Mechanism of Mismatch Repair Induced Mutagenesis in Somatic Hypermutation

by

Darina Frieder

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Graduate Department of Immunology
University of Toronto

© Copyright by Darina Frieder 2010
Mechanism of Mismatch Repair Induced Mutagenesis in Somatic Hypermutation

Darina Frieder

Degree of Doctor of Philosophy
Graduate Department of Immunology
University of Toronto
2010

Abstract

B cells produce a diverse array of antibody specificities that are of low affinity during the initial phase of a humoral immune response. However, somatic hypermutation of the rearranged V region in the immunoglobulin locus generates new antibody affinities, accompanied by the selection of B cells that produce superior antibody affinities. Somatic hypermutation is initiated by the conversion of G:C base pairs to G:U lesions by the enzyme activation induced cytosine deaminase. Left unrepaired, G:U lesions will give rise to transition mutations at G:C base pairs, but are converted to transition and transversion mutations at G:C and A:T base pairs by the paradoxical participation of the base excision repair and mismatch repair pathways. The mismatch repair pathway, which evolved to correct errors produced during DNA replication, is co-opted by hypermutating B cells to produce A:T mutations via the processing of G:U lesions. This process requires the mismatch repair components Msh2, Msh6, and Exo1, but is additionally dependent upon the translesional DNA polymerase η, a known A:T mutator, and on ubiquitinated PCNA, an initiator of translesion synthesis. The presence of certain types of lesions in the template strand during DNA replication leads to the activation of translesion synthesis. I propose
that a similar mechanism operates during somatic hypermutation to activate translesion synthesis and recruit DNA polymerase η. Our model suggests that mismatch repair-generated single-stranded DNA tracts contain abasic sites produced as a result of uracil excision by uracil-\(N\)-glycosylase. Synthesis opposite abasic sites activates translesion synthesis and results in the recruitment of polymerase η and the subsequent production of A:T mutations. In this thesis, I present data from hypermutating murine B cells and the B cell line Ramos to support this model, demonstrating that the base excision repair and mismatch repair pathways cooperate during somatic hypermutation to generate A:T mutations. In addition, I explore the role of the Mre11-Rad50-Nbs1 complex in its contribution to A:T mutations in Ramos cells. Taken together, these studies demonstrate that conversion of classical DNA repair pathways into mutation-generating processes is driven by the unique environment of the V region in hypermutating B cells.
ACKNOWLEDGEMENTS

The pursuit of this degree has been an incredible journey and I am grateful to have had this opportunity. In these past six years I gained so much knowledge, about science and research of course, but also about myself. And through it all, I had the company of wonderful friends. For that reason, I dedicate this thesis to all the people who took this journey with me.

First of all, thank you, Alberto for taking me on as your first student, for supporting me throughout the struggles and victories alike and for encouraging me always to be better, work harder, and most of all, to focus. Your lab has grown tremendously over the years and I have no doubt you will continue to prosper in the years to come. I wish you nothing but happiness both professionally and in your personal life.

Many lab members need to be thanked, some of whom have left to pursue their own goals and some who remain. All of you have contributed laughter and camaraderie and this experience would have been lesser were it not for each and every one of you. Wajiha, Mani, and Jahan, together we formed “the core” and I like to think we helped set the direction for the lab. Mani, your guidance and wisdom helped me through many tough situations and our discussions always provided different perspective. Jahan, I have watched you blossom into someone for whom I have the utmost respect, as a scientist, a friend and fellow “human.” You are one of the most giving individuals I have ever met and I am so glad we became desk buddies and ultimately, friends. Ahmad, there is no need to tell you how funny you are. You should, however, know that your presence in the lab is invaluable and you truly have a great deal to offer. Shaliny, alas, you’ve made me laugh as well, perhaps by providing fodder to Ahmad’s sense of appetite for humor, but you always take it in stride and for that I thank you. Rajeev, our friendship began in the underbelly of MSB, but once the masks came off, I learned what an interesting person you are and I have thoroughly enjoyed our many discussions. Blerta, your confidence and commitment to your goals is remarkable and impressed me from our first meeting. Bryant, I have enjoyed trading opinions about various culinary matters. Maribel, thank you for taking care of my cells while I took care of my post-surgery leg. It was a great pleasure to work with you and to get
to know you in the process. Although not a lab member, I need to also thank Rejeanne for performing miracles and for being a much-needed friend.

To my parents, I know I can always count on you for love and support. I admire and respect the difficult decision you made many years ago to leave your home in pursuit of a new adventure. We will never know how things would have turned out had you chosen a different path, but I am grateful for the opportunities and experiences your decision enabled me.

Richard, your encouragement and support through all of my endeavours has meant the world to me. You manage to believe in me even when I don’t. You encourage me in everything I do and support me every step of the way. I am better with you in my life. Of course I have to thank Clara, our beautiful kitty, for bringing joy to my life and allowing me to take care of her.

Thank you also to Chao, my one and only roommate. I cherish the bond we developed in those two years together on Madison and hope we will continue our friendship for many years. Mireille, you are one of my oldest friends and I find great comfort in sharing my joys and sorrows with you. And as someone who has completed her formal education, you have served as an inspiration in the last couple of years.

Finally, I would like to thank my committee, both past and present, for standing by me, encouraging me and for asking more of me. I respect each and every one of you scientifically and have learned a great deal from all of you. I also feel privileged to have enjoyed your company at conferences and IGSA events. After so many years at this university in this department, I suspect it will always feel a little bit like home to me.
# TABLE OF CONTENTS

Title Page ........................................................................................................................................................................ i
Abstract ..................................................................................................................................................................................... ii
Acknowledgments ...................................................................................................................................................................... iv
Table of Contents ..................................................................................................................................................................... vi
List of Tables ........................................................................................................................................................................ xi
List of Figures .......................................................................................................................................................................... xii
List of Appendices ................................................................................................................................................................... xiv
List of Abbreviations .............................................................................................................................................................. xv

## CHAPTER 1: Introduction ..................................................................................................................................................... 1-83

1.1. B Cells .............................................................................................................................................................................. 2
    1.1.1. The Humoral Immune Response .................................................................................................................. 2
    1.1.2. B Cell in the Immune Response ................................................................................................................... 4
    1.1.3. B Cell Activation .............................................................................................................................................. 5

1.2. Antibody Diversification Mechanisms ..................................................................................................................... 8
    1.2.1. V(D)J Recombination ................................................................................................................................. 8
    1.2.2. Somatic Hypermutation ............................................................................................................................... 12
    1.2.3. Class Switch Recombination ...................................................................................................................... 16
    1.2.4. Immunoglobulin Gene Conversion ............................................................................................................ 18
    1.2.5. Activation Induced Cytidine Deaminase ................................................................................................. 22

1.3. DNA Repair Mechanisms .......................................................................................................................................... 27
    1.3.1. Double Strand Break Repair ....................................................................................................................... 27
        1.3.1.1. Homologous Recombination ............................................................................................................... 28
        1.3.1.2. Non-Homologous End Joining ........................................................................................................... 29
        1.3.1.3. Mre11-Rad50-Nbs1 Complex ............................................................................................................. 31
1.3.2. Base Excision Repair .................................................................36
1.3.3. Nucleotide Excision Repair.........................................................37
1.3.4. Translesion Synthesis ...............................................................38

1.4. Mismatch Repair ...........................................................................44
1.4.1. Overview.................................................................44
1.4.2. Mismatch Repair in Prokaryotes...................................................46
1.4.3. Mismatch Repair in Eukaryotes....................................................50
1.4.4. Molecular Mechanisms of Mismatch Repair – Current Models ..........57
1.4.5. Mismatch Recognition ................................................................59
1.4.6. Mismatch Repair and Antibody Diversification .........................61
   1.4.6.1. Mismatch Repair and Somatic Hypermutation .................62
      i. MMR Knock-Outs and Mutants ..............................................63
      ii. Translesional Polymerases ..................................................68
      iii. UNG Contributes to A:T Mutations ......................................70
      iv. PCNA and the Induction of A:T Mutations .........................71
   1.4.6.2. Mismatch Repair and Class Switch Recombination ...............72
1.4.7. Mismatch Repair and DNA Damage Signaling .............................74
1.4.8. Mismatch Repair and Cancer .....................................................76

1.5. Hypothesis and Objectives .............................................................82

CHAPTER 2: Msh2 and UNG Co-Operate to Induce A:T Mutations in Ramos .................................................................84-112

2.1. Abstract .........................................................................................85
2.2. Introduction ..................................................................................86
2.3. Materials and Methods .................................................................89
   2.3.1. Cell culture, subcloning, and flow cytometry .........................89
   2.3.2. Measurement of MMR activity .............................................89
2.3.3. ELISPOT assays .................................................................................................................... 90
2.3.4. Plasmids and transfections ............................................................................................. 90
2.3.5. DNA extraction, PCR, and sequencing in Ramos ................................................... 90
2.3.6. In vitro UDG assay .............................................................................................................. 91
2.3.7. Statistical analysis .............................................................................................................. 91

2.4. Results ............................................................................................................................................. 92
2.4.1. Transversion mutations at C are linked to A:T mutations ................................ 92
2.4.2. AID does not show a strong stand bias preference ...................................................... 101
2.4.3. Inhibition of UNG abrogates linked C transversion and A mutations ...... 104

2.5. Discussion ................................................................................................................................... 108

CHAPTER 3: A:T Mutations in Murine Immunoglobulin Sequences Are Stimulated by the Concerted Action of UNG and Msh2 ........................................................................................................... 113-134

3.1. Abstract ....................................................................................................................................... 114
3.2. Introduction ............................................................................................................................... 115
3.3. Materials and Methods ......................................................................................................... 117
3.3.1. Amplification and analysis of murine Ig sequences ....................................................... 117
3.3.2. Statistical analysis ............................................................................................................... 117

3.4. Results .......................................................................................................................................... 118
3.4.1. Murine Ig sequences with A:T mutations are enriched for C transversion mutations .................................................................................................................. 118
3.4.2. Adenine mutations on the top strand are associated with cytosine transversion mutations on the bottom strand in murine B cells ............. 121
3.4.3. Bottom strand C transversion mutations are preferentially located upstream of A mutations .................................................................................................................. 126
3.4.4. Deficiency of MMR leads to decreased C transversion mutations........... 127
CHAPTER 4: shRNA-Mediated Silencing of MRE11 Does Not Affect SHM in Ramos Cells

4.1. Abstract

4.2. Introduction

4.3. Materials and Methods
   4.3.1. Cell Culture
   4.3.2. Plasmids and transfections
   4.3.3. Western blot analysis
   4.3.4. Measurement of mismatch repair activity
   4.3.5. ELISPOT assays
   4.3.6. Clonal survival assays
   4.3.7. DNA extraction, PCR and sequencing
   4.3.8. Statistical analysis

4.4. Results
   4.4.1. shRNA reduces MRE11 expression in Ramos
   4.4.2. MRE11\textsuperscript{low} Ramos clones do not display increased IR sensitivity
   4.4.3. MRE11\textsuperscript{low} clones restore MRE11 levels
   4.4.4. Mismatch repair is not compromised in MRE11\textsuperscript{low} clones
   4.4.5. A:T mutation frequency is not affected shMRE11 Ramos clones

4.5. Discussion

CHAPTER 5: Discussion and Future Directions

5.1. Generation of MMR-dependent mutations at A:T base pairs in the absence of UNG
5.1.1. Non-UNG mediated induction of mutagenic MMR.......................... 159
5.1.2. A role for MutLα in somatic hypermutation............................................. 162
5.2. Generation of MMR-independent mutations at A:T base pairs ................... 164
5.3. Translesional polymerases and somatic hypermutation.............................. 166
5.4. Conclusions.................................................................................................. 168

CHAPTER 6: Appendices....................................................................................... 169-186

CHAPTER 7: References ....................................................................................... 187-217
LIST OF TABLES

CHAPTER 1

Table 1.1. Comparison of elevated cancer risk in HNPCC versus general population ................................................................. 77
Table 1.2. Mouse MMR defects and associated tumor spectra ......................................................... 81

CHAPTER 2

Table 2.1. Characteristics of mutations derived from Ramos clones ............................................. 99

CHAPTER 4

Table 4.1. Mutation frequency is not decreased in Mre11<sub>low</sub> Ramos clones ........ 152
LIST OF FIGURES

CHAPTER 1

Figure 1.1. AID-induced lesions initiate somatic hypermutation, class switch recombination and Ig gene conversion............................................................... 9
Figure 1.2. AID-generated G:U lesions give rise to mutations G:C and A:T base pairs ........................................................................................................ 14
Figure 1.3. Organization of the murine C region genes................................................. 19
Figure 1.4. Structural and functional properties of AID................................................. 24
Figure 1.5. The Mre11-Nbs1-Rad50 (MRN) complex senses double strand breaks ........................................................................................................ 34
Figure 1.6. MutH directs mismatch repair to nascent DNA strands in prokaryotes ........................................................................................................ 47
Figure 1.7. Msh2-Msh6 dimers recognize single base-base mismatches................. 52

CHAPTER 2

Figure 2.1. Model for mutagenic repair by MutSα....................................................... 93
Figure 2.2. Sequences with A:T mutations are enriched for transversion mutations at C on the bottom strand in a UNG-dependent manner .................... 97
Figure 2.3. Mutations at G:C basepairs occurs equally on both top and bottom strands in Ramos cells................................................................. 102
Figure 2.4. Expression of UNG inhibitor UGI in R67 cells reduces both A:T mutations and C transversion mutations......................................................... 105

CHAPTER 3

Figure 3.1. Characteristics of mutations derived from wild type and UNG−/− mice ........................................................................................................ 119
Figure 3.2. Sequences with A:T mutations are enriched for G:C transversion mutations .................................................................................................. 122
Figure 3.3. Transversion mutations at C on the bottom strand are associated with mutations at A on the top strand in wild-type mice .................................. 124
Figure 3.4. Ig sequence data for wild type and MMR−/− mice ........................................ 128
Figure 3.5. Bottom, but not top, strand C transversion mutations are reduced in MMR-deficient mice .................................................................................................................. 130

CHAPTER 4

Figure 4.1. Mre11 protein level is reduced in Ramos transfected with shRNA vector ........................................................................................................................................................................ 143
Figure 4.2. Mre11low clones do not display increased sensitivity to ionizing radiation ........................................................................................................................................................................ 147
Figure 4.3. Mismatch repair activity and A:T mutation frequency are not compromised in Mre11low clones ................................................................. 150
LIST OF APPENDICES

Appendix 1. Primary sequence data of IgM+ Ramos 67 subclones that have associated mutations with the reverted nonsense codon ........................................................ 170

Appendix 2. Restricted mutation dataset from 9 wildtype and 5 UNG−/− mice .................. 178
LIST OF ABBREVIATIONS

A
Adenine

ABC
ATP binding cassette

AID
Activation induced cytidine deaminase

APC
Antigen presenting cell

APE1
Apurinic endonuclease 1

APOBEC
Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like

ATCIZ
ATM/ATR-substrate CHK2-interacting zinc finger

ATM
Ataxia telangiectasia mutated

ATP
Adenosine triphosphate

ATR
Ataxia telangiectasia and Rad3 related

BER
Base excision repair

BLM
Bloom syndrome, RecQ helicase-like

BRAF
V-raf murine sarcoma viral oncogene homolog

C
Cytosine

C region
Constant region

CHK1
Checkpoint kinase 1

CHK2
Checkpoint kinase 2

CSR
Class switch recombination

CtlP
C-terminal regions of adenovirus E1A interacting protein

DM
Dimerization motif

DDT
DNA damage tolerance

DNA
Deoxyribonucleic acid

PKCS
Protein kinase – catalytic subunit

DSB
Double stranded break

dsDNA
Double-stranded DNA

Exo
Exonuclease

Fen1
Flap structure specific endonuclease 1

G
Guanine

GFP
Green fluorescent protein

HMGB1
High mobility group box 1

HNPPCC
Hereditary non-polyposis colorectal cancer

HR
Homologous recombination

Ig
Immunoglobulin

INO80
Inositol80

IDL
Insertion-deletion loop

Kb
Kilobase

KRAS
Kirsten rat sarcoma

LP
Long patch

Lys
Lysine

MBD4
Methyl-CpG binding domain protein 4

MHC
Major hitocompatibility complex

Mlh
MutL homologue

MMR
Mismatch repair

MRN
Mre11-Rad50-Nbs1
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Msh</td>
<td>MutS homologue</td>
</tr>
<tr>
<td>MSI</td>
<td>Microsatellite instability</td>
</tr>
<tr>
<td>MUT</td>
<td>Mutation</td>
</tr>
<tr>
<td>NER</td>
<td>Nucleotide excision repair</td>
</tr>
<tr>
<td>NES</td>
<td>Nuclear export signal</td>
</tr>
<tr>
<td>NF1</td>
<td>Neurofibromatosis 1</td>
</tr>
<tr>
<td>NLS</td>
<td>Nuclear localization signal</td>
</tr>
<tr>
<td>NP</td>
<td>Nitrophenyl</td>
</tr>
<tr>
<td>PCNA</td>
<td>Proliferating cellular nuclear antigen</td>
</tr>
<tr>
<td>PIP</td>
<td>PCNA interacting protein</td>
</tr>
<tr>
<td>QM</td>
<td>Quasi-monoclonal</td>
</tr>
<tr>
<td>PKA</td>
<td>Protein kinase A</td>
</tr>
<tr>
<td>RAG</td>
<td>Recombination activating gene</td>
</tr>
<tr>
<td>RecJ</td>
<td>Recombinase J</td>
</tr>
<tr>
<td>RFC</td>
<td>Replication factor C</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>RPA</td>
<td>Replication Protein A</td>
</tr>
<tr>
<td>Ser</td>
<td>Serine</td>
</tr>
<tr>
<td>SHM</td>
<td>Somatic hypermutation</td>
</tr>
<tr>
<td>shRNA</td>
<td>Short hairpin RNA</td>
</tr>
<tr>
<td>SMUG1</td>
<td>Single-strand-selective monofunctional uracil DNA glycosylase 1</td>
</tr>
<tr>
<td>SP</td>
<td>Short patch</td>
</tr>
<tr>
<td>SSB</td>
<td>Single-stranded break</td>
</tr>
<tr>
<td>ssDNA</td>
<td>Single-stranded DNA</td>
</tr>
<tr>
<td>SUMO</td>
<td>Small ubiquitin modifier</td>
</tr>
<tr>
<td>Thr</td>
<td>Threonine</td>
</tr>
<tr>
<td>T</td>
<td>Thymine</td>
</tr>
<tr>
<td>TDG</td>
<td>T:G mismatch thymine glycosylase</td>
</tr>
<tr>
<td>TdT</td>
<td>Terminal deoxynucleotidyl transferase</td>
</tr>
<tr>
<td>TLS</td>
<td>Tranlesion synthesis</td>
</tr>
<tr>
<td>TGF-β</td>
<td>Transforming growth factor β</td>
</tr>
<tr>
<td>TRN</td>
<td>Transition</td>
</tr>
<tr>
<td>TRV</td>
<td>Transversion</td>
</tr>
<tr>
<td>TTD</td>
<td>Thymine-thymine dimer</td>
</tr>
<tr>
<td>Tyr</td>
<td>Tyrosine</td>
</tr>
<tr>
<td>U</td>
<td>Uracil</td>
</tr>
<tr>
<td>Ub</td>
<td>Ubiquitin</td>
</tr>
<tr>
<td>UBM</td>
<td>Ubiquitin binding motif</td>
</tr>
<tr>
<td>UBZ</td>
<td>Ubiquitin binding zinc finger</td>
</tr>
<tr>
<td>UDG</td>
<td>Uracil DNA glycosylase</td>
</tr>
<tr>
<td>UNG</td>
<td>Uracil-N-glycosylase</td>
</tr>
<tr>
<td>UV</td>
<td>Ultraviolet</td>
</tr>
<tr>
<td>V region</td>
<td>V region</td>
</tr>
<tr>
<td>V(D)J</td>
<td>Variable (Diversity) Joining</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction
1.1. B Cells

Immune responses are complex and multi-faceted systems that involve diverse cell types, including B cells. In order to generate a powerful immune response, B cells undergo numerous developmental changes, which rely on the action of DNA repair pathways such as mismatch repair, translesion synthesis, and non-homologous end joining, to name a few. Throughout this thesis, I will explore the network of DNA repair mechanisms that co-operate in the production of effective humoral immune responses.

