ABSTRACT

STONE, JENNIFER LYNN. Feedback Control of Tcrb Gene Recombination. (Under the direction of Dr. Michael L. Sikes).

To contend with a multitude of continually-evolving pathogens, T and B lymphocytes generate de novo the genes encoding billions of distinct antigen receptors by V(D)J recombination. This unique process of gene assembly is driven by promoters and enhancers that regulate antigen receptor chromatin structure to facilitate or restrict recombinase access to individual DNA target sequences. Each developing T cell must produce one successfully-rearranged Tcrb allele and one Tcra allele. The Tcrb gene locus is organized such that two DJ cassettes lie downstream of numerous V segments, and is assembled via sequential D-to-J then V-to-DJ rearrangement. Recombination of the downstream cassette (DJβ2) is less efficient, forcing initial V-to-DJ rearrangements to target the upstream DJβ1. Should the primary Vβ-to-DJβ1 rearrangement be frameshifted and nonfunctional, secondary rearrangement between a more distal Vβ segment and DJβ2 ensures that each allele has two independent chances to form an in-frame rearrangement, thus maximizing pro-T cell survival and TCRβ diversity. Therefore, the epigenetic mechanisms that restrict initial DJβ2 recombinational accessibility provide a central mechanism for regulating TCRβ repertoire.

Previously, our laboratory identified two promoters flanking DJβ2. Prior to recombination, the downstream promoter (3’PDβ2) is activated, ensuring full recombinase access to Jβ2, but only limited access to Dβ2. Though a more effectively-positioned promoter (5’PDβ2) sits immediately upstream of Dβ2, it is repressed prior to D-to-J recombination. In this study, we investigate the epigenetic mechanisms that initially restrict Dβ2 accessibility. We show that repression of 5’PDβ2 is mediated by the binding of the stress-response transcription factor
upstream stimulatory factor 1 (USF1) to a non-canonical E box within the 5’ Dβ2 12-RSS in the absence of recombination. DNA double-strand breaks (DSBs) generated during Tcrb rearrangement in early thymocyte development lead to a loss of USF1-mediated repression and, consequently, a gain of 5’PDβ2 promoter activity in a DNA-dependent protein kinase catalytic subunit (DNA-PKcs)-dependent manner. Additionally, we have defined the signaling pathway through which USF1-mediated repression is resolved. We demonstrate that DNA-PK phosphorylates USF1 at the serine 262 (S262) residue, leading to USF1’s displacement from the repressor. Importantly, the loss of USF1 corresponds with a gain in 5’PDβ2 expression and an increase in recombinational accessibility at Dβ2. Our findings suggest a feedback mechanism where initial rounds of rearrangement activate repair factors that, in turn, increase accessibility for subsequent recombination events. Finally, we show an increase in aberrant D-to-D joins in thymocytes from USF1−/− mice, suggesting a novel role for USF1 in limiting D-to-D recombination. Collectively, our results identify USF1 as an important trans-acting regulator of recombinational accessibility and reveal a functional consequence for USF1 on Tcrb rearrangement.
Feedback Control of Tcrb Gene Recombination

by
Jennifer Lynn Stone

A dissertation submitted to the Graduate Faculty of North Carolina State University in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Microbiology

Raleigh, North Carolina

2015

APPROVED BY:

__________________________________________
Michael L. Sikes, Ph.D.
Chair of Advisory Committee

__________________________________________
Scott M. Laster, Ph.D.

__________________________________________
Frank Scholle, Ph.D.

__________________________________________
Max J. Scott, Ph.D.
DEDICATION

I would like to dedicate this manuscript to Mom, Dad, and Jeff, for your endless love and support, and for always being my biggest cheerleaders. Thank you to my parents for always emphasizing the value of education and a strong work ethic, for pushing me to succeed, and for working incredibly hard to provide me with every opportunity I could ever wish for.

Thank you, Jeff, for being the best brother and friend I could ask for. You really are one of the most special people to me.

I also wish to dedicate this to Stephan, for your constant support and remarkable patience throughout my graduate studies. You helped make this dissertation possible and, for that, I will always be grateful.
BIOGRAPHY

Jenn was born and raised in the quaint New England town of Keene, NH. She attended the University of New Hampshire in Durham and obtained a Bachelor of Science in Biology in 2004. After returning to Keene and working for a year at Smiths Medical, she left snowy New England for Greenville, NC. There she joined the pharmaceutical company Metrics, Inc., where she worked as an analytical chemist and analytical quality auditor from 2005 to 2009. Aspiring to advance her career as a research scientist, she ultimately decided to continue her education at North Carolina State University in Raleigh, where she pursued her Ph.D. in Microbiology. In addition to her love for science, her passions include traveling, photography, outdoor adventure, volunteering, and spending time with family and friends.
ACKNOWLEDGMENTS

I would like to express my sincere appreciation to all of the individuals who have supported me through the doctoral program at NC State. Foremost, I would like to thank Dr. Michael Sikes for his guidance, friendship, and encouragement throughout my graduate studies. I feel extremely lucky to have spent my time at NCSU under his wing. I would also like to thank my committee members, Dr. Scott Laster, Dr. Max Scott, and Dr. Frank Scholle for their valuable advice and support. You have all challenged me to become a better scientist.

Thank you also to my fellow grad student, Susan Gardner, for being a thoughtful and supportive friend during our journey at NC State. Additionally, special thanks to Wiley and Jesse for our regular play dates. I immensely enjoyed our days of drawing, Play-doh™ sculpting, paper airplane flying, car racing, Frisbee™ throwing, and cartoon watching.

Lastly, I would like to express the deepest gratitude to my wonderful family and friends. Thank you to all my extended family members for being such a huge influence in my life, and a special thanks to my grandparents – Nana, Dar, Bummy & Stoney – for your never-ending encouragement and unconditional love. I wish to also acknowledge my dear friends, Kelly, Kara, and Jenny, who have been by my side since our days of hanging out at the Friendly Farm, building sand castles on the Cape, climbing trees, cracking rocks, and ice skating on the pond. I can’t begin to describe how lucky I am to have all of you.
# TABLE OF CONTENTS

**LIST OF TABLES** ........................................................................................................... vii  
**LIST OF FIGURES** ....................................................................................................... viii  

## CHAPTER 1: Literature Review .................................................................................. 1  
1 Introduction ......................................................................................................................... 2  
2 Overview of the Immune System ........................................................................................... 3  
  2.1 Cells of the Adaptive Immune Response ........................................................................ 4  
  2.2 Antigen Receptors ......................................................................................................... 4  
3 Generation of Antigen Receptors ........................................................................................ 7  
  3.1 Germline Organization of TCR and Ig Loci .................................................................... 7  
  3.2 V(D)J Recombination ..................................................................................................... 10  
    3.2.1 Recombinase & RSSs ............................................................................................... 11  
    3.2.2 DNA Repair Machinery ......................................................................................... 16  
4 Accessibility Model of Recombination .............................................................................. 18  
  4.1 Chromatin Environment ............................................................................................... 19  
  4.2 Histone Modifications ................................................................................................... 20  
  4.3 DNA Methylation ......................................................................................................... 21  
  4.4 Epigenetic Regulation within Antigen Receptor Loci ................................................... 21  
5 αβ T Cells .......................................................................................................................... 23  
  5.1 αβ T cell Development & TCRβ Assembly .................................................................... 24  
  5.2 Epigenetic Control at Dβ1 ............................................................................................ 28  
  5.3 Regulation of Dβ2 Rearrangement ............................................................................... 31  
  5.4 Beyond the 12/23 Rule ................................................................................................. 35  
6 Upstream Stimulatory Factor 1 ........................................................................................ 39  
  6.1 USF1 and the Immune Response ............................................................................... 41  
  6.2 Regulation at Antigen Receptor Loci ........................................................................... 43  
7 References ......................................................................................................................... 46  

## CHAPTER 2: DNA double-strand breaks relieve USF-mediated repression of Dβ2 germline transcription in developing thymocytes ......................................................... 66  

Abstract ............................................................................................................................... 67  
1 Introduction ......................................................................................................................... 68  
2 Materials and Methods ...................................................................................................... 72  
  2.1 Cells and antibodies ..................................................................................................... 72  
  2.2 Plasmids and transient transfection ............................................................................ 73  
  2.3 Germline transcription and recombination .................................................................. 74  
  2.4 CpG methylation ......................................................................................................... 74  
  2.5 Chromatin immunoprecipitation ............................................................................... 75  
  2.6 EMSA ....................................................................................................................... 75  
3 Results ............................................................................................................................... 76  
  3.1 Repressed 5’PDβ2 remains accessible in DN thymocytes ........................................ 76  
  3.2 5’PDβ2 repression requires an E box in the Dβ2 12-RSS ........................................... 81
### CHAPTER 3: USF1 regulates feedback control of Dβ2 rearrangement to enforce B12/23 at the level of D-to-D restriction

**Abstract**

1 **Introduction**

2 **Materials and Methods**
   2.1 Cell culture
   2.2 Plasmids and transient transfection
   2.3 Germline transcription
   2.4 Rearrangement analysis in *USF1*−/− mice
   2.5 Southern analysis
   2.6 Sequence analysis of D-to-D joins
   2.7 RIC analysis of potential RSSs

3 **Results**
   3.1 Phosphorylation of USF1 at S262 leads to 5’PDβ2 activation
   3.2 DJβ2 recombination is unchanged in *USF1*−/− adult thymocytes
   3.3 Loss of USF1 increases D-to-D rearrangement in adult thymocytes

4 **Discussion**

5 **References**

6 **Supplemental Data**

---

### APPENDIX I: CTCF is not involved with epigenetic silencing at the TEA promoter in ES cells

1 **Introduction**

2 **Materials and Methods**
   2.1 Cells and antibodies
   2.2 RNAi knockdown of CTCF in ES cells
   2.3 Transcriptional analysis
   2.4 Chromatin immunoprecipitation

3 **Results**
   3.1 TEA promoter is accessible in DP thymocytes
   3.2 CTCF binds the TEA promoter prior to the DP stage
   3.3 CTCF does not direct histone modifications at the TEA promoter

4 **Discussion**

5 **References**
LIST OF TABLES

CHAPTER 2: DNA double-strand breaks relieve USF-mediated repression of Dβ2 germline transcription in developing thymocytes

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>D-to-J recombinant sequence in clones c20 and c22</td>
<td>79</td>
</tr>
<tr>
<td>2</td>
<td>Percent methylation of CpG dinucleotides surrounding Dβ2</td>
<td>80</td>
</tr>
<tr>
<td>S1</td>
<td>PCR primer sets</td>
<td>102</td>
</tr>
<tr>
<td>S2</td>
<td>EMSA primer sets</td>
<td>103</td>
</tr>
</tbody>
</table>

CHAPTER 3: USF1 regulates feedback control of Dβ2 rearrangement to enforce B12/23 at the level of D-to-D restriction

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sequence analysis of D-to-D joins in wild-type and USF1−/− murine thymocytes</td>
<td>124</td>
</tr>
<tr>
<td>2</td>
<td>RIC analysis of putative RSSs</td>
<td>126</td>
</tr>
<tr>
<td>S1</td>
<td>Oligonucleotides for PCR &amp; Southern analysis</td>
<td>136</td>
</tr>
</tbody>
</table>

APPENDIX I: CTCF is not involved with epigenetic silencing at the TEA promoter in ES cells

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>PCR primer sets</td>
<td>156</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

CHAPTER 1: Literature Review

Figure 1 T and B lymphocyte receptors ................................................................. 5
Figure 2 Germline structure of the mouse antigen receptor loci ............................... 8
Figure 3 Mechanism of V(D)J recombination ......................................................... 11
Figure 4 Recombination signal sequences within the antigen receptor loci ................. 12
Figure 5 Non-homologous end joining repair pathway ............................................. 17
Figure 6 Ordered rearrangement of antigen receptor loci during \( \alpha \beta \) T cell thymopoiesis......................................................................................... 25
Figure 7 Mechanism of Tcrb rearrangement ................................................................ 26
Figure 8 Promoter control at D\( \beta \)2 ........................................................................... 33
Figure 9 Protein structure of murine upstream stimulatory factor 1 (USF1) ................. 40

CHAPTER 2: DNA double-strand breaks relieve USF-mediated repression of D\( \beta \)2 germline transcription in developing thymocytes

Figure 1 DJ\( \beta \)2 rearrangements relieve repression at both germline and rearranged DJ\( \beta \)2 clusters ............................................................................................. 78
Figure 2 Repression does not alter histone modifications at D\( \beta \)2 ................................ 81
Figure 3 Repression of 5’PD\( \beta \)2 promoter activity requires cis targets in the D\( \beta \)2 5’RSS .................................................................................................................. 82
Figure 4 USF1/2 bind the repressor site in the D\( \beta \)2 5’RSS spacer .............................. 84
Figure 5 USF1 binds the repressor in the absence of recombination .......................... 85
Figure 6 Genotoxic DSBs relieve 5’PD\( \beta \)2 repression .............................................. 88
Figure 7 Genotoxic DSBs induce USF1 binding and loss of cad expression in P5424 cells .................................................................................................................. 89
Figure 8 5’PD\( \beta \)2 repression is restored in c22 by inhibition of DNA-PKcs ................. 91

CHAPTER 3: USF1 regulates feedback control of D\( \beta \)2 rearrangement to enforce B12/23 at the level of D-to-D restriction

Figure 1 Phosphomimetic and dominant-negative forms of USF increase 5’D\( \beta \)2 expression ........................................................................................................ 116
Figure 2 Loss of USF1 does not impact D\( \beta \)2\( \beta \)2 rearrangement in unsorted adult thymocytes ......................................................................................................................... 118
Figure 3 Expression from 5’PD\( \beta \)2 is increased in thymocytes from USF1\(^{-} \) mice .............................................................. 119
Figure 4 Loss of USF1 increases inappropriate D-to-D rearrangement in unsorted adult thymocytes .................................................................................................................. 121
Figure 5 D-to-D rearrangement is evident in murine adult thymocytes ..................... 122
CHAPTER 4: Summary

Figure 1  Mechanism of Dß2 regulation ................................................................. 140
Figure 2  A role for USF1 in B12/23 regulation .................................................. 142

APPENDIX I: CTCF is not involved with epigenetic silencing at the TEA promoter in ES cells

Figure 1  TEA promoter is associated with enhancer-dependent activating histone marks in DP cells .......................................................... 157
Figure 2  CTCF binds the TEA promoter in ES cells ............................................ 159
Figure 3  Knockdown levels of CTCF in EF1 subclones ...................................... 160
Figure 4  Loss of CTCF does not impact chromatin organization at the TEA promoter .......................................................... 161
CHAPTER 1

Literature Review
1 Introduction

Mammalian immune systems have evolved to defend the host from a vast number of potential pathogens including bacteria, viruses, and parasites. The immune response is comprised of two components – the innate response and the adaptive response. T and B lymphocytes are the fundamental cells of adaptive immunity and provide an exceedingly specific defense against foreign antigens through the generation of billions of unique antigen receptors. These receptors are created by V(D)J recombination, which involves the site-specific recombination of dozens of gene segments organized within T cell- and B cell-specific antigen receptor loci. The focus of this dissertation is on the regulatory mechanisms of V(D)J recombination within the T cell receptor Tcrb locus.

While the process of rearrangement generates an enormous diversity of antigen-specific receptors, genetic restructuring is inherently dangerous and can lead to chromosomal aberrations and, consequently, lymphoid malignancies. Thus, recombination must be tightly regulated to ensure tissue and stage specificity, as well as ordered rearrangement. The Tcrb locus is organized such that two DJ cassettes lie downstream of numerous V segments, and is assembled via sequential D-to-J then V-to-DJ rearrangement. Previous research has shown that recombination of the downstream cassette (DJβ2) is less efficient, forcing initial V-to-DJ rearrangements to target the upstream DJβ1 cluster. This ordering prevents deletion of the DJβ2 cassette during the first round of V-to-DJ joining, ensuring that each allele has two independent chances to form an in-frame rearrangement, and maximizing pro-T cell survival and TCRβ diversity. The focus of my research is on the epigenetic mechanisms that restrict initial DJβ2 recombinational accessibility and likely provide a central mechanism for regulating TCRβ repertoire. The research herein describes a role for the transcription factor
USF1 in regulating ordered assembly within the Dβ2 region. This review describes the epigenetic mechanisms that direct accessibility within the T cell receptor loci, and emphasizes the importance of Dβ2 control in receptor diversity. Emphasis is also placed on the role of USF1 in the mammalian immune system, particularly within the antigen receptor loci.

2 Overview of the Immune System

The immune system defends against a host of invading bacteria, viruses, and parasites that pose immediate or long-term risks to human health. This is accomplished by two distinct branches of the immune system – the innate (natural) system and the adaptive (acquired) system. Innate immunity provides an immediate, non-specific, anti-microbial response against invading pathogens. Within the innate system, professional phagocytic cells including macrophages, dendritic cells, neutrophils, and monocytes express germline-encoded pattern recognition receptors (PRRs) that recognize conserved microbial structures known as pathogen-associated molecular patterns (PAMPs) (1, 2). While this system provides an effective broad response to microorganisms, it lacks specificity to recognize the huge number of unique foreign antigens. This limitation is overcome by the adaptive system. Although initially a more protracted response than the innate counterpart, the adaptive system generates an antigen-specific response that results in long-term protective immunity and immunological memory (3). This acquired response is driven by T and B lymphocytes which, unlike cells of the innate system, must undergo somatic rearrangement of their germline DNA to generate the nearly unlimited receptor diversity needed to distinguish billions of distinct antigens.
2.1 Cells of the Adaptive Immune Response

T and B lymphocytes are the primary cells of the adaptive immune system and are morphologically and functionally similar. Both cell types arise from a common lymphoid progenitor (CLP) of hematopoietic origin in the bone marrow. B cell progenitors (pro-B cells) continue development in the bone marrow, while early T cell progenitors (ETP) migrate to the thymus, where they initiate their developmental program. Immature B and T cells undergo parallel rearrangement of antigen receptor gene segments during subsequent stages of development. Mature B cells present antigen to T cells, but are also responsible for the antibody component of humoral immunity. T lymphocytes, on the other hand, play a central role in cell-mediated immunity. In the final stages of development, thymocytes differentiate into immunocompetent CD8+ cytotoxic T cells, which directly lyse target cells, or CD4+ helper T cells, which are critical for the activation of B cells, cytotoxic T lymphocytes, and innate effector cells. Together, T and B lymphocytes are the foundation of the adaptive immune system and are indispensable for protective immunity.

2.2 Antigen Receptors

T and B lymphocytes are able to mount highly specific responses to a multitude of pathogens through the expression of antigen-specific receptors. T lymphocytes express surface-bound T cell receptors (TCRs), while B cell receptors (BCRs) can either be displayed on the surface of mature B cells or secreted as antibodies by terminally-differentiated plasma cells. The BCR consists of four immunoglobulin (Ig) polypeptide chains covalently linked by disulfide bonds. Antibodies and membrane-bound immunoglobulins differ only with regard to the inclusion of an alternatively-spliced membrane anchor. Each BCR is comprised of two heavy chains (IgH) and two light chains (IgL). The heavy chain contains a single variable
region (V_H) and multiple constant regions (C_H), while the light chain consists of a single variable (V_L) and constant (C_L) domain (Fig. 1). The carboxy-terminal portion of the heavy chain constant region is referred to as the fragment crystallizable (Fc), and defines the isotype of antibody (IgA, IgD, IgE, IgG, or IgM). This region is connected to the amino-terminal antigen-binding fragment (Fab) through a flexible, disulfide-linked hinge. Each of the two F(ab) arms contains a single constant and variable region, with the antigen-binding site at the distal end of the N-terminus.

**Figure 1.** T and B lymphocyte receptors. The T cell receptor (left panel) is comprised of two membrane-bound disulfide-linked heterodimeric chains, each with a constant region and variable region. The hypervariable region at the end of the receptor is the site of antigen recognition. This receptor is analogous in structure to the F(ab) fragment of a B cell receptor (BCR) (right panel). The TCR consists of an α chain (blue) and β chain (light green), analogous to the BCR light chain (blue) and heavy chain (light green), respectively.

The TCR is analogous in structure to the F(ab) portion of an antibody (Fig. 1). It contains two polypeptide chains, most commonly TCRα and TCRβ, linked by disulfide
bonds at the C-terminal transmembrane region. Each α and β chain contains a single constant (Cα or Cβ) and variable (Vα or Vβ) region. Unlike B cell receptors that can be secreted as antigen-binding antibodies, TCRs are exclusively surface-bound, with each individual cell expressing approximately 30,000 – 40,000 clonal receptors (4).

Although BCRs and TCRs are structurally similar, there are key differences in their mechanisms of antigen recognition. BCRs recognize both continuous and discontinuous epitopes – native, folded molecules of proteins, nucleic acids, lipids, or carbohydrates. The TCR, conversely, binds only continuous epitopes, small polypeptide chains. Additionally, while the BCR can directly bind free antigen, the TCR can only recognize peptide antigen complexed with major histocompatibility complex (MHC) molecules displayed on the surface of antigen-presenting cells (5).

The antigen-binding sites of T and B cell receptors interact closely with a host of foreign substances, and diversity within this region is generated by the somatic rearrangement of multiple V, D, and J gene segments. The antigen-binding site of each TCR and BCR polypeptide chain is made up of four framework (FR) regions interspersed with three hypervariable (HV) regions. FR1-4 are β-sheets that provide structural support at the site of antigen interaction, while HV1-3 form loops of concentrated variability termed complementarity-determining regions (CDRs). CDR1, 2, and 3 loops are juxtaposed upon protein folding to form a single hypervariable surface that directly contacts the antigen. CDR1 and CDR2 are composed entirely of the V segment germline sequence, while CDR3 is the most diverse, containing the region encompassing the VJ or VDJ junction of rearranged gene segments (6, 7).
3 Generation of Antigen Receptors

T and B lymphocytes generate de novo the genes encoding billions of distinct antigen receptors by V(D)J recombination. As the human genome contains no more than 25,000 genes (8), we depend on this unique, site-specific recombination process to provide the coding capacity for such enormous receptor diversity. During T and B cell development, antigen receptor genes are assembled from a number of (V), diversity (D), and joining (J) exonal gene segments by V(D)J recombination (9). This unique assembly is driven in part by transcriptional control elements such as promoters and enhancers that alter chromatin structure at recombinase targets to drive assembly of the antigen receptor loci.

3.1 Germline Organization of TCR and Ig Loci

T-cell receptors and B cell immunoglobulins are encoded by seven antigen receptor loci. There are two classes of T lymphocytes – αβ and γδ T cells – encoded by the Tcra/b loci and Tcrg/d loci, respectively. Similarly, B cell immunoglobulins are encoded by the heavy chain locus (Igh), and one of two light chain loci, kappa (Iγk) or lambda (Igl). Germline configurations of the four antigen receptor loci discussed in this review are depicted in Figure 2. γδ T cells, a specialized subset of T lymphocytes involved in mucosal immunity, are beyond the scope of this review and will not be detailed herein.

Each antigen receptor locus is comprised of variable (V) and joining (J) or, in some cases, V, diversity (D), and J exonal gene segments interposed between non-coding DNA sequence. Additionally, each locus contains at least one enhancer element, as well as multiple promoters to drive both germline transcription and transcription following rearrangement (Fig. 2) (10). Although the organization and number of gene segments varies between
species, the antigen receptor loci structure and overall mechanism of V(D)J recombination is evolutionarily well-conserved within the jawed vertebrates (11-13).

Figure 2. Germline structure of the mouse antigen receptor loci. Each locus is made up of various V and J or V, D, and J gene segments. Regulatory elements including promoters (green diamonds with arrows) and enhancers (green circles) are dispersed throughout the vast loci to control transcription and accessibility. Promoters associated with each V segment are not shown.

The BCR is comprised of heterodimers of a heavy chain (IgH) and light chain (Igκ or Igλ). Located on chromosome 12 in the mouse, the IgH locus encompasses ~2.5 Mb and contains 150 V₃H, 9 D₃H, and 4 J₃H segments. Germline promoters are situated immediately upstream of each V₃H gene segment, and the PDQ52 promoter sits just upstream of the most 3’ D₃H segment. Germline promoters likely exist for each of the 9 D₃H segments, though they have not been experimentally defined. The locus also contains two enhancers – Eμ is
positioned upstream of the constant region coding exons, and the 3’RR is at the distal 3’ end of the locus. The mouse Igκ locus spans ~3.5 Mb on chromosome 6 with 140 Vκ gene segments and their associated promoters dispersed over 3.2 Mb of the locus. Two promoters, the distal and proximal, separate the Vκ region from 4 downstream Jκ segments. Downstream of Jκ, two enhancer elements, iEκ and 3’Eκ bracket a single Cκ exon.

The heterodimeric αβ TCR is made up of an α and β chain and is analogous to the BCR in structure. Similar to IgH organization, the Tcrb locus contains V, D, and J segments and extends over 700 kb of mouse chromosome 6. 30 Vβ segments and their corresponding promoters are spread over ~380 kb in the 5’ region. Two DβJβ clusters are positioned 250 kb downstream, each containing one Dβ segment, 6 functional Jβ segments, and one Cβ constant region coding segment. A single promoter, PDβ1 sits immediately 5’ of the upstream DJβ1 cassette. At the downstream DJβ2 cluster, two germline promoters flank the Dβ2 segment. 5’PDβ2 is positioned upstream of Dβ2, while 3’PDβ2 lies between Dβ2 and the Jβ2 elements. Tcrb contains a lone enhancer, Eβ, downstream of the DJβ region. Finally, a single inverted Vβ segment, Vβ14, is located at the distal 3’ end of the locus.

Gene segments encoding the TCRα chain are interlaced with those of TCRδ in a single locus spanning 1.6 Mb of mouse chromosome 14. The Tcra/d locus is composed of 100 Vα/δ segments and associated promoters, some contributing to both αβ and γδ repertoires, and 61 Jα segments. Dδ and Jδ segments and their associated regulatory elements are situated between the Vα and Jα regions. At least two germline promoters, TEA and Jα49 reside within the Jα region, though others are predicted. A single constant-coding region (Cα) is situated downstream of the Jα elements, and a lone enhancer (Eα) occupies the 3’ end of the locus.
3.2 V(D)J Recombination

The V, D, and J gene segments encoding the antigen receptor loci are unique in that, in the germline configuration, the coding portions are entirely non-functional. Each locus must undergo a lineage-specific, stage-specific reordering of multiple gene segments to generate the variable region of the T-cell or B-cell receptor. The first conclusive evidence that these coding segments undergo a specialized type of somatic rearrangement was provided by the Tonegawa group, who used restriction enzyme mapping to reveal recombination products within the Ig\(\lambda\) light chain locus (14). Their findings laid the foundation for understanding the complex DNA recombination program that is carried out during lymphopoiesis.

During B cell development, Ig\(h\) rearranges first in a two-step process where D segments rearrange initially with J elements, followed by rearrangement of a V to a pre-formed DJ complex. As the cell transitions from the pro-B to the pre-B stage, the light chain genes (first Ig\(k\), then Ig\(l\)) undergo V-to-J rearrangement. Recombination follows an analogous pattern in precursor T cells. Tcr\(b\) rearranges first in CD4\(^+\)CD8\(^-\) double negative (DN) thymocytes, where D-to-J joining precedes that of V-to-DJ. After progressing to the CD4\(^+\)CD8\(^+\) double positive (DP) stage, Tcr\(a\) then completes V-to-J rearrangement.

V(D)J recombination (Fig. 3) is initiated by a lymphocyte-specific recombinase, encoded by the recombination activating genes 1 and 2 (RAG-1 and RAG-2), that targets highly-conserved recombination signal sequences (RSSs) flanking each coding segment (15). The RAG complex cleaves DNA at the junction of the coding segment and corresponding RSS, creating hairpin coding ends and blunt signal ends. DNA repair machinery processes and resolves the transient breaks, and the resulting VJ or VDJ coding joint will encode a
distinct receptor variable region. Finally, the novel variable region exon is joined to the constant domain exon by RNA splicing of the primary transcript. This generates an mRNA transcript that will be translated into a single polypeptide, ultimately resulting in a unique, clonal antigen receptor.

Figure 3. Mechanism of V(D)J recombination. Recombination is driven by a lymphocyte-specific recombinase (RAG-1/2) that cleaves DNA at conserved RSSs flanking each gene segment. Cleavage results in hairpin coding ends as well as blunt signal ends containing the RSSs and intervening DNA. Repair machinery opens, processes, and religates hairpin ends to form a functional coding joint, while blunted signal joints are religated and deleted from the genome as circular episomes.

