 1996. **5**(3): p. 241-52.

68. van der Stoep, N., J.R. Gorman, and F.W. Alt, *Reevaluation of 3'Ekappa function in stage- and lineage-specific rearrangement and somatic hypermutation.* Immunity, 1998. **8**(6): p. 743-50.

69. Inlay, M.A., et al., *Roles of the Ig kappa light chain intronic and 3' enhancers in Igk somatic hypermutation.* J Immunol, 2006. **177**(2): p. 1146-51.

70. Xiang, Y. and W.T. Garrard, *The Downstream Transcriptional Enhancer, Ed, positively regulates mouse Ig kappa gene expression and somatic hypermutation.* J Immunol, 2008. **180**(10): p. 6725-32.

71. Kothapalli, N., D.D. Norton, and S.D. Fugmann, *Cutting edge: a cis-acting DNA element targets AID-mediated sequence diversification to the chicken Ig light chain gene locus.* J Immunol, 2008. **180**(4): p. 2019-23.

72. Blagodatski, A., et al., *A cis-acting diversification activator both necessary and sufficient for AID-mediated hypermutation.* PLoS Genet, 2009. **5**(1): p. e1000332.

73. Vuong, B.Q., et al., *Specific recruitment of protein kinase A to the immunoglobulin locus regulates class-switch recombination.* Nat Immunol, 2009. **10**(4): p. 420-6.

74. Quong, M.W., et al., *E2A activity is induced during B-cell activation to promote immunoglobulin class switch recombination.* Embo J, 1999. **18**(22): p. 6307-18.

75. Romanow, W.J., et al., *E2A and EBF act in synergy with the V(D)J recombinase to generate a diverse immunoglobulin repertoire in nonlymphoid cells.* Mol Cell, 2000. **5**(2): p. 343-53.

76. Michael, N., et al., *The E box motif CAGGTG enhances somatic hypermutation without enhancing transcription.* Immunity, 2003. **19**(2): p. 235-42.

77. Conlon, T.M. and K.B. Meyer, *The chicken Ig light chain 3'-enhancer is essential for gene expression and regulates gene conversion via the transcription factor E2A.* Eur J Immunol, 2006. **36**(1): p. 139-48.

78. Schoetz, U., et al., *E2A expression stimulates Ig hypermutation.* J Immunol, 2006. **177**(1): p. 395-400.

79. Shen, H.M., et al., *Mutation of BCL-6 gene in normal B cells by the process of somatic hypermutation of Ig genes.* Science, 1998. **280**(5370): p. 1750-2.

80. Gordon, M.S., et al., *Somatic hypermutation of the B cell receptor genes B29 (Igbeta, CD79b) and mb1 (Igalpha, CD79a).* Proc Natl Acad Sci U S A, 2003. **100**(7): p. 4126-31.

81. Pasqualucci, L., et al., *BCL-6 mutations in normal germinal center B cells: evidence of somatic hypermutation acting outside Ig loci.* Proc Natl Acad Sci U S A, 1998. **95**(20): p. 11816-21.

82. Muschen, M., et al., *Somatic mutation of the CD95 gene in human B cells as a side-effect of the germinal center reaction.* J Exp Med, 2000. **192**(12): p. 1833-40.

83. Liu, M., et al., *Two levels of protection for the B cell genome during somatic hypermutation.* Nature, 2008. **451**(7180): p. 841-5.

84. Arakawa, H., D. Lodygin, and J.M. Buerstedde, *Mutant loxP vectors for selectable marker recycle and conditional knock-outs.* BMC Biotechnol, 2001. **1**: p. 7.

85. Kastner, P. and S. Chan, *PU.1: A crucial and versatile player in hematopoiesis and leukemia.* Int J Biochem Cell Biol, 2007.

86. Pio, F., et al., *Co-crystallization of an ETS domain (PU.1) in complex with DNA. Engineering the length of both protein and oligonucleotide.* J Biol Chem, 1995. **270**(41): p. 24258-63.

87. Pio, F., et al., *New insights on DNA recognition by ets proteins from the crystal structure of the PU.1 ETS domain-DNA complex.* J Biol Chem, 1996. **271**(38): p. 23329-37.

88. Kodandapani, R., et al., *A new pattern for helix-turn-helix recognition revealed by the PU.1 ETS-domain-DNA complex.* Nature, 1996. **380**(6573): p. 456-60.

89. Lloberas, J., C. Soler, and A. Celada, *Repression mechanisms of the I-A beta gene of the major histocompatibility complex.* Immunobiology, 1997. **198**(1-3): p. 249-63.

90. Oikawa, T., et al., *The role of Ets family transcription factor PU.1 in hematopoietic cell differentiation, proliferation and apoptosis.* Cell Death Differ, 1999. **6**(7): p. 599-608.

91. DeKoter, R.P., et al., *Regulation of the interleukin-7 receptor alpha promoter by the Ets transcription factors PU.1 and GA-binding protein in developing B cells.* J Biol Chem, 2007. **282**(19): p. 14194-204.

92. McKercher, S.R., et al., *Targeted disruption of the PU.1 gene results in multiple hematopoietic abnormalities.* Embo J, 1996. **15**(20): p. 5647-58.

93. Scott, E.W., et al., *PU.1 functions in a cell-autonomous manner to control the differentiation of multipotential lymphoid-myeloid progenitors.* Immunity, 1997. **6**(4): p. 437-47.