1.1.1. The Humoral Immune Response

The immune system is a complex organ matrix composed of many different cell types and tissues which work together to protect the host from invading pathogens. The innate arm of the immune system consists of various cell types, including monocytes, macrophages, dendritic cells, natural killer cells, natural killer T cells, and eosinophils, and provides the essential first line of defense against foreign antigen. It does not, however, possess the capacity to provide long-lasting, pathogen-specific immunity. In contrast, the adaptive arm of the immune system, which consists of B and T cells, has the capacity to produce antigen-specific immune responses, to discriminate between self and non-self, and to differentiate into long-lived memory cells that can be rapidly stimulated into action upon re-infection with the same antigen. The adaptive immune system is further divided into humoral and cell-mediated arms, regulated by B and T cells, respectively. Despite these separations, cross-talk between all cells of the immune system is crucial to generate an effective immune response.
The primary role of B cells in the immune response is to produce antigen-specific antibodies upon activation and differentiation into plasma cells. Antibodies are complex protein structures which consist of two distinct chains, a heavy and a light chain, that form an antigen-specific variable (V) region and a constant (C) region that dictates the antibody isotype and function. Secreted antibodies coat the antigen surface and interact with other host cells in order to mediate three immune functions: neutralization, opsonization, and complement activation. Antibodies can neutralize pathogens by coating their surface, which prevents interactions between pathogen and the host cell-surface receptors used for entry. Phagocytosis of pathogens by macrophages is enhanced by interactions between antibody receptors, termed Fc receptors, and antibodies bound to the surface of pathogens. Finally, immune complexes consisting of antibody and antigen activate the complement system by stimulating the cleavage of the C1q complement protein, thereby stimulating an inflammatory cascade that induces cell death. In addition to antibody-mediated functions, activated B cells can also behave as antigen presenting cells that express co-stimulatory molecules and assist in the activation of T cells.

Interactions between B cells and other hematopoietic cells, primarily T helper (T_h) cells, ensure that B cells receive all of the signals necessary for activation and that they are activated exclusively in response to appropriate antigens. Some B cell subsets, including immature B cells, B-1 cells and marginal zone B cells, can respond to certain classes of antigen, termed T-independent antigen, independently of T cell help. However, most B cells respond to T-dependent antigen and require T cell help in order to become activated. T lymphocytes can be divided into minimally three subsets: cytotoxic T cells that contribute to the destruction of pathogen, helper T cells that assist in the activation of other immune cells, and regulatory T cells that attenuate the immune response. T cell activation is
mediated by two distinct and highly regulated signals: recognition of antigen in the context of the major histocompatibility complex (MHC) and engagement of co-stimulatory molecules on antigen presenting cells (APCs) that deliver crucial activation signals only to the appropriate T cells. Dendritic cells and other professional APCs that have processed antigen upregulate co-stimulatory molecules on their cell surface and migrate to peripheral lymphoid organs such as the lymph nodes and spleen in order to interact with and activate T cells, which then acquire the ability to activate B cells.

1.1.2. B Cells in the Immune Response

The predominant role of B cells in the immune response is the production of high affinity antigen-specific antibodies of various isotypes with distinct functions. B cell defects that affect antibody production or secretion result in immunodeficiency syndromes such as X-linked agammaglobulinemia (XLA), common variable immunodeficiency (CVID), or hyper IgM (HIGM) syndrome, characterized by increased susceptibility to infections by encapsulated bacteria (Bacchelli et al., 2007; Conley et al., 2005; Durandy et al., 2007; Revy et al., 2000; Tsukada et al., 1993; Vetrie et al., 1993). B cells recognize antigen through the V region of the B cell receptor (BCR), which has the same specificity as its secreted antibody form and is unique to each B cell clone. A given B cell’s antibody specificity can remain the same while the C region, which determines its function and anatomical localization, can vary (Nossal and Lederberg, 1958; Preud’homme and Seligmann, 1972).

Mice and humans produce five main antibody isotypes, designated IgM, IgG, IgA, IgE, and IgD, with IgG divided into four and IgA divided into two additional isotypes, each one harboring a specialized role. The initial phase of the humoral immune response is
characterized by the production of low affinity IgM antibodies, generally found in the blood and, to a lesser degree, in the lymph. IgM molecules oligomerize into pentameric and hexameric complexes that possess ten or twelve antigen-binding sites, respectively, thereby enhancing the avidity of low affinity clones by increasing the number of potential interactions with antigen. Notably, the hexameric IgM molecule has been shown to be more cytolytic than the pentameric form (Davis et al., 1988; Randall et al., 1990). The IgG class of antibodies consists of small, high affinity, monomeric molecules with long half-lives. This is the principal antibody in the blood and extracellular fluid, but also in tissues into which it diffuses easily because of its small size. The chief function of IgG is in pathogen opsonization and complement system activation. IgA resides in the mucosal epithelium of intestinal and respiratory tracts in mono- or dimeric form and performs mainly neutralization functions. IgE is a monomeric antibody that interacts with Fcε receptors on the surface of mast cells via its C terminus, and activates them to release inflammatory mediators. Finally, IgD, which is found in small amounts in the serum, lacrimal, salivary and mammary secretions, and upper respiratory mucosa, is the least well characterized class of antibody, but has recently been shown to stimulate the humoral immune response by binding to basophils and triggering the release of B-cell stimulatory factors such IL-4 (Brandtzaeg et al., 1999; Chen et al., 2009; Conrad et al., 1990; Plebani et al., 1983; Preud'homme et al., 2000).

1.1.3. B Cell Activation

The majority of B cell responses are induced by T-dependent antigens and require T-cell help before they can be initiated. B cells reside in the follicular region of secondary lymphoid organs, either the lymph nodes or the spleen, and can detect antigen directly via
the lymph or the blood, respectively, or in immune complexes on dendritic cells and macrophages (Carrasco and Batista, 2007; Colino et al., 2002; Junt et al., 2007; Phan et al., 2007; Qi et al., 2006; Wykes et al., 1998). Recirculating B cells enter the secondary lymphoid organs through the T-cell zone, where dendritic cells that picked up antigen in the periphery engage T cells for activation. Although most B cells will migrate on to the B cell follicles, some will recognize and engage antigen through the BCR and become primed to receive help from specialized CD4+ T_{H2} cells (Pape et al., 2007). B cells localized in the follicles can also encounter antigen and become primed to receive T cell help (Bergtold et al., 2005).

Ligation of the BCR and the B cell co-receptor (CD19:CD21:CD81:Leu 13) complex (Rickert, 2005; Sato et al., 1997) induces antigen internalization, processing and expression of antigenic peptides in the context of MHC class II (MHC-II) molecules. Furthermore, B cells express the co-stimulatory molecule CD40, enabling them to receive activation signals from T_{H2} cells. In order to receive T cell help, the peptide-MHC-II complex on the B cell surface must be recognized by cognate T cells that were activated earlier by dendritic cells presenting peptide components of the same antigen. This contact enables B cells to receive the crucial activation signal mediated by the ligation of CD40 on B cells by CD40L (CD154) on T cells and ensures that only antigen specific B cells become activated, thereby maintaining tolerance to self.

Antigen engagement by the BCR drives B cell proliferation and induces transit to the B/T cell boundary (Vinuesa et al., 2005) where B cells seed secondary follicles and set up germinal centres in which isotype switching and affinity maturation occur. Affinity maturation is the process whereby the antibody locus undergoes targeted V-region
mutagenesis, somatic hypermutation, in order to generate amino acid differences in the antibody binding sites that translate to varying levels of antibody affinity. After each round of mutagenesis, B cells are selected based on antibody affinity, with the cells that bear higher affinity BCRs outcompeting those with lower affinity BCRs, such that the overall antibody affinity increases throughout the course of the immune response. B cells compete for survival signals provided by resident T cells and follicular dendritic cells and those that do not receive appropriate signals are eliminated via apoptosis. In some cases, B cell activation drives the direct differentiation into antibody-secreting plasmablasts, an intermediate developmental stage that precedes fully-differentiated, non-proliferating, antibody-secreting plasma cells.

Germinal centres are regions of intensively proliferating B cells, a small proportion of T cells and follicular dendritic cells, surrounded by a mantle of non-proliferating B cells (Batista and Harwood, 2009). The germinal centre is divided into two zones, the dark zone and the light zone. The dark zone contains centroblasts, rapidly-proliferating B cells that are actively undergoing somatic hypermutation. The light zone contains follicular dendritic cells and centrocytes, non-proliferating B cells with mutated V regions that are in the process of being selected based on antigen affinity (Hanna, 1964; Liu et al., 1989). B cells that survive the affinity maturation process exit the germinal centre either as antibody-secreting plasmablasts that undergo further development into plasma cells or as memory B cells. Memory B cells are long lived cells that have the capacity to mount a rapid and efficient response upon re-stimulation with the target antigen (Liu et al., 1991).
1.2. Antibody Diversification Mechanisms

In order to effectively fight pathogens, B cells require the ability to recognize virtually any antigen that invades the organism. This is achieved through the production of a large immunoglobulin (Ig) repertoire via the process of V(D)J gene rearrangement or Ig gene conversion. Somatic hypermutation is used to increase Ig affinity for the target antigen, while class switch recombination generates antibodies with varying functions (Figure 1.1). These primary and secondary antibody diversification mechanisms will be described in detail below.

1.2.1. V(D)J Recombination

B cells originate in the adult/fetal bone marrow or the fetal liver as the common lymphoid progenitor and undergo numerous genetic and phenotypic changes as they progress through different developmental stages. The defining characteristic of B cell development is the genetic rearrangement of the Ig loci to generate a wide array of antibody specificities from a relatively small gene segments. Each antibody molecule is composed of four polypeptides, consisting of two identical heavy and two identical κ or λ light chains each of which is encoded by a separate locus. The Ig locus is partitioned into gene segments designated as variable (V), diversity (D), and joining (J) genes, which undergo genetic rearrangement in pre-B cells mediated by the recombination activating (RAG) genes to produce a rearranged, coding functional VDJ heavy chain exon and a rearranged coding VJκ
Figure 1.1. AID-induced lesions initiate somatic hypermutation, class switch recombination and Ig gene conversion

AID induces (A) somatic hypermutation, (B) class switch recombination and (C) Ig gene conversion by introducing point mutations into the V region, the switch regions, or the V region downstream of the pseudogenes. Figure is not to scale.
Figure 1.1.

A Somatic Hypermutation

B Class Switch Recombination

C Gene Conversion
or V\(\lambda\) light chain exon (Tonegawa, 1983). Heavy and light chains then pair up to produce functional BCRs expressed on the B cell surface. Immature B cells undergo selection based on the ability of their BCRs to recognize foreign antigen without reacting to host proteins and are permitted to exit the bone marrow and undergo further development in the peripheral lymphoid organs.

V(D)J recombination is initiated by a series of cleavage reactions at conserved sequences located in each V, D, and J gene segment, leading to the ligation of V and J genes in the case of Ig light chain loci and D and J followed by V and DJ genes in the case of Ig heavy chain loci. The cleavage reaction is mediated by the RAG1 and RAG2 enzymes, which form a complex unique to developing lymphocytes (Oettinger et al., 1990; Shatz et al., 1989). The RAG1/2 complex cleaves V, D, and J genes at recombination signal sequences, highly conserved sequence motifs characterized by conserved heptamer and nonamer sequences separated by a nonconserved 12 or 23 nucleotide tract. These motifs are located at the 5' termini of J genes, 3' termini of V genes and both the 5' and 3' termini of D genes where cleavage by RAG1/2 generates double strand breaks terminated by a hairpin loop, which are in turn processed by Artemis (de Villartay, 2002). This cleavage occurs asymmetrically and thus generates DNA ends with single strand overhangs which require further processing, including the addition of nucleotides by the enzyme terminal deoxynucleotidyl transferase (TdT). This generates a blunt double strand break at the terminus of each gene segment which can be joined to another gene segment via the non-homologous end joining pathway. The diversity of antibody specificities is generated through a number of mechanisms: 1) the recombinatorial diversity generated by the random selection, cleavage and joining of V, D and J segments; 2) the junctional diversity that occurs as a result of the addition or deletion of nucleotides at the site of the RAG1/2-
mediated double strand break; and 3) the combinatorial diversity provided by the pairing of heavy and light chains. These processes work together to generate a potential $\sim 10^{11}$ unique antibody specificities that are capable of recognizing and binding virtually any three-dimensional structure.

### 1.2.2. Somatic Hypermutation

Immunoglobulin rearrangement generates a large amount of antibody diversity, but the majority of these antibodies possess only low binding affinity for any given antigen. To increase antibody affinity, activated B cells undergo affinity maturation in germinal centres, a process in which the variable region of the Ig locus undergoes somatic hypermutation to generate small modifications in the antigen binding site. B cells expressing high affinity BCRs are then preferentially selected over those with lower affinities. The somatic hypermutation process involves the introduction of point mutations into the rearranged V region at a very high rate, $\sim 10^{-3}$ mutations per base pair per generation (see Figure 1.1.). Iterative cycles of somatic hypermutation and selection during the immune response lead to the progressive production of high affinity antibodies (Li et al., 2004b). Somatic hypermutation is initiated 150 to 200 base pairs downstream of the Ig promoter and extends to a distance of 1.5 to 2 kilobase pairs, while the C region is spared (Both et al., 1990; Gearhart and Bogenhagen, 1983; Lebecque and Gearhart, 1990; Longerich et al., 2006; Rada et al., 1994; Rada and Milstein, 2001; Winter et al., 1997). The mutagenic process occurs on both productively and non-productively rearranged heavy and light chain loci (Pech et al., 1981; Roes et al., 1989). Defects in the somatic hypermutation
process result in a low affinity immune response and increased susceptibility to bacterial infections.

Somatic hypermutation is initiated by the enzyme activation-induced cytidine deaminase (AID), which deaminates cytosine (C) residues, thereby converting them to uracil (U) and generating G:U lesions (Jacobs and Bross, 2001; Martin et al., 2002; Muramatsu et al., 2000; Muramatsu et al., 1999; Petersen-Mahrt et al., 2002; Poltoratsky et al., 2000; Revy et al., 2000). Since U is not tolerated in DNA, the G:U lesion needs be eliminated, a process accomplished by one of three possible mechanisms leading to three different mutational outcomes (see Figure 1.2.). If the cell enters S phase prior to repair of the lesion, the U will be recognized as a thymine (T) by polymerases δ or ε and an adenine (A) will be inserted into the nascent strand, thus generating G:C to A:T transition mutations. However, transition mutations at G:C basepairs account for only ~30% of the total mutation load and AID-induced lesions produce transversion mutations at G:C basepairs as well as mutations at A:T base pairs, which constitute ~20% and ~50% of the total mutation load, respectively. The presence of U in DNA can also trigger its removal by one of four glycosylases, uracil-N-glycosylase (UNG), single-strand-selective monofunctional uracil DNA glycosylase 1 (SMUG1), T:G mismatch thymine glycosylase (TDG), and methyl-CpG binding domain protein 4 (MBD4), which specialize in U excision and generate abasic sites that are repaired by the base excision repair pathway. In hypermutating B cells, U is removed predominantly by UNG, an action that leads to error-prone repair of the abasic site and generates transversion mutations at G:C base pairs (Bardwell et al., 2003; Di Noia and Neuberger, 2002; Di Noia et al., 2006; Rada et al., 2002b).
Figure 1.2. AID-generated G:U lesions give rise to mutations G:C and A:T base pairs

AID converts cytidine to uridine, generating a G:U lesion, which is then subject to one of three possible mechanisms. If the lesion is not repaired prior to entry into S phase, DNA replication across the U will generate G:U to A:T transition mutations (top right). If the U is removed by UNG and processed by the base excision repair pathway in an error-prone mode, G:C to T:A or C:G transversion mutations will be produced (bottom right). Notably, this pathway also gives rise to a small number of mutations at A:T base pairs. Engagement of the G:U lesion by MutSα, Exo1 and other components of the mismatch repair pathway results in the generation of mutations at A:T base pairs (bottom left). Figure is not to scale.
Figure 1.2.

DNA Replication

Transition mutations at G:C base pairs

APE1/BER → Translesional polymerases (REV1)

Transition and transversion mutations at G:C base pairs; some A:T mutations

DNA Replication

Transition and transversion mutations at A:T base pairs
The G:U lesion is also recognized by the mismatch repair pathway, an evolutionarily-conserved mechanism specialized in repairing base-base mismatches caused by DNA polymerase errors. In hypermutating B cells, processing of G:U lesions by the mismatch repair pathway leads to the production of mutations at A:T base pairs. The mismatch recognition complex MutSα (Msh2-Msh6) engages the G:U lesion and recruits other components of the mismatch repair pathway, including the exonuclease Exo1 which degrades one of the DNA strands in the region of the mismatch. While this normally leads to gap-filling by the high fidelity DNA polymerase δ, hypermutating B cells recruit the error-prone DNA polymerase η to the V-region instead. Polymerase η has a high error rate opposite template A and T nucleotides and thus generates a large number of mutations at A:T base pairs. Absence of Msh2, Msh6, Exo1 or polymerase η results in ~80% reduction in these mutations, with the remaining mutations contributed by the UNG/base excision repair pathway (Bardwell et al., 2004; Martin et al., 2003; Phung et al., 1998; Rada et al., 2004; Rada et al., 1998; Wiesendanger et al., 2000; Zeng et al., 2004; Zeng et al., 2001). The preferential usage of error-prone polymerases in the synthesis of mismatch repair-induced DNA gaps will be discussed further in section 1.4.6. as well as in Chapters 2 and 3.

1.2.3. Class Switch Recombination

Class switch recombination or isotype switching is the process whereby the constant region of an antibody is replaced by another constant region, while maintaining the same antigenic specificity dictated by the V region. While all activated B cells begin by expressing IgM, class-switching to IgG, IgA, or IgE occurs during the course of the immune response and is determined by the type of signal received through the BCR and the combination of
cytokines received by the B cell. Each isotype is encoded by its own C region gene located
downstream of the V region, with IgM and IgA flanking the 5′ and 3′ ends, respectively, and
is preceded by a promoter, an intronic enhancer and a switch region (see Figure 1.3). Switch regions are highly repetitive G-rich, non-coding sequences ranging in length from 1
to 10kb and are targets for AID activity, which leads to the production of DNA double
strand breaks that mediate the switching reaction (Chaudhuri and Alt, 2004). Cytokine-
mediated intracellular signaling induces the transcription of sterile, non-coding transcripts
from the appropriate C promoter, which serves to open up the switch locus and make the
DNA accessible to DNA processing enzymes that lead to the switching of one C region gene
for another (Chaudhuri et al., 2007; Manis et al., 2002b; Stavnezer, 1996). Expression of IgD
is normally mediated by alternative splicing of a single VDJ-Cμ-Cδ transcript. However, a
small number (1%) of B cells class switch to IgD using a cryptic switch sequence located in
the Cμ-Cδ intron (White et al., 1990; Yasui et al., 1989). These cells are derived from
germinat centres and have undergone SHM (Liu et al., 1996; Wilson et al., 2000). Analysis of
IgM-IgD+ switched cells led to the identification a recombinational event within the Cμ-Cδ
intron and the presence of G-rich pentameric repeat tracts characteristic of switch
sequences in the μ-δ intronic region (Kluin et al., 1995), suggesting that rare IgD class
switching events do occur. Importantly, secretion of IgD has been implicated in the
regulation B cell homeostasis and activation (Chen et al., 2009) and IgD-deficient mice have
diminished peripheral B cell compartments (Nitschke et al., 1993; Schiemann et al., 2001).
IgD binds to basophils and these cross-linked cells may serve to trigger rapid innate and
adaptive immunity after sensing pathogens in the upper respiratory tract (Chen et al.,
2009).
As with somatic hypermutation, class switch recombination is initiated by AID, which deaminates cytosines in the upstream switch regions of the expressed and targeted C region genes (see Figure 1.1.) (Honjo et al., 2002; Muramatsu et al., 2000; Muramatsu et al., 1999). One of the hallmarks of the switch regions is a high concentration of C residues located within AID’s preferred target sequence, which enables AID to deaminate C’s on opposite strands in close proximity (see Section 1.2.5.) Processing of closely-spaced G:U lesions on opposite DNA strands by UNG and the apurinic endonuclease APE1 or by Msh2-Msh6 and Exo1 generates double strand breaks in both switch regions. This induces the DNA damage response and the recruitment and activation of factors involved in DNA double strand break repair, including γH2AX, Ku70, Ku80, and DNA-PKcs. The DNA ends of distal switch regions are then joined by the non-homologous end joining pathway or other DNA repair pathways by ligating the 5’ terminus of the upstream switch region to the 3’ terminus of the downstream switch region. Concurrently, the intervening DNA region is eliminated, enabling the expression of an Ig locus with a new isotype (Honjo et al., 2002).