3.2.1 Recombinase & RSSs

Recombination is mediated by specific, conserved sequences of DNA, termed recombination signal sequences (RSSs), that border all antigen receptor loci gene segments (9, 16). A palindromic heptamer with the consensus 5’-CACAGTG-3’ is immediately proximal to the coding element. This is separated from an AT-rich nonamer with the consensus motif 5’-ACAAAAACC-3’ by a short spacer sequence consisting of either 12 or
23 base pairs (Fig. 4) (9, 16). While the spacer sequence is comparatively unconserved, spacer length is critical for efficient rearrangement. Because the lengths correlate with one or two turns of the DNA helix, it is likely that maintaining this distance is required to align the heptamer and nonamer on the same helical face for RAG binding. Indeed, mutational experiments using plasmid recombination substrates in a pre-B cell line revealed that spacer lengths had a tolerance of ± 1 base pair, as recombination frequency was drastically reduced when either 12- or 23-bp spacer lengths were altered by >1 nucleotide (17).

**Figure 4.** Recombination signal sequences within the antigen receptor loci. RAG-1/2 recombinase targets RSSs flanking each gene segment. Only gene segments with compatible RSSs (12- and 23-bp spacers) will be able to form a coding join.

Rearrangement only occurs between gene segments containing RSSs with dissimilar spacer lengths, a principle known as the 12/23 rule (9). A gene segment with a 12-RSS, therefore, can only join with one flanked by a 23-RSS. In the heavy chain locus, for example,
a D<sub>H</sub> segment can rearrange to a J<sub>H</sub> segment; likewise a V<sub>H</sub> segment can join with a D<sub>H</sub> segment. A V<sub>H</sub> segment, however, is prohibited from joining directly to a J<sub>H</sub> segment as both are flanked by 23-RSSs (Fig. 4). Rearrangement follows a similar pattern in the other antigen receptor loci, although only one recombination event (V-to-J joining) is required in the light chain and T<sub>cr</sub>a loci.

The cleavage activity of V(D)J recombinase is only expressed in developing T and B cells (18), and is encoded by RAG-1 and -2. Experiments in the Baltimore lab initially demonstrated that genomic DNA (gDNA) fragments from a human B cell line carried the information necessary to activate rearrangement in non-lymphoid NIH 3T3 fibroblasts containing a recombination substrate (19). Using oligonucleotide tagging, the research group was able to specifically identify the RAG-1 gene. Similar follow-up experiments using RAG-1 cDNA clones recapitulated results obtained with gDNA. Rearrangement levels were significantly less than the expected threshold, however, suggesting the involvement of an additional factor (20). Based on these findings, the Baltimore group was then able to similarly identify and isolate the adjacent RAG-2 gene. Recombinase activity assays using co-transfected RAG-1 and RAG-2 increased recombination frequency by 1000-fold, suggesting that closely-linked RAG-1/2 cooperate to activate rearrangement (21). Indeed, both RAG-1 and RAG-2 are absolutely required for lymphocyte development and recombination, as targeted mutations of either gene in mice result in a severe combined immunodeficiency (SCID) phenotype (22, 23).

In order to induce DNA cleavage, the RAG proteins must first recognize and bind to RSS elements. A considerable body of in vitro and in vivo investigations supports a capture model, where RAGs preferentially assemble at a 12-RSS and capture a 23-RSS, forming a
synaptic complex (24-27). Recent oligo-capture assays within the Tcrb locus found evidence further supporting the capture model, although these data suggest that a 23-RSS can equally capture a 12-RSS to form a paired complex (28). RAG-1 is the principal factor involved with DNA binding and cleavage, while RAG-2 functions as an important co-factor to enhance the targeting and catalytic activity of RAG-1. RAG-1 initially binds the nonamer of the RSS by means of an amino-terminal nonamer-binding domain (29-31). This interaction is critical for anchoring the recombinase complex to the RSS. RAG-1 is then able to make contact with the heptamer, albeit only weakly. Occupancy at the heptamer is considerably stabilized by RAG-2, which induces a conformational change in RAG-1 near the site of cleavage (30, 32).

While RAG-2 lacks any independent DNA binding activity (31), the carboxy-terminal region contains a non-canonical plant homeodomain (PHD) that binds hypermethylated lysine 4 residues of histone H3 (H3K4me3), an epigenetic mark associated with transcriptional activation (33, 34). The PHD finger facilitates RAG recruitment to accessible RSSs and also enhances the catalytic activity of the recombinase complex (35). Mutations within the PHD that prevent H3K4me3 recognition by RAG-2 have been shown to severely impair rearrangement (34, 36). Notably, mutation of a highly-conserved tryptophan residue, W453 (37), abrogated binding to H3K4me3 and, consequently, led to a profound decrease in rearrangement on both the endogenous Igh locus and extrachromosomal substrates (34). Moreover, a W453R mutation of RAG-2 has been observed in patients with Omenn’s Syndrome, a severe form of immunodeficiency (36, 38).

Both RAG-1 and -2 proteins undoubtedly contain important regulatory regions and, together, the two recombinase components are the sole factors required for RSS recognition and subsequent cleavage at the heptamer::coding segment junction (39, 40). RAG activity
must be stringently regulated, as the introduction of DNA breaks poses inherent risks to genome integrity (41). To minimize the risk of aberrant recombination events, evidence suggests that RAGs are recruited exclusively to foci of open chromatin and highly accessible RSSs, termed “recombination centers” (15, 42, 43), where cleavage and repair can be tightly synchronized. Chromatin immunoprecipitation (ChIP) analyses reveal that recombination centers of RAG binding include J segments (*Tcra, Igk*) or D-J segments (*Tcrb, Igh*) and are considerably enriched in H3K4me3, acetylated histone H3, and RNA polymerase II, hallmarks of accessible chromatin and active transcription (42). V segments are repositioned near the recombination center by large-scale DNA looping and chromatin reorganization, and are then captured in the synaptic complex (15, 42).

DNA cleavage is catalyzed by RAG-1 through an acidic DDE motif within the active site of the core region (44-46). Recent findings suggest that AMPK-mediated phosphorylation of a serine residue (S528) within its core region enhances the catalytic activity of RAG-1 (47). Although RAG-2 does not contain its own catalytic center, its stabilization of the RAG complex is critical for enhancing cleavage by RAG-1 (48). RAG-mediated cleavage at compatible RSSs occurs in two steps: (1) nicking of the DNA and (2) hairpin formation of coding ends (39, 40). RAG-1, in synergy with RAG-2, introduces a single-strand nick between the coding segment and the RSS heptamer sequence. This is followed immediately by a transesterification reaction, where the free 3’OH initiates a nucleophilic attack on the second strand to generate a double-strand break in the DNA (49). This results in covalently-sealed hairpin intermediates of the coding ends, and blunt, 5’-phosphorylated signal ends containing the RSSs and intervening DNA between the coding segments. After cleavage, the RAG proteins remain associated with the signal sequence in a
post-cleavage synaptic complex. RAG-1 itself, via the N-terminal domain, can facilitate recruitment of initial DNA repair factors, Ku70/Ku80, to the break site (50). Likewise, RAG-2 may potentially recruit classical repair apparatus through its C-terminal domain (51). DNA repair machinery is ultimately able to resolve the coding ends to produce a functional coding joint, while concurrently ligating signal ends to form episomal signal joints that are excised from the genome during cell division.

3.2.2 DNA Repair Machinery

DNA double strand breaks (DSBs) pose one of the most significant threats to cellular fitness (52). Cells are routinely exposed to genotoxic pressures such as UV light, ionizing radiation, and oxidative stress that compromise genome stability through induction of DSBs (53). The consequences of such lesions include apoptotic cell death as well as chromosomal aberrations that can lead to numerous human cancers and disorders (54-57). Although a considerable risk to genomic integrity, the physiological generation of DSBs in developing lymphocytes is imperative to the assembly of antigen receptor genes.

After introduction of DNA DSBs by the RAG recombinase complex, DNA must be promptly and efficiently repaired to prevent deleterious chromosomal translocations. During V(D)J recombination, breaks are resolved via the ubiquitous non-homologous end joining (NHEJ) repair pathway (Fig. 5). Repair is initiated when Ku70/Ku80 heterodimers colocalize at the break site. Evidence suggests that Ku70 and Ku80 physically interact with the RAG:RSS complex at the RAG1 N-terminal domain (50). A direct association between the recombinase and repair proteins provides a convenient mechanism for RAG DSBs to be sequestered and rapidly repaired, preventing translocations that could compromise genomic integrity. Upon recruitment to the DSB site, Ku70/Ku80 are threaded onto free DNA hairpin
ends, providing a scaffold for the DNA-dependent protein kinase catalytic subunit (DNA-PKcs) (58-60). DNA-PKcs is then able to tether the two DNA coding ends together to facilitate repair (61). Importantly, upon association with both DNA and the Ku70/80 scaffolding, DNA-PK activates other repair components through its serine/threonine kinase activity (62-64). A related kinase, ataxia telangiectasia (ATM), colocalizes to the DNA break and cooperates with DNA-PK to phosphorylate factors involved with the DNA damage response (DDR) (65, 66). While DNA-PK and ATM share some downstream targets (66), ATM independently phosphorylates p53 to induce cell cycle arrest at G0/G1 (67, 68). This developmental pause ensures that the cell can completely resolve DSBs prior to proliferation.

**Figure 5.** Non-homologous end joining repair pathway. Following RAG-mediated double strand break induction, DNA repair factors are recruited to the break site. Compatible hairpin coding ends are opened, processed, and ultimately ligated together. Inaccurate repair, including the indiscriminate addition or removal of nucleotide bases, increases junctional diversity at the coding joints.
Ultimately, DNA-PK phosphorylates the protein Artemis, resulting in a gain of endonuclease activity (69). Artemis can then open hairpin ends in cooperation with nuclease activity from the MRN complex (comprised of Mre11-Rad50-Nbs1 homodimers) (70). Finally, DNA end processing is completed by the addition of non-templated nucleotides by terminal deoxynucleotidyl transferase (TdT) (71, 72). While this results in inaccurate repair of the coding joint, the random modifications to the DNA sequence afford increased junctional diversity and, consequently, increased diversity in the overall antigen receptor repertoire. After hairpins have been opened and processed, XRCC4 and DNA Ligase IV finalize DSB repair. XRCC4 recruits DNA ligase to the site, which catalyzes religation of DNA ends, forming the final coding joint between two compatible gene segments (73-76).

4 Accessibility Model of Recombination

Seminal studies using transformed B cell lines provided remarkable insight into the complex regulation of V(D)J recombination during the early stages of lymphocyte development. In 1985, Yancopoulos and Alt used bone marrow-derived cells transformed with Abelson murine leukemia virus (A-MuLV) to show that transcription of the unrearranged (or germline) immunoglobulin VH segments corresponded with their rearrangement (77). From these results, the investigators hypothesized that rearrangement was associated with gene segment accessibility, in that local chromatin structure could restrict or permit access of the recombinase to recombination substrates. Since these initial experiments, a tremendous volume of data has been generated that supports this accessibility model of recombination. Exhaustive research has established that cis-acting regulatory elements, including promoters and enhancers, alter antigen receptor chromatin structure to facilitate RAG access to RSSs.
Modulating the physical accessibility of antigen receptor gene segments provides an additional tier of regulation for RAG-mediated cleavage beyond simply restricting RAG expression to precise stages of lymphocyte development. While RAG-1 binding is constrained solely to RSS elements, ChIP-sequencing analysis revealed that RAG-2 binds more than 24,000 H3K4me3 sites throughout the genome (though no function has been identified for either RAG protein outside of their roles in V(D)J recombination) (42). Regulation of DNA accessibility at the chromatin level, therefore, further ensures that the cleavage reaction is strictly controlled to avoid anomalous rearrangement.

4.1 Chromatin Environment

In eukaryotic cells, genomic DNA is packaged with nucleosome proteins as chromatin, allowing the DNA to be compacted 10,000 – 20,000 fold within the nucleus (78). Each nucleosome octamer is made up of four core histone dimers – H2A, H2B, H3, and H4 – and 147 base pairs of DNA wraps around the protein complex (79-81). An additional linker histone, H1, separates individual nucleosomes every ~20-60 DNA base pairs and allows the flexible structure to be further compacted (82). The chromatin of interphase cells typically adopts two general conformations: heterochromatin, a compact, 30-nm fiber associated with silent genes, and euchromatin, an open configuration associated with actively-transcribed genes (83). A third, intermediate chromatin state, facultative heterochromatin, has also been described, primarily in developmentally-regulated cells (84). Unlike constitutive heterochromatin, facultative heterochromatin is transcriptionally silent but can convert to active euchromatin upon the cell’s receipt of developmental cues. In particular, the temporally-regulated antigen receptor loci exist as facultative heterochromatin, with features of both inactive and permissive chromatin structures (85). To maintain this dynamic
chromatin environment, eukaryotes possess several systems to covalently modify either histones or DNA, resulting in chromatin remodeling and changes in the accessibility of various genes to DNA binding proteins.

4.2 Histone Modifications

The amino-terminal tails of the core histones account for 25-30% of each individual protein and interact closely with the surrounding DNA (86). Covalent modifications to the N-terminal tail domains, including acetylation, methylation, phosphorylation, ubiquitination and ribosylation, collectively form a ‘histone code,’ and directly influence histone::DNA interactions and, consequently, overall chromatin structure (86-88). Several residue-specific modifications have been well-characterized in facilitating the formation of either active, accessible euchromatin or silent, condensed heterochromatin. Lysine acetylation, particularly at histone H3 lysine 9 (H3K9) and histone H4 residues, is a hallmark of actively-transcribed genes (89, 90). Acetylation likely neutralizes the positively-charged lysine, weakening the histone::DNA interaction and rendering a more accessible, open chromatin structure. Histone acetyltransferase (HAT) enzyme complexes (91) catalyze the transfer of acetyl groups to lysine residues and also recruit SWI/SNF, a chromatin remodeling complex that repositions nucleosomes through ATP hydrolysis (92). Conversely, hypoacetylated lysine residues are associated with inactive chromatin, and histone deacetylases (HDACs) work antagonistically to HATs to remove acetyl groups from lysine residues (93, 94).

N-terminal histone tails can also be mono-, di-, or trimethylated by the catalytic activity of histone methyltransferases (HMTs) (95, 96), though the effect on chromatin conformation is more varied than histone acetylation. Histone H3 lysine 4 di- and trimethylation (H3K4me2 or H3K4me3) are correlated with accessible, transcriptionally
active genes (97-99). H3K9me and H3K27me3, on the other hand, are enriched in transcriptionally-silent, heterochromatic regions (100-102). Collectively, these modifications to histone proteins shape the chromatin topography to regulate gene expression.

4.3 DNA Methylation

In addition to post-translational histone modifications, DNA methylation is an important epigenetic mechanism for regulating gene expression. Methylation occurs specifically at cytosine nucleotides situated immediately adjacent to guanine bases (CpG). DNA methyltransferases (DNMTs) catalyze the transfer of methyl groups to the fifth atomic position on the cytosine ring, resulting in 5-methylcytosine (5mC) (103). In vertebrates, CpG dinucleotides are infrequently yet globally distributed, with the vast majority concentrated at CpG islands (104). These CpG-rich regions are typically positioned at the 5’ end of genes near transcription starts sites (TSSs). Early experiments in murine embryonic stem (ES) cells suggested a direct role for DNA methylation on cellular differentiation by inducing changes in gene expression. Using the cytidine analogs 5-azacytidine and FCdR (5-fluoro-2’-deoxycytidine) to inhibit DNMT-mediated CpG methylation, researchers were able to induce differentiation of ES cells into fat and muscle cells (105). Indeed, since these preliminary experiments, a wealth of research has provided convincing evidence that CpG dinucleotides are methylated at transcriptionally-silent genes and fully demethylated at expressed genes (106, 107).

4.4 Epigenetic Regulation within Antigen Receptor Loci

Changes in local chromatin structure are essential for regulating V(D)J recombination during thymopoiesis. Experiments using chromatinized recombination substrates revealed that nucleosome compaction and CpG methylation collectively abolished transcription and
rearrangement, reducing accessibility by at least 10,000-fold, and that displacement of nucleosomes could effectively restore rearrangement (108). Moreover, the Oettinger and Schlissel labs used nucleosomal templates to demonstrate that RAG-mediated cleavage was severely impaired at RSSs associated with histone octamers (109, 110). While the compact chromatin structure occluded RAG access to RSS targets, histone acetylation and SWI/SNF-mediated remodeling of chromatinized templates were able to restore cleavage activity to levels observed with naked DNA substrates (111, 112).

Antigen receptor loci have dynamic marks of accessible or repressive chromatin at various stages of development, corresponding with the onset and cessation of recombination. Epigenetic regulation at H3K9 is central to the stage-specificity of V(D)J rearrangement. Repressive H3K9me within the DβJβ region of the Tcrb locus reduces accessibility and, consequently, inhibits rearrangement (113). Similarly, the VH region of Igh is hypermethylated at H3K9 in non-B cell lineages (Pax5−/− cells). H3K9me is subsequently lost upon Pax5 expression and B lineage commitment, suggesting a role for the repressive mark in preventing rearrangement in non-B cells (114). Indeed, in early B cell development pro-B cells are promptly enriched with histone H3 acetylation at DβJH segments (115). Histone H3/H4 dimers within the neighboring VH region become acetylated in response to DJH joining, indicating proximal VH segments are now poised for rearrangement (116). A similar pattern has been observed in the other antigen receptor loci, in that H3ac levels are elevated during developmental stages concomitant with rearrangement (115, 117-119). Additionally, other activating histone marks are also detected during the precise windows of antigen receptor gene assembly. The Tcrb and Tcra loci are enriched for H3K9ac, H3K4me2, and H3K4me3 in DN and DP thymocytes, respectively, when each locus is undergoing
rarrangement (42, 115, 117, 118). \(Igh\) and \(Igk\) are similarly decorated during recombination in pro-B and pre-B cells, respectively (34, 42, 115, 116, 119).

DNA methylation is also important for modulating accessibility within the antigen receptor loci. Recombinationally-silent loci exhibit high levels of CpG methylation prior to rearrangement. A loss of methylation coincides with the onset of rearrangement, and recombinationally-active loci are generally fully demethylated (120-123). Methyl-sensitive restriction enzyme-dependent (MSRE) PCR experiments have substantiated this alteration in methylation status within the \(Igk\) locus, which is hypermethylated in pro-B cells prior to rearrangement and loses methylation marks upon transitioning to the recombinationally-active pre-B stage (124). Likewise, similar assays have demonstrated that the \(Tcrb\) locus is hypomethylated in DN cells, corresponding with the window of its active rearrangement (125). Perhaps unexpectedly, recent experiments using bisulfite mapping revealed that specific \(D_H\) and \(J_H\) regions are only partially demethylated in pro-B cells (126). Although the \(D_HJ_H\) region was not fully demethylated prior to recombination, CpG methylation was absolutely eliminated at DJ\(_H\) junctions formed during the first step of rearrangement. This suggests a novel regulatory mechanism for the second step of \(Igh\) gene assembly, where demethylated DJ\(_H\) substrates could facilitate successive V\(_H\)-DJ\(_H\) joining (126). Taken together, the findings suggest that changes in DNA methylation and modifications to histones cooperate to strictly regulate chromatin structure during antigen receptor gene assembly.

5 \(\alpha\beta\) T cells

As master regulators of adaptive immunity, T lymphocytes generate a pathogen-specific response to a nearly limitless number of microorganisms through the tightly-regulated somatic rearrangement of TCR gene segments. Greater than 90% of T lymphocytes
in the peripheral blood express αβ receptors, which are encoded by the Tcra and Tcrb loci. The other <10% belong to the γδ lineage (127, 128), a specialized subpopulation of T cells involved with mucosal immunity and wound repair (129, 130). αβ T cells generate their receptors by V(D)J recombination during a finely-tuned program of cellular differentiation that is regulated at the levels of tissue specificity, lineage specificity, stage specificity, and allelic exclusion. Indeed, the Schlissel laboratory used an in vitro cleavage assay on chromatinized gene segments to demonstrate that chromatin organization directs RAG-mediated cleavage of target RSS substrates in a tissue-, lineage-, and developmental stage-specific manner (131). Thus, in addition to regulating RAG expression, modulating gene segment accessibility at precise windows of development is central for ensuring the specificity of the V(D)J reaction during thymopoiesis.

5.1 αβ T cell Development & TCRβ Assembly

The earliest stages of T cell development begin in the bone marrow, where hematopoietic stem cells differentiate into a common lymphoid progenitor (CLP), the precursor to both T and B lymphocytes. Activation of a specific pattern of transcriptional networks and consequent expression of cell surface proteins in CLPs impels those precursors into the T-lineage developmental program. This subset of cells then migrates to the thymus as uncommitted early T-lineage progenitors (ETPs), entering the thymic cortico-medullary junction through the vasculature. Early thymocyte development begins with several double-negative (DN) pro-T stages, which are defined by the up- and down-regulation of cell surface markers (Fig. 6) as cells continue their migration through thymic tissue. In the earliest double negative stages, DN1 and DN2 cells (CD44+CD25− and CD44+CD25+, respectively) successively upregulate CD44 and CD25 and move through the cortex to the sub-capsular
zone. Progenitor cells then commit to the T cell lineage upon upregulating CD25 and transitioning from the DN2 to the DN3 stage (CD44^CD25^) (132). Expression of both surface markers is lost in CD44^CD25^- DN4 thymocytes, as cells migrate back through the cortex. Cells then upregulate both CD4 and CD8 upon reaching the double positive (DP) stage at the cortico-medullary junction, where they complete recombination and undergo positive and negative selection. Selected single-positive (SP) CD4^+ and CD8^+ cells ultimately emigrate from the thymus to peripheral lymphoid tissues, where they will encounter foreign antigens (133).

Figure 6. Ordered rearrangement of antigen receptor loci during αβ T cell thymopoiesis. Tcrb is assembled first in double-negative (DN) thymocytes. Cells test the newly-rearranged β-chain as part of the pre-TCR complex at the DN4 stage. Tcra then undergoes rearrangement in double-positive (DP) cells. Single-positive (SP) thymocytes surviving positive and negative selection ultimately differentiate into mature CD4^+ or CD8^+ T cells.

In addition to changes in CD44 and CD25 expression, developmental progression also coincides with expression of the V(D)J recombinase, RAG-1/2 (Fig. 6). In αβ T cells,
Tcrb rearrangement precedes assembly of Tcra (134), and RAG-1 and RAG-2 are first expressed at the onset of Tcrb rearrangement, in the DN1-DN3 stages. Tcrb rearranges in a two-step process, with Dβ-to-Jβ joining on both chromosomes (135) preceding monoallelic Vβ-to-DJβ rearrangement (Fig. 7). The upstream DJβ1 cassette is preferentially targeted for rearrangement, as Dβ1Jβ1 joins have been detected at the most primitive DN1 stage (136, 137). Experiments in fetal thymocytes and TCRβ transgenic mice have shown that Dβ2-to-Jβ2 joining initiates later, in DN2, and that Dβ2Jβ2 joins accumulate slower than those involving Dβ1 (138-141).

Figure 7. Mechanism of Tcrb rearrangement. Tcrb is assembled via sequential Dβ-to-Jβ then Vβ-to-DJβ rearrangement. The first recombination reaction generates the substrate for rearrangement involving Vβ segments.

Following Dβ-to-Jβ joining, Vβ-to-DJβ rearrangement is activated on only one allele at a time in DN3 cells, a phenomenon known as allelic exclusion. This ensures that each cell expresses a unique, clonal receptor with a single antigen specificity (142). Allelic exclusion is likely enforced by three different mechanisms: (1) developmentally-timed, monoallelic
decontraction of the Tcrb locus (143), (2) localization of the silent allele with the repressive environment of the nuclear periphery (144, 145), and (3) E47-dependent feedback inhibition to prevent persistent Vβ-to-DJβ rearrangement (146). It is also believed that recombinational accessibility (chromatin structure) cooperates in the maintenance of allelic exclusion, though the mechanisms are not currently understood (131). Upon generation of a productive VDJβ join, late DN3 thymocytes express a pre-T-cell receptor complex to test the nascent β-chain. The pre-TCR is comprised of the newly-assembled TCRβ chain, a CD3 co-receptor complex, and a surrogate TCRα chain (pTα) (147, 148). This critical checkpoint, termed β-selection, promotes developmental progression to the DN4 stage, corresponding with the termination of both RAG expression and rearrangement as well as a surge in proliferation that leads to clonal expansion of β-selected cells (149).

Pre-TCR-selected cells then advance to the CD4+CD8+ double positive stage, where RAG expression is restored and the Tcra locus rearranges by means of a one-step assembly (Vα-to-Jα) (150). In addition to rearranging Tcra, DP thymocytes also undergo positive selection to ensure cells are capable of recognizing MHC-presented antigen (151-153). During positive selection, the TCR engages self-peptide-bound MHC with the cooperation of either the CD4 or CD8 co-receptor. Cells with a moderate affinity for MHC class I subsequently upregulate CD8 expression and differentiate into cytotoxic CD8+ T cells, while cells that recognize MHC class II upregulate CD4 expression and become CD4+ helper T cells. Finally, in the thymic medulla, single positive CD4+ and CD8+ T cells undergo negative selection (153-155). Here, cells with TCRs that engage self-antigen presented by MHC on either thymic dendritic cells or medullary thymic epithelial cells are eliminated by apoptosis (156). Additionally, negative selection is mediated by the autoimmune regulator
(AIRE) transcription factor. The AIRE protein activates transcription of numerous tissue-specific genes that are in turn expressed by the thymic epithelium (157). This enables presentation of otherwise inaccessible self-antigens to CD4\(^+\) and CD8\(^+\) T lymphocytes, in order to eliminate self-reactive cells that could elicit an inappropriate immune response (158). It is estimated that greater than 95% of developing cells die within the thymus (159), evidencing the stringent selection processes and checkpoints throughout thymocyte development.

5.2 Epigenetic Control at Dβ1

Assembly of TCRβ during thymopoiesis is driven in large part by cis-acting accessibility control elements (ACEs) such as promoters and enhancers that cooperate to modulate chromatin accessibility in a stage-specific manner. The Tcrb locus contains a lone enhancer element (Eβ) situated downstream of Cβ2 (Figure 2) that was shown to have T cell-specific activity in chloramphenicol acetyltransferase (CAT) reporter assays (160-162). In vivo studies provided evidence that Eβ is critical for global regulation across the entire DβJβ region. In mice lacking Eβ, cleavage, rearrangement and germline transcription were abrogated across both DβJβ clusters and animals exhibited a block in αβ T cell development similar to that observed in RAG-deficient mice (163-165).

Within the Dβ1Jβ1 cassette, Eβ facilitates activation of a single germline promoter, PDβ1 (Fig. 2) (166, 167). Situated immediately upstream of the Dβ1 gene segment, PDβ1 drives transcription in pro-T cells, and maximum promoter activity requires a TATA element within the 5’ Dβ1 RSS as well as the transcription factors SP1 and GATA-3 (166). In addition to its role in directing germline transcription, PDβ1 is indispensable for rearrangements involving Dβ1. Deletion experiments using TCRβ miniloci in a recombinase-
inducible cell line silenced D-to-J assembly (168). These results were corroborated by *in vivo* studies using mice with Cre/loxP-mediated targeted deletions in the PDβ1 promoter region (169). In homozygous mutant mice (PDβ1<sup>−/−</sup>), Dβ1-Jβ1 rearrangements were attenuated 20- to 50-fold relative to WT mice, whereas deletion of PDβ1 had no effect on Dβ2-Jβ2 rearrangement (169).

Collectively, the analyses of the PDβ1 promoter show a strong correlation between activation of germline transcription and rearrangement. To test if rearrangement was a direct consequence of transcriptional activation, the Oltz laboratory used a RAG-inducible system to analyze stably-integrated miniloci in which PDβ1 had been repositioned. In clones where the promoter was relocated between Dβ1 and Jβ1.1 (D-P-J), germline transcription was unchanged relative to WT (P-DJ) while Dβ1-Jβ1 rearrangement was reduced to ~10% of WT levels (170). When the promoter was inverted, transcription was nearly eliminated while recombination was equivalent to WT levels (170). The results suggest that the effects of PDβ1 are highly localized and that transcriptional elongation does not directly drive rearrangement within the Dβ1Jβ1 region, rather changes in chromatin configuration more likely influence RAG access.