94. Scott, E.W., et al., *Requirement of transcription factor PU.1 in the development of multiple hematopoietic lineages.* Science, 1994. **265**(5178): p. 1573-7.

95. Polli, M., et al., *The development of functional B lymphocytes in conditional PU.1 knock-out mice.* Blood, 2005. **106**(6): p. 2083-90.

96. Matsudo, H., et al., *Disruption of the PU.1 gene in chicken B lymphoma DT40 cells and its effect on reported target gene expression.* Gene, 2003. **322**: p. 169-74.

97. Hromas, R., et al., *Hematopoietic lineage- and stage-restricted expression of the ETS oncogene family member PU.1.* Blood, 1993. **82**(10): p. 2998-3004.

98. Klemsz, M.J., et al., *The macrophage and B cell-specific transcription factor PU.1 is related to the ets oncogene.* Cell, 1990. **61**(1): p. 113-24.

99. Su, G.H., et al., *The Ets protein Spi-B is expressed exclusively in B cells and T cells during development.* J Exp Med, 1996. **184**(1): p. 203-14.

100. Eisenbeis, C.F., H. Singh, and U. Storb, *Pip, a novel IRF family member, is a lymphoid-specific, PU.1-dependent transcriptional activator.* Genes Dev, 1995. **9**(11): p. 1377-87.

101. Eisenbeis, C.F., H. Singh, and U. Storb, *PU.1 is a component of a multiprotein complex which binds an essential site in the murine immunoglobulin lambda 2-4 enhancer.* Mol Cell Biol, 1993. **13**(10): p. 6452-61.

102. Pongubala, J.M., et al., *PU.1 recruits a second nuclear factor to a site important for immunoglobulin kappa 3' enhancer activity.* Mol Cell Biol, 1992. **12**(1): p. 368-78.

103. Lu, R., et al., *IRF-4,8 orchestrate the pre-B-to-B transition in lymphocyte development.* Genes Dev, 2003. **17**(14): p. 1703-8.

104. Taniguchi, T., et al., *IRF family of transcription factors as regulators of host defense.* Annu Rev Immunol, 2001. **19**: p. 623-55.

105. Kanno, Y., et al., *Immune cell-specific amplification of interferon signaling by the IRF-4/8-PU.1 complex.* J Interferon Cytokine Res, 2005. **25**(12): p. 770-9.

106. Lazorchak, A.S., M.S. Schlissel, and Y. Zhuang, *E2A and IRF-4/Pip promote chromatin modification and transcription of the immunoglobulin kappa locus in pre-B cells.* Mol Cell Biol, 2006. **26**(3): p. 810-21.

107. Matsuyama, T., et al., *Molecular cloning of LSIRF, a lymphoid-specific member of the interferon regulatory factor family that binds the interferon-stimulated response element (ISRE).* Nucleic Acids Res, 1995. **23**(12): p. 2127-36.

108. Mittrucker, H.W., et al., *Requirement for the transcription factor LSIRF/IRF4 for mature B and T lymphocyte function.* Science, 1997. **275**(5299): p. 540-3.

109. Baba, T.W., B.P. Giroir, and E.H. Humphries, *Cell lines derived from avian lymphomas exhibit two distinct phenotypes.* Virology, 1985. **144**(1): p. 139-51.

110. Buerstedde, J.M., et al., *Light chain gene conversion continues at high rate in an ALV-induced cell line.* Embo J, 1990. **9**(3): p. 921-7.

111. Kim, S., et al., *Ongoing diversification of the rearranged immunoglobulin light-chain gene in a bursal lymphoma cell line.* Mol Cell Biol, 1990. **10**(6): p. 3224-31.

112. Winding, P. and M.W. Berchtold, *The chicken B cell line DT40: a novel tool for gene disruption experiments.* J Immunol Methods, 2001. **249**(1-2): p. 1-16.

113. van Steensel, B. and S. Henikoff, *Identification of in vivo DNA targets of chromatin proteins using tethered dam methyltransferase.* Nat Biotechnol, 2000. **18**(4): p. 424-8.

114. Orian, A., et al., *DamID: A Methylation-Based Chromatin Profiling Approach.* Methods Mol Biol, 2009. **567**: p. 155-69.

115. Horton, J.R., et al., *Structure and substrate recognition of the Escherichia coli DNA adenine methyltransferase.* J Mol Biol, 2006. **358**(2): p. 559-70.

116. Rouet, P., F. Smih, and M. Jasin, *Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease.* Mol Cell Biol, 1994. **14**(12): p. 8096-106.

117. Sung, P. and H. Klein, *Mechanism of homologous recombination: mediators and helicases take on regulatory functions.* Nat Rev Mol Cell Biol, 2006. **7**(10): p. 739-50.

118. Weterings, E. and D.J. Chen, *The endless tale of non-homologous end-joining.* Cell Res, 2008. **18**(1): p. 114-24.

Vita

In 1998, Hong Luo entered the University of Science and Technology of China in Hefei, Anhui, China, majoring in biological sciences and received the degree of Bachelor of Science in July 2003. In September 2004, she entered the school of biological sciences in the University of Texas at Austin and started her graduate study in the section of Molecular Genetics and Microbiology. In December 2007, she received the degree of Master of Arts.

Permanent email address: hongluo09@gmail.com

This dissertation was typed by Hong Luo.