1.2.4. Immunoglobulin Gene Conversion

V(D)J recombination generates a large antibody repertoire in mice and humans, but birds and certain mammals, including cows, pigs, and rabbits, contain a very small number of highly related heavy and light chain V, D, and J segments, which are rearranged to produce a small array of functional, but poorly diversified, antibody specificities. Instead, these species use secondary antibody diversification mechanisms to produce a large antibody repertoire (Ratcliffe, 2006; Reynaud et al., 1991; Reynaud et al., 1989). The rearranged transcribed V region is located downstream of a series of V region pseudogenes, which
Figure 1.3. Organization of the murine C region genes.

(A) The C region genes are located downstream of the V region. Each C region gene, with the exception of Cδ, is preceded by a repetitive switch region (S) and a promoter. (B) The correct combination of signals (squares and circles) activates transcription from the appropriate promoter and opens up the locus, allowing AID to gain access to ssDNA in the switch region. RNA transcript indicated by dashed lines; non-transcribed DNA strand indicated by solid black arched line. (C) The 5’ region of Sμ fuses with the 3’ region of Sγ1, thus permitting expression of the Cγ1 isotype. The intervening DNA forms a switch circle, which is subsequently eliminated. Figure is not to scale.
Figure 1.3.
contain 10 to 20% sequence divergence from the transcribed V region. The pseudogenes lack the conserved sequences recognized by the RAG1/2 complex, thereby inhibiting their participation in Ig rearrangement, lack upstream promoters, or contain stop codons, thus preventing their transcription. Antibody diversity in these organisms is created by the non-reciprocal replacement of a portion of the transcribed V region by a portion of an upstream pseudogene via Ig gene conversion, a form of homologous recombination (HR) (Reynaud et al., 1987; Reynaud et al., 1994; Sale, 2004).

Ig gene conversion requires the introduction of a double strand DNA break within the V region, strand invasion of a homologous V pseudogene and DNA synthesis off the pseudogene template (Helleday, 2003; Tang and Martin, 2006). Repair of the broken DNA strand by copying the pseudogene template DNA diversifies the sequence of the transcribed V gene. Absence of AID or UNG abolishes Ig gene conversion (Arakawa et al., 2002; Di Noia and Neuberger, 2002; Harris et al., 2002), indicating that processing of the AID-generated G:U lesion into an abasic site contributes to this diversification mechanisms, possibly by generating closely spaced single strand breaks on opposite strands which are converted to double strand breaks (Cook et al., 2007; Tang and Martin, 2006). Alternatively, a single strand break created by G:U lesion processing can be converted into a double strand break during DNA replication and replication fork collapse (Saleh-Gohari et al., 2005). Proteins involved in HR such as Rad51 and Rad51 paralogues are also critical for Ig gene conversion (Sale et al., 2001; Sung and Klein, 2006). The involvement of mismatch repair proteins has not been well-documented, but is expected to participate in the diversification process since the annealing of homologous DNA strands generates base-base mismatches which need to be resolved.
1.2.5 Activation-Induced Cytidine Deaminase

AID was first identified in the mouse B cell lymphoma line CH12F3-2 and was subsequently shown to be required for somatic hypermutation, class switch recombination and Ig gene conversion and to induce antibody diversification mechanisms when ectopically expressed in B and non-B cell lines (Arakawa et al., 2002; Bardwell et al., 2002; Martin and Scharff, 2002; Muramatsu et al., 2000; Muramatsu et al., 1999; Nakamura et al., 1996; Okazaki et al., 2002; Revy et al., 2000; Yoshikawa et al., 2002). AID is a member of the cytidine deaminase family and was initially thought to act on RNA based on its amino acid homology to apolipoprotein B RNA editing catalytic component 1 (APOBEC1), an RNA cytidine deaminase. However, genetic and biochemical analyses have unequivocally demonstrated that AID acts directly on DNA (Beale et al., 2004; Bransteitter et al., 2003; Chaudhuri et al., 2003; Dickerson et al., 2003; Petersen-Mahrt et al., 2002; Pham et al., 2003; Ramiro et al., 2003; Sohail et al., 2003). AID deaminates deoxycytidines in single-stranded, but not double-stranded DNA and preferentially acts on the WRG motif, where W=A or T and R=A or G (Larijani et al., 2005; Neuberger and Milstein, 1995; Pham et al., 2003).

The 198 amino acid, 24 kDa AID protein is expressed exclusively in activated B cells and is found predominantly in the cytoplasm, with only a small percent of the protein localized to the nucleus (Rada et al., 2002a). The enzyme is organized into four domains: 1) an N-terminal nuclear localization signal; 2) an α-helical segment important for dimerization; 3) a central catalytic domain; and 4) a C-terminal nuclear export signal (Brar et al., 2004; Ito et al., 2004; McBride et al., 2004). Both the catalytic and dimerization domains are required for AID function (Wang et al., 2006). Truncation of the C-terminus abolishes class switch recombination, but has no effect on somatic hypermutation or Ig gene conversion, while the N-terminus has been shown to be important in somatic
hypermation but not class switch recombination (Barreto et al., 2003; Revy et al., 2000; Shinkura et al., 2004; Ta et al., 2003; Zhu et al., 2003). It has been suggested that the C-terminus mediates interaction with other cofactors involved in class switch recombination, but the identity of these is not known.

AID-mediated cytidine deamination is largely restricted to the Ig locus, but has also been observed at some non-Ig genes, albeit at mutation rates that are 50- to 100-fold lower than those at the V and switch regions (Gordon et al., 2003; Muschen et al., 2000; Pasqualucci et al., 1998; Shen et al., 1998). The misdirected targeting of AID to non-Ig genes has the potential to introduce deleterious mutations and/or translocations that lead to tumorigenesis (Franco et al., 2006; Gordon et al., 2003; Kuppers and Dalla-Favera, 2001; Pasqualucci et al., 2003; Ramiro et al., 2006). Numerous studies have demonstrated using Ig and artificial substrates that transcription is a pre-requisite for AID activity, potentially as a means of opening up the locus (Bachl et al., 2001; Bottaro et al., 1994; Fukita et al., 1998; Peters and Storb, 1996; Winter et al., 1997; Yang et al., 2006) and it has been postulated that AID may be brought to its substrate by the transcription machinery (Longerich et al., 2006; Peters and Storb, 1996). The Ig promoter does not play a unique role in somatic hypermutation since replacement with a heterologous promoter supports the mutational process (Bachl et al., 2001; Betz et al., 1994; Martin et al., 2002; Papavasiliou and Schatz, 2000; Parsa et al., 2007; Tumas-Brundage and Manser, 1997; Yelamos et al., 1995). Nevertheless, not all heterologous promoters that induce high levels of transcription are able to support AID activity, indicating that the Ig promoter provides some as yet unknown support for AID targeting (Yang et al., 2006). AID has been shown to
Figure 1.4. Structural and functional properties of AID.

A schematic representation of the motifs of AID and regions important for antibody diversification. NLS: nuclear localization signal; DM: dimerization Motif; NES: nuclear export signal; SHM: somatic hypermutation; CSR: class switch recombination. Figure is not to scale.
Figure 1.4.
interact with replication protein A (RPA), a single-stranded DNA binding protein and MDM2, a ubiquitin ligase that targets p53 for degradation (MacDuff et al., 2006) and is thought to interact with other unidentified cofactors that may help recruit AID to its target genes.

The regulation of AID activity is also poorly characterized. AID is likely to be restricted at the entrance to the nucleus since the majority of the protein is located in the cytoplasm (Ito et al., 2004; McBride et al., 2004). Moreover, AID is regulated at the post-transcriptional and –translational level (Basu et al., 2005; Chaudhuri et al., 2004; Pasqualucci et al., 2006). Recent evidence suggests that AID is suppressed by the microRNA miR-155 (Dorsett et al., 2008; Teng et al., 2008; Thai et al., 2007; Vigorito et al., 2007). Additionally, AID has been shown to be phosphorylated at serine-38 by protein kinase A (PKA) and that phosphorylated AID is able to interact with RPA. This phosphorylation and association with RPA seems to be necessary for deamination of the switch region and for enabling class switch recombination (Basu et al., 2005; McBride et al., 2006; Pasqualucci et al., 2006; Pham et al., 2008).
1.3. DNA Repair Mechanisms

The DNA of organisms is under constant attack from both exogenous and endogenous DNA damaging agents which cause a variety of DNA errors, including point mutations, single-stranded and double-strand DNA breaks, translocations, and chromosomal deletions. Whether damage is induced by UV-irradiation, chemical mutagens, AID, or the byproducts of oxidative metabolism, cells have evolved sophisticated and elaborate mechanisms to repair the lesions. A brief overview of homologous recombination, non-homologous end joining, base excision repair, and nucleotide excision repair will be provided in the following sections, while the Mre11-Rad50-Nbs1 complex and translesion synthesis will be examined in greater detail as these are pertinent to the discussion on antibody diversification mechanisms presented in this thesis. The mismatch repair pathway will be reviewed in its own section as it will be discussed in greater depth.

1.3.1. Double Strand Break Repair

The deleterious consequences of DNA damaging agents can take many forms, but DNA double strand breaks (DSBs) are possibly the most disastrous to an organism because they can result in large genomic aberrations such as the unequal distribution of chromosomes during cell division or chromosomal translocations. If such catastrophic damage is left unrepaired, it can cause cell death, or worse, the cancerous modification of tumor suppressor genes and oncogenes (Agarwal et al., 2006; Hoeijmakers, 2001). DSBs can be caused by numerous endogenous and exogenous agents. Ionizing radiation, replication fork collapse due to synthesis across damaged DNA template, reactive oxygen species generated
during cellular metabolism, gene rearrangement of the T- and B-cell receptors, and AID-induced lesions all produce DSBs (Dudley et al., 2005; Livak, 2004; Paulsen and Cimprich, 2007; Ward, 1985). When the DNA backbone is severed on both DNA strands in close proximity to one another, a DSB is formed. To circumvent the damage caused by DSBs, organisms have evolved a sophisticated network of cell cycle checkpoint factors and repair mechanisms that work together to repair the DNA damage before the next cell division begins (Valerie and Povirk, 2003; Zhou and Elledge, 2000). The roles of homologous recombination, non-homologous end joining and the Mre11-Rad50-Nbs1 complex in the response to DSBs will be discussed below.

1.3.1.1. Homologous Recombination

Broken DNA strands can be repaired by HR using regions of sequence similarity in DNA such as sister chromatids, homologous chromosomes, or repeated regions on the same or different chromosomes. Some organisms such as yeast repair DSBs using predominantly HR, while mammalian cells utilize HR predominantly in late S/G2 phase when homologous sister chromatid template is available (Barlow et al., 2008; Ward, 1988). The importance of this pathway to Ig gene conversion was discussed earlier in section 1.2.4. Higher eukaryotes possess a more elaborate set of HR enzymes than do yeast, possibly due to the requirement of HR for the restart of blocked or collapsed replication forks, a process required much more in higher eukaryotes due to the larger genome size (Shrivastav et al., 2008). Although HR is much more accurate than other DNA repair processes such as non-homologous end joining, the use of a non-identical homologous template results in reduced heterozygosity, and therefore decreased genetic diversity.
The presence of broken DNA ends is sensed by the Mre11-Rad50-Nbs1 complex which initiates the repair process. First, the end of the broken DNA molecule is resected in the 5'‐to‐3' direction by C-terminal region of adenovirus E1A interacting protein (CtIP), followed by more extensive resection by BLM or Exo1 to produce 3' overhangs (Clerici et al., 2005; Hickson, 2003; Krogh and Symington, 2004; Lengsfeld et al., 2007; Mimitou and Symington, 2008; White and Haber, 1990; Zhu et al., 2008). The newly-formed ssDNA is coated by the ssDNA-binding protein RPA, which signals the presence of DNA damage to cell cycle checkpoint factors. This is followed by interaction with Rad51 to mediate the homology search and strand invasion, leading to formation of a double Holliday junction (Ira and Haber, 2002; Liu et al., 2004; Sugiyama et al., 2006; Sung, 1994; Wold, 1997). The 3' end of the invading strand is used as a primer to begin synthesis off of the template strand. Once synthesis is complete, the double Holliday junction dissolves (Plank et al., 2006; Wu et al., 2006; Wu and Hickson, 2003).

1.3.1.2. Non-Homologous End Joining

DSBs formed during the G1 or early S phases are largely repaired by the non-homologous end joining (NHEJ) pathway. NHEJ is generally less accurate than HR because it involves the direct ligation of compatible ends, which often require end processing, resulting in the loss or addition of genetic material. This pathway involves an extensive network of proteins that bind DSBs and bring distal DNA ends together for ligation. This process is used to ligate DNA ends during V(D)J rearrangement and class switch recombination. Within seconds after a DSB is incurred, the broken DNA ends are bound by the Ku70/Ku80 heterodimer, an open ring structure with high affinity for DNA ends (Walker et al., 2001).
The Ku70/80-DNA complex forms a scaffold for the assembly of other proteins, including the DNA dependent protein kinase catalytic subunit (DNA-PKcs), which forms a bridge between the two broken DNA ends (DeFazio et al., 2002; Rivera-Calzada et al., 2007; Spagnolo et al., 2006). Formation of the Ku70/80-DNA-PKcs complex activates the Ser/Thr kinase activity of DNA-PKcs, resulting in autophosphorylation on minimally 16 amino acid residues, followed by a conformational change that allows DNA processing enzymes to access the broken DNA ends (Chan et al., 2002; Chen et al., 2005; Cui et al., 2005; Ding et al., 2003; Douglas et al., 2007; Soubeyrand et al., 2003; Uematsu et al., 2007; Weterings and Chen, 2007). Activated DNA-PKcs also phosphorylates other substrates, including XRCC4, Ku70/80, and p53.

DNA end processing, either by resection or by fill-in synthesis, is necessary because most DSBs contain 3' or 5' ssDNA overhangs and are therefore not directly compatible for ligation. Resection of 3' or 5' overhangs is performed by Artemis, a 5'-3' exonuclease that also possesses DNA-PKcs-dependent endonuclease activity (Ma et al., 2002). Templated and non-templated nucleotides can be added by TdT and polymerases λ or μ, which interact with Ku70/80 to stimulate NHEJ (Lee et al., 2004; Mahajan et al., 1999; Mahajan et al., 2002; Mickelsen et al., 1999; Nick McElhinny et al., 2005). Before ligation can proceed, polynucleotide kinase (PNK) adds 5' phosphate groups, while AP endonuclease (APE1), PNK, Artemis, or tyrosyl-DNA phosphodiesterase (TDP1) remove 3' phosphoglycolates. Compatible DNA ends are then sealed by the combined action of DNA ligase IV, XRCC4 and Cernunos/XRCC4-like factor (XLF) through interactions with the Ku70/80 complex (Costantini et al., 2007; Dai et al., 2003; Hsu et al., 2002; Lee et al., 2000; Nick McElhinny et al., 2000; Pacher et al., 2007; Sekiguchi and Ferguson, 2006).
1.3.1.3. **Mre11-Rad50-Nbs1 Complex**

Before the decision to repair DSBs using HR or NHEJ is made, the presence of DSBs is detected by the Mre11-Rad50-Nbs1 (MRN) complex. The MRN complex is a sensor of DNA damage and activator of the cell cycle checkpoint response and is involved in numerous vital cellular pathways, including DNA repair, DNA replication, telomere maintenance, and cell cycle signaling (Bosch et al., 2003; D'Amours and Jackson, 2002; Mirzoeva and Petrini, 2003). Null mutations in any of the components of the MRN complex result in early embryonic lethality in mice, while humans with hypomorphic mutations in Mre11 or Nbs1 are affected with ataxia telangiectasia-like disorder (ATLD) or Nijmegen breakage syndrome (NBS), respectively, characterized by immunodeficiency, radiation sensitivity, cell cycle checkpoint defects, chromosomal translocations involving the immunoglobulin locus, and increased cancer susceptibility (Carney et al., 1998; Luo et al., 1999; Shiloh, 1997; Stewart et al., 1999; Taylor et al., 2004; Varon et al., 1998; Xiao and Weaver, 1997; Young and Painter, 1989; Zhu et al., 2001). In undamaged cells, the MRN complex is uniformly distributed throughout the nucleus (Mirzoeva and Petrini, 2003). However, upon induction of DNA damage, the MRN complex rapidly recognizes and binds DNA DSBs, after which it recruits ATM, which in turn phosphorylates the MRN complex as well as more than 30 downstream substrates involved in the response to DSBs (Bakkenist and Kastan, 2003; Kozlov et al., 2003; Kurz and Lees-Miller, 2004; Matsuoka et al., 2007). Failure to recruit ATM results in impaired cell cycle arrest in response to DNA damage.

Mre11 forms the core of the MRN complex and interacts with DNA, other Mre11 molecules, Rad50, and Nbs1. It possesses endonuclease activity that is stimulated by Rad50 or Nbs1, exonuclease activity that is stimulated by Rad50, and limited DNA unwinding
activity when associated with both Rad50 and Nbs1 (Assenmacher and Hopfner, 2004; Furuse et al., 1998; Moncalian et al., 2004; Paull and Gellert, 1998; Paull and Gellert, 1999; Trujillo and Sung, 2001). The nuclease and DNA binding activities of Mre11 are structure-specific, with high DNA-binding affinity for the ends of linear DNA molecules. The ability of Mre11 to mediate various nuclease activities led to suggestions that it might participate in mismatch repair and SHM. This topic will be explored further in Chapter 4. Rad50 is an ATPase-containing protein that assumes an elongated architecture in order to bridge the gap between two broken DNA molecules. Nbs1 is the regulatory subunit of the complex and is instrumental in the recruitment and activation of ATM.

In response to a DSB, MRN rapidly associates with broken DNA ends, spanning up to 10kb from the break site, in order to bridge the gap between the broken DNA molecules (de Jager et al., 2001; Hopfner et al., 2001; Hopfner et al., 2000; Lisby et al., 2004; Maser et al., 1997; Shroff et al., 2004). Each DNA end is bound directly by two globular Mre11 dimers, which also form contacts with two Rad50 molecules such that DNA is bound by an M2R2 tetramer complex (see Figure 1.5) (Bosch et al., 2003; de Jager et al., 2001). The base of each Rad50 molecule contains two ATPase domains (Walker A and Walker B motifs), Mre11-binding sites and an elongated coiled-coil domain that extends away from the base and possesses a Zn2+-binding hook structure at its terminus. This hook structure is tethered to the end of another Rad50 molecule which participates in a second M2R2 complex located at the other DNA end (Cahill and Carney, 2007; Hopfner et al., 2002; Lobachev et al., 2004; Wiltzius et al., 2005). Shortly after the MRN complex binds DNA, an ATM dimer is recruited to the break site, autophosphorylates and dissociates into a monomeric form (Bakkenist and Kastan, 2003; Berkovich et al., 2007; Kozlov et al., 2006; Lee and Paull, 2004; Lee and Paull, 2005; Shiloh, 2006; You et al., 2005). Activated ATM can then phosphorylate the MRN
complex, which is necessary for the MRN- and ATM-mediated recruitment and activation of various cell cycle checkpoint factors, including p53, H2AX, BRCA1, 53BP1, Chk2, and many others. Mre11 and the chromatin remodeling protein INO80 evict H2B and H3 histones from nucleosomes, thereby exposing ssDNA (Shen et al., 2000; Usui et al., 2006). Although Mre11 foci co-localize with γ-H2AX in the nucleus, γ-H2AX molecules are located further from the break site. In summary, the MRN complex participates in the DNA damage response by rapidly sensing the presence of DSBs, forming a bridge in between DNA ends, and recruiting and activating ATM. ATM in turn activates downstream DNA damage response factors that result in cell cycle arrest at the G1/S, intra-S, or G2/M phases.
Figure 1.5. The Mre11-Nbs1-Rad50 (MRN) complex senses double strand breaks.