Indeed, the Eβ enhancer and PDβ1 promoter contribute to extensive chromatin remodeling to regulate Dβ1 accessibility. The Krangel laboratory demonstrated that germline transcription from PDβ1 leads to nucleosome repositioning and a consequential gain in accessibility across the Dβ1Jβ1 cassette without impacting the downstream Dβ2 region (171). *Rag<sup>−/−</sup>* mice harboring deletions of either PDβ1 or Eβ exhibit more densely-packed nucleosomes from Dβ1 to Jβ1.3., suggesting both ACEs are required for nucleosome repositioning (171). Moreover, thymocytes from *Rag<sup>−/−</sup>*Eβ<sup>−/−</sup> knockout mice display an
increase in DNA methylation, a reduction in H3ac, and greater resistance to restriction endonuclease-mediated cleavage across both DβJβ clusters (85). Eβ+/− alleles also show enrichment for heterochromatic H3K9me marks, whereas Eβ+/+ alleles are marked with activating, euchromatic H3K4me (172). A loss of local accessibility is also seen in mice harboring a deletion of the PDβ1 promoter (ΔPD). ΔPD mice exhibit increased DNA methylation and reduced H3ac within the Dβ1 region but not at Dβ2, corresponding with the impairment of Dβ1 transcription and rearrangement (173). Ultimately, the ability of promoters and enhancers to regulate accessibility was confirmed by chromatin immunoprecipitation (ChIP) experiments, which revealed that the cis-elements can directly shape RAG binding. In ΔEβ knockout mice, RAG-1 binding and H3ac levels were eliminated at both DβJβ cassettes, whereas mice with the PDβ1 promoter deletion (ΔPDβ1) showed reductions in RAG-1 binding and H3ac only within the Dβ1Jβ1 region (117).

Taken together, knockout studies provide compelling evidence that transcription, chromatin remodeling, and recombination are regulated globally and locally by Eβ and PDβ1, respectively. Because promoter activity is enhancer-dependent, and the ACEs are separated by a considerable distance, it was proposed that Eβ and PDβ1 could cooperate via a direct physical interaction facilitated by DNA looping. Indeed, ChIP studies in Eβ−/− mice provided preliminary evidence of a promoter-enhancer complex. Ferrier and colleagues showed that RNA polymerase II, part of the transcription apparatus, and CREB-binding protein (CBP), a coactivator with HAT activity, bind Eβ and PDβ1 in an Eβ-dependent manner (172). Their findings suggest that Eβ associates with PDβ1 to activate transcription within the Dβ1Jβ1 region. This model was further supported by ChIP and chromosome conformation capture (3C) data from the Oltz laboratory. In ΔPDβ1 thymocytes, promoter
deletion eliminated binding of SP1, a transcription factor required for tissue-specific activation of PDβ1 (166), at both PDβ1 and Eβ (174). 3C analysis using an Eβ-specific primer revealed that the enhancer interacts specifically with the Dβ1 and Dβ2 regions, as no interaction was seen between Eβ and either Cβ2 or the region upstream of Dβ1. Furthermore, thymocytes from mice harboring a deletion of the enhancer or promoter (ΔEβ or ΔPDβ1) exhibit a decrease in 3C amplification products by >10-fold (174). Altogether the findings support a regulatory mechanism where Eβ and PDβ1 form an enhancer-promoter holocomplex in DN thymocytes to direct accessibility and, consequently, rearrangement within the Dβ1Jβ1 region.

5.3 Regulation of Dβ2 Rearrangement

While the mechanisms that drive accessibility within the DJβ1 region have been well-characterized, less emphasis has been placed on the intricacies at DJβ2. Northern analysis in ΔPDβ1 mice revealed that while promoter deletion abolished germline transcription at Dβ1, Dβ2 germline transcripts persisted (173). This suggested that accessibility at Dβ2 was regulated independently of Dβ1. Supporting this model, analysis of nucleosome organization revealed that deletion of PDβ1 did not influence nucleosomal positioning at Dβ2, while deletion of Eβ resulted in tightly-packaged histone octamers surrounding Dβ2 (171). Furthermore, deletion of the PDβ1 promoter had no effect on H3K9 acetylation or rearrangement within the DJβ2 cluster (169, 174). Finally, 3C analysis revealed that Dβ2 associates independently with the Eβ enhancer, likely through protein-protein interactions by RUNX1 and TATA-binding protein (TBP) (174). Dβ2-Eβ interactions, as well as RUNX1 and TBP binding, were lost in ΔEβ mice but not in ΔPDβ1 animals (174). Together these
findings suggest that a distinct Dβ2 promoter regulates accessibility at the downstream cassette and that activity is likely enhancer-dependent.

In order to gain a better understanding of the regulatory mechanisms at Dβ2 and to determine their influence on receptor repertoire, the Sikes laboratory sought to characterize cis-acting elements specific to the DJβ2 cluster. The research group first examined germline transcription at Dβ2 during different stages of thymocyte development. Northern analysis revealed that germline transcripts were limited to 3’ of Dβ2 in DN thymocytes from Rag2−/ mice. In DP thymocytes from WT C57BL/6 mice, however, germline transcripts were detected both 5’ and 3’ of Dβ2 (175). To pinpoint a potential promoter element, the researchers then used 5’-RACE (rapid amplification of cDNA ends) to map transcription start sites (TSSs). Interestingly, germline transcripts in all DN P5424 clones (Rag1−/ p53−/ thymoma cell line) initiated 400-600 bp 3’ of Dβ2, while all DP clones containing Dβ2-Jβ2.1 rearrangements contained start sites within ~200 bp 5’ of Dβ2 (175). Putative promoter activity was then assessed using luciferase reporters. A construct containing a deletion of the ~150 bp containing transcription start sites 3’ of Dβ2 resulted in a reduction in activity below the promoterless control. Conversely, 5’ deletions of all but the 150 bp containing TSSs 3’ of Dβ2 led to a two-fold increase in activity relative to the full-length construct. Interestingly, a construct containing a deletion from Dβ2 to Jβ2.1, that mimics rearrangement, demonstrated a 7.5-fold increase in promoter activity relative to the promoterless control (175).

Collectively, these results suggested that two independent promoters regulate Dβ2 – a downstream promoter that is initially active in early DN cells, and an upstream promoter which is likely repressed prior to rearrangement (Fig. 8).
Figure 8. Promoter control at Dβ2. Two promoters flank the Dβ2 gene segment. The upstream promoter, 5′PDβ2 (blue arrow), is repressed in early DN cells, forcing transcriptional activity to the downstream promoter, 3′PDβ2 (green arrow). Upon deletion of 3′PDβ2 during D-to-J joining, transcriptional activation shifts to 5′PDβ2.

The laboratory next sought to characterize trans-acting regulatory factors that control activity of the dual 5′PDβ2 and 3′PDβ2 promoters. Sequence analysis revealed two predicted NF-κB binding sites (κB-1 and κB-2) within the 150 bp span of maximum 3′PDβ2 activity (175). Luciferase reporter assays revealed that destruction of either NF-κB site reduced 3′ promoter activity to 38% or 49% of the WT fragment, respectively, and that the effect was additive, as 3′PDβ2 activity was merely 20% of WT when both sites were mutated (175). In vivo binding by NF-κB was verified by ChIP analysis in PMA-treated P5424 cells, suggesting the downstream 3′PDβ2 promoter is driven by NF-κB (175).

Experiments with luciferase reporter vectors were also used to identify the location of the upstream 5′PDβ2 promoter. Sequential deletions 5′ of Dβ2 suggested that the promoter was likely situated within a 500-600 bp region immediately upstream of the Dβ2 gene segment, as deletions of ~140, 510, and 660 bp reduced luciferase activity by 63, 82, and 98%, respectively, relative to the full-length fragment (175). Several transcription factor
binding sites and promoter elements were identified within the span of peak 5’ promoter activity, including GATA and RUNX binding sites, two E boxes, a consensus TATA sequence (TATAAA), and overlapping initiator elements (inr) (176). Electrophoretic mobility shift assays (EMSA) established that the GATA, RUNX, and E box sites bound GATA3, RUNX1, and the E protein E47, respectively, and ChIP analysis confirmed that the transcription factors bound the 5’PDβ2 promoter in vivo in the DN P5424 cell line. E47 was strongly enriched at 5’PDβ2 (80-fold over IgG control), while GATA3 and RUNX1 were also detected at modest levels (2-fold and 7-fold over IgG, respectively) (176).

Additional deletion constructs were subsequently generated to distinguish the core promoter. The region of greatest promoter activity was localized within 220 bp directly upstream of the 5’ Dβ2 12-RSS, containing a single RUNX site (R-II), the two E boxes (E-I and E-II) and the initiator elements. Removal of the 220-bp core reduced luciferase activity to ~40% of the full-length fragment, while a construct containing 5’ deletion of all but the 220-bp core, conversely, retained full activity (176). Finally, mutational analyses were performed to determine the involvement of each site on 5’PDβ2 activity. Mutation of the most 3’ GATA site (G-IV), immediately upstream of the TATA box, and the R-II and E-I sites similarly reduced activity to ~20-30% of the full-length promoter fragment (176). Destruction of either the three upstream GATA sites or the most 3’ E-II site had minimal impact on activity, while mutation of the initiator elements resulted in ~88% reduction in promoter activity. In total, the findings suggest that GATA3, RUNX1 and, particularly, E47 synergize to drive activity of a 220-bp core 5’PDβ2 promoter only after initial Dβ2-to-Jβ2 rearrangements delete an NF-κB-dependent 3’PDβ2 promoter (Fig. 8). The unique configuration of cis-elements within the DJβ2 cluster could rationalize differences in
recombinational efficiencies between the two DJβ cassettes. Promoter location has direct consequences on recombination, as repositioning PDβ1 downstream of the Dβ1 segment severely attenuated rearrangement (170). Forcing initial Dβ2 transcriptional activity to the 3’PDβ2 promoter would similarly be predicted to reduce the efficiency of DJβ2 joining relative to DJβ1 (138-141).

What elements limit activity of the upstream Dβ2 promoter and how is repression imposed? It is intriguing to consider the implications of repression on restricting Dβ2 rearrangement. Forcing initial recombination events to the upstream DJβ1 cassette could provide a rescue mechanism for non-productive V-to-DJβ1 joins, as the DJβ2 cluster is not deleted during recombination events involving DJβ1. By restricting DJβ2 usage, each allele would ultimately have two opportunities to assemble a productive β-chain, maximizing both pro-T cell survival and receptor diversity. The research described herein seeks to elucidate the precise regulatory mechanisms of Dβ2 repression, and determine their impact on rearrangement and overall receptor repertoire.

5.4 Beyond the 12/23 Rule

According to the 12/23 rule, only RSSs with dissimilar spacer lengths can recombine (9). The immunoglobulin heavy chain (Igh) and T-cell receptor β (Tcrb) loci rearrange in an ordered, two-step process where D-to-J rearrangements occur first, followed by V-to-DJ joining. In the Igh locus, V_H and J_H segments are flanked with 3’ 23-RSSs and 5’ 23-RSSs, respectively, while D_H segments are flanked with 5’ and 3’ 12-RSSs (Fig. 4). This configuration ensures ordered rearrangement and D_H segment usage during the two-step assembly. In Tcrb, however, V_β segments are flanked with 3’ 23-RSSs, D_β segments with 5’ 12-RSSs and 3’ 23-RSSs, and J_β segments with 5’ 12-RSSs (Fig. 4). Thus, direct V_β-to-J_β
joining is permitted within *Tcrb* based on the 12/23 rule. Direct Vβ-to-Jβ rearrangement without inclusion of a Dβ segment rarely occurs, however (139, 177-179), suggesting that additional constraints prevent such joining. Having such mechanisms in place to ensure Dβ segment utilization is critical, as the Dβ segment encodes the majority of the third complementarity-determining region (CDR3) of the T-cell receptor (7). As the CDR3 loop is the site that cooperates most intimately with the antigen:MHC complex, Dβ gene segment inclusion is imperative to maximizing repertoire diversity (7).

Restrictions “beyond the 12/23 rule” (B12/23) are imposed either by *cis*-acting limitations of the RSS sequence or by *trans*-acting factors. Indeed, a significant body of work provides evidence for the impact of sequence determinants on B12/23 restriction. Experiments using extrachromosomal recombination substrates demonstrate that B12/23 is enforced, in part, at the level of the 5’ Dβ1 12-RSS (178, 180-182). Using chimeric RSSs, where components of the 5’ Dβ1 12-RSS were swapped with those of the Jβ1.2 and Jβ1.4 12-RSSs, investigators revealed that the nonamer and spacer sequences (180) or the heptamer/nonamer pair (181) of the 5’ Dβ1 RSS were essential for maintaining B12/23 restriction, with Vβ segments significantly biased to the 5’ Dβ1 RSS. Consistent with these findings, a study using transgenic mice (in which the entire DJβ2 cassette and the Dβ1 segment were removed by targeted deletion) confirmed that Vβ segments preferentially target the 5’ Dβ1 12-RSS over the Jβ1.2 12-RSS (183). In Jβ1<sup>MS</sup> mice, where the Jβ1.2 12-RSS was replaced with the 5’ Dβ1 12-RSS, there was nearly exclusive Vβ-to-Jβ1.2 rearrangement on mutant alleles (183). The suggestion that Jβ RSSs are markedly less efficient than 5’ Dβ1 12-RSSs provides a convenient mechanism for preventing direct Vβ-to-Jβ rearrangement and ensuring Dβ segment inclusion.
In addition to the restriction that limits Vβ rearrangement to a 5’ Dβ RSS, B12/23 is also imposed in cis at the level of the 3’ Dβ1 23-RSS. A study using transgenic mice revealed that replacing the Vβ14 23-RSS with the 3’ Dβ1 23-RSS drastically skewed the thymocyte population toward Vβ14+ cells (184). Jβ1M3/M7 mice harboring the alteration contained 7-fold more Vβ14+ thymocytes compared to control animals, suggesting that the 3’ Dβ1 23-RSS is more efficient than 23-RSSs flanking Vβ segments (184). Additionally, sequence analysis of the Dβ1 and Dβ2 23-RSSs identified a conserved, non-canonical AP-1 binding site (185). In vitro experiments determined that c-Fos, a component of AP-1, binds to this site and interacts with the RAG-1/2 complex, facilitating RAG recruitment to the Dβ 23-RSS and thereby augmenting Dβ-to-Jβ rearrangement (185). In Fos−/− mice, Dβ-to-Jβ recombination was strikingly reduced. Total thymocyte and splenocyte counts were approximately 5% of wild-type controls, emphasizing the importance of c-Fos in ensuring ordered assembly (185). Jointly, these results suggest that the efficiency of the Dβ 23-RSSs relative to Vβ RSSs is due to constraints both in cis and in trans.

While these results provide evidence for preventing direct Vβ-to-Jβ rearrangement at the level of the RSS, they do not specifically address the role of RAG in RSS recognition and cleavage beyond 12/23. Franchini and colleagues used oligonucleotide capture and in vitro cleavage assays to dissect the molecular mechanism of B12/23 at the levels of chromatin accessibility and RAG-mediated DNA cleavage (28). Their results demonstrated that the RAG-1/2 complex favors Dβ RSSs for cleavage and paired complex formation over those flanking Vβ and Jβ segments, with preference given to Dβ 23-RSSs over 12-RSSs. Additionally, cleavage was Eβ-dependent, supporting the model that chromatin opening is required for accessibility of individual RSS elements to the recombinase. Perhaps most
interestingly, the investigators discovered that RAG-mediated nicking at the 3’ Dβ 23-RSS inhibited cleavage at the adjoining 5’ Dβ 12-RSS. DNA nicks at the Dβ 12-RSSs were not detected on germline alleles; rather 5’Dβ cleavage products were formed only after Dβ-to-Jβ joining and simultaneous excision of the Dβ 23-RSS (28). By impeding RAG-mediated breaks at the upstream RSS until after Dβ-to-Jβ rearrangement, illegitimate Vβ-to-Jβ rearrangement without Dβ segment inclusion is largely prevented.

Taken together, these studies demonstrate that intrinsic properties of RSSs predominantly shape the efficiency of RAG-mediated cleavage beyond the 12/23 rule to prevent direct Vβ-to-Jβ joining. According to the 12/23 rule, however, D-to-D rearrangement in Tcrb would also be permitted based on the organization of Dβ RSSs (Fig. 4). Of note, evidence suggests that RAG-1/2 preferentially targets the Dβ-flanking RSSs for cleavage; hence efficient signal sequences at both Dβ segments could, in theory, facilitate D-to-D rearrangement. Furthermore, the Dβ1 23-RSS rearranges with the Dβ2 12-RSS with the equivalent efficiency as all six Jβ1 RSSs in plasmid substrates (181). Indeed, Dβ1-Dβ2 joins have been described in both human and murine TCRβ assemblies, albeit infrequently (138, 186-188). Intriguingly, findings from the Alt laboratory suggest that Dβ1-Dβ2 restriction is not imposed in cis at the level of flanking RSS sequences (181). What element(s) could effectively prohibit Dβ1-Dβ2 joining? RAG nicking at the Dβ1 23-RSS was shown to inhibit cleavage at the adjacent 12-RSS (28). Does RAG binding at the downstream RSS directly occlude binding at the proximal site? Could a distinct trans-acting factor function similarly at Dβ2 to sterically hinder RAG access to the 5’ Dβ2 12-RSS? The research described herein presents a model where the transcription factor USF1 binds an E box element within the 5’ Dβ2 12-RSS prior to rearrangement. It is appealing to consider that binding at this site could,
in turn, obstruct RAG access and unequivocally enforce B12/23 at the level of Dβ1-Dβ2 restriction.

6 Upstream Stimulatory Factor 1

Upstream stimulatory factor 1 (USF1) belongs to a large family of basic helix-loop-helix (bHLH) proteins involved with transcriptional regulation (189-192). These transcription factors contain two evolutionarily-conserved domains – a carboxy-terminal leucine zipper domain that facilitates homo- or hetero-dimerization with other bHLH factors, and a basic amino-terminal motif that mediates DNA binding (Fig. 9) (193, 194). Through the basic region, bHLH proteins interact with a hexameric DNA sequence known as an E box. E boxes have a consensus sequence of \( \text{CANNTG} \), with a canonical, palindromic conformation of \( \text{CACGTG} \) (195, 196). As indiscriminate \( \text{cis} \)-elements, however, E box motifs frequently are non-canonical, and the more than 240 transcription factors in the bHLH family (197) bind different versions with varying affinities (198).
Figure 9. Protein structure of murine upstream stimulatory factor 1 (USF1). The transcription factor codes for 310 amino acids and is comprised of three functional domains: (1) a basic region (green) that mediates DNA-binding, (2) a helix-loop-helix structural motif (blue), and (3) a leucine zipper (red) that facilitates dimerization. 3-dimensional structure of the DNA-binding region and HLH was generated using the Protein Picture Generator (199) v1.21 (2005), with PDB Id. 1AN4 as a search model.

E box elements serve as indispensable docking stations within promoter regions throughout the genome (200), allowing bHLH proteins to bind and directly activate or repress transcription, or to recruit basal transcriptional machinery to the promoter region (201-205). USF1 has been shown to regulate transcription in each of these distinct capacities. It is constitutively expressed in nearly all cell types (206), and participates as a stress-response regulator at a broad range of gene promoters, including those involved with pigmentation in melanocytes (207), lipid metabolism (208), and cell proliferation and differentiation (209-213). Importantly, USF1 also regulates genes of the DNA damage response (DDR) (214) as well as numerous genes involved with immune function (215).
6.1 USF1 and the Immune Response

The mammalian immune system is constantly challenged with a host of bacterial, viral, and parasitic pathogens. In order to respond effectively to infections caused by these organisms, we need the ability to respond rapidly and specifically through upregulation of various immune genes. It is reasonable to predict that a transcriptional regulator associated with a universal response to cellular stress could play a critical role in modulating immune responses to invading pathogens. Indeed, USF1 has been established as a key participant in innate and adaptive immune responses, upregulating transcription in genes involved with the complement cascade, MHC classes I & II, Fc receptors, lymphocyte differentiation, and antigen receptor assembly.

In the innate response, C4 is the first complement protein to be activated in the classical and lectin complement pathways (216, 217), and is expressed primarily in macrophages (218). Using cell-free transcription assays and mutational analysis, Miyagoe and collaborators identified a functional E box required for complete activation of the C4 promoter (219). DNA footprinting further revealed a predicted USF1-binding motif, suggesting a role for the transcription factor in promoter activation (219). Additional work showed that not only did USF1 interact positively with the E box element, but that phosphorylation enhanced this interaction, emphasizing the significance of this post-translational modification in USF1 activity (220).

Substantial evidence has also elucidated the role of USF1 as a transcriptional activator in the adaptive response. MHC class I and II are imperative for presenting processed antigenic peptides to cytotoxic CD8+ and helper CD4+ lymphocytes, respectively (221-225). MHC I complexes are stabilized by the β2 microglobulin chaperone (226-228). At
the β₂m promoter, USF1/2 heterodimers bind an upstream non-canonical E box (CACGAG) to positively regulate β₂m transactivation (229). Overexpression of USF1 in a human carcinoma cell line (Tera-2) resulted in a 10-fold increase in promoter activity, while E box mutations in human lymphocyte cell lines (Jurkat, Raji) reduced activity by as much as 50% (229). In MHC II, the class II transactivator (CIITA) controls IFN-γ-induced MHC II expression (230, 231), with the PIV promoter mediating expression in non-APCs that do not constitutively express MHC II (232). Extensive in vitro analyses by Bernard Mach’s group revealed a functional E box and GAS (gamma interferon activation site) element at the PIV promoter that require cooperative binding of USF1 and STAT1, respectively, to the sites for complete transactivation (233). Mutational and insertional analyses as well as off-rate experiments revealed that site integrity, site proximity, and stable USF1:STAT1 interactions are all necessary for PIV promoter activation in response to IFN-γ (233).

Further demonstrating its importance as a critical regulator of adaptive immunity, USF1 directs transcription within promoters of Fc receptors. The high-affinity IgE receptor FcεRI is expressed primarily in mast cells and basophils (234, 235). Upon cross-linking with IgE-bound antigen, FcεRI receptor activation ultimately results in the production of allergy-associated mediators, such as histamines, leukotrienes, and prostaglandins (236-239). The α chain portion of FcεRI directly contacts IgE, suggesting that receptor expression on the cell surface is imposed at the level of α chain transcription (240). USF1/USF2 heterodimers were shown to bind a non-canonical E box (CAGCTG) to activate transcription of the FcεRI α chain promoter (240). Mutations within the E box resulted in a 50% decrease in promoter activity (240). USF1 was similarly implicated in augmenting transcriptional activity of the polymeric Ig receptor (pIgR) promoter (241). pIgR facilitates the transport of dimeric IgA
and pentameric IgM secretory antibodies across mucosal epithelia (242-244). Because each molecule of antibody requires one molecule of pIgR for transcytosis, pIgR expression is likely the rate-determining factor of IgA or IgM transport into the lumen (245). Extensive in vitro and in vivo analyses have detailed the role of USF1 in pIgR promoter activation. Studies in mouse and human epithelial cell lines have established that USF1 binds an E box element at the pIgR promoter and, consequently, upregulates transcriptional activity (241, 246-248). Reporter assays revealed that destroying the core E box (CACATG) motif reduced promoter activity to the level of the promoterless control (241).

Finally, USF1 has been identified as an activator of RORγT transcription in cooperation with USF2. The retinoic acid-related orphan receptor γT (RORγT) is the central transcription factor that directs differentiation of the Th17 lymphocyte subset and expression of the pro-inflammatory cytokine IL-17 (249), both of which are critical to defense against extracellular pathogens (250, 251). Binding of USF1 to an E box motif upstream of the RORγT promoter is likely essential for transcription, as binding site mutations and knockdown of USF1 significantly diminish promoter activity and RORγT expression (252). Furthermore, in vitro differentiation experiments showed that Th17 differentiation from CD4+ lymphocytes corresponded with a significant increase in RORγT and USF1 expression (252). Collectively, the results from the various studies suggest that USF1 is indispensable for transcriptional regulation of numerous immune genes.

6.2 Regulation at Antigen Receptor Loci

USF1 plays a critical role in transcriptional activation or repression in nearly all cell types (206), including those of the immune response. In particular, USF1 has been identified as a fundamental regulatory element within the antigen receptor loci. A role for USF1 in
transcriptional regulation of the antigen receptor genes was first demonstrated in the immunoglobulin light chain (Igλ) locus by Chang and colleagues. The investigators showed that USF1 binds the hexanucleotide sequence CACGTG to activate the λ2 promoter in vitro (253). USF-depleted nuclear extracts exhibited a reduction in λ2 promoter activity of 4- to 6-fold relative to control extracts, while λ2 transcription was restored upon the addition of exogenous USF (253). Similarly, in the murine Tcrd locus, ChIP analysis revealed that USF1 binds an E box immediately upstream of the Dδ2 promoter in vivo (254). In reporter assays, a construct that removed the USF1 binding site and TSS effectively eliminated promoter activity. Similarly, mutation of the USF1 binding site alone decreased activity by 80% (254). Because the Dδ2 promoter lacks both a TATA box and initiator element, USF1 could facilitate recruitment of transcription apparatus for transactivation (254-256).

The immunoglobulin heavy chain (IgH) locus also requires a USF1::E box interaction for regulation of the tissue-specific enhancer (Eμ), however USF1 negatively regulates transcription in this system. Carter and colleagues reported that USF1 binds an E box (μE3) at the IgH enhancer, and demonstrated USF1-mediated repression using an enhancer-based reporter assay. In their findings, USF1 binding to the μE3 site inhibited reporter activity five-fold relative to control (257). Expanding on these findings, a subsequent report revealed that USF1 binding at the μE3 enhancer E box prevented TFE-3-mediated activation of the IgH enhancer in vivo (258). In this study, the investigators deleted the basic region of USF1 to generate a dominant-negative mutant that was defective for DNA binding. In reporter experiments where the dominant-negative USF1 could not effectively bind the μE3 site, transactivation by TFE-3 was enabled (258). The results suggest that occupancy by USF1 at the μE3 E box in vivo likely prevents the access of the transcription factor TFE-3, thereby
inhibiting improper activation of the enhancer in non B-cells and ensuring tissue-specific expression of IgH (258).

The research presented herein provides evidence for an additional role of USF1 within the antigen receptor loci, and describes a repressive role for the transcription factor within the TCR β-chain (Tcrb) locus. This work identifies USF1 as the primary factor that binds an E box element within the 5’ Dβ2 12-RSS, resulting in repression of the 5’PDβ2 promoter. DNA DSBs generated during Tcrb rearrangement in DN thymocytes relieve USF1-mediated repression in a DNA-PK-dependent manner. Over-expression studies using phosphomimetic and dominant-negative mutants suggest a model where USF1 is phosphorylated by DNA-PK, leading to USF1’s displacement from the repressor E box and, consequently, a gain in 5’PDβ2 promoter activity. The findings suggest that USF1 plays a critical role in regulating TCRβ repertoire by restricting initial accessibility to Dβ2. This likely preserves the downstream DJβ2 cassette as a rescue mechanism for non-productive joins involving the favored upstream DJβ1 cluster. Furthermore, as USF1 binds precisely within the 5’ Dβ2 12-RSS prior to rearrangement, the transcription factor could serve as an indispensable regulatory element for restricting inappropriate D-to-D joining beyond 12/23. To address this possibility, this work additionally assesses Tcrb recombination in thymocytes from USF1−/− mice. Indeed, D-to-D rearrangement was increased in USF1−/− mice relative to WT control animals, suggesting a novel role for USF1 in maintaining D-to-D restriction beyond the 12/23 rule.
7 References


CHAPTER 2

DNA double-strand breaks relieve USF-mediated repression of Dβ2 germline transcription in developing thymocytes
Abstract

Activation of germline promoters is central to V(D)J recombinational accessibility, driving chromatin remodeling, nucleosome repositioning and transcriptional read-through of associated DNA. We have previously shown that of the two TCRβ locus (Tcrb) D segments, Dβ1 is flanked by an upstream promoter that directs its transcription and recombinational accessibility. In contrast, transcription within the DJβ2 segment cluster is initially restricted to the J segments and only redirected upstream of Dβ2 after D-to-J joining. The repression of upstream promoter activity prior to Tcrb assembly correlates with evidence that suggests DJβ2 recombination is less efficient than that of DJβ1. Because inefficient DJβ2 assembly offers the potential for V-to-DJβ2 recombination to rescue frameshifted V-to-DJβ1 joints, we wished to determine how Dβ2 promoter activity is modulated upon Tcrb recombination. In this study, we show that repression of the otherwise transcriptionally primed 5’Dβ2 promoter (5’PDβ2) requires binding of upstream stimulatory factor (USF)-1 to a noncanonical E-box within the Dβ2 12-recombination signal sequence spacer prior to Tcrb recombination. USF binding is lost from both rearranged and germline Dβ2 sites in DNA-dependent protein kinase, catalytic subunit-competent thymocytes. Finally, genotoxic dsDNA breaks lead to rapid loss of USF binding and gain of transcriptionally primed 5’Dβ2 promoter activity in a DNA-dependent protein kinase, catalytic subunit-dependent manner. Together, these data suggest a mechanism by which V(D)J recombination may feed back to regulate local Dβ2 recombinational accessibility during thymocyte development.
1 Introduction

Lymphocytes express a diverse array of Ag-specific receptors. The genes that encode these receptors are uniquely assembled in developing lymphocytes through a series of somatic rearrangements termed V(D)J recombination after the Variable, Diversity and Joining gene segments that are recombined (1, 2). B and T cell Ag receptor genes are each assembled by a single enzymatic complex centered on the lymphocyte-specific recombination activating gene (RAG) 1/2 proteins that target conserved recombination signal sequences (RSSs) flanking each V, D, and J segment. Despite the singular nature of enzyme and substrate, proper lymphocyte development and function requires that V(D)J recombination follow a precise program of ordered gene assembly imposed in part by RSS genetic variation (3-5), and in part by epigenetic regulation of promoters that populate each Ag receptor gene (3, 6, 7). Activation of promoters associated with D or J segments (so-called germline transcription, reflecting the unrearranged nature of the transcribed template) augments the accessibility of transcribed segments to recombinase. Chromatin remodeling (8), nucleosome repositioning (9) and transcriptional elongation associated with germline promoter activation (10) facilitate the recombinational accessibility of individual gene segments. However, the mechanism by which promoter-mediated accessibility is modulated during lymphocyte development is unclear.

T cell development begins when early thymocyte progenitors emigrate from the bone marrow to the subcapsular region of the thymus cortex. Progression of early thymocytes from early thymocyte progenitors through CD4-CD8- double negative (DN) development is coincident with rearrangement of the Tcrb, Tcrg and Tcrd genes. If cells assemble functional Tcrg and Tcrd joints before completing Tcrb assembly, they commit to the γδ lineage (11).
Conversely, expression of a rearranged \textit{Tcrb} gene triggers the silencing of additional \textit{Tcrb} recombination and drives the cell forward in development to the CD4\textsuperscript{+}CD8\textsuperscript{+} double positive (DP) stage in which \textit{Tcra} rearrangement occurs (12).

\textit{Tcrb} assembly proceeds in a stepwise manner that involves independent D-to-J recombination at two DβJβCβ gene segment clusters, followed by V rearrangement to a newly formed DJβ joint. Though Dβ RSS sequence strongly influences the order of gene segment assembly (5, 13, 14), the recombinational accessibility of individual RSSs is dependent on their chromosomal location (15) and the activity of associated germline promoters. Deletion of the Dβ1-associated promoter, PDβ1, alters nucleosomal phasing across the Dβ1 5’RSS (9) and specifically impairs \textit{Tcrb} Dβ1-to-Jβ recombination (16, 17) without affecting recombination at the downstream DJβ2 gene segment cluster (17). Though the mechanism of PDβ1’s influence over DJβ1 assembly is unclear, the promoter’s position immediately upstream of Dβ1 (18), and its recruitment of switch/sucrose nonfermentable chromatin remodeling complexes are critical for efficient DJβ1 assembly (8). Indeed, moving PDβ1 progressively downstream of Dβ1 increasingly impairs its ability to direct DJβ1 assembly of chromosomal \textit{Tcrb} transgenes (19).

Although both DJCβ clusters are transcriptionally active at the start of thymopoiesis (20), unrearranged DJβ2 clusters persist in the endogenous loci of thymocytes from \textit{Tcrb} transgenic mice, as well as from wild-type fetal thymocytes (21-24). Unlike germline transcription at Dβ1, transcription in the germline DJβ2 cluster predominantly initiates 400-600 bp downstream of Dβ2 (20). However, DJβ2 rearrangement, which deletes the germline promoter, results in the activation of a second promoter upstream of Dβ2. The role of promoter activity in DJβ2 recombination is unknown. Based on our understanding of DJβ1
assembly and the conserved role of promoter activity in driving recombination accessibility at other Ag receptor loci (7), it is likely that the downstream location of the germline Dβ2 promoter may contribute to the persistence of unrearranged DJβ2 clusters during thymocyte development (20). By extension, transcription from the upstream transcriptionally primed 5’Dβ2 promoter (5’PDβ2), which passes through the Dβ2 coding sequence and flanking RSSs, would then be predicted to enforce DJβ2 accessibility during V-to-DJ recombination. Separate DJβ cassettes offer each Tcrb allele the potential for two attempts at assembling an in-frame V(D)J rearrangement, provided V elements initially target DJβ1. Repression of 5’PDβ2 until after DJβ2 recombination might offer a potential mechanism to limit the initial accessibility of Dβ2 RSSs and thereby increase the frequency with which Vβ elements target DJβ1. However, the process by which 5’PDβ2 repression is first imposed and then relieved in a timely manner after DJβ2 recombination is unknown.

Upstream stimulatory factor (USF)-1 and -2 are ubiquitously expressed stress-response regulators that belong to the E protein family of basic helix-loop-helix leucine zipper transcription factors (25). USF1 and -2 bind as either homo- or heterodimers to E-box targets (CANNTG) (26) at promoters across the mammalian genome (27). USF proteins serve as master transcriptional regulators capable of interacting with a variety of transcription factors and chromatin modifiers to regulate such stress responses as UV-induced melanin production and insulin-dependent lipogenesis (25). DNA damage following UV treatment of keratinocytes and melanocytes induces phosphorylation of USF1 by the mitogen-activated protein kinase (MAPK) p38, which in turn alters USF’s gene regulatory properties (28). During periods of fasting, the fatty acid synthase promoter is repressed by USF1 associated with histone deacetylase 9. Upon feeding, USF1 is phosphorylated by the DNA-dependent
protein kinase, catalytic subunit (DNA-PKcs), leading to disassociation of histone deacetylase 9 and activation of the fatty acid synthase promoter (29). V(D)J recombination also critically depends on DNA-PKcs, as well as the related kinase, Ataxia telangiectasia mutated (ATM), which are activated as part of a broader response to the dsDNA break (DSB) intermediates of recombination (30). Indeed, DSBs generated during V(D)J recombination alter the expression of a wide array of genes including cell-type-specific genes not directly linked to the canonical DNA damage response (31).

In this study we show that 5’PDβ2 repression in DN thymocytes is mediated by USF1 bound to a noncanonical E box within in the spacer sequence of the Dβ2 12-RSS. Developmental activation of 5’PDβ2 correlates with loss of USF1 from the repressor element of both rearranged and germline DJβ2 clusters in DNA-PKcs-competent thymocytes. Finally, we show that 5’PDβ2 activity and loss of USF1 can be induced in RAG2-deficient thymocytes and cell lines following treatment with the chemical genotoxin, etoposide, or with sublethal doses of ionizing radiation to generate DSBs. Downregulation of USF1 binding and 5’PDβ2 repression is blocked by the DNA-PKcs inhibitor, Nu7026. Given the general role of promoter activity in regulating recombinational accessibility, our data suggest a model in which Dβ2 promoter activity both instructs and is instructed by V(D)J recombination.
2 Materials and Methods

2.1 Cells and antibodies

The $\text{Rag}^1^-$, $\text{P53}^-$ P5424 pro-T cell line was cultured as previously described (20). CD44+/CD25− (DN1) and CD44+/CD25+ (DN3) C57BL/6 cell populations were isolated from DP-depleted thymocytes using a three-laser MoFlo cell sorter (DakoCytomation) as previously described (20). P5424 subclones harboring rearrangements of their DJβ1 and DJβ2 gene segment clusters were isolated after repeated transient transfection of P5424 parental cells with p-phosphoglycerate kinase (PGK)-RAG1 as described below. Transfectants were subcloned and screened for Ag receptor recombinations by PCR, and specific DJβ recombinations were confirmed by sequencing. Thymi were isolated from 4- to 5-wk-old mice including: wt C57BL/6, $\text{Rag2}^-$, $\text{Prkdc}^{scid}$, $\text{Lat}^-$, and $\text{Rx}^-$ (Rag2/− mice that express a functionally rearranged TCRβ transgene (32)). Harvest procedures were reviewed and approved by the institutional animal care and use committee at North Carolina State University.

To induce dsDNA breaks in P5424 cells or primary thymocytes, cells were plated in fresh RPMI and exposed to increasing doses of ionizing radiation using a Gammacell 220 cobalt-60 irradiator (MDS Nordion), or plated in RPMI supplemented with 3 μM etoposide for 4 h. After genotoxic insult, cells were replated in normal RPMI and allowed to recover for 18-24 h prior to chromatin and RNA extraction. For kinase inhibition studies, cells were pretreated with 0.1% DMSO, SB203508 (10 μM, Calbiochem), Nu7026 (10 μM, Cayman Chemical), or Ku55933 (15 μM, Selleck Chemicals) for 40 min at 37°C before genotoxic insult.
Antibodies to CD44 (PE-labeled IM7; BD-Pharmingen), CD25 (FITC-labeled 7D4; BD-Pharmingen), CD117 (allophycocyanin-labeled 2B8; BD-Pharmingen), H3K9ac (Ab10812; Abcam), H3K4me2 (07-030; Millipore), H3K4me3 (39159; Active Motif), H3K27me3 (mAb6002; Abcam), as well as the following Abs from Santa Cruz Biotechnology: USF1 (sc-229), USF2 (sc-862), E47 (sc-763), HEB (sc-357), Myc (sc-764) and Max (sc-197). Rabbit IgG (10-4102) was from Rockland Immunochemicals.

2.2 Plasmids and transient transfection

For all transfections, 10^7 P5424 cells in log-phase growth were electroporated and luciferase reporter assays were conducted as previously described (20). Luciferase transfections were performed four or more times using independent plasmid preparations. For repeated PGK-RAG1 transfections, 7 x 10^6 cells were electroporated (300 μl serum free RPMI, 260 V/950 μF), with 10 μg RAG plasmid and 5 μg pMACS4.1 (Miltenyi Biotech), and allowed to recover overnight. Transfectants were enriched using Dynabeads FlowComp Mouse CD4 magnetic beads (Invitrogen), and retransfected as above prior to subcloning. Luciferase reporter plasmids were generated by cloning individual restriction fragments or PCR amplification products of p5’D2JJ-BS (20) into the SmaI site of pGL3-Eβ (20). Tiled site-specific mutations (TTCCA) were introduced into individual reporter constructs using Quikchange II (Stratagene) according to the manufacturer’s recommendations. The integrity of all reporter constructs was confirmed by sequencing.
2.3 Germline transcription and recombination

RNA was extracted using TRI Reagent (Sigma-Aldrich), according to the manufacturer’s instructions. DNA contaminants were removed using RNase-free DNase I (Fermentas) according to the manufacturer’s instructions, and 1-3 μg DNA-free RNA was reverse transcribed as previously described (20). The resultant cDNAs were amplified using standard (30-35 cycles) or quantitative PCR (50 cycles) reaction mixes (10 mM Tris-Cl (pH 9), 50 mM KCl, 2 mM MgCl₂, 200 mM dNTPs and 1 U Taq, or 1X SensiMix Plus (Quantace), respectively) as noted. Primer sequences are as shown (Supplemental Table 1). The relative abundance of Dβ or USF1 cDNAs was quantified following QPCR by ΔΔ threshold cycle (C_{T}) normalization to matched untreated controls, and standardized for loading variations by comparison with values obtained for β-actin. Genomic DNA PCR extracts were prepared as described (16), and DJβ rearrangements were assessed using primers and conditions as shown (Supplemental Table 1).

2.4 CpG methylation

Sodium bisulfite modification of DNA was performed using EpiTect Bisulfite Kits (Qiagen) according to the manufacturer’s instructions. Methylation was quantitated by the Sequenom MassARRAY platform with EpiTYPER analysis software (Sequenom). EpiDesigner software (Sequenom) was used to design T7-tagged and matched primers to CpG-deficient targets across Dβ2. PCR was performed using HotStarTaq (Qiagen), and products were processed using MassCLEAVE as per the manufacturer’s protocol (Sequenom). Resulting fragmented transcripts were spotted onto SpectroCHIPS for mass spectrometry analysis on a MassARRAY instrument (Sequenom) to quantify the methylated fraction in each amplicon.
2.5 Chromatin immunoprecipitation

Chromatin was prepared from formaldehyde cross-linked P5424 or the indicated thymocytes and assayed by chromatin immunoprecipitation (ChIP) as described (33). Bound and input samples (4 μl) were subjected to QPCR with 1X SensiMix Plus (Quantace) in triplicate reactions. Primers and annealing temperatures for chromatin immunoprecipitation are shown (Supplemental Table 1). Cycling parameters for 20 μl reactions were 95°C for 10 min., followed by 50 cycles of 95°C for 20 s; appropriate annealing temp for 30 s; and 72°C for 30 s. Average fold enrichment in bound fractions was calculated for triplicate amplifications as previously described (34). Where indicated, enrichment signals were further normalized to that obtained for isotype-matched control antisera.

2.6 EMSA

P5424 nuclear extracts and radioactive probes were prepared and EMSA reactions performed as previously described (20). Sequences of wild-type and mutant oligonucleotide EMSA primers are as shown (Supplemental Table 2).
3 Results

3.1 Repressed 5’PDβ2 remains accessible in DN thymocytes

Transcriptional promoters embedded in the germline sequences of antigen receptor genes drive localized recombinational accessibility of proximal gene segments (7). We have previously shown that both of the DJβ gene segment clusters in Tcrb contain germline promoters immediately upstream of their respective D segments (18, 20). However, Dβ2 germline transcription differs significantly from Dβ1 in that transcription of the unrearranged DJβ2 cluster initiates from a promoter positioned downstream of Dβ2 and proximal to Jβ2.1 (20). Following Dβ2-to-Jβ2 recombination, transcription is redirected to a promoter that sits upstream of Dβ2 (20, 35), suggesting that the upstream promoter is initially repressed prior to DJβ2 recombination.

To test the possibility that the 5’PDβ2 repressor is located downstream of Dβ2, and is deleted upon DJβ2 recombination, we sought to determine if transcription in DP thymocyte populations is restricted to rearranged DJβ2 sequences (Fig. 1). Using quantitative RT-PCR (QRT-PCR) primer pairs specific for either unrearranged Dβ2 (Fig. 1A, primers a and b) or total germline Jβ2-Cβ2 spliced message (primers c and e), we assessed levels of DJβ2 transcription in thymocytes from wt C57BL/6 mice (primarily DP cells) or mice deficient for RAG-1 (DN cells). As expected, germline transcription was readily detected in both DN and DP thymocytes, as well as in the Rag1−/−p53−/− DN cell line, P5424, while significant levels of transcription through Dβ2 were only apparent in DP cells (Fig. 1B, black bars). Because RT-PCR measures steady state transcription levels, it remains possible, if unlikely, that Dβ2 transcripts are inherently less stable in DN cells than those initiating further downstream.
Regardless, the abundance of germline Dβ2 transcription in DP cells strongly suggests that activation of 5’PDβ2 does not require DJβ2 recombination in cis.

To more directly address the role of recombination in 5’PDβ2 activity, we transiently transfected P5424 via repeated rounds of PGK-RAG1 electroporation, and identified multiple subclones that harbored biallelic DJβ1 and monoallelic DJβ2 rearrangements, as well as rearrangements in their Tcrd and Tcrg loci. Dβ sequences of one such clone (Fig. 1C, c22), as well as a representative control that maintained unrearranged DJβ2 segments on both alleles (Fig. 1C, c20) are shown in Table 1. Both c20 and c22, as well as parental P5424 contained germline transcripts downstream of Dβ2 that spliced from Jβ2 segments to Cβ2 (Fig. 1D, middle panel). However, germline transcription across Dβ2 was limited to c22 (top panel), which also expressed the rearranged (Dβ2)Jβ2.5 segment (data not shown). Similar results were obtained for three additional subclones that carried monoallelic DJβ2 rearrangements with Jβ2.1, 2.3, and 2.5, respectively (data not shown). Together with the analysis of thymocyte transcription, these data suggest that relief of 5’PDβ2 repression during DJβ2 recombination occurs at both rearranged and germline DJβ2 gene segment clusters.
Figure 1. DJβ2 rearrangements relieve repression at both germline and rearranged DJβ2 clusters. (A) Schematic representation of the DJCβ2 cluster and spliced transcripts from 5’PDβ2 and 3’PDβ2 (Jβ2.1), respectively. The positions of oligonucleotide primers used for expression and recombination assays are indicated (black arrows). (B) QRT-PCR of spliced Jβ2Cβ2 germline transcripts (gray bars, primers c and e), versus transcripts of unrearranged Dβ2 (black bars, primers a and b). Means (± SD, n = 3) are shown for DJβ2 signals relative to signals obtained in the absence of reverse transcriptase, and normalized to β-actin loading controls. (C) PCR of DJβ1 (upper panel) and DJβ2 rearrangements (lower panel, primers a and d) in the P5424-c20 and c22 subclones. (D) RT-PCR of germline transcription across Dβ2 (upper panel, primers a and b), versus total Cβ2 mRNA spliced from Jβ1 or Jβ2.2 (middle panel, primers c and e) in the P5424 subclones. PCR of the unrelated β-actin message (lower panel) served as a loading control. (E) QRT-PCR of spliced Dβ2Jβ2Cβ2 germline transcripts (primers a and e) in the P5424-c20 and c22 subclones. Means (± SD, n = 3) for each subclone relative to P5424 parental cells were calculated by ΔΔCt, and normalized to β-actin loading control signals. B cell, M12 B cell line; c20 and c22, RAG-transfected P5424 subclones; P5424, Rag1−/p53−/− DN thymocyte cell line; R1-Thy, unsorted Rag2−/− thymocytes; wt Thy, unsorted C57BL/6 thymocytes.
Table 1. D-to-J recombinant sequence in clones c20 and c22

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Sequence</th>
<th>Clone</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dβ1Jβ1 cluster:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P5424</td>
<td>Dβ1—GGGACAGGGGGCCACGTGATTCAATTCTATGG</td>
<td>a) Dβ1—GGGACAGGGG* ---t--- TTTCAACGAAA—Jβ1.4</td>
<td>b) Dβ1—GG********* ------- **AATCCGACT—Jβ1.2</td>
</tr>
<tr>
<td>P5424-c20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P5424-c22</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dβ2Jβ2 cluster:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P5424</td>
<td>Dβ2—GGGACTGGGGGGCCACATGATTCAACTGGAA</td>
<td>a) Dβ2—GGGACTGGGGGGCCACATGATTCAACTGGAA</td>
<td>b) Dβ2—GGGACTGGGGGGCCACATGATTCAACTGGAA</td>
</tr>
<tr>
<td>P5424-c20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P5424-c22</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Dβ coding sequence in parental P5424 is underlined.

We have previously shown that repressed 5′PDβ2 is bound by a variety of transcription factors including E47, Runx-1 and GATA-3 (35), suggesting that 5′PDβ2 chromatin remains accessible prior to promoter activation. To directly measure chromatin accessibility in the P5424 subclones, we next used bisulfite conversion to map the methylation state of CpG dinucleotides near Dβ2 (Table 2). Of the eight CpG dinucleotides found within 500 bp upstream and downstream of Dβ2, all were strongly demethylated in P5424, c20 and c22, correlating with the general hypomethylation of the DJβ2 cluster in DN
and DP thymocytes (36). In contrast, CpGs at -453, -373, +422 and +464 (relative to the first coding base of Dβ2) were methylated in >30-80% of screened amplicons from the Balb3T3 fibroblast cell line. Methylation in Balb3T3 was not universal however, declining markedly proximal to Dβ2. Indeed, methylation was essentially undetected at -15 and +178 CpGs in fibroblasts, suggesting that Dβ2 is protected from methylation irrespective of Tcrb accessibility. ChIP analyses of Rag2−/− thymocytes also showed that sites across Tcrb are marked by histone modifications consistent with accessible promoter regions including histone H3 lysine 9 acetylation (Fig. 2A) and H3 lysine 4 di- and trimethylation (Fig. 2B, 2C), and lacked H3 lysine 27 trimethylation found at silent promoters (Fig. 2D), corroborating previous analyses of Dβ2 accessibility (37-39). Moreover, QPCR primers that selectively amplified either the germline or (Dβ2)Jβ2.5 gene segment of c22 found equivalent levels of histone marking, again suggesting that 5’Pβ2 repression does not involve epigenetic silencing of the germline Dβ2 chromatin.

Table 2. Percent methylation of CpG dinucleotides surrounding Dβ2

<table>
<thead>
<tr>
<th>Cell line</th>
<th>-453</th>
<th>-374</th>
<th>-15</th>
<th>+178</th>
<th>+251</th>
<th>+422</th>
<th>+464</th>
<th>+529</th>
</tr>
</thead>
<tbody>
<tr>
<td>B3T3</td>
<td>0.33</td>
<td>0.48</td>
<td>0.07</td>
<td>0.05</td>
<td>0.18</td>
<td>0.44</td>
<td>0.80</td>
<td>0.21</td>
</tr>
<tr>
<td>P5424</td>
<td>0.07</td>
<td>0.04</td>
<td>0</td>
<td>0.02</td>
<td>0.04</td>
<td>0</td>
<td>0.08</td>
<td>0.04</td>
</tr>
<tr>
<td>P5424-c20</td>
<td>0.09</td>
<td>0.05</td>
<td>0.08</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.07</td>
<td>0.01</td>
</tr>
<tr>
<td>P5424-c22</td>
<td>0.14</td>
<td>0.05</td>
<td>0</td>
<td>0.01</td>
<td>0.05</td>
<td>0.05</td>
<td>0.13</td>
<td>0.05</td>
</tr>
</tbody>
</table>

aNumbering relative to the first coding nucleotide of Dβ2.
Figure 2. Repression does not alter histone modifications at Dβ2. Chromatin from RAG2-deficient C57BL/6 thymocytes or the c22 subclone of P5424 was immunoprecipitated with antibodies to: H3K9ac (A), H3K4me2 (B), H3K4me3 (C), and H3K27 (D). Resultant DNAs were analyzed by QPCR for histone modifications at the indicated Tcrb and control cad promoter (CAD) amplicons. Enrichment was calculated relative to pre-immunoprecipitation input control levels and was normalized against signals obtained with nonspecific IgG control Abs. Bars indicate means (± SD, n = 3), and are representative of 2 experiments with independent chromatin preparations.

3.2 5'PDβ2 repression requires an E box in the Dβ2 12-RSS

A mechanistic understanding of 5'PDβ2 repression necessitated the identification and characterization of the repressor element. Promoter activation in our rearranged subclones was restricted to cell lines that harbored DJβ2 joints. However, our transcriptional analyses (Fig. 1) excluded the possibility that 5'PDβ2 activity requires deletion of downstream repressor by DJβ2 recombination. Rather, the data suggest that either deletion of the repressor on one allele leads to loss of repression on the second allele, or the repressor is not deleted by Dβ recombination. To define the repressor’s location, we used luciferase reporter analyses (Fig. 3). Serial 3’ deletion of all downstream sequence, the Dβ2 coding sequence
and the 12-RSS heptamer failed to relieve 5’PDβ2 repression (Fig. 3A, compare -1104/+230 through -1104/-7). In sharp contrast, 3’ deletion of an additional 21 bp fully restored promoter activity (compare -1104/-28 to the full-length 5’PDβ2 -1104/+230 construct), suggesting that repressor activity was localized to the Dβ2 12-RSS nonamer (-28 to -20) and/or spacer (-19 to -8).

Figure 3. Repression of 5’PDβ2 promoter activity requires cis targets in the Dβ2 5’RSS. (A) The indicated PCR fragments were inserted upstream of the luciferase cassette in pGL3-Eβ. Numbering is relative to the first base of the Dβ2 coding sequence (+1). Protein extracts were assayed for luciferase activity 24 h after transfection with each plasmid and normalized to co-transfected renilla. Bars represent mean normalized luciferase activity ± SEM of at least 6 transfections, and are expressed as fold activity over the fully repressed -1104/+230 5’PDβ2 fragment. (B) pGL3-Eβ constructs containing the wt -1104/+13 fragment (top) or the indicated TTCCA substitutions were assayed for luciferase activity 24 h after co-transfection with renilla plasmid. Bars represent mean normalized luciferase activity ± SEM and are expressed as percent activity of wt -1104/+13.

To identify potential repressor elements, we screened a panel of 5’PDβ2 repression reporters (-1104/+13) into which we had introduced tiled five-base TTCCA substitutions (Fig. 3B). Whereas promoter activity was repressed in the wt construct and mutants that harbored substitutions in either the nonamer or Dβ2 coding sequence, two contiguous
mutations that spanned the spacer sequence induced promoter activity 2.5 and 3.5-fold over wt, respectively. These two mutations altered a noncanonical heptameric E box (CACGATG) that included the strongly demethylated CpG at -15, suggesting that 5’PDβ2 repression may be localized to a single cis-acting element that is upstream of Dβ2.

3.3 USF1 binds the 5’PDβ2 repressor element in DN thymocytes

We next used EMSA to determine if the element identified in our reporter assays could function as a bona fide E box (Fig. 4). Indeed, a radiolabeled probe spanning the putative E box strongly bound a single specific protein complex in nuclear extracts from the P5424 cell line (Fig. 4, lanes 1 and 8). Excess unlabeled probe readily competed for protein binding (Fig. 4, lane 2), while an oligonucleotide that carried the 10-bp repressor substitutions (-17 to -8) identified in our reporter assay failed to compete for protein binding (Fig. 4, lane 3). Mutation of the upstream CACGA sequence to ttCca, in which lower case indicates substituted bases) was sufficient to abolish competition by the unlabeled primer (Fig. 4, lane 4), whereas primers carrying the TGTA to TtccA mutation remained efficient competitors (Fig. 4, lane 5). Mutation of the critical CpG dinucleotide in the center of the putative E box also abolished competition (Fig. 4, lane 6), though its methylation on unlabeled primers had little impact on their ability to compete for protein binding (Fig. 4, lane 7). Finally, the specificity of the protein complex was confirmed by its supershift in the presence of antibodies to USF1 and USF2 (Fig. 4, lanes 9 and 10), whereas antibodies to other E proteins including E47, HEB, Myc and Max all failed to alter binding activity (Fig. 4, lanes 11-14).
Figure 4. USF1/2 bind the repressor site in the Dβ2 5’RSS spacer. Nuclear extracts from the P5424 cell line were incubated with a radiolabeled double-stranded oligonucleotide probe to the putative repressor site. Probes were incubated with nuclear extract alone (lanes 1 and 8), in the presence of 100-fold molar excess of unlabeled wt (lane 2), mutant (lanes 3-6) or methylated competitors (lane 7), or in the presence of the indicated Abs (lanes 9-15). Specific nucleoprotein (filled arrows) and Ab-supershifted complexes (empty arrows) are indicated.

We had previously shown that although the more distal E boxes within 5’PDβ2 specifically bound E47, USF1 binding upstream of Dβ2 was nonetheless detected in vivo when either Rag2−/− thymocyte or P5424 cell line chromatin was assayed by ChIP, suggesting the presence of nearby USF-binding elements (35). Our EMSA findings now suggested that our previous ChIP assays were detecting USF1 bound to the repressor E box. Specifically, USF1 but not USF2 was strongly enriched at Dβ2 in chromatin from either Rag2−/− DN thymocytes or Rag2−/− thymocytes that express a rearranged Tcrb transgene and progress to
the DP stage of development (Rxβ), but was absent in DP thymocytes from recombination-competent C57BL/6 mice (Fig. 5A).