Shortly after a double strand break is induced, Mre11 (M) binds directly to broken DNA ends together with Nbs1 and the ATP-binding domains (A, B) of Rad50. The coiled coil domains of Rad50 bridge and tether two broken DNA molecules, while Nbs1 (N) assists in the phosphorylation and activation of ATM. Activated ATM is able to recruit downstream factors involved in cell cycle checkpoint signaling and double strand break repair. Figure is not to scale.
Figure 1.5.

DNA Repair Factors
\( \gamma\)-H2AX, INO80, Ku70/80, DNA-PK\(_{c}\)  

Cell Cycle Checkpoint Factors
Chk2, p53, 53BP1, BRCA1
1.3.2. Base Excision Repair

DNA damage produced by oxidation, alkylation, or deamination generates modified bases that need to be repaired because they form non-Watson-Crick base pairs and ultimately lead to point mutations. Oxidated guanine, for example, preferentially pairs with A rather than with C, yielding G:C to T:A transversion mutations, while cytosine deamination, either spontaneous or AID-induced, produces G:U base pairs that lead to G:C to A:T transition mutations. The base excision repair (BER) pathway repairs this type of damage using one of nine glycosylases, each of which is uniquely adapted to excise a particular type of damaged base. BER is initiated by the recognition and excision of a damaged or modified base, which is accomplished by extrahelical flipping of the base into a lesion-specific recognition pocket (Huffman et al., 2005; Parker et al., 2007; Slupphaug et al., 1996). This generates an abasic site, which is cleaved by APE1 (Friedberg et al., 2006; Mitra et al., 1997). The excision of AID-generated uracils is required for the production of G:C transversion mutations in the Ig locus (Di Noia and Neuberger, 2002; Rada et al., 2002b). Its additional participation in A:T mutagenesis will be explored in Chapters 2 and 3.

Subsequent processing of the abasic site proceeds via either single-nucleotide (SN) or long patch (LP) BER, as determined by the characteristics of the APE1-cleavage product. If the cleavage product contains a 3’ hydroxyl (OH) and 5’ deoxyribose phosphate (dRP) terminus, BER will proceed via the SN pathway (Matsumoto and Kim, 1995; Piersen et al., 1996; Prasad et al., 1998). This involves cleavage of the dRP group by the dRP lyase activity of polymerase β to generate a 5’ phosphate and filling in of the single nucleotide gap by DNA polymerase β. If the cleavage product contains a 5’ terminus that cannot be removed by polymerase β, such as a 3’ phospho α,β-unsaturated aldehyde or 3’ and 5’ phosphates at the strand break, it will proceed along LP-BER (Frosina et al., 1996; Klungland and Lindahl,
During LP-BER the lesion is removed by the 5’ flap endonuclease activity of Flap structure specific endonuclease 1 (FEN-1). FEN-1 is also a 5’ exonuclease (Guo et al., 2008; Singh et al., 2007) that degrades a short region of the DNA backbone, thereby generating a 2- to 8-nucleotide gap that is subsequently filled in by a DNA polymerase, most likely polymerase β. The final step in BER involves ligation of the nick and is mediated by ligase I (Liu et al., 2005).

1.3.3. **Nucleotide Excision Repair**

DNA lesions that distort the DNA helix such as UV-induced photolesions, bulky DNA adducts produced by chemicals, or DNA-protein complexes induced by topoisomerase I inhibitors are repaired by the nucleotide excision repair (NER) pathway (Desai et al., 2003). The NER pathway is divided into two subpathways: global genome NER and transcription coupled NER. Global genome NER repairs lesions throughout the genome and its repair efficiency is determined by the chromatin environment rather than the DNA strand (Feng et al., 2003; Mullenders and Berneburg, 2001). On the other hand, transcription coupled NER repairs DNA lesions in transcribed strands and enables the resumption of stalled transcription machinery at sites of DNA damage (Bohr et al., 1985; Mellon et al., 1987). After the initial damage recognition step, NER proceeds via a “cut-and-patch” mechanism that involves the excision and removal of a short patch of DNA containing the lesion - typically 25-30 nucleotides - resynthesis using the non-damaged strand as the template, and ligation (O’Donovan et al., 1994; Gillet et al., 2006). Gap-filling and ligation involve DNA polymerases δ and ε, PCNA, RFC, and DNA ligase I or the DNA ligase III-XRCC1 complex (Aboussekhra et al., 1995; Araujo et al., 2000; Kelman, 1997;
Moser et al., 2007). These steps are the same whether global genome NER or transcription couple NER is being used.

1.3.4. Translesion Synthesis

The human genome consists of $3 \times 10^9$ nucleotides, all of which need to be accurately replicated in each cell cycle. High fidelity in replication is achieved through the combined action of accurate DNA polymerases, DNA proofreading, and the mismatch repair pathway. The accuracy of replicative DNA polymerases $\alpha$, $\delta$, $\varepsilon$, and $\gamma$, which are responsible for the bulk of DNA synthesis in eukaryotes, is rooted in their stringent base selectivity and is brought about by a restrictive nucleotide binding pocket that limits insertion/deletion or base substitution errors to a frequency of $10^{-4}$ nucleotide insertion events (Einolf and Guengerich, 2000; Hashimoto et al., 2003; Kunkel et al., 1987; Thomas et al., 1991). Replicative polymerases possess 3′-to-5′ exonuclease activity, giving them the ability to carry out proofreading functions and increasing the fidelity of replication to $\sim 10^{-7}$ nucleotide insertion events. The mismatch repair pathway contributes to replication accuracy by another 50- to 1000-fold, bringing the total error frequency to $\sim 10^{-10}$ nucleotides. Each day, DNA polymerases encounter up to 30,000 damaged or altered bases that block the replication machinery and stall replication (Friedberg, 2005; Lindahl and Barnes, 2000). These lesions have the potential to become DSBs if stalled replication forks cannot restart due to DNA licensing constraints once the replication machinery has moved on. Thus, cells have evolved a DNA damage tolerance (DDT) mechanism which they use to repair or move past replication-blocking lesions, sometimes with associated genomic instability (Andersen et al., 2008).
The regulation of DDT is mediated by PCNA, the DNA processivity clamp and scaffolding protein involved in DNA unwinding, synthesis, cell cycle progression, and chromatin structure maintenance (Moldovan et al., 2007). In response to DNA damage, PCNA undergoes post-translational modification on lysine residues by ubiquitin (Ub) or the Ub-like protein SUMO (small ubiquitin modifier), leading to three potential outcomes: error-prone translesion synthesis, error-free DDT, or suppression of inappropriate homologous recombination. Ub is a 76 amino acid protein that is covalently bound to the ε amino group of lysine on target proteins via a three-step process, involving thioester linkage to a Ub-activating enzyme (E1 or Uba), transfer to the active site Cys residue of a Ub-conjugating enzyme (E2 or Ubc), and conjugation of Ub to the target protein Lys residue and isopeptide bond formation by a Ub-ligase (E3). The first Ub linkage is often followed by the formation of poly-Ub chains, the best characterized of which are K63-linked poly-Ub chains that target the proteasome-dependent degradation of substrate proteins (Hochstrasser, 1996a; Hochstrasser, 1996b; Pickart and Cohen, 2004; Pickart and Eddins, 2004; Pickart and Fushman, 2004). SUMO is part of a class of Ub-like molecules, adopts a Ub-like fold with conserved positioning of C-terminal residues for isopeptide bond formation and employs its own E2-E3 conjugating-ligase complex for target conjugation, thereby altering the activity of substrate proteins (Bayer et al., 1998; Sheng and Liao, 2002).

In *Saccharomyces cerevisiae*, proteins in the Rad6 epistasis group, including Rad6, Rad18, Rev1 and DNA polymerase ζ (Rev3 and Rev7) are involved in PCNA modification and DDT. Yeast with mutations in these proteins exhibit increased sensitivity to various DNA damaging agents, often with reduced mutagenesis, suggesting that they are involved in the error-prone DNA damage bypass pathway (Prakash et al., 1993). Rad6 is an E2
conjugating enzyme that forms a complex with Rad18, an E3 ligase with ssDNA-binding and ATPase activities that interacts with and modifies PCNA (Bailly et al., 1994; Hoege et al., 2002; Lovering et al., 1993; Saurin et al., 1996). The Rad6-Rad18 complex monoubiquitinates PCNA on K164 and mutation of the K164 residue results in defects in post-replication repair (Gay and Gillespie, 2005; Hoege et al., 2002; Torres-Ramos et al., 1996). An error-free subpathway was identified with the characterization of the Mms2-Ubc13-Rad5 complex which catalyzes the formation of K63-linked poly-Ub chains on PCNA and mutations of either Mms2 or Ubc13 result in a massive increase in spontaneous mutagenesis (Brusky et al., 2000; Hofmann and Pickart, 1999; Johnson et al., 1992; Johnson et al., 1994). This error-free DDT mechanism is thought to occur through template switching or replication fork regression (Blastyak et al., 2007; Branzei and Foiani, 2007a; Branzei and Foiani, 2007b; Broomfield et al., 2001; Cobb and Bjergbaek, 2006; Klein, 2006). The conjugation of SUMO onto PCNA K164 is dependent on the Ubc9-Siz3 complex (Hoege et al., 2002; Stelter and Ulrich, 2003). SUMOylated PCNA recruits SRS2, a 3’-to-5’ DNA helicase that disrupts Rad51-ssDNA filaments and prevents inappropriate homologous recombination, thereby suppressing DDT defects. (Aboussekhra et al., 1989; Aguilera and Klein, 1988; Broomfield and Xiao, 2002; Chanet et al., 1996; Haracska et al., 2004; Kaytor et al., 1995; Milne et al., 1995; Papouli et al., 2005; Pfander et al., 2005; Rong and Klein, 1993; Rong et al., 1991; Schild, 1995; Ulrich, 2001). PCNA can also be SUMOylated on K127 and this modification is required for sister chromatid cohesion during S-phase (Moldovan et al., 2006).

A class of specialized polymerases termed translesional polymerases has the capacity to bypass bulky or uninformative bases by incorporating nucleotides opposite these lesions, often at the expense of overall accuracy. Well before their discovery, error-
prone synthesis was proposed to participate in the SHM process (Brenner and Milstein, 1966). It is now established that a number of translesional polymerases contribute to the mutation spectrum, with polymerase η taking centre-stage in the generation of A:T mutations (Zeng et al., 2001). Translesion polymerases in mammals are classified into two families: the Y family contains polymerases η, κ, ι, and Rev1, while the B family contains polymerase ζ in addition to replicative polymerases. The error rates of translesional polymerases are much higher than those of replicative polymerases and range from $10^{-1}$ to $10^{-4}$ nucleotides. The higher error rates are due to the lack of 3’-to-5’ proofreading activity and unrestrained nucleotide selectivity, achieved through a non-restrictive nucleotide binding pocket that makes few contacts with the template base and incoming nucleotide, the use of alternative, non-Watson-Crick basepairing, and the failure to undergo an induced fit conformation upon nucleotide binding (Goodman, 2002; Kunkel, 2004; Nair et al., 2005a; Nair et al., 2005b; Prakash et al., 2005; Yang, 2003; Yang, 2005). As a result, translesional polymerases efficiently bypass damaged DNA bases and extend from mismatched or damaged primer termini, but contribute to elevated error rates on undamaged template.

Some Y family polymerases are specialized in synthesizing opposite specific types of damage. Polymerase η, one of the least accurate Y family polymerases, efficiently inserts dAMP opposite thymine-thymine dimers (TTDs) generated by exposure to ultraviolet (UV) radiation (Washington et al., 2001). However, polymerase η frequently inserts dGMP opposite the template T at the 3’ position of the TTD as well as opposite undamaged template at a rate of $\sim 10^{-2}$ nucleotides (McCulloch et al., 2004; McCulloch et al., 2007). Rev1 is a dCMP transferase and inserts dCMP opposite template Gs and abasic sites, while
polymerase ζ is efficient at extending from distorted basepairs and as such makes a significant contribution to the overall level of mutagenesis (Lawrence, 2004; Prakash et al., 2005). The contribution of polymerases η and Rev1 to SHM will be discussed in Chapter 2, while the roles of other polymerases will be explored in Chapter 5.

Translesion polymerases possess PCNA-, Ub-, and Rev1-interacting domains, serving to enhance the interaction with ubiquitinated PCNA as well as Rev1, which is thought to play a structural role in assembling translesional polymerases at sites of DNA damage. All Y family polymerases and polymerase ζ contain PCNA-interacting peptide (PIP) boxes and Ub-binding domains, in the form of Ub-binding motifs (UBMs) or Ub-binding zinc fingers (UBZs), while Rev1 interacts with PCNA and ubiquitinated PCNA through its BRCT domain and two UBMs (Guo et al., 2006a; Guo et al., 2006b; Kannouche and Lehmann, 2004; Kannouche et al., 2004). They also have Rev1-binding domains, while Rev1 mediates interactions with other polymerases through its C-terminal region (Guo et al., 2003; Murakumo et al., 2001; Ohashi et al., 2004; Tissier et al., 2004). The importance of Rev1 as a structural component is demonstrated by the decreased mutagenesis in cells with reduced Rev1 protein levels, but not in cells with the catalytically-inactive mutant, suggesting that Rev1 enzymatic activity is not required for translesion synthesis (Guo et al., 2003; Haracska et al., 2001; Otsuka et al., 2005; Ross et al., 2005).

Translesion synthesis is thought to proceed through one of two, not necessarily mutually exclusive models. The polymerase switch model acts at the replication fork after DNA damage induces fork stalling and enables replication to continue, while the gap-filling model operates outside the context of the replication fork to seal gaps that have resulted from replication fork repriming downstream of the lesion (Rupp and Howard-Flanders,
In the polymerase switch model, the presence of a replication fork blocking lesion generates ssDNA, which is coated by Rad18 and is followed by monoubiquitination of PCNA by the Rad6/Rad18 complex (Bailly et al., 1994; Bailly et al., 1997; Hoege et al., 2002). PCNA monoubiquitination results in a switch from a replicative polymerase to an appropriate translesion polymerase and is proceeded by a second switch to the “extender” polymerase ζ or perhaps κ (Prakash et al., 2005; Prakash and Prakash, 2002). A final polymerase switch reinstates the replicative polymerase at the primer terminus. The gaps left behind by the replication machinery are likely to be relatively short, and as such a polymerase switch may not be necessary in the gap-filling situation. Nevertheless longer gaps, as have been observed after UV irradiation, would require two to three polymerase switch events (Lopes et al., 2006). The role of PCNA ubiquitination in these two models is currently under debate. In experiments performed in the DT40 cell line, Rev1 was required to maintain fork progression on damaged DNA, whereas PCNA-ubiquitin was dispensable (Sale et al., 2009). On the other hand, PCNA-ubiquitin was required for post-replicative gap filling, suggesting that a spatio-temporal separation of PCNA- or Rev1-assisted translesion synthesis may occur (Edmunds et al., 2008). Moreover, repair of MMR-generated gaps during SHM also resembles the gap-filling model in that ubiquitinated PCNA is required to generate A:T mtutations (see Section 1.4.6. and Chapter 2). In addition to PCNA and Rev1, a third mechanism involving the alternative 9-1-1 DNA clamp may regulate the DNA damage response, as indicated by the decreased level of mutagenesis observed in the absence of the 9-1-1 clamp and the alternative clamp loader (Kai and Wang, 2003; Paulovich et al., 1998; Sabbioneda et al., 2005). In *S. pombe*, the 9-1-1 clamp interacts with polymerase κ and in *S. cerevisiae* it interacts with Rev1 and is monoubiquitinated by Rad6/Rad18, suggesting that
it may be involved in recruiting these factors to sites of DNA damage and in inducing the DNA damage response (Fu et al., 2008; Kai and Wang, 2003; Sabbioneda et al., 2005).
1.4. Mismatch Repair

1.4.1. Overview

The mismatch repair (MMR) pathway is conserved in all organisms ranging from archaea to metazoa (Lin et al., 2007). The primary role of this DNA repair mechanism is to correct DNA errors produced during DNA replication, which increases the fidelity of DNA replication by up to 1000-fold (Iyer et al., 2006; Modrich and Lahue, 1996; Umar et al., 1996). Incorrectly incorporated nucleotides and insertion-deletion loops (IDLs) produced by DNA polymerase slippage are the most common types of DNA lesions processed by the MMR pathway. MMR deficiency leads to a mutator phenotype with an increased frequency of point and frameshift mutations. In addition to reducing the mutation load, the MMR pathway also participates in the damage-induced cell cycle checkpoint response (Fishel, 1999), which contributes to an increased cancer incidence when MMR is defective (Jiricny and Marra, 2003). In contrast to the conventional role in mutation avoidance, MMR in activated, centroblast-stage B cells contributes to secondary antibody diversification mechanisms and as such, MMR deficiencies result in alteration and reduction in somatic hypermutation and class switch recombination. Since this DNA repair pathway is the cornerstone of this thesis, the next sections will provide a detailed overview of its mechanism, followed by an examination of its involvement in secondary antibody diversification processes.

The MMR system was first studied in *Escherichia coli* and subsequently in eukaryotes (see Sections 1.4.2. and 1.4.3.). Although there are substantial differences between these different organisms, the fundamental aspects of the eukaryotic and
prokaryotic MMR systems are the same and therefore inferences about eukaryotic MMR can be made from the prokaryotic system. In general, MMR proceeds via the following four steps: 1. Mismatch recognition and binding, 2. Communication with and signaling to downstream factors, 3. Degradation of the DNA strand containing the incorrect nucleotide in the region of the mismatch, and 4. DNA resynthesis and ligation. The following sections will highlight the mechanism of mismatch repair and its role in antibody diversification, cancer development, and other processes.

1.4.2. Mismatch Repair in Prokaryotes

Much of what is known about the MMR pathway stems from work in *E. coli* making this an appropriate starting point for the discussion on MMR as well as to emphasize how *E. coli* MMR differs from MMR in other organisms. In *E. coli* and other Gram-negative bacteria, the MMR pathway consists of three unique components – MutS, the mismatch recognition factor, MutL, the molecular matchmaker, and MutH, the strand discrimination factor – while the remaining factors, such as DNA helicases and polymerases, are common to many other DNA-processing pathways. MutS is a DNA-binding protein, contains an ATPase domain and a protein-protein interaction domain that enables two MutS molecules to pair up in order to form dimers (Lamers et al., 2000; Obmolova et al., 2000). MutS dimers scan DNA molecules in search of DNA mismatches and IDLs and upon binding to these lesions, undergo a conformational change which produces asymmetry at the mismatch binding site (Lamers et al., 2000; Modrich, 1991; Obmolova et al., 2000). Mismatch binding also translates into changes at the ATPase domain, resulting in modification of the occupancy state at the ATP-binding site (Haber and Walker, 1991). After binding a mismatch site,
**Figure 1.6. MutH directs mismatch repair to nascent DNA strands in prokaryotes.**

(A) DNA is scanned for mismatches by the asymmetrical MutS homodimer. (B) Mismatch recognition triggers ATP hydrolysis and binding of the MutL homodimer. MutH scans DNA for hemi-methylated GATC motifs. (C) MutS and MutL activate the endonuclease activity of MutH, which cleaves the unmethylated DNA strand at the GATC motif, located 5’ or 3’ to the mismatch. (D) UvrD unwinds the DNA helix, while one of four exonucleases (ExoI, ExoX, ExoVII, or RecJ) degrades the nicked strand. RPA coats the opposite DNA strand to protect it from nuclease activity. (E) The gap-filling reaction is mediated by DNA polymerase holoenzyme, the processivity clamp, the clamp loader, and DNA ligase. CL: clamp loader; POL: polymerase holoenzyme; EXO: ExoI, ExoX, ExoVII, RecJ. Figure is not to scale.
Figure 1.6.
MutS associates with MutL, which facilitates communication between MutS and downstream effectors (Grilley et al., 1989). The primary role of MutL is to modulate the activity of other MMR proteins, including assisting in mismatch recognition, supporting the assembly of the MMR complex, and restricting DNA degradation to the mismatch region. Like MutS, MutL also contains DNA binding, protein-protein interaction, and ATPase domains, all of which are critical for MMR activity as mutations in any of these domains in MutS or MutL result in defective MMR.