Figure 5. USF1 binds the repressor in the absence of recombination. (A) Unsorted thymocyte chromatin from Rag2<sup>−/−</sup>, Rxβ, and wt C57BL/6 mice was immunoprecipitated with Ab to USF1 (black bars) or USF2 (grey bars), and analyzed by QPCR for binding proximal to Dβ2. (B) USF1 binding at Dβ2 (Fig. 1A, primers a and b) was analyzed by ChIP of chromatin from sorted DN1 and DN3 subsets of wt C57BL/6 thymocytes, as well as unsorted thymocytes from wt, Lat<sup>−/−</sup>, and Prkdc<sup>−/−</sup> (SCID) mice and from the P5424 and P5424-c22 cell lines. Ab-dependent enrichment over input control is expressed relative to nonspecific IgG as mean ± SD (n = 3), and is representative of two independent experiments. (C) QRT-PCR of 5′Dβ2 mRNA (Fig. 1A, primers a and e) in thymocytes from Rag2<sup>−/−</sup>, wt C57BL/6, Rxβ, and Prkdc<sup>−/−</sup> (SCID) mice. (D) QRT-PCR of USF1 mRNA in Rag2<sup>−/−</sup>, wt C57BL/6, and Rxβ thymocytes and in the P5424 subclones. Bars represent means (± SD, n = 3). Relative signals were calculated by ΔΔC<sub>T</sub> and normalized to β-actin controls.

Because both Rxβ and wt C57BL/6 thymocytes are predominantly DP, USF1 binding in the Rxβ mice suggested that loss of USF1 binding at Dβ2 is not strictly dependent on DN
to DP development. However, it remained possible that USF1 was retained in Rxβ DP cells because of accelerated DN development in the presence of the Tcrb transgene (15, 40). To exclude this possibility, we assessed USF1 binding (Fig. 5B) in sorted DN1 and DN3 subpopulations of C57BL/6 thymocytes, as well as in the DN thymocytes from mice that lack DNA-PKcs or the pre-TCR signaling molecule, linker for activation of T cells (LAT), and are consequently prevented from maturing to DP cells (41). Although USF1 binding was modestly reduced in wt DN3 cells relative to DN1, it was abolished in LAT-deficient cells that support normal Tcrb assembly but cannot complete β-selection. In sharp contrast, USF1 was strongly enriched at Dβ2 in the recombinationally-impaired DN cells of DNA-PKcs-deficient SCID mice (Fig. 5B), and this enrichment correlated with the absence 5’Dβ2 transcription (Fig. 5C). USF1 was similarly enriched in P5424, but was lost from both the germline and rearranged DJβ2 clusters of c22 (Fig. 5B and data not shown).

Despite the loss of USF1 from Dβ2 in wt DP cells, USF1 RNA levels were equivalent between Rag2−/−, Rxβ and wt thymocytes (Fig. 5D). Steady state USF1 RNA levels were also similar between P5424 and the c20 and c22 subclones, though <100-fold lower than USF1 levels in primary thymocytes. The persistent expression of USF1 in DN and DP thymocytes is consistent with its ubiquitous distribution in mammalian tissues (42), and argues against a mechanism in which loss of USF1 from the 5’PDβ2 repressor in DP cells is due to downregulation of USF1 expression. Indeed, we also found that USF1 was absent from Dβ2 sequences in chromatin isolated from either the Balb3T3 fibroblast or M12 B cell lines despite USF1 expression in both (data not shown). Taken together with our luciferase and EMSA findings, our in vivo analyses strongly suggest that USF1 binding at the Dβ2 12-RSS
is sufficient to repress 5′PDβ2 activity, and that loss of USF1 binding is triggered by DJβ2 recombination rather than by developmental progression.

3.4 Genotoxin-induced DSBs lead to loss of USF1 binding and relieve 5′PDβ2 repression

Unlike RAG-2 deficiency, lymphocytes that lack DNA-PKcs accumulate DSB intermediates of V(D)J recombination, leading to the activation of a variety of transcriptional programs via the related PI3K, ATM (31). However, Dβ2 remains bound by USF1 in DNA-PKcs-deficient thymocytes, despite intact ATM signaling (Fig. 5). Given that DNA-PKcs directly regulates USF1-dependent expression of fatty acid synthase in response to insulin signaling (29), we wished to determine if USF1 binding at Dβ2 is similarly regulated by DNA-PKcs. Treatment of Rxβ thymocytes with either ionizing radiation or etoposide, both of which induce DSBs, led to loss of USF1 and a reciprocal increase in 5′PDβ2 expression (Fig. 6A, 6B). However, this genotoxin-induced derepression was blocked in cells pretreated with the DNA-PKcs inhibitor, Nu7026 (Fig. 6C, 6D).
**Figure 6.** Genotoxic DSBs relieve 5’PDβ2 repression. (A and C) ChIP-QPCR analysis of USF1 enrichment at Dβ2. (B and D) QRT-PCR of 5’Dβ2 mRNA. (A and B) Chromatin and mRNA signals in untreated Rxβ thymocytes (-) or in Rxβ thymocytes 1 d after treatment with DMSO carrier, increasing doses of ionizing radiation, or 3 μM etoposide. (C and D) Chromatin and mRNA signals in Rxβ (black bars) and P5424 (gray bars) 1 d after treatment with etoposide alone, or after pretreatment with Nu7026. In each case, bars represent means ± SD (n = 3) for each sample. Fold enrichment of USF1 and relative gene expression were calculated as described in Fig. 5.

USF1 is a pleiotropic stress response transcription factor that has been implicated in the activation or repression of many genes across a broad spectrum of tissues (43). Previous studies have shown that activation of the carbamoyl-phosphate synthetase 2, aspartate transcarbamylase, and dihydroorotase (cad) gene promoter in cycling cells is dependent on displacement of USF1 by Myc/Max heterodimers (44). As predicted, USF1 was absent from the transcriptionally active cad promoter in P5424 cells (Fig. 7). Genotoxic insult led to an enrichment of USF1 and inhibition of cad expression. However, USF1 binding and cad expression were not altered by addition of the DNA-PKcs inhibitor, Nu7026. These data
suggest that USF1 binding is regulated by different mechanisms at the cad and 5’PDβ2 promoters. Additionally, the loss of USF1 from both promoters suggests that DNA damage may alter the regulation of multiple USF1 target genes.

Our finding that genotoxic DSBs can lead to activation of 5’PDβ2 suggests that development-dependent promoter activation is mediated by DNA-PKcs in response to physiologic RAG DSBs. Such a mechanism would account for derepression at either germline or rearranged DJβ2 clusters. However, 5’PDβ2 activity in the rearranged subclones

Figure 7. Genotoxic DSBs induce USF1 binding and loss of cad expression in P5424 cells. Shown are representative ChIP-QPCR of USF1 binding at the cad promoter (A) and cad mRNA levels (B) in P5424 1 d after treatment with ionizing radiation or etoposide ± Nu7026. Bars represent means ± SD (n = 3) for each sample. Fold enrichment of USF1 and relative gene expression were calculated as described in Fig. 5.
was restricted to those that harbored DJβ2 joints. Consequently, it remained unclear whether loss of 5’PDβ2 repression during V(D)J recombination requires specific DJβ2 DSBs, or results from the general accumulation of RAG DSBs. If loss of repression is regulated in trans by DSB-induced activation of DNA-PKcs, then perhaps the constitutive activation of 5’PDβ2 in c22 reflected the presence of extensive and/or persistent DSBs generated during the repeated transfections of p53-deficient P5424 cells. To test this possibility, we assessed the impact on promoter activity of inhibiting DNA-PKcs. In contrast to controls exposed to DMSO carrier or inhibitors of p38 MAPK or ATM, when c22 cells were cultured with Nu7026, both USF1 binding and 5’PDβ2 repression were induced (Fig. 8). The ability to restore repression in c22 suggests that 5’PDβ2 activity is sensitive to DSB signals, and may not specifically require breaks in the DJβ2 cluster. More generally, our data suggest a model in which promoter contributions to Dβ2 recombinational accessibility are in turn regulated by the DSB-sensitive repressive actions of USF1.
**Figure 8.** 5’PDβ2 repression is restored in c22 by inhibition of DNA-PKcs. Shown are representative ChIP-QPCR of USF1 binding at Dβ2 (A) and Dβ2 mRNA levels (B) in P5424-c22 1 d after treatment with the indicated kinase inhibitors. Bars represent means ± SD (n = 3) for each sample. Fold enrichment of USF1 and relative gene expression were calculated as described in Fig. 5.
4 Discussion

Recruitment of chromatin-modifying proteins to germline promoters and the subsequent transcriptional read-through of downstream RSSs contribute to gene segment recombinational accessibility (7). How such promoter-mediated accessibility shifts during lymphocyte development to target individual gene segments or clusters remains unclear. The Tcrg recombination bias toward Vγ3 and Vγ4 that is observed in fetal thymocytes is overcome in adult thymocytes by E2A-dependent repression of the Vγ3 and Vγ4 promoters (45, 46). Similarly, repression of distal germline promoters allows initial Vα-to-Jα joints assembled during Tcra recombination to target proximal Js (47). We have shown that repression of germline promoter activity upstream of Dβ2 redirects germline transcription downstream of Dβ2 RSSs (20), which may account for the persistence of unrearranged DJβ2 sequences relative to DJβ1 in fetal thymocytes (21-23).

PDβ1 is required to displace histones from the Dβ1 23-RSS, augmenting its accessibility for RAG protein binding. Conversely, 5’PDβ2 repression may account for nucleosome occlusion of the Dβ2 23-RSS (9). We now show that this repression of 5’PDβ2 activity is mediated by binding of USF1 to a noncanonical E box within the Dβ2 12-RSS spacer sequence, and that DSBs can induce a DNA-PKcs-dependent loss of USF1 that relieves repression. Based on the relative inefficiency with which PDβ1 directs DJβ1 recombinational accessibility when repositioned downstream of Dβ1 (19), 5’PDβ2 repression prior to recombination would be expected to limit DJβ2 accessibility. DSBs generated during recombination would be expected to subsequently increase accessibility by inducing a DNA-PKcs-dependent loss of 5’PDβ2 repression.
RAG DSBs impact the regulation of a wide range of genes in developing lymphocytes, principally through activation of the PI3 kinase, ATM (31). By extension, initial steps in V(D)J recombination could induce signals that feedback signals to regulate subsequent steps in Ag receptor assembly, perhaps acting to influence the order of Tcr gene assembly beyond an otherwise stochastic process in DN thymocytes. Indeed, the transcriptional regulation of Dβ2 suggests separate pathways by which recombination could feed back to regulate Tcrb assembly. Prior to recombination, 3’PDβ2 activity is dependent on low levels of constitutively nuclear NFκB (48). However, given that NFκB is activated by ATM (31), we speculate that RAG DSBs may simultaneously induce an ATM-dependent increase in activity of 3’PDβ2 and a DNA-PKcs-dependent activation of 5’PDβ2. As such, the downstream promoter may ensure Jβ2 accessibility until the onset of V(D)J recombination extends accessibility to the Dβ2 RSSs. However, the location of the repressor box in the Dβ2 12-RSS spacer suggests an alternate model in which USF1 could allosterically limit RAG access to the Dβ2 12-RSS prior to DJβ2 recombination, and thereby contribute to beyond 12-23 regulation (5, 13, 49). The presence of transcription factor binding sites within an RSS is not unprecedented. The AP-1 protein c-Fos binds sites present in the 23-RSSs of both Dβ1 and Dβ2, and may enhance RAG deposition at the Dβ1 23-RSS while impeding RAG deposition at the 12-RSS (14). RAG1 and RAG2 are strongly enriched at both Dβ gene segments (39). However, a more detailed examination that distinguishes RAG occupancy between the closely spaced 12- and 23-RSSs of each Dβ will be necessary to test the potential of USF1 to specifically limit RAG access to the Dβ2 12-RSS. Although regulation of Dβ2 promoter activity may impact the usage of individual DJβ segments in Vβ-to-DJβ recombination, DSB-inducing signaling could similarly impact ongoing
rearrangements of other Tcr loci. For example, if USF1 is lost from the Dδ2 promoter where it activates Dδ2 transcription (50), RAG DSBs could theoretically impact Tcrd assembly.

USF1 is a ubiquitously expressed stress-response protein that plays a critical role in lipid metabolism, cell cycle regulation, proliferation control, tumor suppression, and response to UV damage (25). It has also been linked to immune system development and function, regulating genes such as CIITA, β2-microglobulin, Igh and Igλ (25) and germline Dδ2 promoter activity in Tcrd (50). In each of the latter cases, USF1 acts as a transcriptional activator, whereas it appears to function as a repressor of the 5’PDβ2 and cad promoters. In addition, USF1 is critical to the chromatin barrier function of the chicken β-globin insulator (51). We found no evidence that USF1 regulates Dβ2 chromatin accessibility. Indeed, multiple studies have shown that the Dβ2 sequence is accessible in DN and DP thymocytes (36-39). Rather, our data are consistent with a narrower role for USF1 in limiting germline transcription across Dβ2, which may in turn be necessary to enhance the recombinational accessibility of the Dβ2 RSSs.

Despite considerable overlap in the DNA-binding properties of various E proteins, our data suggest that 5’PDβ2 repression is uniquely mediated by USF1. EMSA data (Fig. 4) suggest that the repressor site can be bound by USF1 and/or USF2. Nonetheless, we did not detect significant USF2 binding in vivo. This absence, together with the DNA-PKcs-sensitivity of 5’PDβ2 repression, suggests that USF2, which lacks the phosphorylation sites that regulate USF1 function (43), does not regulate the 5’PDβ2 repressor. The apparent inability of other E proteins to bind the repressor may owe to its noncanonical heptameric structure. Though Myc-Max heterodimers can bind synthetic heptameric sites, they display a clear preference for canonical hexameric E-boxes (52). Conversely, the UV-responsive
regulation of laminin-5 expression in epithelial cells is mediated by USF1 binding to a heptameric E box in the *lama3* promoter (53). Although USF1/2 double knockout mutations are lethal to embryonic development, no defects in thymocyte development or TCR repertoire diversity were reported for *USF1*−/− mice (54). Indeed, our findings would predict that loss of USF1 would manifest in a TCRβ repertoire subtly skewed toward inclusion of DJβ2 joints at the expense of DJβ1. However, it should be noted that thymocyte development and *Tcrb* recombination are essentially normal even in the complete absence of either the DJβ1 or DJβ2 gene segment clusters of mutant (17, 55) or New Zealand White mice.

Given P5424’s p53 deficiency and relatively poor transfectability, it is unclear whether initial USF1 loss from the 5′PDβ2 repressor in the P5424 subclones was triggered by RAG DSBs in general, DJβ2 DSBs specifically, or DNA stress accumulated after multiple transfections with PGK-RAG1. Indeed, similar derepression of 5′PDβ2 was observed in thymocytes treated with UVB radiation (data not shown), which induces pyrimidine dimer and DNA adduct formation, and leads to USF1 phosphorylation by p38 MAPK (28). The ability of Nu7026 to restore USF1-mediated repression in c22 suggests that sustained derepression in the P5424 subclones may derive from persistent DNA damage or RAG DSBs accrued during repeated transfection. We speculate that DNA-PKcs recruited for the repair of new DJβ2 joints would, by dint of its proximity to the Dβ2 12-RSS, be in position to ensure USF1 removal and 5′PDβ2 activation for subsequent V-to-DJβ2 recombination. However, a DNA-PKcs-dependent modulation of USF1 would not appear to require specific DJβ2 rearrangements *per se* to achieve 5′PDβ2 derepression.

In conclusion, our findings suggest that in addition to functional differences between the various Dβ and Jβ RSSs (5), differential promoter usage at Dβ2 and its attendant
epigenetic modulations may account for the longstanding observations of enhanced DJβ1 recombination efficiency relative to DJβ2 (21-24). Future studies will be necessary to determine the precise mechanism by which USF1 mediates repression of 5’PDβ2, how repression is resolved, and the extent to which this repression impacts DJβ2 recombinational efficiency. However, the remarkable detail to which DJβ transcriptional control has now been mapped renders Tcrb assembly an ideal system to tease apart the relative contributions of RSS- and promoter-driven control to recombinational accessibility.
5 References


### Supplemental Data

#### Table S1. PCR primer sets

<table>
<thead>
<tr>
<th>Amplicon</th>
<th>Strand</th>
<th>Sequence</th>
<th>Tm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Djβ1</td>
<td>S</td>
<td>GAT CCA GAA TGC TTT CAC G</td>
<td>63°C</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>CTG CAT CCT TTG CTG CTA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>S</td>
<td>CCC TCT CAG TCA GAC AAA CCT CTC</td>
<td></td>
</tr>
<tr>
<td>Djβ2</td>
<td>A (gl)</td>
<td>CTT CCA GTT GAA TCA TTG TGG CCC C</td>
<td>63°C</td>
</tr>
<tr>
<td></td>
<td>A (DJ2.5)</td>
<td>CAA AGT ACT GGG TGT CTT GGT TAC</td>
<td></td>
</tr>
<tr>
<td>Jβ2.2</td>
<td>S</td>
<td>CAG GAC TGT GCA AAC ACC GGG CAG</td>
<td>63°C</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>TCC CAG CAT CCA ATC CAG AAA TCC</td>
<td></td>
</tr>
<tr>
<td>Jβ2.4</td>
<td>S</td>
<td>GCA TTT CCC AGG CTA AGG</td>
<td>56°C</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>CCC CAG AAC CCA ACA CAA</td>
<td></td>
</tr>
<tr>
<td>Eβ</td>
<td>S</td>
<td>GAA TGG TAG TGG GGA AAT CT</td>
<td>56°C</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>TCG GTC AGG GAA ACA AGT AT</td>
<td></td>
</tr>
<tr>
<td>cad</td>
<td>S</td>
<td>GTC TGC GTG CTT GCC CTG TCT CAG C</td>
<td>71°C</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>CGG GCT TGC TTA CCC ACT TCC CCA GC</td>
<td></td>
</tr>
<tr>
<td>3’Djβ1 RNA</td>
<td>S</td>
<td>GGC TAC CTC ACT TTG ATG</td>
<td>63°C</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>CCC CAG GCC TCT GCA TG ATG TTC TGT GTG</td>
<td></td>
</tr>
<tr>
<td>3’Djβ1 RNA</td>
<td>S</td>
<td>GGT CTT ATA ACA TCT ATG CAT C</td>
<td>63°C</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>CTG TGT GAC AGG TTT GGG TGA GCC CTC TG</td>
<td></td>
</tr>
<tr>
<td>β-actin</td>
<td>S</td>
<td>GTG GGC CGC CCT AGG CAC CAG</td>
<td>56°C</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>CTC TTT GAT GTC ACC CAC GAT TTC</td>
<td></td>
</tr>
<tr>
<td>DJβ2 REC</td>
<td>S</td>
<td>GTA GAC ACC TGT GGA GAA GAA ACT</td>
<td>63°C</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>TGA GAG CTG TCT CCA ACT ATC GAT</td>
<td></td>
</tr>
<tr>
<td>Amplicon</td>
<td>Strand</td>
<td>Sequence</td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>--------</td>
<td>-------------------------------</td>
<td></td>
</tr>
<tr>
<td>wt</td>
<td>S</td>
<td>TTT TTT GTA TCA CGA TGT AAG ATT GTG G</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>AAT GTT ACA TCG TGA TAC AAA AAA GTT</td>
<td></td>
</tr>
<tr>
<td>m</td>
<td>S</td>
<td>TTT TTT GTA TTT CCT TTC CTG ATT GTG G</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>AAT GAG GAA AGG AAA TAC AAA AAA GTT</td>
<td></td>
</tr>
<tr>
<td>m(a)</td>
<td>S</td>
<td>TTT TTT GTA TTT CCT TGT AAG ATT GTG G</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>AAT GTT ACA AGG AAA TAC AAA AAA GTT</td>
<td></td>
</tr>
<tr>
<td>m(b)</td>
<td>S</td>
<td>TTT TTT GTA TCA CGA TTC CTG ATT GTG G</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>AAT GAG GAA TCG TGA TAC AAA AAA GTT</td>
<td></td>
</tr>
<tr>
<td>mCG</td>
<td>S</td>
<td>TTT TTT GTA TCA TGA TGT AAG ATT GTG G</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>AAT GTT ACA TGA TAC AAA AAA GTT</td>
<td></td>
</tr>
<tr>
<td>CmeG</td>
<td>S</td>
<td>TTT TTT GTA TCA meCGA TGT AAG ATT GTG G</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>AAT GTT ACA TmeCG TGA TAC AAA AAA GTT</td>
<td></td>
</tr>
</tbody>
</table>
CHAPTER 3

USF1 regulates feedback control of Dβ2 rearrangement to enforce B12/23 at the level of D-to-D restriction
Abstract

The T-cell receptor Tcrb gene is assembled by V(D)J recombination in a tightly-ordered process during thymocyte development, where Dβ-to-Jβ joining at two distinct clusters precedes Vβ-to-DJβ rearrangement. Joins involving the Dβ1 segment initiate earlier than those incorporating Dβ2, and this disparity in rearrangement potential is likely the consequence of differences in promoter positioning. While a promoter upstream of Dβ1 drives its recombinational accessibility, Dβ2 is flanked by two promoters, the downstream of which is initially active, limiting accessibility at Dβ2. We recently showed that repression of the 5’PDβ2 promoter is mediated by the stress-response regulator, upstream stimulatory factor 1 (USF1). DNA double-strand breaks (DSBs) generated during initial rounds of rearrangement lead to a loss of USF1 binding at an E box within the 5’ Dβ2 RSS, and activation of 5’PDβ2, in a DNA-PK dependent manner. Because USF1 binds within the 5’ Dβ2 RSS, we wanted to assess the impact of USF1 binding on Dβ2 recombination. We also wished to determine the DNA-PK-dependent mechanism by which USF1-mediated repression is resolved. In this study, we show that both DNA-PK-mediated phosphorylation of USF1 at serine 262 (S262) as well as a dominant-negative USF that abrogates DNA binding increase transcription from 5’PDβ2. Together, the data suggest that USF1 and the DSB repair protein DNA-PK cooperate to regulate Dβ2 recombinational accessibility through a unique feedback mechanism. Moreover, we show that levels of D-to-D rearrangements are increased in thymocytes from USF1<sup>−/−</sup> mice, suggesting a novel role for USF1 in preserving ordered Tcrb assembly by enforcing recombinational control beyond the 12/23 rule.
1 Introduction

T and B lymphocytes generate de novo the genes encoding the variable region of the T-cell receptor (TCR) or B cell immunoglobulin (Ig) through the somatic rearrangement of variable (V), diversity (D), and joining (J) coding segments (1, 2). V(D)J recombination is mediated by a lymphocyte-specific recombinase, encoded by recombination-activating genes 1 (RAG-1) and 2 (RAG-2), that targets and cleaves DNA at conserved recombination signal sequences (RSSs) flanking each gene segment (3). The cleaved coding ends are then promptly resolved by ubiquitous DNA repair machinery of the non-homologous end joining (NHEJ) pathway (4), as the persistence of double-strand breaks (DSBs) is inherently dangerous to genome stability (5).

The specificity of recombination is imposed, in part, at the level of RAG recognition and cleavage at individual RSSs. Each RSS consists of a palindromic heptamer, situated immediately adjacent to the coding segment, and an AT-rich nonamer. The two motifs are separated by a less well-conserved spacer sequence consisting of either 12 or 23 base pairs (referred to as a 12-RSS or 23-RSS, respectively) (1), and physiological rearrangement proceeds only between gene segments flanked by RSSs with dissimilar spacer lengths, a principle known as the 12/23 rule (1). Due to the intrinsic risks associated with DSB generation, RAG targeting to RSSs must be tightly controlled. Yancopoulos and Alt first proposed that rearrangement was associated with gene segment accessibility, in that local chromatin structure could restrict or permit access of the recombinase to RSS targets (6). Indeed, the precisely-timed programs of TCR and Ig gene assembly have been linked to alterations in chromatin organization that direct RAG-mediated cleavage of target RSSs in a tissue-, lineage-, and stage-specific manner (7). Such changes in accessibility are driven
largely by cis-acting promoters and enhancers that remodel chromatin to facilitate RAG access to RSS targets (8).

The majority of T lymphocytes express a heterodimeric αβ TCR, encoded by the TCRα (Tcra) and TCRβ (Tcrb) loci. Tcrb rearranges first at the CD4−CD8− double negative (DN) stage of development, followed by Tcra assembly at the CD4+CD8+ double positive (DP) stage. The Tcrb locus is comprised of 30 Vβ gene segments positioned ~250 bp upstream of two Dβ-Jβ-Cβ cassettes, each containing a single Dβ and six functional Jβ segments. Tcrb rearranges in a two-step assembly with simultaneous Dβ-to-Jβ joining at both DJβ cassettes on both alleles preceding monoallelic Vβ-to-DJβ rearrangement (9, 10). Recombinational accessibility at each DJβ cluster is directed by activation of germline promoter elements that complex independently with a single enhancer (Eβ) (11). Transcription and rearrangement of the DJβ1 region is controlled by a single promoter, PDβ1, that is situated immediately upstream of Dβ1 (12). Its activation in early DN cells directs chromatin changes at both the Dβ1 and Jβ1 RSS elements that facilitate RAG binding and rearrangement. While PDβ1 is indispensible for rearrangements involving the upstream cassette, its deletion does not impact recombination at the DJβ2 cluster (13). Unlike Dβ1, two independent promoters flank the Dβ2 gene segment. Prior to recombination, the downstream promoter (3’PDβ2) is activated, ensuring full recombinase access to Jβ2 but only limited access to Dβ2. A second promoter (5’PDβ2) sits immediately upstream of Dβ2, but is repressed until DJβ2 rearrangement deletes 3’PDβ2, shifting transcription upstream (14). Differences in the location of promoter activity between the two cassettes could account for differences in Dβ segment usage, as primary rearrangements favor inclusion of Dβ1 (15-17). Analysis of PDβ1 revealed that repositioning the promoter downstream of Dβ1 severely
attenuated DJβ1 rearrangement in transgenic substrates, suggesting that promoter activity upstream of the DJβ1 segment is critical for full recombinational accessibility (18). It is plausible, therefore, that preferential recombination of DJβ1 is enforced by initial repression of the 5’PDβ2 promoter. Because the two DJβ cassettes are in line and rearrange by deletion, targeting DJβ1 for primary Tcrb recombination ensures that the downstream DJβ2 cassette would remain available for secondary rearrangements to rescue nonproductive joins involving DJβ1, thereby maximizing Tcrb diversity.

* Tcrb is organized such that Vβ segments are flanked with 23-RSSs, DJβ segments with 5’ 12-RSSs and 3’ 23-RSSs, and Jβ segments with 12-RSSs. Thus, direct joining of Vβ segments with unrearranged DJβ or Jβ sequences, as well as DJβ1-to-DJβ2 rearrangements, are permitted based on the 12/23 rule. Direct Vβ-to-Jβ joining, however, is rarely detected within Tcrb (2, 17, 19, 20), suggesting that additional constraints beyond the generic 12/23 restriction (so-called beyond 12/23 or B12/23) must limit Tcrb assembly to the canonical pattern of D-to-J followed by V-to-DJ joining. Preserving ordered Tcrb rearrangement is critical, as exclusion of the DJβ segment from the Tcrb CDR3 hypervariable domain would substantially reduce receptor diversity (21).

Extensive research has focused on understanding the mechanisms that enforce beyond 12/23 restriction. Experiments using extrachromosomal recombination substrates have shown that Vβ segments preferentially target the 5’ DJβ1 12-RSS over those flanking Jβ segments (20, 22-24). Corroborating these findings, studies in transgenic Jβ1<sup>M5</sup> mice (in which the entire DJβ2 cassette and the DJβ1 segment were removed by targeted deletion) demonstrated that replacement of the Jβ1.2 RSS with the DJβ1 5’ RSS resulted in nearly exclusive Vβ-to-Jβ1.2 usage on mutant alleles (25). Separate mouse experiments revealed that replacing the
Vβ14 23-RSS with the 3’ Dβ1 RSS considerably skewed the thymocyte population toward Vβ14+ cells (26), suggesting that the Dβ RSS elements are more effective recombination targets than either their Vβ or Jβ counterparts. Recently, c-Fos, a component of the transcription factor AP-1, was shown to bind a highly-conserved AP-1 binding site within the 3’ 23-RSSs of both Dβ1 and Dβ2 gene segments and to interact directly with the RAG proteins (27). In Fos−/− mice, not only was Dβ-to-Jβ rearrangement attenuated, but aberrant joins were detected between Vβ segments and germline Dβ1, violating ordered assembly within Tcrb. These findings suggest that c-Fos regulates ordered rearrangement and ensures Dβ usage by facilitating RAG recruitment to the 3’ Dβ RSSs. Interestingly, although the Dβ2 3’ RSS also binds c-Fos, no joins between Vβ segments and unrearranged Dβ2 were detected in Fos−/− mice, suggesting additional factor(s) preclude direct Vβ-to-Dβ2 joining (27).

Upstream stimulatory factor 1 (USF1) is a ubiquitously-expressed transcriptional regulator involved with responses to cellular stress and DNA damage (28, 29). USF1 belongs to a large family of basic helix-loop-helix (bHLH) proteins and binds E box targets (CANNTG) genome-wide, either as homodimers or as heterodimers with USF2 (30-32) or related E proteins (28). We recently showed that USF1 binds an E box element within the 5’ Dβ2 12-RSS in recombinase-deficient thymocytes, resulting in repression of the 5’PDβ2 promoter. When RAG-deficient thymocytes were treated with genotoxic agents to induce DSBs, however, we observed a loss of USF1 enrichment at Dβ2 and consequent activation of 5’PDβ2 in a DNA-PKcs-dependent manner (33).

To evaluate the role of USF1 phosphorylation in 5’PDβ2 activation, we created a series of mutant USF1 overexpression constructs containing single amino acid substitutions at S262. We also examined the impact of overexpressing a recombinant USF2 that lacks its
DNA-binding domain. Overexpression of either the phosphomimetic S262D mutant USF1 or the dominant-negative USF2 increased 5’PDβ2 transcription in the DNA-PK-deficient CD4+CD8− DN thymocyte cell line, SCID-Tac. Given the location of USF1 binding within the Dβ2 5’ RSS, we asked whether USF1 occupancy might contribute to the B12/23 restriction that prevents Dβ1-to-Dβ2 recombination. While the Dβ1 23-RSS pairs with the Dβ2 12-RSS with the equivalent efficiency as all six Jβ1 RSSs on plasmid substrates (23), D-to-D joins are only infrequently detected in vivo (16, 34-36). To test the possibility that USF1 contributes to B12/23 restriction against noncanonical D-to-D joins, we analyzed the Tcrb rearrangement patterns in thymic tissue from USF1−/− mice. The analyses revealed a considerable increase in inappropriate D-to-D joining in USF1−/− animals relative to wild-type controls. Collectively, the findings in this study suggest a feedback mechanism of Dβ2 recombination, where initial DSBs generated during recombination activate repair factors that, in turn, regulate successive rearrangements. Our data also suggest a novel role for USF1 in enforcing B12/23 in trans, at the level of D-to-D restriction.
2 Materials and Methods

2.1 Cell culture

The DNAPKcs−/− SCID-Tac CD4−CD8− DN thymocyte cell line was kindly provided by David Wiest, Fox Chase Cancer Center (37). Cells were cultured at 37°C/5% CO₂ in RPMI 1640 medium supplemented with 10% FBS, 2 mM L-glutamine, 0.01% penicillin/streptomycin, and 50 μM β-mercaptoethanol.

2.2 Plasmids and transient transfection

For all transient transfections, 10⁷ SCID-Tac cells in log-phase growth were electroporated (300 μL serum-free RPMI, 300 V/950 μF) with 6 μg expression plasmid and 1 μg phCD4cDNA (human CD4 plasmid), and allowed to recover 24 hours. Transfectants were enriched using Dynabeads Protein A magnetic beads (Life Technologies) conjugated to hCD4 antibody (Invitrogen, MHCD0400), and allowed to recover 24 hours.

USF1 expression plasmids were created by cloning a mouse USF1 cDNA amplicon into the EcoRI/EcoRV sites of pcDNA6 (Invitrogen). Site-specific mutations were introduced into individual constructs using Quikchange II (Stratagene) according to the manufacturer’s recommendations. Sequences for the oligonucleotide primers used to introduce each mutation (shown in bold) are listed in Table S1. The integrity of all expression vectors was confirmed by sequencing. Additional expression plasmids (pCMV4, pN4, and pUSF2-VP16) were generously provided by Howard Towle, University of Minnesota. pPΔB was generously provided by Karl Olson, Michigan State University.
2.3 *Germline transcription*

RNA was extracted using RNAzol RT (Sigma), according to the manufacturer’s instructions. RNAs (300 ng) were reverse transcribed using 100U Moloney murine leukemia virus reverse transcriptase (Fermentas) and oligo d(T) primers. The resultant cDNAs were amplified by QRT-PCR (40 cycles) using 2X SensiMix (Bioline). Primer sequences are as shown (Table S1). Following QRT-PCR, the relative abundance of Dβ2 or USF-1 cDNAs was quantified by ΔΔC_T normalization to matched vector controls, and standardized for loading variations by comparison to values obtained for β actin expression.

2.4 *Rearrangement analysis in USF1−/− mice*

Thymic and splenic tissues from 5-week-old male and female WT C57BL/6J and *USF1−/−* knockout mice (38) were kindly provided by Marie-Dominique Galibert, CHU Rennes. Tissues were homogenized, digested overnight (10 mM Tris 8.0, 100 mM NaCl, 0.5% SDS, 25 mM EDTA supplemented w/ 50 μg/mL proteinase K), and genomic DNA was purified by phenol-chloroform extraction. PCR amplification of D1J1 and D2J2 rearrangement products was performed in 20 μL reactions using 2x MyFi Mix (Bioline) and 20 μM of each primer (Table S1). Initial denaturation was performed at 95°C for 1 minute, followed by 30 cycles consisting of 15 seconds at 95°C, 15 seconds at 63°C and 1.5 minutes at 72°C, with a final extension of 2 minutes at 72°C. D-to-D rearrangement was assessed by PCR in 20 μL reactions using 2x MyFi Mix (Bioline) and 20 μM of each primer (Table S1), and amplified for 35 cycles (95°C for 15 sec, 63°C for 15 sec, 72°C for 15 sec). Cλ loading control reactions (25 cycles of 95°C for 15 sec, 64°C for 15 sec, 72°C for 15 sec) were amplified in 20 μL reactions with Taq polymerase.
2.5 Southern analysis

Rearrangement PCR products were separated by electrophoresis using 1% (D1J1 and D2J2) or 2% (D-D REC) agarose gels, and transferred to Zeta- Probe membranes (Bio-Rad) overnight by alkaline transfer. Oligonucleotide probes (Table S1) were labeled with $[\gamma^{-32}P]ATP$ in 10 μL reactions at 37°C for one hour using 10 units T4 PNK (Thermo Scientific), and hybridized to Zeta-Probe membranes at 60°C for at least 6 hours. Membranes were washed twice at room temperature and visualized by autoradiography.

2.6 Sequence analysis of D-to-D joins

Genomic DNAs from control and USF1−/− knockout mice were amplified by PCR, using primers to detect D-to-D recombination (Table S1). 50 μL reactions using Q5 high-fidelity polymerase (NEB) were run for 35 cycles (98°C for 10 sec, 63°C for 30 sec, 72°C for 30 sec). PCR amplification products were cloned into pBluescript (Stratagene), and individual clones were isolated and sequenced.

2.7 RIC analysis of potential RSSs

The RSSsite web server (39) was used to assign recombination information center (RIC) scores to 130 bp of sequence between Dβ2 and Jβ2.1 to identify potential cryptic 12-RSS sites. Input sequence was compared to the mouse RIC profile for sites containing a 12-bp spacer using the DNAGrab algorithm. A 12-RSS with a RIC value $\geq$-38.81 is scored as pass (functional); values below the threshold are scored as fail (non-functional).
3 Results

3.1 Phosphorylation of USF1 at S262 leads to 5’PDβ2 activation

We previously showed that USF1 binds an E box within the 5’ Dβ2 12-RSS to repress the 5’PDβ2 promoter in the absence of recombination, and that DSBs induce a loss of USF1 to relieve repression. Interestingly, though, when we analyzed thymocytes from SCID mice, which are RAG-competent but have a defective DNA-PKcs gene, we found that USF1 remained bound at Dβ2 and repressed 5’PDβ2 activity. Likewise, we found that pretreating RAG-deficient thymocytes or cultured cells with a DNA-PK inhibitor (Nu7026) prior to genotoxic insult prevented USF1 displacement and promoter activation, strongly suggesting a role for DNA-PK in regulating USF1-mediated repression of Dβ2 activity (33). An analysis of lipogenic gene regulation previously showed that DNA-PK phosphorylated USF1 at the serine 262 (S262) residue, leading to activation of the fatty acid synthase (FAS) promoter under feeding conditions (40).

To test the possibility that DNA-PK similarly modifies USF1 at Dβ2, we evaluated the impact of mutating USF1 S262 on 5’PDβ2 transcription in the DNA-PKcs−/− pro-T cell line, SCID-Tac. Because SCID-Tac cells are RAG-competent but lack the repair protein DNA-PK, RAG-DSBs accumulate. DNA-PK deficiency not only blocks religation of broken DNA ends, but also prevents phosphorylation of USF1 at S262. SCID-Tac cells were transiently transfected with expression vectors harboring wildtype or mutant USF1 cDNAs (Fig. 1A), along with a plasmid expressing human CD4. After enriching for CD4+ cells, we measured USF1 and Dβ2 expression by QRT-PCR. As expected, USF1 expression was approximately 1100- to 2200-fold higher in cells that received USF1 cDNA than in those that were transfected with empty vector (Fig. 1B). Despite the dramatic increase in USF1 levels,
cells transfected with wild-type USF1 (pUSF1-WT) showed no increase in Dβ2 transcription relative to vector control, nor did cells that were transfected with a nonphosphorylatable mutant (pUSF1-S262A), where serine 262 was replaced with alanine (Fig. 1C). Indeed, Dβ2 levels were modestly reduced in the S262A transfectants. In contrast, when we transfected SCID-Tac with a phosphomimetic mutant (pUSF1-S262D), containing a single amino acid substitution from serine to aspartate at position 262, we observed a roughly four-fold increase in 5’PDβ2 expression. Like the FAS promoter, these data suggest that DNA-PK regulates the 5’PDβ2 promoter by phosphorylating USF1 at S262.

Chromatin immunoprecipitation further suggests that this phosphorylation leads to USF1 displacement from 5’PDβ2 (33), paralleling the role of USF1 at the apolipoprotein A5 promoter, where phosphorylation of USF1 by phosphatidylinositol 3-kinase (PI3K) abolished its binding (41). Given that USF1 binds DNA and functions as either a homo- or heterodimer, we next transfected SCID-Tac cells with a dominant-negative form of USF designed to abolish both USF1 and USF2 binding activity. The pCMV4-PΔB encodes a murine USF2 that lacks amino acids 228–247 of the basic region (42), and consequently cannot bind DNA whether as homo- or heterodimeric assemblies. When we transfected SCID-Tac cells with overexpression vectors for wild-type USF2 (pN4) or USF2 fused to the constitutively-active VP16 transactivation domain (pUSF2-VP16), we did not observe an increase in 5’PDβ2 expression relative to vector control (Fig. 1D). Conversely, cells transfected with pPΔB showed a robust increase in transcription levels, approximately 40-fold above cells transfected with empty vector. Taken together, these studies suggest that repression of Dβ2 transcription is relieved when USF1 is phosphorylated at serine 262 by DNA-PK, and is then displaced from the promoter.
Figure 1. Phosphomimetic and dominant-negative forms of USF increase 5′Dβ2 expression. (A) Schematic representation of USF expression vectors. Missense and dominant-negative (pPΔB) mutants were derived from murine USF1 or USF2 cDNA, respectively. pPΔB contains a deletion of amino acids 228–247 of the USF2 basic region. All proteins were expressed from the cytomegalovirus (CMV4) promoter. (B) QRT-PCR of USF1 transcripts. Means (± SD, n=3) are shown for USF1 signals relative to empty vector control and normalized to β-actin loading controls. (C, D) QRT-PCR of 5′Dβ2 mRNA in SCID-Tac transfectants. Bars represent means (± SD, n=3) relative to empty vector controls and normalized to β-actin.

3.2 DJβ2 recombination is unchanged in USF1−/− adult thymocytes

The role of promoter activity in directing the recombinational accessibility of downstream gene segments is well-established (8). In particular, we and others have shown that the location and activation of PDβ1 immediately upstream of Dβ1 is essential for recombination of the DJβ1 cassette, but does not influence recombination at DJβ2 (43). To
determine if activity of the two Dβ2 promoters similarly regulates DJβ2 recombinational accessibility, we assessed the relative levels of DJβ1 and DJβ2 joins in thymocytes from adult C57BL/6J mice deficient for USF1 (38). Embryonic deletion of USF1 results in a mild phenotype that includes inappropriate proliferation and cell fate decisions by melanocytes in response to genotoxic stress (44). Despite the physiological induction of RAG-DSBs during thymocyte development, our initial analyses of recombination in USF1−/− thymocytes revealed no apparent differences in Dβ2-to-Jβ2 (Fig. 2) recombination relative to wildtype controls. We amplified two-fold serial dilutions of purified genomic DNA by PCR, using primers immediately upstream of Dβ2 and downstream of Jβ2.7 to detect all possible rearrangement products. An analogous PCR to detect Dβ1-to-Jβ1 joins was run as a control, as USF1 has no recognized regulatory function at the upstream DJβ1 cluster. Following electrophoresis, PCR products were hybridized with 32P-labeled oligonucleotide probes and visualized by autoradiography. As expected, there was no difference in Dβ1-to-Jβ1 rearrangement between USF1−/− and wild-type mice (Fig. 2A). Similarly, no differences in either loss of Dβ2 germline sequence or accumulation of DJβ2 joins were observed in USF1−/− mice relative to wild-type controls (Fig. 2B).
Figure 2. Loss of USF1 does not impact Dβ2Jβ2 rearrangement in unsorted adult thymocytes. (A,B) Southern analysis of Dβ1Jβ1 and Dβ2Jβ2 joins, respectively. Genomic DNA was purified from thymic tissue of 5-week-old USF1−/− or wild-type (WT) control mice, serially diluted 1:2 (100 ng, 50 ng, and 25 ng), and amplified by PCR. Dβ1Jβ1 (A) was run as a control PCR for Dβ2Jβ2 rearrangement (B). Unrearranged germline DNA is abbreviated gl in each panel.

Given the well-established role for promoter activity in directing recombinational accessibility, why was even a modest increase in DJβ2 recombination not seen in USF1−/− mice? Perhaps the loss of USF1 failed to effectively activate 5′PDβ2. Alternatively, our use of whole adult thymus, composed predominantly of more mature, clonally-expanded and selected cells, masked the impact of increased recombinational accessibility that might derive from premature activation of the upstream promoter. To address the first possibility, we analyzed Dβ2 germline transcription in USF1−/− thymocytes. Unlike rearrangement levels, USF1−/− mice showed a two- to four-fold increase in expression from 5′PDβ2 relative to wild-type controls (Fig. 3), corroborating our findings in SCID-Tac cells that expressed USF1-S262D phosphomimetic proteins (Fig. 1C). As with the USF1-S262D-expressing cell lines, increased Dβ2 transcription in USF1−/− mice was not as pronounced as in cells transfected with the dominant-negative PΔB, which may indicate the ability of USF2
homodimers to bind the repressor site in the absence of USF1. Indeed, our previous ChIP and mobility shift assays detected USF2 binding at Dβ2, albeit at significantly lower levels than USF1 (33). Such limited redundancy between USF1 and USF2, which lacks a DNA-PK phosphorylation site, might also account for the apparent lack of impact on DJβ2 recombination in USF1−/− animals.

Figure 3. Expression from 5’PDβ2 is increased in thymocytes from USF1−/− mice. 5’Dβ2 mRNA levels from murine USF1−/− thymocytes were measured by QRT-PCR. Results were calculated relative to mRNA from wild-type control thymocytes, and normalized to β-actin for loading variations. The data represent the mean ± SD of triplicate QRT-PCR reactions.

3.3 Loss of USF1 increases D-to-D rearrangement in adult thymocytes

The Dβ RSSs are arranged such that, according to the 12/23 rule, Dβ1 could join with the Dβ2 segment. D-to-D joins are not frequently detected, however, suggesting that constraints beyond 12/23 prohibit such rearrangements (16, 34-36). Experiments using extrachromosomal recombination substrates demonstrated that D-to-D restriction is not enforced in cis at the level of the 3’ Dβ1 RSS, as the Dβ1 23-RSS rearranged equivalently
with the Dβ2 12-RSS and all six Jβ1 RSSs (23). This finding suggested that a \textit{trans}-acting factor may be involved in restricting physiological D-to-D joins. Because USF1 binds precisely within the 5’ Dβ2 12-RSS, we speculated that it may interfere with RAG targeting of the Dβ2 5’ RSS at the onset of DJβ recombination. This interference would impair unfavorable D-to-D rearrangements that would limit alleles to a single attempt at assembling a functional \textit{Tcrb} gene, and limit \textit{Tcrb} diversity by deleting the Jβ1 gene segments.

To determine if USF1 limits rearrangements between the Dβ1 3’RSS and the compatible Dβ2 5’ RSS, we used primers upstream of Dβ1 and downstream of Dβ2 to amplify thymic gDNA, and analyzed amplification products for D-to-D rearrangement. \textit{USF1}⁻/− mice demonstrated a marked increase in D-to-D joins relative to wild-type controls (Fig. 4A), suggesting USF1 may, in fact, normally limit D-to-D rearrangement. We used QPCR to quantitate D-to-D rearrangement in \textit{USF1}⁻/− and wild-type mice relative to \textit{Rag1}⁻/− P5424 control cells. Although V(DD)Jβ joins are exceedingly rare, we found that thymocytes from wild-type animals displayed levels of D-to-D joining 8-fold above the recombinase-deficient control (Fig. 4B). Indeed, equivalent levels of D-to-D joining were observed in thymocytes from a variety of inbred and wild-derived mouse lines (45) (Fig. 5), suggesting that while the Dβ1 3’ RSS may not join with the Dβ2 5’ RSS to the same extent that it joins with Jβ RSSs, such rearrangements are not entirely prohibited (15-17). Strikingly, \textit{USF1}⁻/− thymocytes showed a 24- to 31-fold increase in inappropriate D-to-D recombination relative to the P5424 control, approximately 3- to 4-fold above levels observed in wild-type thymocytes. As such, these findings suggest higher levels of D-to-D rearrangement are normally prevented by USF1 binding to the Dβ2 5’ RSS.
Figure 4. Loss of USF1 increases inappropriate D-to-D rearrangement in unsorted adult thymocytes. (A) Southern analysis of two-fold serial dilutions of thymocyte DNA from wild-type (WT) or USF1−/− mice to detect D-to-D joins (top panels). 100 ng or 50 ng of genomic DNA was amplified by PCR, and amplification products were detected with a radiolabeled Dβ1 oligonucleotide probe. The Rag1−/−, P53−/− P5424 pro-T cell line was used as a negative control. Cλ (bottom panels) was amplified as a DNA loading control. (B) QPCR analysis of D-to-D joins using 100 ng genomic DNA from wild-type (WT) or USF1−/− thymic tissue. Results were calculated relative to gDNA from P5424 cells (negative control), and normalized to Cλ for loading variations. The data represent the mean ± SD of triplicate QPCR reactions.
Figure 5. D-to-D rearrangement is evident in murine adult thymocytes. Genomic DNA was purified from thymic tissue of adult DBA, WSB, A/J, and PWK mice, and amplified by QPCR to detect D-to-D joins. Results were calculated relative to gDNA from C57BL/6J mice, and normalized to Cλ for loading variations. The data represent the mean ± SD of triplicate QPCR reactions. Each number (1-7) corresponds to an individual animal.

Because D-to-D joins are atypical Tcrb rearrangements, we wanted to characterize the recombination products to determine if there was any impact on junctional diversity. Repair of DSBs generated during V(D)J recombination is highly inaccurate. Insertion of palindromic (P) and non-templated (N) nucleotides at coding segment junctions results from the resolution of closed hairpin intermediates and the activity of terminal deoxynucleotidyl transferase (TdT), respectively (46-48). These random modifications to the DNA sequence are advantageous for host immunity, however, as they result in a more diverse antigen receptor repertoire. To evaluate the junctional diversity within aberrant Dβ1-to-Dβ2 coding
joins, we amplified genomic DNA from wild-type and USF1−/− mice, and sequenced individual clones. There was no obvious difference between the two genotypes in terms of diversity at the D-to-D junction (Table 1). The addition of N and P nucleotides was observed in approximately half of D-to-D joins in both wild-type and USF1−/− thymocytes. Interestingly, though, we did observe large sequence deletions in the majority of individual clones. In particular, the region encompassing the Dβ2 segment, 3’ Dβ2 RSS, and adjacent downstream sequence was most frequently deleted. 63% (10/16) of wild-type and 60% (9/15) of USF1−/− clones contained a complete deletion of the Dβ2 gene segment. Of those where Dβ2 was deleted, 38% (6/16) of wild-type and 27% (4/15) of USF1−/− clones contained deletions extending from Dβ2 to a site roughly 70 bp downstream of the 3’ Dβ2 RSS nonamer. By contrast, analyses of both productive and nonproductive rearrangements in human T cells showed that the average length of nucleotide excision at Jβ segments was less than 5 bases (49). Our finding of much more extensive deletion suggests that either D-to-D joins are slow to resolve and, consequently, more vulnerable to end processing, or that many of these aberrant joins are between Dβ1 and a cryptic RSS downstream of Dβ2.
Table 1. Sequence analysis of D-to-D joins in wild-type and USF1\(^+\) murine thymocytes

<table>
<thead>
<tr>
<th>Sample</th>
<th>5'D(\beta)1 RSS</th>
<th>D(\beta)1 coding</th>
<th>P</th>
<th>N</th>
<th>P</th>
<th>D(\beta)2 coding</th>
<th>3'D(\beta)2 RSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>germline</td>
<td>CTTTTTTGTAATAAGCTGTAACATTGTG</td>
<td>GGGACAGGGGGCC</td>
<td></td>
<td></td>
<td></td>
<td>GGGACTGGGGGGGC</td>
<td>CACAATGATTCAACTGGAAGAGGTGCTTTTACAAGAC</td>
</tr>
<tr>
<td>WT 1</td>
<td>CTTTTTTGTAATAAGCTGTAACATTGTG</td>
<td>GGGACAGGGGG*</td>
<td></td>
<td></td>
<td></td>
<td>*********GGGC</td>
<td>CACAATGATTCAACTGGAAGAGGTGCTTTTACAAGAC</td>
</tr>
<tr>
<td>WT 2</td>
<td>CTTTTTTGTAATAAGCTGTAACATTGTG</td>
<td>GGGACAGGGG**</td>
<td>A</td>
<td></td>
<td></td>
<td>*********</td>
<td>**************</td>
</tr>
<tr>
<td>WT 3</td>
<td>CTTTTTTGTAATAAGCTGTAACATTGT</td>
<td>************</td>
<td></td>
<td></td>
<td></td>
<td>**********</td>
<td>CACAATGATTCAACTGGAAGAGGTGCTTTTACAAGAC</td>
</tr>
<tr>
<td>WT 4</td>
<td>CTTTTTTGTAATAAGCTGTAACATTGT</td>
<td>GGGACAG***</td>
<td></td>
<td></td>
<td></td>
<td>GGGACTGGGGGGGC</td>
<td>CACAATGATTCAACTGGAAGAGGTGCTTTTACAAGAC</td>
</tr>
<tr>
<td>WT 5</td>
<td>CTTTTTTGTAATAAGCTGTAACATTGT*</td>
<td>************</td>
<td></td>
<td></td>
<td></td>
<td>****CTGGGGGGGC</td>
<td>CACAATGATTCAACTGGAAGAGGTGCTTTTACAAGAC</td>
</tr>
<tr>
<td>WT 6</td>
<td>CTTTTTTGTAATAAGCTGTAACATTGT</td>
<td>G***********</td>
<td>T</td>
<td></td>
<td></td>
<td>**********</td>
<td>CACAATGATTCAACTGGAAGAGGTGCTTTTACAAGAC</td>
</tr>
<tr>
<td>WT 7</td>
<td>CTTTTTTGTAATAAGCTGTAACATTGTG</td>
<td>GGGACAGGGG**</td>
<td>A</td>
<td></td>
<td></td>
<td>*********</td>
<td>**************</td>
</tr>
<tr>
<td>WT 8</td>
<td>CTTTTTTGTAATAAGCTGTAACATTGTG</td>
<td>GGGACAGGGG**</td>
<td>AGA</td>
<td></td>
<td></td>
<td>************</td>
<td>CACAATGATTCAACTGGAAGAGGTGCTTTTACAAGAC</td>
</tr>
<tr>
<td>WT 9</td>
<td>CTTTTTTGTAATAAGCTGTAACATTGTG</td>
<td>GGGACAGGGG**</td>
<td></td>
<td></td>
<td></td>
<td>**********</td>
<td>CACAATGATTCAACTGGAAGAGGTGCTTTTACAAGAC</td>
</tr>
<tr>
<td>WT 10</td>
<td>CTTTTTTGTAATAAGCTGTAACATTGT</td>
<td>G************</td>
<td>T</td>
<td></td>
<td></td>
<td>**********</td>
<td>CACAATGATTCAACTGGAAGAGGTGCTTTTACAAGAC</td>
</tr>
<tr>
<td>WT 11</td>
<td>CTTTTTTGTAATAAGCTGTAACATTGTG</td>
<td>GGGACAGGGG**</td>
<td>A</td>
<td></td>
<td></td>
<td>**********</td>
<td>CACAATGATTCAACTGGAAGAGGTGCTTTTACAAGAC</td>
</tr>
<tr>
<td>WT 12</td>
<td>CTTTTTTGTAATAAGCTGTAACATTGTG</td>
<td>************</td>
<td></td>
<td></td>
<td></td>
<td>****CTGGGGGGGC</td>
<td>CACAATGATTCAACTGGAAGAGGTGCTTTTACAAGAC</td>
</tr>
<tr>
<td>WT 13</td>
<td>CTTTTTTGTAATAAGCTGTAACATTGT</td>
<td>GGGACAGGGG**</td>
<td>A</td>
<td></td>
<td></td>
<td>**********</td>
<td>CACAATGATTCAACTGGAAGAGGTGCTTTTACAAGAC</td>
</tr>
<tr>
<td>WT 14</td>
<td>CTTTTTTGTAATAAGCTGTAACATTGTG</td>
<td>GGGACAGGGG**</td>
<td>A</td>
<td></td>
<td></td>
<td>**********</td>
<td>CACAATGATTCAACTGGAAGAGGTGCTTTTACAAGAC</td>
</tr>
<tr>
<td>WT 15</td>
<td>CTTTTTTGTAATAAGCTGTAACATTGTG</td>
<td>************</td>
<td>AT</td>
<td></td>
<td></td>
<td>**********</td>
<td>CACAATGATTCAACTGGAAGAGGTGCTTTTACAAGAC</td>
</tr>
<tr>
<td>WT 16</td>
<td>CTTTTTTGTAATAAGCTGTAACATTGTG</td>
<td>G************</td>
<td></td>
<td></td>
<td></td>
<td>**********</td>
<td>CACAATGATTCAACTGGAAGAGGTGCTTTTACAAGAC</td>
</tr>
</tbody>
</table>

*The sequence 3’ of D\(\beta\)2 is shown below. The D\(\beta\)2 3’RSS nonamer (italics), junction sites (º), and putative c-Fos binding site (underlined) are indicated:

```plaintext
ACAAAAAGCTCTTACCCAAAAAACACCAACTCTCAGCCCACAAAAATGCATTATACATGTGAGGAGGATCTATGTGAGTGACTACAAGG
```
Table 1. (continued)