The MMR pathway is largely responsible for correcting errors produced during DNA replication, and as such must be appropriately targeted to the nascent DNA strand since this is the strand that contains the incorrectly incorporated nucleotides. In *E. coli*, strand recognition is facilitated by MutH, an endonuclease with specificity for unmethylated dGATC sites in hemimethylated DNA – that is, DNA methylated on one strand only (Dohet et al., 1985; Lahue et al., 1987). The DNA of *E. coli* and other Gram-negative bacteria is normally methylated at the N6 position of dA in the dGATC motif (G-Me\textsubscript{N6}A-TC), but because Dam methylase, the methylating enzyme, lags behind the replication fork, the newly synthesized DNA strand is transiently unmethylated. The unmethylated DNA strand is a beacon to the MMR complex and directs repair to the newly replicated DNA strand (Langle-Rouault et al., 1987; Lu et al., 1983; Pukkila et al., 1983). Mismatch binding by MutS and interaction with MutL activates MutH to cleave the DNA at an unmethylated dGATC site, creating an incision that serves as the entry point for DNA helicase II (MutU or UvrD in *E. coli*), which unwinds the DNA helix. Single strand binding (SSB) protein then coats the unwound DNA to protect it from nucleolytic activity, while the nicked strand is excised beginning at the cleavage site and ending ~150 nucleotides beyond the mismatch (Kunkel and Erie, 2005). In *E. coli*, four different exonucleases are involved in the degradation step.
of MMR and the choice of exonuclease is in part determined by the location of the single-stranded nick relative to the location of the mismatch. That is, when MutH produces a nick 5' of the mismatch, a 5’-3’ exonuclease (ExoVII or RecJ) is utilized. In contrast, when the nick is 3' of the mismatch, a 3’-5’ exonuclease (ExoI or ExoX) is used (Burdett et al., 2001; Viswanathan et al., 2001). Following the DNA excision step, the resulting ssDNA gap is resynthesized by DNA polymerase III holoenzyme and then sealed by DNA ligase (Burdett et al., 2001; Lahue et al., 1989).

1.4.3. Mismatch Repair in Eukaryotes

Eukaryotic organisms share a number of common elements with E. coli MMR, in particular the MutS and MutL proteins, whereas the MutH proteins are found exclusively in Gram-negative bacteria. The basic structure and function of the eukaryotic counterparts of MutS and MutL, termed MutS homologue (Msh) and MutL homologue (Mlh), are similar to their prokaryotic counterparts. As in prokaryotes, MutS and MutL homologues are ATPases that function in dimeric form and contain DNA-binding and protein-protein interaction domains. There are five Msh proteins in humans (Msh2 to Msh6) and six in S. cerevisiae (Msh1 to Msh6), however only Msh1 (S. cerevisiae), Msh2, Msh3, and Msh6 participate in MMR. In contrast, Msh4 and Msh5 are involved in meiotic cross-over reactions and are required for efficient progression through meiosis (Ross-Macdonald, 1994; Hollingsworth, 1995). Human cells possess two mismatch recognition dimers: MutSα and MutSβ. MutSα (Msh2-Msh6) recognizes single base-base mismatches and IDLs up to 2 nucleotides in length, while MutSβ (Msh2-Msh3) recognizes larger IDLs ranging from 3 to 16 nucleotides (McCulloch et al., 2003).
Humans and budding yeast possess four MutL homologues which form three functionally distinct heterodimers – Mlh1, Mlh3, Pms1, and Pms2 in humans and Mlh1, Mlh2, Mlh3, and Pms1 in *S. cerevisiae*. MutLα (Mlh1-Pms2 in humans, Mlh1-Pms1 in *S. cerevisiae*) participates directly downstream of MutSα or MutSβ to facilitate MMR and is the sole MutL dimer that is required for successful MMR. MutLγ (Mlh1-Mlh3) primarily interacts with Msh4-Msh5 to facilitate crossovers in meiosis, but can also provide a backup MMR function for MutLα (Cannavo et al., 2005; Flores-Roas and Kolodner, 1998; Wang et al., 1999). The biological role of MutLβ, (Mlh1-Pms1) is not understood and a role in MMR has not been demonstrated. As in *E. coli*, eukaryotic MutLα is termed the molecular matchmaker because it acts as a liaison between the mismatch recognition proteins and downstream factors. Although its requirement for MMR in vivo (Kunkel and Erie, 2005; Li and Modrich, 1995) and for 3'-directed MMR in vitro (Dzantiev et al., 2004; Zhang et al., 2005) is well-documented, its precise biochemical function was elucidated only recently with the identification of a latent endonuclease motif in MutL homologues (Kadyrov et al., 2006).

The recognition of a mismatch or IDL by MutSα/β produces a conformational change at the ATP-binding site, resulting in a switch from ADP to ATP and a reduced affinity for mismatched DNA, allowing the MutS complex to translocate away from the mismatch site in order to interact with downstream MMR components (see Section 1.5.5.). Interaction with MutLα is required on several levels, including in strand discrimination, strand excision initiation and termination, and in the stimulation of other MMR factors. Subsequent to mismatch recognition by MutSα or MutSβ and interaction with MutLα, the
Figure 1.7. Msh2-Msh6 dimers recognize single base-base mismatches in eukaryotes.

(A) DNA errors such as single base-base mismatches or insertion-deletion loops are produced by incorrect nucleotide selection or polymerase slippage during DNA replication. (B) Single base-base mismatches are identified by Msh2-Msh6 dimers (MutSα) based on the degree of DNA flexibility. The dimer forms a ring around the DNA molecule. Affinity for the mismatch is modulated by the occupancy state at the ATP-binding site. (C) ATP hydrolysis enables the recruitment of other factors, including the Mlh1-Pms2 complex (MutLα) and PCNA. The multi-protein complex can search for a strand break to direct repair to the appropriate strand. A nick located on the 3' side of the mismatch directs strand cleavage to the 5' side of the mismatch by the MutLα endonuclease. (D) The nick is used by Exo1 to initiate strand degradation in the 5'-to-3' direction while RPA coats the opposite DNA strand to protect it from nuclease activity. (E) The gap-filling reaction is mediated by DNA polymerase δ, PCNA, and RFC and DNA ligase. Figure is not to scale.
Figure 1.7.
region of DNA in the vicinity of the lesion is degraded. As alluded to in the section on prokaryotic MMR, the nick that initiates excision can be located either 5’ or 3’ to the mismatch or IDL. However, this stage of MMR differs from the prokaryotic system on the following three crucial points: 1) DNA in eukaryotic cells following synthesis is not hemimethylated as in Gram-negative bacteria and thus a different strand-discrimination mechanism is required. In contrast to bacteria where DNA methylation is used primarily to protect the genome from foreign DNA, in mammals, DNA methylation takes place mainly on the cytosine of CpG dinucleotides and plays crucial roles in X-chromosome inactivation, genomic imprinting, genome stability, transcriptional silencing, and early development (Sasai and Defossez, 2009). 2) Eukaryotes lack the endonuclease MutH; and 3) exonucleolytic degradation proceeds exclusively in the 5’-to-3’ direction. To address the first point, it has been suggested that in the absence of methylation-directed strand discrimination, the relevant eukaryotic strand discrimination signal is provided by the strand breaks between adjacent Okazaki fragments on the lagging strand and perhaps by the free 3’ hydroxyl end located at the replication fork on the leading strand (Kunkel and Erie, 2005). In addition, the orientation of PCNA molecules has also been proposed to aid in this process (Lau and Kolodner, 2003; Shell et al., 2007). Thus, one would envisage that the frequent single-stranded breaks and abundance of PCNA molecules on the lagging strand would lead to more efficient MMR on the lagging strand than on the leading strand. Indeed, it has been observed that MMR in yeast is ~100-fold more efficient on the lagging strand than on the leading strand (Kow et al., 2007; Pavlov et al., 2003). The answer to points 2) and 3) was provided by Kadyrov et al. who demonstrated that MutLα possesses latent endonuclease activity in vitro, which enables DNA cleavage near the mismatch site (Kadyrov et al., 2006). The study further revealed that MutLα-mediated DNA cleavage is
restricted to the DNA strand that contains a pre-existing nick, thus targeting MMR to the newly-replicated DNA strand. This targeting may be assisted by the interaction between MutLα and PCNA, which could direct MutLα to the appropriate strand (Dzantiev et al., 2004; Kadyrov et al., 2006; Lau and Kolodner, 2003; Shell et al., 2007). The cryptic endonuclease motif has thus far been identified in human Pms2 and Mlh3 and yeast Pms1 (Kadyrov et al., 2006; Kadyrov et al., 2007) and is required not only for MMR, but also for the suppression of homeologous recombination, DNA damage signaling, and in the case of Mlh3, meiotic cross-over reactions (Deschenes et al., 2007; Erdeniz et al., 2007; Nishant et al., 2008). The sole MMR-specific exonuclease demonstrated to mediate the degradation reaction is Exo1, a 5’-to-3’exonuclease. This was a source of further confusion since in vitro MMR reactions were shown to proceed from both directions. That is, provided with a nick located 5’ or 3’ to the mismatch, degradation always appeared to be initiated at the nick closest to the mismatch, irrespective of orientation (Dzantiev et al., 2004). While it was suggested that Exo1 may possess cryptic 3’-to-5’ hydrolytic activity (Genschel et al., 2002), the aforementioned study by Kadyrov et al. revealed that MutLα introduces a nick on the distal side of the mismatch relative to the pre-existing nick, such that if the nearest mismatch-proximal nick is located 3’ to the lesion, a nick introduced by MutLα 5’ to the mismatch would be used as the Exo1-loading dock for the degradation step of MMR.

Using a nick located 5’ to the mismatch, MutLα stimulates DNA helicase II to unwind the DNA in the 5’-to-3’ direction (Robertson et al., 2006). Subsequent Exo1 degradation is stimulated by MutSα (Dzantiev et al., 2004) and proceeds until the mismatch has been degraded (Genschel et al., 2002; Tishkoff et al., 1997). In vitro experiments indicate that the length of the excision tract can be up to 1000 nucleotides (Genschel and Modrich, 2003).
The exact length of the excision tract \textit{in vivo} is not known, but has been postulated to be \(~200\) nucleotides based on electromicroscopy, cell line, and murine data (Unniraman and Schatz, 2007) (Chapters 2 and 3). \textit{MutLα} and \textit{MutSα} have also been postulated to play a role in limiting the excision tract length, but only on homoduplex DNA (Nielsen et al., 2004; Zhang et al., 2005). A more recent report indicates that RPA is instrumental in limiting Exo1-mediated degradation, with \textit{MutLα} playing only a secondary role (Genschel and Modrich, 2009). Degradation of the nascent strand is accompanied by coating of the template strand by RPA to protect it from nucleolytic attack. Gap-filling by DNA polymerases \(\delta\), which is enhanced by phosphorylated RPA (Constantin et al., 2005; Guo et al., 2006c; Longley et al., 1997) and ligation by ligase I complete the MMR reaction (Karthikeyan et al., 2000; Nick McElhinny et al., 2008; Zhang et al., 2005).

In addition to the core MMR proteins, several other factors have been implicated in the MMR reaction, including PCNA (Gu et al., 1998; Umar et al., 1996). PCNA interacts with Msh3, Msh6, Mlh1, and Exo1 and mutations that abolish these interactions between PCNA and any one of these MMR factors result in defective MMR and a mutator phenotype (Clark et al., 2000; Flores-Rozas et al., 2000; Gu et al., 1998; Johnson et al., 1996; Kleczkowska et al., 2001; Lee and Alani, 2006; Umar et al., 1996). While involvement of the DNA polymerase processivity clamp at the resynthesis step is intuitive, its requirement in earlier MMR steps is more surprising. Some studies have suggested that PCNA enhances mismatch binding by \textit{MutSα} or \textit{MutSβ}, while others propose that PCNA, in fact, facilitates dissociation of the mismatch recognition complex from the DNA lesion (Bowers et al., 2001). PCNA and the clamp loader replication factor C (RFC) also stimulate the endonuclease activity of \textit{MutLα} (Kadyrov et al., 2006). It is clear that PCNA plays more than one role in the MMR...
process and its absence prevents both mismatch-provoked DNA excision and resynthesis (Bowers et al., 2001; Gu et al., 1998; Longley et al., 1997; Umar et al., 1996). High mobility group box 1 (HMGB1) is a non-histone chromatin protein that has been shown in some studies to be involved in MMR (Fleck et al., 1998; Yuan et al., 2004). However, since other studies suggest that HMGB1 protein is dispensable for MMR, its role is yet to be fully characterized (Dzantiev et al., 2004; Genschel and Modrich, 2009). Not all of the above components are required for MMR and the requirements differ for 5’- and 3’-directed repair. In vitro reconstitution of mammalian MMR yielded some conflicting results. While Genschel et al. demonstrated that MutSα, RPA, and Exo1 are sufficient for 5’-directed repair, a study by Dzantiev et al. revealed that 3’-directed repair requires the addition of MutLα, PCNA, and RFC (Dzantiev et al., 2004; Genschel and Modrich, 2003). In contrast, Zhang et al. suggest that MutSα/β, MutLα, Exo1, RPA, PCNA, RFC, and HMGB1 participate collectively in efficient MMR (Zhang et al., 2005).

1.4.4. Molecular Mechanisms of Mismatch Repair – Current Models

A successful MMR reaction requires that the initiating mismatch recognition step be communicated to downstream repair factors such as Exo1, which can be located some distance away from the mismatch. The available data on this topic points to three models that can be used to describe this process. The first of these models is a trans or stationary model, while the other two are cis or moving models. The trans model posits that the MutS mismatch recognition proteins remain bound at the mismatch site, while distal DNA ends are brought together via DNA looping or bending, which is induced by protein-protein interactions between MutS, MutL and Exo1. ATP hydrolysis by MutS is required to verify
mismatch binding and to authorize excision (Junop et al., 2001; Schofield et al., 2001b; Wang and Hays, 2004). Evidence for this model stems from the observation that a physical barrier placed in between the mismatch site and the strand nick at which DNA excision is initiated does not inhibit the MMR reaction, thus indicating that MMR proteins do not migrate between the two sites (Wang and Hays, 2003; Wang and Hays, 2004). In contrast to the stationary model, the two cis models suggest that MutS proteins move away from the mismatch in search of strand discrimination signals. The first of the cis models is the translocation model and suggests that when MutS is bound to the mismatch with ATP at the nucleotide binding site, the affinity of MutS for the mismatch is reduced. Subsequent ATP hydrolysis stimulates bidirectional translocation away from the mismatch, which proceeds until a nick is reached. Evidence for the translocation model came from the observation that the presence of MutS, a mismatch, and hydrolysable ATP induces α-shaped DNA loops, where MutS is located at the base of the loop and the mismatch is found within the loop, some distance away from MutS (Allen et al., 1997). Furthermore, DNA loop size increased with time, arguing against the trans signaling model. The second of the cis models is the Molecular Switch or Sliding Clamp model and it proposes that the MutS complex interacts with the mismatch in an ADP-bound state, but mismatch binding induces a conformational change that triggers the exchange of ATP for ADP. ATP binding induces a second conformational change which allows MutS to dissociate from the mismatch, form a sliding clamp and translocate away from the mismatch, possibly in concert with MutL, in search of a strand break. Additionally, this model involves the loading of multiple MutS complexes onto the mismatch and their translocation along the DNA molecule in search of the strand break, in the absence of ATP-hydrolysis (Fishel, 1998; Gradia et al., 1997; Jiang et al., 2005; Mendillo et al., 2005). Evidence for this model stems from studies
demonstrating that multiple MutS molecules localize to a single MMR event and that a physical barrier along the DNA molecule significantly reduced MMR in *E. coli* (Pluciennik and Modrich, 2007; Smith et al., 2001).

### 1.4.5. Mismatch Recognition

MutSα complexes recognize and process DNA lesions that consist of mispaired bases rather than grossly modified nucleotides and as such, mismatch recognition relies on subtle structural changes in the DNA double helix. Crystallographic and biochemical data suggest that MutS proteins differentiate between paired and mispaired DNA based on the flexibility and ease of kinking of the DNA molecule (Lamers et al., 2000; Obmolova et al., 2000). Indeed, mispaired bases weaken base stacking interactions, the predominant energy force for double helix formation (Crothers and Zimm, 1964). This induces modifications in the minor groove and leads to an increased propensity toward DNA bending, thus facilitating mismatch recognition. Paradoxically, there is an inverse correlation between the efficiency of repair and the degree of bending induced by distinct mismatches in the DNA double helix (Brown and Jiricny, 1988; Kramer et al., 1984). Although the G:T mismatch causes the least amount of double helix distortion, it is the most efficiently repaired mismatch out of the eight possible mismatch combinations (Dohet et al., 1985). Conversely, the C:C mismatch is the most destabilizing mismatch, but is virtually refractory to repair (Salaro et al., 1993). The efficiency of mismatch recognition is also context specific (Mazurek et al., 2009), however alterations in base stacking interactions play a pivotal role in mismatch recognition.
Crystal structures of MutS-DNA complexes from *E. coli* and human cells revealed that each MutS protein consists of two DNA-interacting domains (I and IV), one ATP-binding domain (V), and two connecting domains (II and III) (Lamers et al., 2000; Natraj et al., 2003; Obmolova et al., 2000; Warren et al., 2007). As noted earlier, the two MutS subunits bind asymmetrically at the mismatch site, with the majority of protein-DNA interactions arising from Msh6 or its MutS equivalent in *E. coli* (Lee et al., 2007; Warren et al., 2007). Most of the MutS-DNA interactions are with the DNA backbone and therefore not nucleotide-specific, with the exception of Phe432 and Glu434 of human Msh6, which are critical for MMR function (Bowers et al., 1999; Das Gupta and Kolodner, 2000; Drotschmann et al., 2001; Schofield et al., 2001a; Yamamoto et al., 2000). The Phe side chain is wedged inside the minor groove and causes mismatch displacement (Dufner et al., 2000; Malkov et al., 1997). Concurrently, the major groove is narrowed, thus accommodating a $60^\circ$ kink in the DNA helix. The carboxyl group of the Glu residue is hydrogen-bonded to a mismatched base and contributes to mismatch recognition (Lebbink et al., 2006). The distortion of the DNA double helix also arises from base-base interactions in heteroduplex DNA where wobble basepairing between mismatched basepairs causes base displacement, which interferes with optimal base stacking.

The frequency of replication-induced mismatches is $\sim$1 mismatch per $10^8$ base pairs. The required specificity and accuracy of mismatch recognition is accomplished with the high energy cofactor ATP. MutS proteins are members of the ATP-binding cassette (ABC) transporter ATPase superfamily, with each subunit of the MutS dimer contributing a critical domain of the ATP-binding site. As a result, loss of dimerization also abolishes ATPase activity. When bound to perfectly matched DNA, ATP is quickly hydrolyzed by MutS, while binding to heteroduplex DNA inhibits ATP hydrolysis (Bjornson et al., 2000;
Gradia et al., 1999; Junop et al., 2001). The ATP-bound dimer has greatly reduced affinity for homoduplex DNA, but only marginally reduced affinity for heteroduplex DNA (Junop et al., 2001; Lamers et al., 2004). Thus, the MutS-ATP complex is released from normal DNA, but not from heteroduplex DNA, which, together with the inhibited ATPase activity, enhances the specificity of MutS for mismatched DNA. The stabilized MutS-DNA-ATP complex is then able to recruit MutL and activate subsequent mismatch processing events.

The current mismatch recognition model suggests a three-phase process for mismatch identification and repair initiation. First, MutSα scans the DNA, during which time it is bent. Evidence for this comes from the observation that bent DNA molecules are observed with both homo- and heteroduplex DNA (Wang et al., 2003). Upon mismatch recognition, MutSα undergoes a conformational change that converts it into the Initial Recognition Complex and causes kinking of the DNA. Finally, MutSα undergoes a second conformational change to the Ultimate Recognition Complex in which the DNA is unbent with the mismatched base possibly flipped out. MutSα-homoduplex DNA complexes were found strictly in the bent conformation, while MutSα-heteroduplex DNA complexes were found in both bent and unbent arrangements, indicating that only heteroduplex DNA will occur in the unbent state (Tessmer et al., 2008). The final, unbending step, potentially provides a “double-check” mechanism before proceeding with the repair reaction.