<table>
<thead>
<tr>
<th>Sample</th>
<th>5'Dβ1 RSS</th>
<th>DJβ1 coding</th>
<th>P</th>
<th>N</th>
<th>DJβ2 coding</th>
<th>3'Dβ2 RSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>germline</td>
<td>CTTTTTTGTATAGGCTGTAACATTTG</td>
<td>GGGACAGGGGGGC</td>
<td></td>
<td></td>
<td>GGGACAGGGGGGC</td>
<td>CACAATGATTCAGACTGGAGATGGCTTTTACAAAAAGC</td>
</tr>
<tr>
<td>KO 1</td>
<td>CTTTTTTGTATAGGCTGTAACATTTG</td>
<td>GGGACAG****</td>
<td></td>
<td></td>
<td>**************</td>
<td>***************************************</td>
</tr>
<tr>
<td>KO 2</td>
<td>CTTTTTTGTATAGGCTGTAACATTTG</td>
<td>***********</td>
<td></td>
<td></td>
<td>**************</td>
<td>CACAATGATTCAGACTGGAGATGGCTTTTACAAAAAGC</td>
</tr>
<tr>
<td>KO 3</td>
<td>CTTTTTTGTATAGGCTGTAACATTTG</td>
<td>GGGACAGGGGG*</td>
<td></td>
<td></td>
<td>**********</td>
<td>CACAATGATTCAGACTGGAGATGGCTTTTACAAAAAGC</td>
</tr>
<tr>
<td>KO 4</td>
<td>CTTTTTTGTATAGGCTGTAACATTTG</td>
<td>GGGACAGGGG**</td>
<td></td>
<td></td>
<td>**********</td>
<td>CACAATGATTCAGACTGGAGATGGCTTTTACAAAAAGC</td>
</tr>
<tr>
<td>KO 5</td>
<td>CTTTTTTGTATAGGCTGTAACATTTG</td>
<td>GGGACAGGGG***</td>
<td></td>
<td>AG</td>
<td>*****</td>
<td>CACAATGATTCAGACTGGAGATGGCTTTTACAAAAAGC</td>
</tr>
<tr>
<td>KO 6</td>
<td>CTTTTTTGTATAGGCTGTAACATTTG</td>
<td>GGGACAG****</td>
<td></td>
<td></td>
<td>**************</td>
<td>***************************************</td>
</tr>
<tr>
<td>KO 7</td>
<td>CTTTTTTGTATAGGCTGTAACATTTG</td>
<td>***********</td>
<td></td>
<td>C</td>
<td>**************GC</td>
<td>CACAATGATTCAGACTGGAGATGGCTTTTACAAAAAGC</td>
</tr>
<tr>
<td>KO 8</td>
<td>CTTTTTTGTATAGGCTGTAACATTTG</td>
<td>GGGACA*****</td>
<td></td>
<td>A</td>
<td></td>
<td>***************************************</td>
</tr>
<tr>
<td>KO 9</td>
<td>CTTTTTTGTATAGGCTGTAACATTTG</td>
<td>GGGACAG*****</td>
<td></td>
<td>A</td>
<td></td>
<td>***************************************</td>
</tr>
<tr>
<td>KO 10</td>
<td>CTTTTTTGTATAGGCTGTAACATTTG</td>
<td>G**********</td>
<td></td>
<td>CGG</td>
<td></td>
<td>CACAATGATTCAGACTGGAGATGGCTTTTACAAAAAGC</td>
</tr>
<tr>
<td>KO 11</td>
<td>CTTTTTTGTATAGGCTGTAACATTTG</td>
<td>GGGACAGGGG**</td>
<td></td>
<td>AG</td>
<td>*****</td>
<td>CACAATGATTCAGACTGGAGATGGCTTTTACAAAAAGC</td>
</tr>
<tr>
<td>KO 12</td>
<td>CTTTTTTGTATAGGCTGTAACATTTG</td>
<td>***********</td>
<td></td>
<td>GA</td>
<td></td>
<td>CACAATGATTCAGACTGGAGATGGCTTTTACAAAAAGC</td>
</tr>
<tr>
<td>KO 13</td>
<td>CTTTTTTGTATAGGCTGTAACATTTG</td>
<td>***********</td>
<td></td>
<td>ATAA</td>
<td></td>
<td>CACAATGATTCAGACTGGAGATGGCTTTTACAAAAAGC</td>
</tr>
<tr>
<td>KO 14</td>
<td>CTTTTTTGTATAGGCTGTAACATTTG</td>
<td>***********</td>
<td></td>
<td>ACA</td>
<td>**********C</td>
<td>CACAATGATTCAGACTGGAGATGGCTTTTACAAAAAGC</td>
</tr>
<tr>
<td>KO 15</td>
<td>CTTTTTTGTATAGGCTGTAACATTTG</td>
<td>***********</td>
<td></td>
<td></td>
<td></td>
<td>CACAATGATTCAGACTGGAGATGGCTTTTACAAAAAGC</td>
</tr>
</tbody>
</table>

*The sequence 3’ of DJβ2 is shown below. The DJβ2 3’RSS nonamer (italics), junction sites (*), and putative c-Fos binding site (underlined) are indicated:

```
ACAAAAAGCTCTCCCAAAAAACACCAACTCTCGCCCCAATGTGCACTTACATGTGAGAGAGCTACAGACCACCTTCTG
```
Recombination information content (RIC) analysis (39) of sequences between Dβ2 and Jβ2.1 failed to identify potential RSS sites (Table 2). A single A-rich, nonamer-like sequence was positioned within 20 bp of the 3’ rearrangement sequence of these cryptic joins, but was not associated with any sequence that matched known RSS heptamers. By contrast, the 3’ rearrangement sequence (see footnote in Table 1) contains a near-perfect match to the canonical binding site for c-Fos (TGACTCA) (50). Given the ability of c-Fos to recruit RAG to the Dβ1 and Dβ2 RSSs (27), it is possible that the sterile Dβ1-3’Dβ2 joins we detected in wildtype and USF1−/− thymocytes stem from c-Fos:RAG complexes at Dβ1 that capture a 3’ Dβ2 c-Fos target rather than the Dβ2 5’RSS.

Table 2. RIC analysis of putative RSSs

<table>
<thead>
<tr>
<th>start-end</th>
<th>12-RSS sequence</th>
<th>strand</th>
<th>RIC score</th>
<th>RIC pass/fail</th>
</tr>
</thead>
<tbody>
<tr>
<td>7 - 34</td>
<td>CAAAAAAACACCACTCAGCCAAAAAA</td>
<td>+</td>
<td>-49.678</td>
<td>FAIL</td>
</tr>
<tr>
<td>10 - 37</td>
<td>CATTGGCTGAGGAGTTGGTTTTT</td>
<td>-</td>
<td>-57.416</td>
<td>FAIL</td>
</tr>
<tr>
<td>15 - 42</td>
<td>CAACACTCTCAGCCAAAAATGCATTTT</td>
<td>+</td>
<td>-59.268</td>
<td>FAIL</td>
</tr>
<tr>
<td>18 - 45</td>
<td>CAACTTCTAGCCAAAAATGCATTTTA</td>
<td>+</td>
<td>-57.778</td>
<td>FAIL</td>
</tr>
<tr>
<td>21 - 48</td>
<td>CATGTTAAATGCACTCTTTTGCTGAGAG</td>
<td>-</td>
<td>-65.975</td>
<td>FAIL</td>
</tr>
<tr>
<td>23 - 50</td>
<td>CAACATGGTTAATGCACTCTTTTGCTGAG</td>
<td>-</td>
<td>-69.954</td>
<td>FAIL</td>
</tr>
<tr>
<td>25 - 52</td>
<td>CAACGGTTAAATGCACTCTTTTGCTGAG</td>
<td>+</td>
<td>-54.747</td>
<td>FAIL</td>
</tr>
<tr>
<td>29 - 56</td>
<td>CAACAAAATGCACTCTTTTGCACTTTTACATGAGGG</td>
<td>+</td>
<td>-59.098</td>
<td>FAIL</td>
</tr>
<tr>
<td>37 - 64</td>
<td>CATGACTCCTCCTACATGTTAAAATGCA</td>
<td>-</td>
<td>-54.345</td>
<td>FAIL</td>
</tr>
<tr>
<td>38 - 65</td>
<td>CATTGAACATTGAGGAGGGAGTCTATG</td>
<td>+</td>
<td>-69.442</td>
<td>FAIL</td>
</tr>
<tr>
<td>39 - 66</td>
<td>CAACATAGACTCCTCCTACATGTTAAAATGCA</td>
<td>-</td>
<td>-47.356</td>
<td>FAIL</td>
</tr>
<tr>
<td>43 - 70</td>
<td>CACTCAACTAGACTCCTCCTCCTACATGTTTT</td>
<td>-</td>
<td>-53.854</td>
<td>FAIL</td>
</tr>
<tr>
<td>45 - 72</td>
<td>CATGTTGAGGAGGAGCTATGTTAGTGAGGGAGG</td>
<td>+</td>
<td>-58.959</td>
<td>FAIL</td>
</tr>
<tr>
<td>70 - 97</td>
<td>CATAGATGTTAATAGACCTTGTGAGTCC</td>
<td>-</td>
<td>-57.613</td>
<td>FAIL</td>
</tr>
<tr>
<td>75 - 102</td>
<td>CACCAAGCTTTAATACCATCTATGCACTCA</td>
<td>+</td>
<td>-48.869</td>
<td>FAIL</td>
</tr>
<tr>
<td>77 - 104</td>
<td>CAAGGGTTTAAATACCATCTATTGGGCACATCAAC</td>
<td>+</td>
<td>-44.864</td>
<td>FAIL</td>
</tr>
<tr>
<td>80 - 107</td>
<td>CAAGGTTGAGCAGTAGGCTTTATAGGACC</td>
<td>-</td>
<td>-39.426</td>
<td>FAIL</td>
</tr>
<tr>
<td>90 - 117</td>
<td>CATCTATGCATCAACTTGGCCTAGCAAG</td>
<td>+</td>
<td>-59.045</td>
<td>FAIL</td>
</tr>
<tr>
<td>98 - 125</td>
<td>CAATCAACTTGGCCTAGCAAGTTCTCCAC</td>
<td>+</td>
<td>-76.284</td>
<td>FAIL</td>
</tr>
<tr>
<td>101 - 128</td>
<td>CAACCTTGGGAGCAGCTTTGCACTTTTCCACAGGAG</td>
<td>+</td>
<td>-55.412</td>
<td>FAIL</td>
</tr>
<tr>
<td>103 - 130</td>
<td>CACTCGTTGGAAACTTGGTCTAGGGCAGGCACTT</td>
<td>-</td>
<td>-64.722</td>
<td>FAIL</td>
</tr>
</tbody>
</table>

*Start-end numbering is relative to the last base of the Dβ2 3’ RSS nonamer.*
4 Discussion

In this study, we describe an expanded role for USF1 in modulating Dβ2 transcription and rearrangement. Our findings offer new insight into Tcrb regulation by defining the mechanism of USF1-mediated repression of the 5′PDβ2 germline promoter. Previously, we showed that USF1 binding at an E box within the 5′ Dβ2 12-RSS repressed activation of 5′PDβ2. As such, germline transcription is driven by the downstream 3′PDβ2 promoter, situated between Dβ2 and Jβ2.1, a position predicted to limit DJβ2 recombination. Genotoxin-induced DNA DSBs leads to the loss of USF1 from its binding site, and consequential 5′PDβ2 activation. Moreover, our data suggested that repression was relieved in a DNA-PK-dependent manner (33). Because DNA-PK is recruited to RAG-DSBs as part of the NHEJ complex, its proximity to the 5′ Dβ2 RSS might facilitate interaction with USF1, similar to that seen in regulation of the lipogenic FAS promoter, where USF1 is phosphorylated by DNA-PK at S262 (40).

Our in vitro overexpression assays in the DNAPKcs−/− pro-T cell line, SCID-Tac, showed that introduction of a phosphomimetic USF1 mutant (S262D) led to a ~3-fold increase in 5′PDβ2 expression, that was not seen following introduction of either wildtype USF1 or a nonphosphorylatable USF1 mutant (S262A). Transfection of SCID-Tac with a dominant-negative USF2 that blocks all USF binding resulted in a significantly more pronounced increase in 5′PDβ2 transcription. Together, our results suggest that DNA-PK phosphorylates USF1 at the S262 residue, presumably reducing its stability at Dβ2 and resulting in activation of the upstream promoter. Such posttranslational modification of USF1 can indeed alter USF1’s DNA-binding activity at E box target sites. For example, PI3K-mediated phosphorylation abolishes USF1 binding at the APOA5 promoter (28, 41). Whether
phosphorylation at S262 residue alters the three-dimensional structure of USF1, thereby reducing its DNA-binding ability, has yet to be determined (51). Displacement of phospho-USF1 from the Dβ2 5’ RSS could also be influenced by the noncanonical, heptameric nature of the 5’ Dβ2 E box (CAGCATG), similar to the differential binding of USF1 to a heptameric E box within the lama3 promoter (52).

The loss of USF1 from the 5’ Dβ2 RSS coincides with a gain in transcription from the 5’PDβ2 promoter (33, Fig. 1D and Fig. 3). Activation of germline promoters is associated with increased accessibility of recombinase to individual gene segments, and deletion of promoter elements has been shown to impair local rearrangement by limiting the chromatin accessibility of embedded RSS elements (8). D-to-J rearrangements involving Dβ1 appear to initiate earlier in thymocyte development or are otherwise favored relative to Dβ2 rearrangement (53, 54). Given that germline activation of 3’PDβ2 results in transcription initiation downstream of the Dβ2 3’RSS, we predicted that germline activation of 5’PDβ2, which does transcribe through the Dβ2 RSSs, would increase accessibility at the Dβ2 gene segment and, in turn, lead to higher levels of DJβ2 rearrangement. However, we saw no such changes in the levels of germline Dβ2 or rearranged DJβ2 DNA in adult thymocytes from USF1−/− mice, despite a 3-fold induction in 5’PDβ2 promoter activity relative to wildtype thymocytes. Perhaps our analysis of adult thymocytes, which are roughly 95% DP (CD4+CD8+) and SP (CD4+ or CD8+) cells that have passed through multiple rounds of clonal selection and expansion, masked a potentially subtle change in Dβ2 recombination. Alternatively, binding at the Dβ2 E box and promoter repression in our USF1−/− thymocytes may have been at least partially restored by USF2, which we have previously shown can bind at Dβ2 (33). Indeed, USF2 levels are elevated in USF1-null mice, whereas USF1/USF2
double-knockout mutations are embryonic lethal (55). Consistent with the possibility that USF2 may at least partially replace USF1 to repress 5’PDβ2 activity, we saw much more robust promoter activation in cultured cells that expressed a dominant-negative USF2 (defective for DNA binding) than we did in cells expressing a USF1 phosphomimetic mutant. To test the potentiality that USF2 may temper the sensitivity of thymocytes to USF1 depletion, and to exclude potential complications from clonal selection during thymocyte development, studies are underway to assess USF1/USF2 binding, promoter activity, and Dβ2 rearrangement levels in sorted USF1/− DN thymocytes.

While we did not detect an increase in Dβ2-to-Jβ2 joining in USF1/− thymocytes, we did observe a considerable increase in noncanonical rearrangement of Dβ1-to-Dβ2 in the absence of USF1. Such D-to-D joins are not uncommon in the Tcrd locus, and give rise to functional V(DD)Jδ-containing receptors (56, 57). However, V(DD)Jβ joins, whether functional or non-functional, are rare within Tcrb (16, 34-36). Given the extensive homology between Dβ1 and Dβ2, scoring unambiguous V(DD)Jβ rearrangements is extremely difficult, and the frequency of such joins may be somewhat higher than predicted by previous studies (58). In fact, we could detect rearrangements between the germline Dβ1 and Dβ2 segments in five different laboratory mouse strains, though such joins are far less frequent than Dβ-to-Jβ rearrangements (59, 60). Therefore, the finding that D-to-D rearrangement was increased in thymocytes of USF1/− mice suggests that USF1 contributes to B12/23 at the level of D-to-D restriction. Previous studies have indicated that there is no sequence-dependent constraint that prevents aberrant D-to-D joining within the Tcrb locus and, moreover, that the 3’ Dβ1 23-RSS rearranges equivalently with the 5’ Dβ2 12-RSS as well as with all 6 Jβ1 RSSs in plasmid substrates (23). Whether USF1 binding in the Dβ2 5’ RSS spacer may interfere with
RAG access to the RSS, or whether repression of 5’PDβ2 limits recombinational accessibility of the Dβ2 5’ RSS remains to be resolved. The location of the USF1 binding site also raises the possibility that occupancy prevents premature Vβ-to-Dβ2 joining. Interestingly, in Fos⁻/⁻ mice, joins were detected between Vβ segments and germline Dβ1, violating ordered Tcrb assembly by impairing efficient RAG recruitment to the Dβ1 3’ RSS (27). However, inappropriate Vβ joins with Dβ2 (which also binds c-Fos at its 3’ RSS) were not detected in Fos⁻/⁻ mice, suggesting additional restrictions prevent direct recombination between Vβ and unrearranged Dβ2 (27).

In conclusion, our findings suggest a feedback mechanism for regulating initial Tcrb assembly beyond a strictly stochastic process during DN thymocyte development. Our model suggests that DNA-PK recruited to DSBs generated during initial rounds of Dβ-to-Jβ rearrangement phosphorylates USF1 at S262, destabilizing the transcription factor at the 5’ Dβ2 RSS. Displacement of USF1 relieves repression of the 5’PDβ2 promoter, though it remains unclear whether this then increases Dβ2 recombinational accessibility for successive rounds of rearrangements. Additionally, our in vivo data suggest a novel role for USF1 in imposing B12/23 within Tcrb at the level of D-to-D restriction. Future studies will be necessary to determine the impact of USF1 occupancy on RAG binding to the 5’ Dβ2 RSS. USF1 may enforce B12/23 by sterically hindering RAG access to the 5’ Dβ2 12-RSSs, or it may limit recombinational accessibility of the Dβ2 RSSs through its repression of 5’PDβ2. Indeed, these two potential mechanisms, which may also contribute to B12/23 restriction of direct Vβ-to-Dβ2 recombination, are not mutually exclusive. Overall, the identification of USF1 as a trans-acting regulator of Dβ2 transcription and B12/23 provides further detail into the mechanisms that impose ordered assembly and maximize αβ T cell receptor diversity.
5 References


49. Manfras, B.J., D. Terjung, and B.O. Boehm, Non-productive human TCR beta chain genes represent V-D-J diversity before selection upon function: insight into biased


## Supplemental Data

**Table S1.** Oligonucleotides for PCR & Southern analysis

<table>
<thead>
<tr>
<th>Amplicon</th>
<th>Strand</th>
<th>Sequence</th>
<th>Tm</th>
</tr>
</thead>
<tbody>
<tr>
<td>USF1cDNA</td>
<td>S</td>
<td>GGA ATT CAC AAA GTC CTC CGA AAA AAG ACC</td>
<td>66°C</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>GGA TAT CCT GAG TCC CCA GAG TTA GTT CG</td>
<td></td>
</tr>
<tr>
<td>USF1S262D</td>
<td>S</td>
<td>AAC CAC CGG CTG <strong>GAT</strong> GAA GAG CTG CAG</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>CTG CAG CTC TTC ACT CAG CCG GTG GTT</td>
<td></td>
</tr>
<tr>
<td>USF1S262A</td>
<td>S</td>
<td>AAC CAC CGG CTG <strong>GCT</strong> GAA GAG CTG CAG</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>CTG CAG CTC TTC AGC CAG CCG GTG GTT</td>
<td></td>
</tr>
<tr>
<td>β-actin</td>
<td>S</td>
<td>AGA GCT ATG AGC TGC CTG ACG GCC</td>
<td>63°C</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>AGT AAT CTC CTT CTG CAT CCT GTC</td>
<td></td>
</tr>
<tr>
<td>Dβ2</td>
<td>S</td>
<td>CCC TCT CAG TCA GAC AAA CCT CTC</td>
<td>63°C</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>CTT CCA GTT GAA TCA TTG TGG CCC C</td>
<td></td>
</tr>
<tr>
<td>USF1</td>
<td>S</td>
<td>ACA AAG TCC TCC GAA AAA AGA CCG</td>
<td>63°C</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>GGG TCC TCT CCA GTA GCC ACT GCG C</td>
<td></td>
</tr>
<tr>
<td>D1J1</td>
<td>A</td>
<td>CCA AGA CCA TGG TCA TCC AAC</td>
<td>63°C</td>
</tr>
<tr>
<td>Probe</td>
<td>S</td>
<td>AAG CTG TAA CAT TGT GGG GAC AGG</td>
<td></td>
</tr>
<tr>
<td>D2J2</td>
<td>A</td>
<td>TGA GAG CTG TCT CCT ACT ATC GAT T</td>
<td>63°C</td>
</tr>
<tr>
<td>Probe</td>
<td>S</td>
<td>AAT GTT ACA TCG TGA TAC AAA AAA GTT</td>
<td></td>
</tr>
<tr>
<td>D-D REC</td>
<td>A</td>
<td>GGG GTA GAC CTA TGG GAG GG</td>
<td></td>
</tr>
<tr>
<td>Probe</td>
<td>S</td>
<td>AAG CTG TAA CAT TGT GGG GAC AGG</td>
<td>63°C</td>
</tr>
<tr>
<td>C lambda</td>
<td>S/Probe</td>
<td>CTT TTT CCA CCT TCC TCT GAA GAG</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>CAC GGG ACA AAC TCT TCT CCA C</td>
<td>64°C</td>
</tr>
</tbody>
</table>
CHAPTER 4

Summary
Summary

In this study, I have described a role for the stress-response transcription factor, USF1, in regulating V(D)J recombination at the level of Dβ2 usage through a novel feedback mechanism. Previous research from our laboratory identified two promoter elements flanking Dβ2. 3’PDβ2 is a germline promoter that is active in early DN thymocytes to drive Dβ2-to-Jβ2 joining, while 5’PDβ2, which is initially repressed, becomes active only after rearrangement deletes 3’PDβ2. This thesis research sought to characterize the mechanism of repression of the upstream 5’PDβ2 promoter, and to determine the impact of repression on recombination.

We first focused on understanding how 5’PDβ2 repression was imposed. To pinpoint the location of the repressor, we performed serial 3’ deletions of sequence spanning 230 bp downstream of Dβ2 to just upstream of the 5’PDβ2 promoter. Interestingly, results from reporter assays suggested the cis-acting repressor element was not situated downstream of Dβ2 and, therefore, was not simply deleted during rearrangement. Rather, an increase in promoter activity was observed in our reporter studies only when the region spanning the 5’ Dβ2 12-RSS nonamer and spacer was removed. Sequence analysis of the RSS revealed a putative E box situated within the 12-bp spacer, and successive mutational analyses showed that destruction of this E box element enhanced promoter activity. Indeed, EMSA verified that the E box was functional. After screening a panel of known E box binding proteins, we identified the ubiquitous stress-response transcription factor, USF1, as binding to the repressor element in vitro. We then used chromatin immunoprecipitation (ChIP) to confirm that USF1 effectively bound the repressor element in vivo.
Utilizing ChIP and QRT-PCR assays to determine protein binding and transcription levels, respectively, we found that USF1 bound the repressor in recombinase-deficient thymocytes and, as a result, limited transcription from 5’PDβ2. Moreover, genotoxin-induced DNA double strand breaks (DSBs) that mimicked RAG-DSBs led to a loss of USF1 from the repressor element, and a consequential gain in transcription from 5’PDβ2. Intriguingly, thymocytes from SCID mice showed a strong enrichment for USF1 at Dβ2 and almost no activity from the 5’PDβ2 promoter (even though these cells accumulate large numbers of unresolved RAG-DSBs, owing to the lack of a functional DNA-PKcs gene). This finding suggested that DNA-PK could be involved in regulating USF1-mediated repression of 5’PDβ2. Indeed, we found that 5’PDβ2 repression could be relieved in a DNA-PK-dependent manner. Cells treated with etoposide or ionizing radiation alone, to induce DSBs, gained activity from 5’PDβ2 while USF1 enrichment was lost. Thymocytes pretreated with the DNA-PK inhibitor Nu 7026 prior to exposure to genotoxic stress, on the other hand, maintained a high level of USF1 enrichment and corresponding absence of Dβ2 transcription. Together, the results suggest a model where RAG-mediated DSBs activate the repair protein DNA-PK which, accordingly, acts on USF1, leading to a loss of transcription factor binding and subsequent derepression of the 5’PDβ2 promoter (Fig. 1).
Figure 1. Mechanism of Dβ2 regulation. In early DN thymocytes, USF1 binds an E box within the 5’ Dβ2 12-RSS spacer. USF1 binding represses transcription of the 5’PDβ2 promoter, forcing transcription to target the downstream 3’PDβ2 promoter. Upon induction of RAG-mediated DSBs, DNA-PK is activated and phosphorylates USF1 at the S262 residue. USF1 is lost from the E box repressor and, consequently, 5’PDβ2 gains transcriptional activity to drive Vβ-to-DJβ2 rearrangement in DN3 thymocytes.

We next sought to evaluate the signaling pathway for DNA-PK-mediated phosphorylation of USF1 and, ultimately, to determine if USF-mediated repression at Dβ2 had a functional consequence on rearrangement. Because DNA-PK had previously been shown to phosphorylate USF1 at the serine 262 (S262) residue, we generated a series of mutant USF1 expression vectors containing single amino acid substitutions of S262. We then assessed the impact of the phosphorylation mutants, as well as a dominant-negative DNA binding-mutant of USF, on Dβ2 transcription. Consistent with previous observations of USF1 signaling, our phosphomimetic mutant (S262D) led to an increase in Dβ2 transcription.
when transfected into the DNA-PK-deficient cell line, SCID-Tac, whereas a non-phosphorylatable mutant (S262A) had no effect on promoter activity. Cells transfected with a dominant-negative USF2 expression vector, which readily dimerizes with endogenous USF1 and USF2 but lacks a DNA binding domain, showed an even greater increase in transcription at Dβ2 compared to cells transfected with S262D. Together the data suggest that DNA-PK acts specifically on USF1 by phosphorylating S262. Phosphorylation likely leads to a loss of USF1 binding at the repressor and, consequently, an increase in Dβ2 transcriptional read-through.

Lastly, to determine if USF1 binding within the 5’ Dβ2 12-RSS impacts recombination, we analyzed thymic tissue from adult USF1−/− mice. We found that rearrangement at the downstream DJβ2 cassette was equivalent in knockout mice and wild-type animals and, as expected, levels of DJβ1 joins were also unchanged. Rearrangements involving DJβ2 were likely not increased at a discernible level in adult USF1−/− mice due to clonal expansion, where individual cells that had completed β rearrangement and selection expanded their specific populations. Fascinatingly, though, D-to-D recombination was measurably increased in USF1−/− mice compared to wild-type animals. This in vivo finding not only demonstrates a functional role for USF1 in regulating recombination events involving Dβ2, but moreover suggests that USF1 enforces the beyond 12/23 rule (B12/23) at the level of D-to-D restriction (Fig. 2). Because Dβ1 and Dβ2 have compatible RSSs, the two gene segments should effectively rearrange according to the 12/23 principle. However, such rearrangements are rare and are prevented by the so-called beyond 12/23 rule—a series of poorly-defined genetic and epigenetic structures that constrain TCRβ assembly to canonical V-D-J joints. Previous studies using non-chromatinized recombination substrates
demonstrated that D-to-D restriction was not enforced in cis at the level of the 3’ Dβ1-23-RSS, implying that an independent trans-acting factor could be involved. Our rearrangement analyses in USF1−/− mice support this notion, and implicate USF1 in maintaining the constraint on D-to-D joining.

**Figure 2.** A role for USF1 in B12/23 regulation. A) The Dβ1 and Dβ2 gene segments are flanked by a 3’ 23-RSS and 5’ 12-RSS, respectively. According to the 12/23 rule, the RAG complex should be able to induce cleavage and hairpin formation at the compatible RSSs, resulting in a non-canonical D-to-D joint within Tcrb. B) USF1 binding within the 5’ Dβ2 12-RSS impedes local RAG access, thereby preventing inappropriate D-to-D rearrangement.

The work described herein presents one of the first detailed analyses of the impact of a specific trans-acting factor in regulating accessibility and rearrangement at Dβ2. While the mechanisms governing V(D)J recombination within the DJβ1 region have been carefully dissected, less is known about control at the downstream cassette. It has been well-established that joins involving DJβ1 are more likely to accumulate than those utilizing DJβ2.
coding segments. The epigenetic mechanisms that restrict initial DJβ2 recombinational accessibility, however, had not been previously explored. Our analyses suggest a model where the transcription factor USF1 binds to a repressor element at Dβ2, limiting its rearrangement potential and forcing initial V-to-DJ recombination to target the upstream cassette.

The results also suggest a feedback mechanism at Dβ2, where initial rearrangements activate local DNA repair factors that, in turn, regulate subsequent rearrangements (Fig. 1). Importantly, this study is the first to provide evidence for a factor that directly regulates B12/23 at the level of D-to-D restriction. Because USF1 binds precisely within the 5' Dβ2 12-RSS, it could sterically hinder RAG access and prevent paired complex formation between compatible Dβ RSSs. Collectively, the findings are significant as they provide a reappraisal of accessibility control within the Tcrb locus, and suggest the presence of additional mechanisms that ensure rearrangement events maximize pro-T cell survival and receptor diversity.

Future studies will be necessary to determine the exact relationship between USF1 and the RAG complex at the 5’ Dβ2 RSS. It is conceivable that USF1 binding within the RSS spacer sequence directly occludes the RAG binding sites within the nonamer and heptamer. To test this possibility, chromatin immunoprecipitations for RAG-1/2 could be performed. One limitation of this assay, however, is the proximity of the 5’ and 3’ Dβ2 RSSs. Because the two signal sequences are separated by a mere 14 nucleotide bases, it would make RAG binding impossible to resolve in vivo without separately destroying each RSS via genome editing (e.g. CRISPR/Cas9 targeting). A more straightforward approach would be to perform in vitro cleavage assays, to assess levels of RAG-mediated cleavage at
the 5’ Dβ2 12-RSS in the presence or absence of USF1. Alternatively, TCRβ transgenic substrates harboring mutations of the USF1 binding site could be stably introduced into a RAG-inducible lymphocyte cell line. Rearrangement levels could then be assessed by PCR to determine if destruction of the site, which would preclude USF1 binding and allow RAG access, increased gene segment utilization.

In addition to traditional molecular techniques, it would be interesting to perform high-resolution atomic-force microscopy (AFM) to directly visualize DNA:protein interactions at Dβ2. Analysis using AFM would be advantageous as it would provide three-dimensional, in vivo images of protein complexes bound to the 5’ Dβ2 RSS. We hypothesize that in wild-type animals, USF1 homodimers or, to a lesser extent, USF1:USF2 heterodimers would occupy the 5’ Dβ2 12-RSS, while RAG proteins would not be detected. In USF1−/− mice, conversely, we would expect to observe RAG binding in the absence of USF1 complexes. While USF1 has been shown to be the primary factor in regulating Dβ2 transcription and usage, our EMSA and ChIP analyses did detect modest USF2 binding at the repressor. To ultimately evaluate any contributions from USF2 in cooperation with USF1, we would need to generate a conditional knockout mouse, as embryonic double deletion of USF1 and USF2 is lethal. To accomplish this, USF1−/− mice would first need to be crossed with mice harboring a floxed USF2 gene. Subsequently crossing progeny with a transgenic Lck-CRE mouse would result in targeted deletion of USF2 at the onset of thymocyte development, and afford a direct assessment of the contributions from both proteins. While the double knockout could augment Dβ2 expression and rearrangement, we anticipate it would have a minimal impact, as our data point to USF1 as the primary binding factor.
In our *in vivo* experiments, we were able to show an increase in D-to-D joining in 
*USF1*−/− mice, suggesting that USF1 blocks RAG access to the 5’ Dβ2 12-RSS, thereby
enforcing B12/23 of otherwise compatible RSS sequences. However, we were unable to
detect an increase in Dβ2-to-Jβ2 joins in knockout animals. This is likely because we
analyzed the thymic compartment of adult mice, in which ~80% of the total population is
double positive (DP) cells that have completed β rearrangement. To more accurately assess
the impact of USF1 binding on Dβ2-to-Jβ2 rearrangement, we intend to analyze sorted DN
thymocytes from 5-week-old wild-type and *USF1*−/− mice. We expect that a loss of USF1
would increase joins within the DJβ2 cassette, as the upstream Dβ2 promoter would be
derepressed and augment rearrangements involving the Dβ2 gene segment. However, it is
possible that USF1 binding at the repressor contributes to Vβ-to-DJβ2 rather than Dβ2-to-Jβ2
joining. To test this possibility, we also intend to perform PCR analysis of Vβ-to-DJβ2
rearrangement levels in sorted DN thymocytes from *USF1*−/− and wild-type control mice.
Additionally, we also expect to see a more pronounced increase in D-to-D rearrangement in
sorted DN thymocytes from *USF1*−/− mice relative to wild-type controls.

Sequence analysis of the D-to-D rearrangement products we observed in *USF1*−/− and
wild-type animals unexpectedly revealed large-scale deletions of the Dβ2 gene segment, the
3’ Dβ2 23-RSS, and additional 3’ sequence extending to ~70 bp downstream of the 3’ Dβ2
RSS nonamer. We therefore considered the possibility that a cryptic RSS element could exist
within this region. While our subsequent analysis ruled out this likelihood, we did identify a
putative c-Fos binding site exactly 70 bp downstream of the 3’ Dβ2 RSS nonamer. As c-Fos
has previously been shown to facilitate RAG recruitment to both the Dβ1 and Dβ2 23-RSSs,
we wondered if this site could be functional and, consequently, if RAG could be
inappropriately recruited, leading to pairing of the Dβ1 23-RSS with sequence downstream of Dβ2, rather than the Dβ2 12-RSS. Such a recombination event would result in a sterile join, preventing Dβ-to-Jβ rearrangement and thereby limiting the opportunity for generating a functional coding join. To determine if this putative c-Fos site is indeed functional, we could perform EMSA to assess protein binding (noting that more powerful in vivo analysis by ChIP would be impractical due to the short distance between the putative site and the recognized c-Fos site within the Dβ2 23-RSS). Additionally, we could repeat our D-to-D rearrangement PCR in thymocytes from Fos−/− mice. If c-Fos encouraged RAG deposition to this site, we would expect to see a decrease in large-scale deletions 3’ of Dβ2 in Fos−/− animals relative to wild-type controls.

Finally, it is interesting to consider the broader implications of USF1 binding at Dβ2. The Tcrb RSSs are evolutionarily well-conserved between mice and humans. Moreover, the human 5’ Dβ2 12-RSS contains a non-canonical, heptameric E box that differs from the murine sequence by only two nucleotides (GATGGTG, human vs. CACGATG, mouse). Because of the conservation of the USF1 site, we anticipate that the model we have defined in the mouse is extensible to humans, and that the effects of USF1 binding on overall TCRβ repertoire would be similar. Presently, we intend to repeat our rearrangement assays in sorted DN thymocytes from USF1−/− mice, and to perform one of the aforementioned in vitro experiments to determine if USF1 sterically interferes with RAG access to the 5’ Dβ2 RSS. Together, the results should strengthen our current findings that USF1 regulates Dβ2 rearrangement and D-to-D restriction by means of a unique feedback mechanism that is mediated by the DNA repair protein DNA-PK.
APPENDIX I

CTCF is not involved with epigenetic silencing at the TEA promoter in ES cells
1 Introduction

T and B lymphocytes generate a diverse repertoire of antigen receptors by the somatic rearrangement of variable (V), diversity (D), and joining (J) segments of the T-cell receptor (Tcra, Tcrb, Tcrg, and Tcrd) and immunoglobulin (Igh, Igk, Igl) loci. V(D)J recombination is driven by a lymphocyte-specific recombinase complex (RAG-1/2) that targets conserved recombination signal sequences (RSSs) flanking each gene segment for rearrangement (1). This unique DNA rearrangement process is regulated in a tissue- and stage-specific manner at the level of chromatin accessibility, where germline transcription and chromatin opening facilitate access of RAGs to targeting sequences that flank each V, D, and J gene segment (2). Because antigen receptor loci coding segments are dispersed over considerable distances, large-scale chromatin contraction is essential for juxtaposing distal regulatory elements and RSSs (3).

CCCTC-binding factor (CTCF) is a highly-conserved, 11-zinc finger transcription factor that mediates long-range chromatin interactions within numerous developmentally-regulated loci including β-globin (4), MHC-II (5), IFN-γ (6), and the imprinted Igf2/H19 locus (7). By managing chromatin looping, CTCF is able to regulate gene expression through several distinct mechanisms. Localization of CTCF to promoter and enhancer regions facilitates a direct, physical interaction between the regulatory elements, resulting in a gain of expression (8, 9). Conversely, CTCF can also function as a chromatin insulator by looping out cis-elements, preventing promoter:enhancer interactions (10, 11), or by acting as a boundary element to prevent the spread of repressive heterochromatin to nearby regions of actively-transcribed genes (12, 13). In its role as a barrier element, CTCF also regulates DNA methylation, and binding at target sites strongly correlates with CpG hypomethylation (14, 15). Indeed, CTCF binding at the imprinted Igf2/H19 locus establishes differential
methylation that is required for appropriate expression of the genes in maternally- and
paternally-inherited alleles (11, 16, 17). Finally, CTCF has also been shown to direct
epigenetic regulation by influencing histone modifications, potentially by recruiting histone
deacetylases (HDACs) or histone acetyltransferases (HATs), to repress or activate gene
expression (4, 18-20).

Importantly, CTCF modulates the chromatin architecture within antigen receptor loci
to shape developmentally-timed accessibility and rearrangement (21). In B cells, CTCF
mediates dynamic looping of the Igh locus to limit or facilitate proper enhancer interactions
in a stage-specific manner (22-24), and also regulates distal V segment usage within both Igh
and Igk to maximize receptor diversity (25, 26). CTCF has been shown to fulfill a dual role
in the Tcra locus during thymopoiesis – by exerting enhancer-blocking activity to sequester
the tissue-specific Tcra enhancer, Ea, from the adjacent, ubiquitously-expressed Dad1 gene
(27), and in mediating long-range interactions between the T early α (TEA) promoter and Ea
enhancer to drive rearrangement (9).

The Tcra/d locus is distinct in that the gene segments encoding TCRα and δ are
arranged in a single locus, with the Dδ-Jδ-Cδ region interposed between a 5’ pool of mixed
Vα and Vδ segments, and the 3’ Jα region. The two loci are independently regulated, with
Tcrd rearranging first, at the DN (CD4−CD8−) stage, and Tcra rearrangement proceeding
later, at the DP (CD4+CD8+) stage (28, 29). Furthermore, because of the position of the Tcrd
region, cells that commit to the αβ lineage will delete the Vδ, Dδ, Jδ, and Cδ segments during
Tcra recombination. This added complexity of the Tcra/d locus requires precisely-timed
regulatory mechanisms to direct the two distinct rearrangement programs. Recombination at
Tcrd is driven by the Eδ enhancer, and upon transitioning to the DP stage activity is
redirected to the Ea enhancer (30), which is indispensible for Tcra rearrangement (31). In addition, the TEA germline promoter, which sits at the boundary between the Tcrd and Tcra gene segments, is important for imposing Jα accessibility within Tcra. Indeed, analysis of TEA−/− mice revealed that deletion of TEA led to a local loss of acetylation at histones H3 and H4 in DP cells, extending from TEA to the first block of Jα segments (32). Moreover, TEA deletion attenuated local RAG binding (33) and rearrangements involving the 5’ Jα segments (Jα61 to Jα52), and also skewed Jα segment usage toward the most 3’ gene segments (Jα34 to Jα2) (34). TEA, therefore, likely directs primary rearrangements to target 5’ Jα segments, while secondary joins involving 3’ Jα substrates are controlled by additional regulatory elements.

TEA is transcriptionally active solely in DP thymocytes, during a narrow window of development in a single cell lineage. Transcriptional activation at this stage coincides with the appearance of activating histone modifications at the Jα chromatin (35), as well as with DNA looping between TEA and Ea (36, 37). Interestingly, however, Tcra is already partially marked for rearrangement prior to the DP stage. Chromatin analyses identified transcription factor binding at TEA, including occupancy by CTCF, as well as enrichment for H3K4me2, H3K4me3, and H3ac at the preceding DN stage of development ((9, 30, 35) and Bradshaw et al., in preparation). We therefore sought to determine how TEA is primed for rearrangement. To answer this, we initially examined embryonic stem (ES) cells. Surprisingly, we found that TEA was already primed in the murine ES cell line, EF1. Additionally, chromatin immunoprecipitation revealed that CTCF bound the TEA promoter in ES cells. CTCF was previously shown to mediate chromatin looping at the DP stage, enabling formation of a promoter:enhancer complex that is required for transcriptional activation and primary
rearrangement (9). However, we found no such loops in ES cells (Bradshaw et al., in preparation). Because CTCF was enriched at TEA particularly early in cell development, before the formation of DNA loops, we speculated that the transcription factor could serve as an epigenetic regulator at TEA in an additional capacity.

In this study, ChIP analyses of CTCF-knockdown clones showed no change in histone marks relative to cells transfected with a non-targeting control. The data suggest that CTCF is dispensable for establishing the embryonic histone profile of TEA. Intriguingly, however, analysis of DNA methylation using bisulfite sequencing revealed that TEA is fully demethylated in the earliest stages of development (ESCs), even though the promoter is only active in the DP stage of thymopoiesis. Moreover, a loss of CTCF results in extensive methylation of CpG dinucleotides within the TEA promoter. Together, our data suggest that CTCF may act to insulate TEA from default repressive DNA methylation, thereby ensuring the promoter remains primed for activation at the proper developmental window.
2 Materials and Methods

2.1 Cells and antibodies

The EF1 murine embryonic stem (ES) cell line was cultured at 37°C/5% CO$_2$ in
DMEM (1X) + GlutaMAX-I medium (Life Technologies) supplemented with 15% FBS,
0.01% penicillin/streptomycin, 0.1 mM non-essential amino acids, 50 μM
β-mercaptoethanol, and 1 unit/mL ESGRO-LIF.

Antibodies to H3K9ac (ABE18; Millipore), H3K4me2 (ab32356; Abcam), H3K4me3
(39159; Active Motif, ab12209; Abcam), H3K27me3 (ab6002; Abcam), CTCF (07-729;
Millipore), and IgG (sc-2027; Santa Cruz) were used according to the manufacturer’s
recommendations.

2.2 RNAi knockdown of CTCF in ES cells

EF1 cells were stably transduced with a CTCF-specific shRNA lentiviral vector or
non-targeting, scrambled control as described (38). Transfectants were subjected to 1 μg/mL
puromycin selection 24 h later, and drug-resistant colonies were subcloned and screened for
CTCF knockdown by QRT-PCR.

2.3 Transcriptional analysis

RNA was extracted using RNAzol RT (Sigma), according to the manufacturer’s
instructions. 300 ng of RNA was reverse transcribed using 100U Moloney murine leukemia
virus reverse transcriptase (NEB) and oligo d(T) primers. The resultant cDNAs were
amplified by QRT-PCR (40 cycles) using 2X SensiMix (Bioline) and primers for CTCF
(Table 1). Following QRT-PCR, the relative abundance of CTCF cDNAs was quantified by
ΔΔCₜ normalization to a matched scrambled vector control, and standardized for loading variations by comparison to values obtained for β-actin.

2.4 Chromatin immunoprecipitation

Chromatin was prepared from formaldehyde cross-linked EF1 ES cells, or the indicated thymocytes (Rag²⁻, Rxβ, Eα⁻ Rxβ) (39, 40), using a Covaris 52 ultrasonicator to shear chromatin to ~200-400 bp DNA size. Chromatin immunoprecipitation (ChIP) was performed as previously described (41) using 15 μL chromatin/IP. Bound and input samples (4 μL) were analyzed in triplicate reactions by QPCR (40 cycles) using 2X SensiMix (Bioline). Primers are listed in Table 1. Average fold enrichment in bound samples at TEA relative to input was calculated for triplicate reactions using the ΔΔCₜ method and normalized to a heterochromatic, intergenic region of chromosome 1 (chr1).
3 Results

3.1 TEA promoter is accessible in DP thymocytes

During thymopoiesis, the Tcra locus undergoes multiple rounds of Vα-to-Jα rearrangement to form a productive coding join (29). 3’ Vα segments and the most 5’ Jα segments are targeted during initial cycles of recombination, while secondary rearrangements join progressively more 5’ Vα and 3’ Jα segments. This ordering of rearrangement events ensures that each thymocyte has numerous opportunities to form a functional α chain and survive positive selection (29). The TEA promoter is required for local accessibility and primary rearrangements involving the most 5’ Jα segments (Jα61 to Jα52) (34). Previous analyses of TEA reveal that the promoter is accessible during the DP stage of thymocyte development, becoming hyperacetylated at histones H3 and H4 upon the DN to DP transition (32, 42). Moreover, deletion of TEA results in a local decrease in H3ac as well as RAG-1 binding (32, 33). Transcriptional elongation from TEA is also required for accessibility and recombination across the 5’ Jα region, as insertion of a transcriptional terminator downstream of Jα56 was sufficient to inhibit rearrangement of Jα segments downstream of the terminator (43).

To confirm the accessibility of TEA in DP thymocytes and to establish an epigenetic profile for TEA during the earliest stages of cell development, we performed chromatin immunoprecipitations (ChIPs) using antibodies for histone modifications associated with open, active chromatin (H3K9ac, H3K4me2, H3K4me3) or a silent, heterochromatic configuration (H3K27me3). QPCR was then used to quantitate the enrichment of each histone mark at the TEA promoter over the course of cell development, and results were reported relative to input and normalized to an intergenic region of chromosome 1.
As expected, epigenetic marks associated with transcriptional activation are enriched at TEA only during the DP stage of development when *Tcra* is actively rearranging. At the earliest stage of development, embryonic stem (ES) cells are enriched for low levels of activating H3K4me2 and H3K4me3 (6- and 10-fold, respectively) (Fig. 1). Conversely, the heterochromatic H3K27me3 mark shows a nearly 40-fold enrichment in ES cells. The repressive chromatin profile in ES cells is consistent with the lack of tissue-specific promoter expression in the undifferentiated cells. Upon committing to the T-cell lineage and progressing to the DN stage of thymocyte development (Fig. 1: *Rag2*−/− cells are unable to rearrange their TCR genes, arresting cells at the DN stage of development), cells lose the silent H3K27me3 mark at TEA, but do not gain levels of H3K4me2 and H3K4me3, hallmarks of transcriptional activation. Loss of the repressive tri-methylated H3K27 suggests that the cell is fully poised for TEA promoter activation at the ensuing developmental window. Corroborating previous findings (32, 33, 42), our results demonstrate a substantial increase in activating histone modifications in DP cells (Fig. 1: Rxβ, these *Rag2*−/− cells progress to the DP stage of development owing to the presence of a rearranged TCRβ transgene, but then arrest at the DP stage), corresponding with transcriptional activation of

---

**Table 1. PCR primer sets**

<table>
<thead>
<tr>
<th>Amplicon</th>
<th>Strand</th>
<th>Sequence</th>
<th>Tm</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTCF</td>
<td>S</td>
<td>CCT GCT GTA GAA ATT GAA CCT GAG CC</td>
<td>66°C</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>ATA GTC CTG GTG CCG AGC AGA GCC CC</td>
<td></td>
</tr>
<tr>
<td>β-actin</td>
<td>S</td>
<td>AGA GCT ATG AGC TGC CTG ACG GCC</td>
<td>59°C</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>AGT AAT CTC TCT CTG CAT CCT GTC</td>
<td></td>
</tr>
<tr>
<td>TEA</td>
<td>S</td>
<td>TCT GTC TCC ATT CCC CAT CAT CC</td>
<td>63°C</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>TCT CTT ATC CTT GAA GCC CGT CC</td>
<td></td>
</tr>
<tr>
<td>chr1</td>
<td>S</td>
<td>GTG GAT AGG GGG TCT GAG G</td>
<td>63°C</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>CCA CTA ACC TGA CTA TCA AGC C</td>
<td></td>
</tr>
</tbody>
</table>
the TEA promoter and the onset of Tcra rearrangement. Our ChIP analyses reveal that H3K4me2 and H3K4me3 are enriched 118- and 201-fold, respectively, upon progression to the DP stage.

Figure 1. TEA promoter is associated with enhancer-dependent activating histone marks in DP cells. Histone H3K9ac (light gray), H3K4me2 (white), H3K4me3 (black) and H3K27me3 (dark gray) were quantitated in chromatin prepared from ES cells (EF1), DN thymocytes (Rag2−/−), DP thymocytes (Rxβ), and Eα−/− DP thymocytes (Eα−/− Rxβ). Enrichment at TEA was calculated relative to input samples and normalized to chr1. The data represent the mean ± SD of triplicate QPCR reactions. Analysis of two independent experiments (n=2) gave similar results.
Additionally, to determine the epigenetic landscape at TEA in the absence of the Eα enhancer, we analyzed chromatin prepared from thymocytes from Eα<sup>−/−</sup> Rxβ mice (40). These mice lack the Eα enhancer, and carry a rearranged TCRβ transgene on a Rag2<sup>−/−</sup> background, to arrest thymocyte development at the DP stage. We found that deletion of Eα resulted in a profound loss of activating histone marks, reducing H3K4me2 and H3K4me3 by 5- and 14-fold, respectively, relative to wild-type DP cells. We also detected a 6-fold increase in the repressive H3K27me3 modification in Eα<sup>−/−</sup> DP thymocytes relative to wild-type, approaching the level observed in ES cells. Interestingly, throughout development we did not see a dramatic change in the level of H3K9ac, a modification associated with open euchromatin. Acetylation at H3K9 appeared early, in ES cells, and was maintained throughout development. This result suggests that even prior to differentiation, chromatin at TEA is not facultative heterochromatin, but is instead maintained in a primed but inactive euchromatic state.

3.2 CTCF binds the TEA promoter prior to the DP stage

Previous chromatin analyses have suggested that transcription factors are loaded at the TEA promoter prior to the transcriptionally-active DP stage of thymopoiesis (30). ChIP-sequencing data in murine embryonic stem cells (Bruce4) specifically detected moderate binding of CTCF at this primitive stage of development (ENCODE GSM918748, (44, 45)). To confirm CTCF binding in our ES cell line, EF1, and to evaluate the presence of CTCF throughout development, we performed ChIP-QPCR assays of EF1 chromatin using a CTCF antibody, and reported results relative to input and normalized to an IgG control. Indeed, we found that CTCF was modestly enriched (approximately 3-fold over control) in ES cells (Fig. 2). CTCF binding increased considerably in DN and DP thymocytes (69- and 45-fold,
respectively), corroborating previous data from the Krangel laboratory showing that CTCF facilitates DNA looping between TEA and Eα in DN and DP cells (9). That the elevated CTCF signals in thymocytes relative to ES cells at least partially reflect stabilization of promoter:enhancer complexes is supported by the 5-6 fold reduction of CTCF levels in Eα-deficient DP cells (12-fold over IgG control) to levels more similar to those seen in ES cells. Although enrichment in ES cells was less pronounced than in adult, differentiated cells, we speculated that CTCF’s presence at the onset of development suggested that it may function as an epigenetic regulator of TEA, prior to its role mediating in chromatin looping (9).

Figure 2. CTCF binds the TEA promoter in ES cells. CTCF enrichment was quantitated from chromatin prepared from ES cells (EF1), DN thymocytes (Rag2<sup>−/−</sup>), DP thymocytes (Rxβ<sup>−/−</sup>), and Eα<sup>−/−</sup> DP thymocytes (Eα<sup>−/−</sup> Rxβ<sup>−/−</sup>). Enrichment at TEA was calculated relative to input samples and normalized to an IgG control. The data represent the mean ± SD of triplicate QPCR reactions.
3.3  **CTCF does not direct histone modifications at the TEA promoter**

Because TEA is already occupied by CTCF at the earliest stages of cell development (Fig. 2), we wondered if the transcription factor could direct histone modifications to prime the promoter for full activation in DP cells. To address this, we performed shRNA knockdown of CTCF in the ES cell line, EF1. Following stable transduction of shRNA knockdown virus, cells were subcloned and mRNA levels of selected clones were assessed by QRT-PCR. Five EF1 subclones displaying the highest knockdown are shown in Figure 3. CTCF mRNA levels for clones 1-2, 1-15, 2-5, 2-12, and 2-15 were knocked down 84%, 69%, 72%, 76%, and 63%, respectively, relative to cells transfected with a non-targeting control plasmid. Clone 2-12 was then selected for further analysis, as knockdown approached 80% at the mRNA level.

![CTCF knockdown clone](image)

**Figure 3.** Knockdown levels of CTCF in EF1 subclones. ESCs were transduced with either a CTCF-targeting shRNA or non-targeting control virus and subcloned under 1 μg/mL puromycin selection. CTCF mRNA levels were then measured by QRT-PCR. Results were calculated relative to scrambled control (CTL6) and normalized to β-actin for loading variations. The data represent the mean ± SD of triplicate QRT-PCR reactions.
We next used ChIP-QPCR to assess histone modifications at TEA in the CTCF knockdown clone 2-12, relative to ES cells transfected with the non-targeting control (CTL6). We saw no difference in histone modifications between CTCF knockdown and control ES cells (Fig. 4). Results for knockdown and control samples were comparable to those seen in wild-type ES cells (Fig. 1). Both knockdown (2-12) and control (CTL6) EF1 clones were enriched for the activating H3K9ac mark (42- and 45-fold enrichment, respectively), as well as repressive H3K27me3 (26- and 28-fold, respectively). As expected, no enrichment was detected for histone modifications associated with transcriptional activation, as H3K4me2 and H3K4me3 were detected at levels equivalent to IgG controls in both cell types.

Figure 4. Loss of CTCF does not impact chromatin organization at the TEA promoter. Chromatin from CTCF knockdown subclone 2-12 (black bars) and control subclone CTL6 (white bars) were immunoprecipitated with antibodies to H3K9ac, H3K4me2, H3K4me3, H3K27me3 or IgG. Resultant DNAs were analyzed by QPCR for histone modifications at the TEA promoter. Enrichment was calculated relative to input samples and normalized to chr1. The data represent the mean ± SD of triplicate QPCR reactions and are representative of two independent experiments (n=2).
The results from knockdown experiments rule out a potential role for CTCF in modulating histone-dependent accessibility of the TEA promoter. The results agree with our observation in wild-type ES cells that chromatin is not facultative heterochromatin in ES cells, but rather open euchromatin. Subsequent studies by our group showed that, while CTCF is dispensable for deposition of euchromatic H3K9ac, it is essential to protect TEA from repressive DNA methylation (Bradshaw et al., in preparation). Specifically, CTCF knockdown clones showed rapid and complete methylation of TEA. Together, our data suggest that embryonic deposition of CTCF helps to ensure that TEA is maintained in a primed state until its window of activation.
4 Discussion

The results from this study argue against a role for CTCF in directing embryonic histone modifications at the TEA promoter, in that no difference in epigenetic marks was seen between CTCF knockdown and control ES cells. However, several surveys suggest that TEA is primed for activation prior to the DP stage of thymocyte development ((9, 30, 35), Bradshaw et al., in preparation). Because CTCF is also known to influence DNA methylation (11, 14-17), we speculated that the transcription factor could serve in such a capacity at TEA. Thus, we performed bisulfite sequencing of the CTCF knockdown clones from Figure 3 to determine the methylation profile of the region surrounding the TEA promoter. Interestingly, we found that TEA is fully demethylated in control ES cells, while knockdown of CTCF results in extensive methylation at TEA. Why might evolution invest in constitutive CTCF binding at a promoter that is used in only one transient developmental stage of one discrete cell type? The answer might lie in the regulatory architecture of Tcra/d. Because TEA sits only 5 kb downstream of Tcrd, which is active earlier in development, TEA must be insulated from the Tcrd enhancer. Indeed, though unmethylated, TEA sits within a 2-3 kb stretch of fully methylated DNA that has been linked to this insulator. Therefore, constitutive binding of CTCF at TEA may be essential to prevent expansion of the flanking methylation. Based on this finding, we hypothesized that CTCF could serve as an insulator, to protect TEA from methylation and ensure prompt activation of the promoter within the appropriate window of thymocyte development. To further address this possibility, we intend to first perform methylated DNA immunoprecipitation (MeDIP), to establish the global methylation pattern of the entire 1.6 Mb Tcra locus in wild-type versus CTCF knockdown ES cells. Additionally, to determine if CTCF can induce hypomethylation at repressed sites, we will
use a human CTCF plasmid to restore expression in our knockdown clones, and then re-evaluate methylation levels at TEA. While binding of CTCF is known to prevent de novo methylation at numerous regulatory regions, it is unclear if CTCF can remove established methylation marks to regulate gene expression (15). Similar studies in DN cells will then establish whether constitutive CTCF binding at TEA might derive from the need to protect the promoter’s activation capacity during differentiation.

Epigenetic priming is fundamental to the precisely-timed expression of lineage-specific genes during cellular differentiation from pluripotent progenitor cells. Many developmentally-regulated promoter regions are poised for activation at the onset of development in ES cells. Bivalent H3K4me3/H3K27me3 marks as well as H3K9ac, DNA hypomethylation, and enrichment of RNA polymerase II have each been suggested to prime numerous genes for expression (46-48). Moreover, CTCF has been established as an important epigenetic regulator, in particular for preventing epigenetic silencing by protecting against abnormal promoter methylation in ES cells (17). Because deletion of CTCF results in early embryonic lethality (49), this presents a challenge for studying its role in priming various tissue-specific genes in early development. Thus, the Tcra locus, expressed exclusively in T lymphocytes, provides an ideal model for further understanding the dynamic relationship between CTCF and DNA methylation. We anticipate that our future experiments focusing on the role of CTCF in TEA promoter methylation will provide additional insight into the mechanisms of epigenetic priming.
5 References