1.4.6. Mismatch Repair and Antibody Diversification

The somatic hypermutation process generates a highly elevated DNA mutation frequency that is six orders of magnitude above the background mutation rate. The majority of these
mutations arise from DNA mismatches, while DNA insertions and deletions contribute only slightly. Thus, it is not surprising that the DNA MMR pathway performs a dominant function in the somatic hypermutation (discussed in section 1.5.6.1) and class switch recombination (discussed in Section 1.5.6.2) processes.

1.4.6.1. Mismatch Repair and Somatic Hypermutation

As early as 1959, Lederberg et al. suggested that antibody-producing cells are subject to high rates of somatic mutations (Lederberg, 1959). This theory was later refined by Brenner and Milstein who proposed a model in which somatic mutations are localized to a small region of the antibody gene and arise when DNA strand breaks in the vicinity of the Ig locus are converted into point mutations during the resynthesis steps of DNA repair (Brenner and Milstein, 1966). Although it is now clear that a substantial proportion of somatic mutations in the Ig locus are due to the direct targeting of AID, in the years leading up to its discovery, numerous laboratories explored the roles of DNA repair pathways in the SHM process. Even before the identification of AID, Rada et al. correctly postulated that the MMR pathway is instrumental in extending mutations beyond those that occur at WRČY hot spot motifs (Rada et al., 1998). However, it was not until the discovery that AID acts on DNA to generate G:U mismatches that this model could be justified. The following sections will outline how the MMR system acts to generate mutations at A:T basepairs and G:C mutations outside the WRČY hot spot motif as well as some of the early studies and models that emerged from them regarding the involvement of MMR in SHM. Briefly, all of the studies utilized mice that were either immunized or unimmunized, and deficient in one or more components of the MMR pathway – that is, Msh2, Msh6, Msh3, Mlh1, Pms2, and Exo1.
to examine mutation frequencies, mutation spectra and mutation distributions in different areas of the rearranged V region.

i. MMR Knock-Outs and Mutants

The first study to report on the involvement of the MMR pathway in SHM used Pms2−/− quasi-monoclonal (QM) mice (Cascalho et al., 1996; Cascalho et al., 1998). QM mice are characterized by limited antibody diversity as a result of the introduction of a rearranged VHDJH exon on one IgH allele and disruption of the JH exon on the other IgH allele in addition to the deletion of both Jκ alleles. The result is that the pre-rearranged IgH chain can only pair with the Igλ chain, which is rearranged using only three Vλ’s and three Jλ’s. The limited rearrangement and combinatorial possibilities yield a small number of antibody identities, which can be identified by idioype-specific antibodies and have the advantage of a known immunoglobulin sequence against which somatic mutations can be compared. Using Pms2−/− QM mice, Cascalho et al. observed V region mutation frequencies reduced by 20-fold relative to the wildtype control (Cascalho et al., 1998). Although later studies of mice with MMR deficiencies have not been able to reproduce this phenotype, making the relevance of these findings unclear, Cascalho et al. suggested that the MMR pathway is somehow involved in fixing mutations in the genome, thereby contributing to, rather than minimizing, the mutation frequency. Subsequent analyses of Pms2−/− mice suggested only a limited effect on the mutation frequency at the Ig locus, with a small, at most 2-fold, reduction in frequency (Frey et al., 1998; Kong and Maizels, 1999; Winter et al., 1998). Moreover, mice deficient in Mlh1, which is the common binding factor for Pms2, Pms1, and Mlh3, did not differ from wildtype controls (Kim et al., 1999; Phung et al., 1999)
and thus, the authors of these studies postulated that MMR does not participate in the SHM process.

A number of studies investigated the role of the Msh2 in SHM at heavy and light chain loci. The studies compared the mutation frequencies of Msh2-deficient mice, either Msh2−/− or Msh2G674A mutant mice, to those of wildtype controls and found mutation frequencies that were either marginally or dramatically decreased compared to wildtype (Martin et al., 2003; Phung et al., 1998; Rada et al., 1998). Despite the discrepancies in mutation frequency, most investigators observed an altered mutation spectrum and distribution. The altered mutation spectrum was manifested by an increased proportion of mutations at G:C basepairs, the majority of which were transition mutations, and a reduced proportion of mutations at A:T basepairs, which make up 50 to 60% of the mutation load in wildtype mice. The distribution of mutations was likewise altered in Msh2-deficient mice, with an increased number of mutations restricted to the mutational hotspot WRCY (RGYW reverse complement) (Martin et al., 2003; Phung et al., 1998; Rada et al., 1998).

Additionally, Msh2−/− mice also displayed a defect in the serum antibody response to T-dependent and T-independent antigens, characterized by a reduction in total antigen specific IgG antibodies and high affinity antigen-specific IgG antibodies, a phenotype attributed to an impaired affinity maturation reaction, but also a potential defect in isotype switching (Ehrenstein and Neuberger, 1999; Martin et al., 2003; Rada et al., 1998; Vora et al., 1999). Collectively these data suggested that Msh2 somehow influences mutation accumulation at the Ig locus.

In contrast to data obtained from Msh2−/− mice, the mutation characteristics in the V region were not dramatically altered in Pms2−/−, Mlh1−/−, or Msh3−/− mice (Phung et al., 1999;
Wiesendanger et al., 2000; Winter et al., 1998). Msh3−/− mice were also identical to controls, an expected result since Msh2-Msh3 form the recognition complex for large (3-15 nucleotide) IDLs, a lesion that is not observed in the somatic mutation spectrum (Wiesendanger et al., 2000). The latter study by Wiesendanger et al. also examined the role of Msh6 in SHM and found that the phenotype of Msh6−/− mice was similar to, albeit somewhat milder than that of Msh2−/− mice, while the Msh3−/− Msh6−/− mice behaved much like Msh2−/− mice (Wiesendanger et al., 2000). Bardwell et al. examined the role of Exo1 (Bardwell et al., 2004), the only exonuclease shown to be active during MMR, and demonstrated that Exo1 deficiency results in diminished high affinity antibody titers to a T-dependent antigen, increased hot spot focusing and decreased A:T mutations in the V region (Bardwell et al., 2004). Although these phenotypes resemble those of Msh2-deficient mice, the mutation frequency was not reduced in the absence of Exo1.

The studies described above demonstrated that Msh2, Msh6 and Exo1 are involved in the generation of mutations at the immunoglobulin locus, but substantial differences among the data, particularly with respect to mutation frequencies, led to diverse interpretations of the results. The varying observations may have been the result of differences in experimental protocols as well as exposure of animals to different environmental antigens in each facility. The most striking difference is observed in comparing V region mutation frequencies between immunized and unimmunized mice, with the former method yielding only minimal effects on mutation frequencies (Phung et al., 1998). On the other hand, those studies that examined intronic V region sequences from chronically stimulated Peyer’s patches revealed mutation frequencies in Msh2−/− mice that were decreased relative to wildtype controls. This underscores the importance of analyzing sequence regions that are not under selective pressure such that the mutations are not
skewed by outside forces. Indeed, for this reason, the ex vivo results presented in Chapter 3 of this thesis were generated from intronic V region sequences.

The varied results obtained from these studies led to equally diverse interpretations of the role for the MMR pathway in SHM. Since MMR’s role is primarily to correct DNA mismatches, it was expected that the absence of this pathway would result in increased mutation frequencies in the V region. Instead, decreased mutation frequencies and an increased percentage of WRCRY hot spot G:C mutations led Rada et al. to postulate that SHM operates as a two-phase process (Rada et al., 1998). That is, the first mutations to be introduced into the locus are G:C mutations located at hot spot motifs and these are acted on by the MMR pathway, which generates a second round of mutations. In attempting to repair the mismatches, MMR engages one or more error-prone polymerases that create A:T mutations during the “repair” phase and thus extends mutations beyond those at the original mismatch. It should be noted, however, that at the time of its proposition, several other explanations for the effect of MMR-deficiency were proposed. These include a role for MMR in fixing mutations in the genome (Cascalho et al., 1998) or in maintaining the mutation load at a viable level, the absence of which results in the elimination of cells with grossly elevated mutation loads (Frey et al., 1998; Phung et al., 1998). The latter hypothesis was disputed by Msh2−/− Bcl2 transgenic mice which demonstrated that enhanced B cell survival does not abolish the increased G:C mutation bias observed in Msh2−/− mice (Alabyev and Manser, 2002). Ultimately, the identification of AID and the accumulation of evidence suggesting that the enzyme acts directly on DNA led to the general acceptance of, and expansion upon, the two-phase model in which the processing of AID-generated G:C mutations by Msh2, Msh6 and Exo1 generates mutations at A:T basepairs (Rada et al.,
1998). The importance of this is highlighted in Chapters 2 and 3 where the relationship between different types of mutations is discussed.

AID generated G:U mismatches are good substrates for MMR (Wilson et al., 2005). During conventional MMR, these mismatches would be recognized by the Msh2-Msh6 complex, which, in coordination with the Mlh1-Pms2 complex and Exo1 leads to excision of one of the DNA strands in the region of the mismatch. This is followed by the resynthesis step during which a replicative polymerase, such as polymerase δ, performs the gap filling reaction in an error-free manner. The choice of strand excision depends on appropriate strand discrimination signals and also determines whether the process results in restoration of the original G:C basepair or in the production of G:C to A:T transition mutations. The latter scenario would occur if the G-containing strand was excised and the U became the template, while the former would occur if the U-containing strand was excised and the G was used as the template. In contrast, Msh2-Msh6-directed repair of G:U lesions during SHM leads to error-prone gap-filling and yields mutations at A:T basepairs in addition to the possible transition mutations described above. That is, following G:U mismatch binding by Msh2-Msh6, Exo1 degrades the DNA backbone in the region of the mismatch, allowing the gap-filling reaction to take place. Resynthesis is achieved by an error-prone translesional polymerase which, by virtue of possessing high error rates on undamaged DNA, produces mutations primarily at A:T basepairs as well as some G:C basepairs.
ii. Translesional Polymerases

The relatively recent identification and increased knowledge of the error rates and mutation spectra of translesion polymerases in mammals enabled significant insights into the error‐prone repair processes that operate during SHM. The contribution of the translesional polymerases \( \eta, \imath, \zeta, \lambda, \kappa, \mu, \) and Rev1 to the mutation load in the V region have been examined both *in vitro* and *in vivo* (Bertocci et al., 2002; Faili et al., 2002a; Faili et al., 2009; Martomo et al., 2008; Masuda et al., 2005; McDonald et al., 2003; Schenten et al., 2002; Shimizu et al., 2003; Zan et al., 2001; Diaz, 2001 #389; Zan et al., 2005; Zeng et al., 2004; Zeng et al., 2001). An examination of polymerase \( \eta \) activity on undamaged template DNA revealed a high error rate, approximately \( 3 \times 10^{-2} \) mutations/bp, and an error spectrum that favours misincorporations opposite template T, with misincorporations opposite template A occurring with a 2-fold lower frequency and those opposite G and C being even lower (Matsuda et al., 2001). These findings were extended by Rogozin *et al.* who demonstrated a high level of correlation between the A:T mutation spectrum in the V region *in vivo* and the polymerase \( \eta \) error spectrum *in vitro*, that is, a preference for A mutations at WA hot spot motifs, where W is A or T, thereby implicating polymerase \( \eta \) in the generation of A:T mutations in hypermutating B cells (Rogozin et al., 2001). Additional data showing that mice and humans harboring polymerase \( \eta \) deficiencies generate V regions with fewer A:T mutations as well as an overall reduction in mutation frequency provided further support for polymerase \( \eta \) as a candidate A:T mutator during the SHM process (Faili et al., 2004; Yavuz et al., 2002; Zeng et al., 2004; Zeng et al., 2001). The preferred mutation spectrum of polymerase \( \eta \) was used extensively in the analyses in Chapters 2 and 3.
The absence of polymerase η severely reduces, but does not eliminate A:T mutations, suggesting that a back-up pathway for A:T mutagenesis exists in hypermutating B cells (Zeng et al., 2001). Numerous other candidate polymerases have been tested for their contribution to A:T mutations at the V region using knock-out and knock-down approaches. Studying the V regions of polymerase η⁻/⁻ mice, Delbos et al. observed that the residual A:T mutations correlated with the error-spectrum of polymerase κ, suggesting that pol κ may serve as a substitute pathway in the absence of polymerase η (Delbos et al., 2005). Although the polymerase κ⁻/⁻ mouse did not display an altered A:T mutation spectrum (Schenten et al., 2002; Shimizu et al., 2003), the polymerase η and polymerase κ double knock-out resulted in a greater reduction in A:T mutations than in the polymerase η single knock-out (Faili et al., 2009). These data reveal that polymerase κ provides backup in the absence of polymerase η, that polymerase κ participates minimally if at all in SHM when polymerase η is active, and finally that at least one other redundant polymerase generates A:T mutations in polymerase η and polymerase κ-deficient mice. Possible reason for the preferential usage of polymerase η over other translesional polymerases will be explored in Chapter 5.

In contrast, the mutation spectra from mice deficient in both polymerase η and polymerase ι resembled those of polymerase η single knock-out mice, indicating that polymerase ι does not provide a back up activity in the absence of polymerase η (McDonald et al., 2003). The phenotypes of mice deficient in polymerase μ or polymerase λ did not differ from those of wildtype (Bertocci et al., 2002), suggesting that these polymerases are not involved in the generation of A:T mutations, although the double knock-outs lacking polymerase η would be required to unequivocally rule out this possibility. Polymerase ζ is
required for embryonic development and is therefore technically challenging to study \textit{in vivo}. Nevertheless, a reduction in polymerase $\zeta$ expression in a human cell line using a knock-down approach resulted in reduction of the overall mutation frequency without any alteration in mutation spectra compared to controls (Zan et al., 2001). Although residual polymerase $\zeta$ protein expression and the fact that the experiments were conducted in a cell line rather than \textit{in vivo} makes these results difficult to interpret, the fact that polymerase $\zeta$ is a mismatch extender and functions with other translesional polymerases suggests that it might have a role in overall mutagenesis in the V region (Guo et al., 2001; Haracska et al., 2001; Johnson et al., 2000). Collectively, these data demonstrated that after AID-mediated deamination of G:C basepairs, G:U lesions are recognized and engaged by the Msh2-Msh6 complex, followed by excision of one of the DNA strands by Exo1 and subsequent gap-filling by polymerase $\eta$, or by polymerase $\kappa$ and one or more translesional polymerases in situations of polymerase $\eta$ deficiency.

\textbf{iii. UNG Contributes to A:T Mutations}

The discussion on MMR and its role in generating A:T mutations would be incomplete without noting that the absence of Msh2, Msh6, Exo1, or polymerase $\eta$ reduces, but does not abolish A:T mutations. Therefore an alternate pathway to polymerase $\eta$ (described above) and Msh2, Msh6 and Exo1 has been explored. The absence of any of these mismatch repair factors dramatically reduces the number of A:T mutations, while G:C mutations are altered but slightly; it has been suggested that Msh2-Msh6 and Exo1 contribute to approximately 10% of all mutations at G:C base pairs. In contrast, UNG$^{-/-}$ mice lose most G:C transversion mutations, while maintaining wildtype levels of A:T mutations (Rada et al.,
Despite the lack of effect on A:T mutations, UNG was suggested to be an alternate candidate for inducing A:T mutations in the absence of the Msh2-Msh6-Exo1 pathway, a reasonable suggestion since UNG is the only other factor known to engage AID-generated G:U lesions. This alternative pathway was confirmed when the V regions from mice lacking both Msh2 and UNG were shown to be entirely deficient in A:T mutations (Rada et al., 2004). Likewise, A:T mutations were reduced to near absent levels in mice lacking polymerase η and Msh2 (Delbos et al., 2007), indicating that polymerase η and UNG act together to generate a small number of A:T mutations when the Msh2-Msh6-Exo1 pathway has been inactivated. Collectively, these studies provided evidence that engagement of the G:U lesions by Msh2-Msh6, followed by degradation of the DNA surrounding the mismatch by Exo1 and error-prone resynthesis by polymerase η mediated by ubiquitinated PCNA generates the bulk of A:T mutations in hypermutating V regions. The possibility of UNG and Msh2 cooperation in A:T mutagenesis will be explored in Chapters 2 and 3.

iv. PCNA and the Induction of A:T Mutations

Although the establishment of polymerase η as the primary A:T mutator during SHM provided significant insight into the generation of these mutations, it did not, however, shed light on the reasons why Msh2-Msh6-mediated processing of G:U lesions in V regions results in the induction of error-prone synthesis, rather than the conventional, error-free mode of repair. In an attempt to gain insight into this process, Wilson et al. demonstrated that Msh2 physically interacts with and stimulates the catalytic activity of polymerase η (Wilson et al., 2005), but this did not explain why the error-prone polymerase is recruited instead of the conventional replicative polymerases δ or ε. Since translesional polymerases are required to achieve the full spectrum of mutations in the V region, Langerak et al.
investigated the role of translesion synthesis and the importance of PCNA ubiquitination by generating mice carrying the PCNA\textsuperscript{K164R} mutation, which abrogates the ability of PCNA to become ubiquitinated on Lys164 (Langerak et al., 2007). Ubiquitination at Lys164 is required for the induction of translesion synthesis, the recruitment of translesional polymerases and the bypass of stalled replication forks during DNA replication, although other mechanisms have also been shown to participate in this process. PCNA\textsuperscript{K164R} mice were characterized by a reduction in A:T mutations, similar to the reduction seen in Msh2- or polymerase η-deficient mice, demonstrating that ubiquitination of the PCNA clamp is imperative for the recruitment of translesional polymerases and A:T mutagenesis in hypermutating B cells. However, the abrogation of A:T mutations was incomplete, suggesting that another mechanism operates either in concert with, or in the absence of, ubiquitinated PCNA to mediate the recruitment of translesional polymerases to the mutating V region. Nevertheless, this finding indicates that the mutational processes that operate at the V region produce similar lesions to those at stalled replication forks and begs the question of what kind of AID-mediated lesions might induce PCNA ubiquitination during SHM. This will be discussed further in Chapters 2 and 3.

1.4.6.2. **Mismatch Repair and Class Switch Recombination**

CSR and SHM are distinct processes that are linked by their mutual requirement for AID, UNG, and MMR. AID and UNG deficiencies completely ablate and severely reduce the CSR reaction, respectively (Muramatsu et al., 2000; Rada et al., 2002b). CSR involves the creation of DSBs in switch regions, which are located upstream of the C region genes, and the ligation of distant DNA ends, which occurs primarily through the NHEJ pathway. Not
surprisingly, NHEJ- and DSB repair-specific factors such as Ku70, Ku80, Nbs1, and many others, are active participants in CSR, but not in SHM (Manis et al., 2002a; Manis et al., 1998; Reina-San-Martin et al., 2005). The role of MMR in this process will be discussed below.

As mentioned earlier, cells lacking Msh2, Msh6, and Exo1 were shown to be defective in the affinity maturation response (Bardwell et al., 2004; Martin et al., 2003; Rada et al., 1998; Wiesendanger et al., 2000). Moreover, these cells produced higher IgM, but lower IgG antibody titers, suggesting that the isotype switching reaction is also compromised. Indeed, CSR is less efficient and switch junction sequences are altered in the absence of MMR. Unlike UNG deficiency, which results in 90 to 95% reduction in CSR (Rada et al., 2002b), absence of Msh2, Msh6, Mlh1, Pms2, or Exo1 reduces switching by only 50 to 70% (Bardwell et al., 2004; Li et al., 2004a; Martin et al., 2003; Martomo et al., 2004; Schrader et al., 1999). The most pronounced effect is seen in Msh2−/− and Msh6−/− mice with a 70% drop in switching (Ehrenstein and Neuberger, 1999; Martomo et al., 2004). Using Msh2-UNG double knock-out mice, Rada et al. demonstrated that Msh2 is required to generate the residual CSR reactions observed in the absence of UNG (Rada et al., 2004).

Deficiencies in the UNG or MMR pathways result in varying reductions in CSR, suggesting that each pathway plays a unique role in this process. In both cases, switch junction sequence products were reduced in mice lacking these DNA repair pathways (Rada et al., 2002b; Schrader et al., 1999), indicating that the CSR mechanism is perturbed at the DNA level. In the absence of Msh2, a greater number of Sμ-containing switch junctions are derived from the central portion of Sμ where WRCY sequences, and therefore opportunities to produce closely spaced SSBs, are most prevalent (Min et al., 2005). This
suggests that MMR is involved in processing more distantly located DNA nicks, which occur at the boundaries of switch regions where WRČY motifs are less concentrated. Deletion of the Sμ tandem repeat region in SμTR−/− mice (Luby et al., 2001) in Msh2−/−, Mlh1−/−, or Exo1−/− backgrounds results in 95% or greater reduction in CSR (Eccleston et al., 2009; Min et al., 2003), indicating that MMR generates DSBs when closely-spaced SSBs are not possible. Thus, while UNG and APE might be sufficient to generate DSBs in the central WRČY-rich region, SSBs occurring outside this region require processing by the MMR pathway.

MMR proteins also regulate the mode of ligation employed during CSR. Most wild type switch junctions consist of blunt joints or joints that contain short, 1- to 2-nucleotide, microhomologies, implicating NHEJ in the process. However, a small percentage of joints contain longer microhomologies, suggesting that alternative mechanisms, such as microhomology-mediated end joining, are also used. Msh2−/− cells contain a greater number of switch junctions with blunt ended joints, while those with long microhomologies are virtually absent, suggesting that Msh2 is required to facilitate microhomology-mediated end joining (Martin et al., 2003; Schrader et al., 2002). Paradoxically, switch junctions from Mlh1−/− or Pms2−/− B cells show the opposite effect: switch junctions with blunt ends are reduced, while those with long microhomologies are dramatically increased (Schrader et al., 2002). Undoubtedly, CSR proceeds via a dynamic and complex mechanism that will require more detailed investigations before it can be fully elucidated.

1.4.7. Mismatch Repair and DNA Damage Signaling

In addition to sensing and repairing DNA mismatches and IDLs, the MMR system is also involved in cell cycle checkpoint regulation and apoptosis signaling in response to various
DNA damaging agents. Cells deficient in MMR display 2-, 10-, and 100-fold greater tolerance to death induced by cisplatin, 6-thioguanine, and methylating agents, respectively, compared to MMR-proficient cells, indicating that an active MMR pathway is required to efficiently signal cell death in response to these agents (Jiricny, 2006; Papouli et al., 2004). Methylating agents and 5-thioguanine induce similar types of DNA damage since both agents generate modified nucleotides in the template strand, O6-methylguanine (MeG) for the former and 6-methylthioguanine (Me6-TG) for the latter, which can pair with either C or T in the nascent strand (Jiricny, 2006). Cisplatin induces intrastrand crosslinks between adjacent purines, which form a physical barrier to DNA replication. Each of these lesions has been shown to interact with MutSα (Duckett et al., 1996; Karran, 2001; Swann et al., 1996; Waters and Swann, 1997; Yamada et al., 1997), suggesting that the MMR pathway processes and attempts to repair these lesions. This finding gave rise to the *futile repair* model in which MMR proteins recognize and process these lesions, but since the damaged or altered base is located in the template strand, MMR is unable to repair the lesion. This leads to repeated rounds of unsuccessful repair attempts, eventually resulting in replication fork arrest, cell cycle checkpoint signaling and cell death. Evidence for this model comes from the observation that cell cycle checkpoint signaling in response to these agents involves the phosphorylation of checkpoint kinase 1 (CHK1) by ataxia telangiectasia Rad3 related (ATR), which in turn is activated by replication fork arrest (Caporali et al., 2004; Kaina, 2004; Stojic et al., 2004; Yamane et al., 2004; Yan et al., 2004).

There is, however, evidence that DNA damage can trigger cell cycle arrest and apoptosis in an MMR-dependent manner without the need to process DNA lesions. The *direct signaling* model suggests that upon sensing DNA damage, MutSα complexes transit along DNA and interact directly with cell cycle checkpoint factors to signal cell cycle arrest.
and apoptosis. First, cells from mice harboring mutations in the Msh2 or Msh6 ATPase domains are unable to bind mismatches, but retain the ability to signal apoptosis. Indeed, cells from these mice are sensitive to DNA damaging agents and respond by inducing apoptotic cell death at levels comparable to wildtype controls (Lin et al., 2004; Yang et al., 2004), suggesting that processing of DNA lesions is not required for apoptotic signaling. The second line of evidence comes from the observation that Msh2 forms complexes with ATR, CHK1, and CHK2 and nuclear foci with ATM/ATR-substrate CHK2-interacting zinc finger (ATCIZ) phosphoprotein in cells treated with methylating agents, suggesting that MMR proteins can bypass lesion processing and signal DNA damage directly to cell cycle checkpoint factors (Adamson et al., 2005; McNees et al., 2005; Wang and Qin, 2003). Although the mode of apoptosis signaling by MMR proteins is not fully resolved, it is possible that a combination of both models will account for the decreased levels of apoptosis in response to DNA damaging agents in MMR-deficient cells and that the exact mode of signaling will differ depending on the type of DNA damaging agent encountered.

1.4.8. Mismatch Repair and Cancer

MMR defects have also been identified in a number of malignancies, most notably hereditary non-polyposis colorectal cancer (HNPCC), also known as Lynch Syndrome. HNPCC accounts for ~3% of all colorectal cancers and is inherited as an autosomal dominant trait manifesting early at ~45 years of age. It is also characterized by an increased risk of malignancies in other tissues, most notably the endometrium, but also ovarian, gastric, cervical, breast, skin, lung, prostate, and bladder tissues. In addition, there is an increased risk for development of glioma, leukemia, and lymphoma. Table 1.1.
Table 1.1. Comparison of elevated cancer risk in HNPCC versus general population

<table>
<thead>
<tr>
<th>Tissue</th>
<th>Cancer Risk (%)</th>
<th>HNPCC</th>
<th>General population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colon</td>
<td>80</td>
<td>5.5</td>
<td></td>
</tr>
<tr>
<td>Endometrium</td>
<td>20-60</td>
<td>2.7</td>
<td></td>
</tr>
<tr>
<td>Stomach</td>
<td>11-19</td>
<td>&lt;1</td>
<td></td>
</tr>
<tr>
<td>Ovary</td>
<td>9-11</td>
<td>1.6</td>
<td></td>
</tr>
</tbody>
</table>

HNPCC – Hereditary non-polyposis colorectal cancer
Data obtained from (Hsieh and Yamane, 2008; Watson et al., 2001)
outlines the risk factors for various tissue malignancies in HNPCC versus the general population.

One of the hallmarks of HNPCC tumors is elevated microsatellite instability (MSI), observed as the expansion or contraction of one to four nucleotide repetitive sequences due to failed repair of IDLs by a defective MMR pathway (de la Chapelle, 2003; Lindor et al., 1998; Thibodeau et al., 1993). Indeed, MSI-positive tumor cells are impaired in the repair of base-base mismatches and IDLs and purified MutSα or MutLα complement this defect. Ninety percent of HNPCC cases are due to germline inactivating mutations in Msh2 or Mlh1, with slightly more mutations in Mlh1 than in Msh2 (Bronner et al., 1994; Fishel et al., 1993; Leach et al., 1993; Papadopoulos et al., 1994). Mutations in Msh2 usually lead to complete loss of the protein, while Mlh1 mutations are usually of the missense type. Approximately 10% of tumors have germline mutations in Msh6, while Pms2 mutations occur rarely, and the incidence of Msh3 mutations is uncertain (Hendriks et al., 2004; Nicolaides et al., 1994; Papadopoulos et al., 1994; Wu et al., 2001a; Wu et al., 2001b; Wu et al., 2001c). HNPCC tumors with Exo1 mutations have not been observed.

Defective MMR and elevated MSI (MSI⁺) are also observed in some sporadic, non-inherited colorectal tumors, the majority of which are caused by methylation-induced silencing of Mlh1 (Cunningham et al., 1998; Kane et al., 1997; Marcus et al., 1999). Biallelic mutations in one of the MMR genes result in a severe phenotype marked by neurofibromatosis type I (NF1) – a disorder that is manifested in childhood and is associated with skin discoloration, non-malignant tumors (neurofibromas), and malignancies in various tissues, including brain, colon and the blood, usually non-Hodkin lymphoma or acute lymphoblastic leukemia (Dunlop et al., 1997; Vasen et al., 1996). One of
the consequences of MMR deficiency in colorectal cancer patients is the resistance to commonly used cancer therapeutics, including cisplatin and 5-fluorouracil. In fact, only a minority of HNPCC sporadic colorectal cancer patients with high MSI respond to 5-fluorouracil-based therapy (Aebi et al., 1996; Carethers et al., 1999). The use of other drug-based regimens that are toxic to MMR-deficient tumors, such as topoisomerase inhibitors, is being investigated (Fallik et al., 2003; Jacob et al., 2001).

Impaired MMR manifests itself as shortening and lengthening of microsatellites, some of which are used as markers to classify MSI+ colorectal cancer. The MMR mutations that occur in these colorectal cancers affect protein-protein interactions, DNA and nucleotide binding, and ATPase activity of Msh2, Mlh1, or Msh6. Failure to repair IDLs leads to frameshift mutations in critical microsatellite-containing genes involved in the control of cell growth, differentiation, and cell cycle progression, which results in the dysregulation of these pathways. The epidermal growth factor pathway is frequently affected, with mutations occurring in downstream factors such as KRAS or BRAF and resulting in constitutive activation of the pathway and ultimately leading to proliferation, metastasis and anti-apoptotic signaling. KRAS is commonly mutated in HNPCC, with mutations in 48, 32, and 83% of tumors harboring Msh2, Mlh1, and Pms2 mutations, respectively, while only ~20% of sporadic colorectal cancers carry KRAS mutations (Oliveira et al., 2004). On the other hand, BRAF, which is activated downstream of KRAS, is more often mutated in sporadic tumors (Rajagopalan et al., 2002). Another pathway targeted by transformed cells is the transforming growth factor β (TGF-β) pathway, which suppresses tumor growth by inhibiting cell proliferation and promoting apoptosis (Wang et al., 1995). Inactivation of the TGF-β pathway frequently occurs via frameshift mutations in a 10-bp polyadenine repeat tract (Lu et al., 1995; Parsons et al., 1995). Similarly, an 8-bp
polyguanine tract in the BAX gene is subject to frameshift mutations, which lead to inactivation of the tumor suppressor.

Various knock-out mouse models are used to gain insight into the cancer predisposition of individuals harboring mutations in MMR genes. Some of the phenotypes, including MSI and tumor development are common to both species. However in contrast to humans in whom the inheritance of HNPCC occurs in an autosomal dominant pattern, mice require the inactivation of both MMR alleles in order to display increased cancer susceptibility while heterozygotes are phenotypically normal. Moreover, MMR-deficient mice develop predominantly lymphomas of the B or T cell type depending on the specific genotype, rather than colorectal cancers (Hsieh and Yamane, 2008). Tumors in other tissues such as the skin or the gastrointestinal tract are secondary to the dominant lymphoma phenotype. Furthermore, some, but not all, MMR-compromised mice are infertile. Table 1.2. summarizes the tumor, MSI, and fertility phenotypes associated with various MMR pathway knock-outs. The roles of Mlh1 and Mlh3 proteins in meiosis have been explored, but the reason for infertility associated with other MMR factors is not understood.
Table 1.2. Mouse MMR defects and associated tumor spectra

<table>
<thead>
<tr>
<th>Genotype</th>
<th>MSI</th>
<th>Tumor Spectrum</th>
<th>Fertility Male/Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Msh2</td>
<td>High</td>
<td>T-cell lymphoma, skin, GI</td>
<td>+/+</td>
</tr>
<tr>
<td>Msh3</td>
<td>Moderate-High</td>
<td>GI</td>
<td>+/+</td>
</tr>
<tr>
<td>Msh6</td>
<td>None-Low</td>
<td>B-cell lymphoma, skin, GI</td>
<td>+/+</td>
</tr>
<tr>
<td>Mlh1</td>
<td>High</td>
<td>T-cell lymphoma, skin, GI</td>
<td>-/-</td>
</tr>
<tr>
<td>Pms1</td>
<td>Low</td>
<td>None</td>
<td>+/+</td>
</tr>
<tr>
<td>Pms2</td>
<td>Low</td>
<td>Lymphoma, sarcoma</td>
<td>-/+</td>
</tr>
<tr>
<td>Exo1</td>
<td>Low-High</td>
<td>Lymphoma</td>
<td>-/-</td>
</tr>
</tbody>
</table>

MSI = Microsatellite instability; + = Fertile
Adapted from (Hsieh and Yamane, 2008)
1.5. Hypothesis and Objectives

**Hypothesis:**
The MMR pathway has been shown to interact with polymerase η to produce mutations at A:T base pairs, but the reason for this synergy is unknown. An important clue derives from the requirement for PCNA ubiquitination in this process. This modification induces translesion synthesis when normal DNA replication is blocked. Thus, I suggest that the presence of non-instructive DNA lesions in hypermutating B cells stimulates translesion synthesis in MMR tracts and leads to the production of mutations at A:T base pairs at the Ig locus. The bulk of this thesis (Chapters 2 and 3) investigates the association between different types of mutations (i.e. A:T and G:C mutations) in order to identify the lesion(s) that triggers mutagenic MMR. In Chapter 4, I explore the involvement of the MRN complex in the generation of MMR-dependent A:T mutations.

**Objectives:**

**Chapter 2:** Examine the relationship between mutations at A:T and G:C base pairs in Ramos cells, focusing on the distance between and relative strand location of A:T mutations and G:C transition or G:C transversion mutations. Inhibit the UNG/BER pathway to determine the effect on A:T mutations.

**Chapter 3:** Examine the relationship between mutations at A:T and G:C base pairs in hypermutating murine B cells using criteria similar to those used in Chapter 2. Explore the effect on G:C transition and transversion mutations in the absence of the MMR pathway.
Chapter 4: Study the effect of Mre11 knock-down on cell viability, MMR function and A:T mutation frequency in Ramos cells in order to determine whether the MRN complex contributes to A:T mutations in this Burkitt’s lymphoma cell line.
Chapter 2

Msh2 and UNG Co-Operate to Induce A:T Mutations in Ramos

Chapter 2 was published as:


M.L. performed experiments in Fig. 2.4A.
2.1. Abstract

Mismatch repair plays an essential role in reducing the cellular mutation load. Paradoxically, proteins in this pathway produce A:T mutations during the somatic hypermutation of immunoglobulin genes. Although recent evidence implicates the translesional DNA polymerase η in producing these mutations, it is unknown how this or other translesional polymerases are recruited to immunoglobulin genes, since these enzymes are not normally utilized in conventional mismatch repair. In this chapter, I demonstrate that A:T mutations were closely associated with transversion mutations at a deoxycytidine. Furthermore, deficiency in uracil-N-glycolase (UNG) or mismatch repair reduced this association. These data reveal a previously unknown interaction between the base excision and mismatch repair pathways and indicate that an abasic site generated by UNG within the mismatch repair tract recruits an error-prone polymerase, which then introduces A:T mutations. Our analysis further indicates that repair tracts typically are ~200 nucleotides long and that polymerase η makes ~1 error per 300 T nucleotides. The concerted action of Msh2 and UNG in stimulating A:T mutations also may have implications for mutagenesis at sites of spontaneous cytidine deamination.
2.2. Introduction

The affinity maturation of the antibody response depends on the somatic hypermutation (SHM) process. The enzyme activation-induced cytidine deaminase (AID) initiates SHM in germinal center B cells by deaminating C within immunoglobulin (Ig) genes, yielding a G:U lesion that is resolved by several mechanisms (Peled et al., 2008). Replication across the U generates G:C to A:T transition mutations, while the removal of the U by uracil-N-glycolase (UNG) leads to transversion and transition mutations at the original G:C base pair (Rada et al., 2002b). The AID-generated G:U lesion is also a substrate for the mismatch repair (MMR) proteins Msh2, Msh6, and Exo1. Unlike their normal role in DNA repair, the processing of this lesion by these MMR proteins during SHM paradoxically leads to the production of mutations at A:T base pairs (see below).

MMR is a DNA repair process utilized by prokaryotes and eukaryotes (Modrich and Lahue, 1996). This pathway repairs DNA errors caused by the misincorporation of nucleotides during DNA synthesis. The initial mismatch is detected by MutSα, which consists of Msh2 and Msh6 in mammalian cells. The ability of MMR to discriminate between the mutated and unmutated strands of DNA is thought to be dictated by nicks or gaps on the newly synthesized lagging strand between Okazaki fragments or by strand ends on the leading strand at the replication fork (Kunkel and Erie, 2005). The MutLα endonuclease (Mlh1-Pms2) uses the DNA nick or end as a marker of the newly synthesized, and therefore mutated, strand to introduce a new nick on either side of the mismatch (Kadyrov et al., 2006). This nicked strand is then excised by the 5’-to-3’ exonuclease Exo1, and the ensuing gap is repaired by the replicative polymerase δ. However, since AID acts primarily during G1 of the cell cycle (Faili et al., 2002b; Schrader et al., 2007), it is unclear
whether Msh2-Msh6 is capable of distinguishing between the AID-mutated and unmutated strands prior to strand excision. Consistently with their role in DNA repair, deficiency in Msh2, Msh6, or Exo1 generally leads to an increase in mutation frequencies in different tissues (Wei et al., 2003). However, in the case of SHM of Ig genes, the loss of these MMR proteins reduces the frequency of mutations at A:T base pairs (Bardwell et al., 2004; Cascalho et al., 1998; Ehrenstein and Neuberger, 1999; Kim et al., 1999; Martin et al., 2003; Phung et al., 1998; Rada et al., 2004; Shen et al., 2006; Wiesendanger et al., 2000; Wilson et al., 2005). One possible difference between conventional and mutagenic MMR is the involvement of the error-prone DNA polymerase η in the latter process. Indeed, both mice and humans lacking polymerase η resemble Msh2-deficient mice, in that mutations at A:T base pairs in the V region are less frequent (Delbos et al., 2007; Delbos et al., 2005; Zeng et al., 2001). Moreover, the error spectrum of polymerase η on undamaged DNA correlates with the mutation spectrum of A:T mutations in the V region (Rogozin et al., 2001). While it is now well established that mutations at A:T base pairs are produced largely by proteins involved in the MMR pathway, it is not known how DNA polymerase η is recruited during SHM. One possible explanation for the use of error-prone polymerases is the occurrence of replication-blocking lesions, such as an abasic site or a modified nucleotide, in the V region of Ig genes. Evidence that a replication block leads to mutagenic MMR comes from recent studies showing the requirement of ubiquitinated PCNA for mutagenic MMR (Arakawa et al., 2006; Langerak et al., 2007; Roa et al., 2008). Monoubiquitination at the K164 residue of PCNA in response to DNA damage leads to translesional synthesis (Arakawa et al., 2006), and SHM at A:T base pairs is reduced in PCNA^{K164R/K164R} mice to levels observed in MMR-deficient mice (Langerak et al., 2007; Roa et al., 2008). In addition, the finding that translesional DNA polymerases are involved in SHM (Diaz et al., 2001; Poltoratsky et al.,
2000; Zan et al., 2001; Zeng et al., 2004; Zeng et al., 2001) suggests that replication-blocking lesions are common at the Ig locus during SHM. Taken together, these observations suggest a model in which replication-blocking lesions recruit error-prone polymerases, which then generate mutations at nearby A:T base pairs. As reported here, we have tested this model by examining the correlated mutations in V region sequences from hypermutating Ramos cells.
2.3. **Materials and Methods**

2.3.1. **Cell culture, subcloning, and flow cytometry**

Ramos 67 cells were maintained as previously described (Zhang et al., 2001), and Abelson pre-B cell lines 15-63 (Msh2+/−) and 8-58 (Msh2−/−) were maintained in RPMI medium (Invitrogen) with 10% bovine calf serum (HyClone), penicillin (100 U/ml), and streptomycin (0.1 mg/ml; Sigma). For subcloning, Ramos 67 cells were plated at 0.1 cell per well into 96-well plates. After ~15 cell divisions (~32,000 cells), cells were harvested and prepared for flow cytometry cell sorting or enzyme-linked immunospot (ELISPOT) assays. To isolate IgM-positive and IgM-negative clones, Ramos cells were stained with fluorescein isothiocyanate- or biotin-conjugated anti-IgM Fab fragment antibody (Jackson ImmunoResearch Laboratories), with the latter followed by staining with allophycocyanin-conjugated-streptavidin (eBioscience), and single IgM-positive and IgM-negative cells were sorted (FacsAria; BD) directly into 96-well plates. Based on the previously measured mutation rate in these cells (10⁻⁵ mutation/bp/generation) (Zhang et al., 2001), we estimate that under this protocol, 15% of IgM-reverted cells had undergone multiple independent mutational events in the 1,000-bp V region (10⁻⁵ mutation/bp/generation X 1,000 bp X 15 generations = 0.15 mutation per V region).

2.3.2. **Measurement of MMR activity**

To quantify MMR activity, pCA-OF-expressing clones were harvested and washed in phosphate-buffered saline (PBS) (Gibco, Invitrogen), and the green fluorescent protein (GFP) reversion frequency was determined using a flow cytometer (FACSCalibur; BD) as
previously described (Vo et al., 2005). Flow cytometry data were analyzed using FlowJo software.

### 2.3.3. ELISPOT assays

The ELISPOT assay for IgM secretion was performed as previously described (Martin et al., 2002).

### 2.3.4. Plasmids and transfections

To measure MMR activity, the microsatellite-like plasmid pCA-OF was used. To inhibit UNG activity, pEF (control) and pEF-UGI plasmids were used. Plasmids were linearized with MluI (pEF and pEF-UGI) and BglII (pCA-OF). For cell transfections, ~4 X 10^6 log-phase cells were mixed with 10 μg of linearized plasmid DNA in 4-mm cuvettes and electroporated (Gene Pulser Xcell) at 250 V and 950 F for Ramos cells and at 450 V, 950 F, and 150 Ω for the pre-B cells. Cells were diluted in appropriate media and plated in 96-well plates. After incubation at 37°C for 24 h, stable clones were selected with puromycin (pEF and pEF-UGI; 0.8 μg/ml) or blasticidin (pCA-OF; 2.5 μg/ml for Ramos and 25 μg/ml for pre-B cells).

### 2.3.5. DNA extraction, PCR, and sequencing in Ramos

Genomic DNA was extracted as previously described (Martin et al., 2002). For V region amplification from IgM-positive and IgM-negative clones, Taq polymerase was used with the following cycling parameters: 95°C for 2 min for 1 cycle, and then 35 cycles of 95°C for 45 s, 58°C for 30 s, and 72°C for 90 s. The forward and reverse primers were 5’RamV5316 (5’ ACAGCCAGCATACACCTCCC) and 3’RamV6209 (5’ CAACCTGAGTCCCATTTTCC), respectively. PCR products were purified using the Wizard PCR Preps purification system.
(Promega) according to the manufacturer’s specifications and sequenced using 5’ RamV_Inner (5’ CACCAACTACAACCCGTCCC) and 3’ RamV_Inner (5’GTGGCCATTCTTACCTGAGG). To measure V region mutation rates in unselected clones by PCR, amplifications were performed on DNA from unselected Ramos clones using PFU Ultra II (Stratagene) as previously described (Martin et al., 2002).

2.3.6. **In vitro UDG assay**

The inhibition of UNG by uracil-DNA glycosylase inhibitor (UGI) was confirmed using the uracil DNA glycosylase assay as previously described (Di Noia and Neuberger, 2002), with minor modifications. Briefly, a double-stranded oligonucleotide containing a single U:G mismatch was 5’ labeled with $[\gamma-32P]$ dATP. The labeled substrate was incubated with 1 U of uracil DNA glycosylase (NEB) or with serially diluted Ramos nuclear extracts (1 to 10 $\mu$g) in uracil glycosylase buffer (NEB) for 3 h at 37°C, followed by incubation with sodium hydroxide (100 mM) for 10 min at 98°C. Samples were electrophoresed on a 20% denaturing acrylamide gel with a running buffer of 1X TBE (Tris-borate-EDTA) at 300 V for 3 h and visualized using a PhosphorImager (Molecular Dynamics). Quantitation was performed using ImageQuant software, version 5.0 (Molecular Dynamics).

2.3.7. **Statistical analysis.**

Ramos data were graphed using GraphPad software (Prism), and statistical analyses were performed using the unpaired t-test, Fisher’s exact test, and Mann-Whitney test.
2.4. Results

2.4.1. Transversion mutations at C are linked to A:T mutations

As noted above, SHM of both G:C and A:T base pairs depends wholly on AID, while A:T mutations depend additionally on mismatch repair and potentially on replication-blocking lesions. These observations suggest the specific model illustrated in Fig. 2.1.A. As in other models, the G:U mismatch generated by AID is detected by the MMR system. Since translesional synthesis is the final stage of MutS\(\alpha\)-mediated repair during SHM, I have investigated the mutational outcomes of MutS\(\alpha\) when it targeted either the AID-mutated or unmutated DNA strand. Assuming that MutS\(\alpha\) targets both strands equally for repair, 50% of the time the mutated (U-containing) strand is degraded by Exo1. The resynthesis of the ensuing gap restores the original G:C base pair. If, however, the unmutated (G-containing) strand is excised, as illustrated in Fig. 2.1.A., the degradation of the unmutated strand by Msh2/Exo1 exposes U in the opposite strand. The excision of the U by UNG, which is three-fold more active on single-stranded DNA (ssDNA) than on double-stranded DNA, generates an abasic site (Krokan and Wittwer, 1981). The resynthesis of the degradation tract begins faithfully but stalls at the abasic site, thus inducing the ubiquitination of PCNA and the assembly of translesional polymerases that can then bypass this lesion. Random nucleotide insertion opposite the abasic site leads to a G:C transversion mutation 50% of the time; further extension 3’ by the translesional polymerase \(\eta\) then introduce A:T mutations into the V region. The scenario described above predicts that sequences with mutations at A:T base pairs will be enriched for mutations at G:C base pairs and that the G:C mutations will be predominantly transversions. To test these predictions, we examined sequences from...
Figure 2.1. Model for mutagenic repair by MutSα

(A) A cytidine in the representative DNA sequence is deaminated by AID, thus generating a G:U lesion, which stimulates the MMR-directed Exo1 excision of the unmutated DNA strand. This yields a U in the opposite strand within an ssDNA region. The removal of the U by UNG generates an abasic site. The resynthesis of the excised strand begins with an error-free replicative polymerase until the abasic site is encountered, causing the polymerase to stall. This leads to translesional synthesis, with the recruitment of error-prone polymerases to bypass the abasic site as well as polymerase η to generate mutations at A:T base pairs. (B) On the top is a schematic of the IgH locus of Ramos 67 (R67) cells with the position of the TAA nonsense codon represented by an arrow. On the bottom is a diagram describing the enrichment for Ramos clones containing A:T mutations, including subcloning, sorting based on IgM expression, and the sequencing of the V region. FITC, fluorescein isothiocyanate. Figure is not to scale.
Figure 2.1.
hypermutating Burkitt’s lymphoma Ramos cells. We used an IgM-negative variant Ramos 67 that harbors a TAA nonsense codon in the endogenous Ig heavy chain (IgH) locus, truncating translation in the variable domain. Consequently, sorting for IgM-positive Ramos 67 cells was expected to select for cells that had undergone a mutation at an A:T base pair. Therefore IgM-positive cells were isolated and the associated mutations in an ~1-kb V region containing segment of the IgH locus were examined.

In these experiments, it was important to minimize the occurrence of unrelated mutations. The mutation rate of the V region in Ramos 67 is known to be ~10^{-5} mutations/bp/generation (Zhang et al., 2001). Therefore, Ramos 67 was subcloned and allowed to expand only ~15 cell divisions before sorting. This protocol was expected to yield an IgM-positive cell population in which only ~15% of cells had undergone multiple independent mutations within the V region being analyzed (see Materials and Methods for the calculations). Subclones of Ramos 67 were stained with fluorescein isothiocyanate-labeled anti-IgM, and IgM-positive and -negative clones were sorted directly into individual wells on a 96-well plate. Sorted cells were then expanded, genomic DNA was extracted, and an ~1-kb region encompassing the V region was amplified by PCR (Fig. 2.1.B.).

To minimize PCR errors, the PCR products were sequenced directly. To assess the likelihood that G:C mutations were associated with A:T mutations, the excess of G:C mutations in the IgM-positive and IgM-negative clones was measured. Under conditions in which the proliferation was limited to ~15 cell divisions, most unselected IgM-negative Ramos clones did not harbor a mutation. Indeed, there was an ~5-fold increase in mutation frequency (not including the reverted nonsense codon) in selected IgM-positive clones compared to that of the unselected IgM-negative clones (Fig. 2.2.A.), indicating that selecting for mutations at A:T base pairs enriched for other types of mutations.
Of 51 independent IgM-positive clones, 21 clones (41%) contained mutations at G:C base pairs, whereas of 68 unselected clones only 8 clones (12%) had a mutated G:C (Fig. 2.2.B. and Table 2.1.), supporting the model shown in Fig. 2.1.A. that mutation at A:T enriches for mutations at G:C. Sequences derived from Ramos 67 IgM-positive clones with associated G:C mutations are shown in Appendix 6.1. Of note, the TAA nonsense codon in all IgM-positive clones was reverted exclusively to either TAC or TAT (see Appendix 6.1.), which encodes tyrosine. Two mechanisms that cause this limited set of revertants might be due to the mutagenic preference of polymerase η (Mayorov et al., 2005; Rogozin et al., 2001) or the assembly of IgM. That is, the TAA codon lies opposite the sequence 5’-TTA-3’, and polymerase η preferentially misincorporates nucleotides opposite a T in the motif TT or TA (the underlined nucleotide is mutated in the preferred sequence). The reversion pattern most probably reflects the fact that the nonsense codon is located within the conserved Tyr-Tyr-Cys motif in the FR3 region and indicates that other amino acids in this motif do not allow the assembly of membrane-bound IgM.

A comparison of the types of mutations in the IgM-positive and IgM-negative clones was conducted (Fig. 2.2.C. Table 2.1.). The mutations in the unselected IgM-negative clones were predominantly transition mutations at G:C base pairs, similarly to that reported earlier (Zhang et al., 2001). However, the mutation spectrum of the IgM-positive pool differed significantly from that of the IgM-negative pool, in that there were more A:T mutations ($P = 0.0204$) and more G:C transversion mutations ($P < 0.0001$). Seventy-five to 80% of G:C mutations in both populations occurred in AID hot spot motifs (i.e., WRC), suggesting that these mutations arose from a processed AID deamination event. Moreover, 65% of mutations at A occurred at polymerase η WA motifs (40% of A in the Ramos V region are in WA motifs), as expected for mutations due to the SHM process.
Figure 2.2. Sequences with A:T mutations are enriched for transversion mutations at C on the bottom strand in an UNG-dependent manner.

(A) Mutation frequency at the V region in R67 clones is shown for sorted IgM+ and IgM- clones. Only unique mutations from independent clones were scored. The reverting mutation at the TAA nonsense codon was excluded. The data represent 51 IgM+ and 68 IgM- R67 clones and 18 IgM+ and 52 IgM- UGI-expressing R67 clones. Statistical analyses were conducted using the Fisher's exact test (***, P < 0.0001; *, P = 0.0284). (B) The percentage of sequences containing associated A:T and G:C mutations was determined for reverted IgM+ untransfected (left) and UGI-expressing (right) R67 clones. (C) Mutation characteristics in the V region of IgM+ and IgM- clones are shown. Individual mutations are represented by white, gray, or black circles. The location of C transition (TRN) and C transversion (TRV) mutations and A mutations (MUT) relative to the reverted mutation at the TAA nonsense codon (vertical line at x = 0) also is depicted. Mutations on the top strand (above the midline) and bottom strand (below the midline) are shown for R67 IgM+ (top), R67 IgM- (middle), and R67 IgM+ UGI-expressing clones (bottom). Statistical analyses to compare top and bottom strand mutations were conducted using the t-test (***, P < 0.0001 for transversion mutations at C [top panel]; **, P = 0.0052 for mutations at A [top panel]; *, P = 0.0486 for transition mutations at C [bottom panel]).
Figure 2.2.
Table 2.1. Characteristics of mutations derived from Ramos clones.

|                                | IgM+  | IgM-  | Ugi-IgM+ | Ugi-IgM- |
|                                |       |       |          |          |
| Independent clones             | 51    | 68    | 18       | 52       |
| Nucleotides sequenced          | 44128 | 59394 | 12079    | 33808    |
| Total mutations\(^a\)          | 36    | 10    | 13       | 13       |
| Mutation Frequency (per bp x 10\(^{-4}\)) | 8.16  | 1.68  | 10.76    | 3.85     |
| Clones with C transition mutation (%) | 5 (10) | 5 (7)  | 5 (28)   | 9 (17)   |
| Clones with C transversion mutation (%) | 16 (31) | 3 (5)  | 3 (17)   | 1 (2)    |
| Clones with A mutation (%)     | 9 (18) | 0 (0)  | 2 (11)   | 2 (4)    |
| G:C mutations (%)              | 26 (70) | 10 (100) | 11 (85)  | 11 (85)  |
| G:C transversion mutations/total G:C (%) | 21/26 (81) | 3/10 (30) | 3/11 (27) | 1/11 (9) |
| G:C mutations at WRC/GYW motif (%) | 17 (65) | 8 (80)  | 7 (64)   | 7 (64)   |
| A:T mutations at WA/TW motif (%) | 7 (70)  | NA     | 1 (50)   | 2 (100)  |

\(^a\) Does not include the mutation at the TAA nonsense codon in IgM+ clones
We also examined the location of these mutations with respect to the TAA nonsense codon (Fig. 2.2.C.). Mutations at G:C base pairs are initiated by altering the C. If the C is in the top (or bottom) strand, mutations at G:C base pairs correspondingly are deemed to be mutations in the top (or bottom) strand. In the case of the unselected IgM-negative clones, 30% of mutations in these clones were found within a distance of 100 bp either upstream or downstream of the TAA codon. In contrast, in the IgM-positive clones, 71% of all mutations were localized within 100 nucleotides on either side of the TAA codon. These data further corroborate the notion that most mutations in the IgM-positive clones were associated with mutations at the TAA codon. Moreover, these results suggest that the typical repair tract is ~200 nucleotides.

Strikingly, C transversion mutations clustered on the bottom strand in IgM-positive cells ($P < 0.0001$), while no bias was seen in the unselected IgM-negative clones ($P = 0.6874$). These transversion mutations in the bottom strand were found equally upstream and downstream of the A:T mutation. This pattern of clustering is not predicted by the model in Fig. 2.1.A. That is, this model hypothesizes that polymerase η is recruited when replication is blocked at an abasic site and then proceeds to introduce mutations at A:T base pairs. This hypothesis therefore predicts that A:T mutations will be associated with an upstream C in the bottom strand and a downstream C in the top strand. The implications of this observation will be considered in the Discussion (Section 2.5.).

The reversion of the TAA codon was also associated with the occurrence of other A:T mutations. As considered in the Discussion, the frequency of these A:T mutations can be used to estimate the error rate of the translesional polymerase. Eight out of these 10 A:T mutations occurred with A on the top strand ($P = 0.005$). This is consistent with observations in wild-type mice, which also have been interpreted to reflect the preferential
excision of the top strand by Msh2/Exo1, followed by preferential misincorporation opposite T (Mayorov et al., 2005; Xiao et al., 2007) (see Fig. 3.1. C.). In keeping with the strong preference for misincorporation opposite T, in this thesis, mutations at A:T base pairs will be referred to as A mutations.

### 2.4.2. AID does not show a strong strand bias preference.

I next tested whether the strand bias of the C transversion mutations reflects a preferential specificity for AID mutating the bottom strand in Ramos cells. For this analysis, I assembled previously obtained sequence data from Ramos clones that were carried in culture but not subjected to any form of selection (e.g., IgM expression). Only unique mutations at G:C base pairs were included in this analysis, unless genealogies indicated that the mutation was unique or that the same mutation occurred in different clones. As suggested by the data in Fig. 2.4., after correcting for base composition, AID marginally prefers to mutate the top strand by a ratio of 0.48:0.44 (63 mutations/131 C on the top strand, 67 mutations/151 C on the bottom strand), supporting previous findings (Martomo et al., 2005; Xiao et al., 2007). In addition, using this data set, I found that mutations at AID hot spot motifs (i.e., WRC) occur on both strands at approximately similar frequencies (Fig. 2.3.), which is consistent with the notion that AID mutates both strands approximately equally. This suggests that the bias for C transversion mutations on the bottom strand in IgM-positive clones is not due to AID directly but likely reflects a strand preference of the repair process (see below).
Figure 2.3. Mutations at G:C basepairs occur equally on both top and bottom strands in Ramos cells.

(A) Data assembled from previous publications (Martin et al., 2002; Martin and Scharff, 2002) and from unpublished work showing unique mutations at G:C basepairs in the V region of Ramos clones. (B) The total number of unique mutations at each nucleotide is depicted. (C) The same data showing the distribution of the deoxycytidine mutations on the top (top) or bottom (bottom) strands.
Figure 2.3.

A

TTCTCCCTCC TCTCTGTGTC AAGTTTCTGA GTGGATGTC TCGATATGC 50
GCTATGCTAG ATAGAGATGT GCTGCTGTGA TCCTAGCTAT CACTGTGGGT 100
TTTTCGCTTC ACAGGGGTCG TGTCCAGAG TCTTCCTAGG GAAGGCGGGG 150
ACAAAGTTTT GAGGCCCTTC GAGACCTCTT CACCTCCCTG GGCTCTTTAT 200
A A T C A T A G A G A A T T T A 250
GCTGGTCCCT TCACTGGTTA CTAATGACC TGAATCCGAG AAGCTCAGG
G A A A T A T A T A A T 300
C A C A G T T A A G A A T G A G
A A T T C T C A C T C A A T A A 350
ACAAAGCTCT CTTCTGATCT CTGCTTTCTA TACTGCTCTA ACGACCAACT
T T C T A C T A C T C T A A A 400
A A A A C A A T G T G T G T G T
A G T G C T G A A T A 450
GTTATCTTGG GCGAAGATTA TTACTAGGCG GAGTCTGTGC AGCAACAGGA
C C G T T A A A G C 500
C C G T T G A T A T A 550
GGTGACATAT GGAACCTCTG GGGCCAGAGA CGCAAGTCTC CCGTCCCTCA
A A G T T T T 600
GTTGACATAT GGAACCTCTG GGGCCAGAGA CGCAAGTCTC CCGTCCCTCA
A A G T T T T 650

B

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>C</th>
<th>G</th>
<th>T</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>-</td>
<td>0</td>
<td>6</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>C</td>
<td>16</td>
<td>15</td>
<td>30</td>
<td>63</td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>39</td>
<td>17</td>
<td>11</td>
<td>67</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>5</td>
<td>2</td>
<td>4</td>
<td>11</td>
<td></td>
</tr>
</tbody>
</table>

V-region nucleotides:
A: 104
C: 131
G: 151
T: 114

C

Top strand

Bottom strand
2.4.3. **Inhibition of UNG abrogates linked C transversion and A mutations.**

An abasic site produced by a DNA glycosylase is likely an intermediate for C transversion mutations within AID hot spot motifs. To test whether these transversions depend on UNG, I inhibited UNG by stably transfecting Ramos cells with a UGI-expressing plasmid (Di Noia and Neuberger, 2002) and examined whether the inhibition of UNG altered the frequency or spectrum of mutations. UGI is an UNG inhibitor protein from the *B. subtilis* bacteriophages PBS1/2 that mimics DNA and is designed to complement UNG’s surface (Parikh et al., 2000). I screened transfected clones and measured residual UNG activity in nuclear extracts using a previously described assay (Di Noia and Neuberger, 2002). As shown in Fig. 2.5.A., nuclear extracts from UGI-expressing Ramos 67 had no detectable residual glycosylase activity relative to that of empty vector controls, indicating that UNG is the major glycosylase in Ramos cells. By sequencing the V region in unselected Ramos clones, I found that transversion mutations at G:C base pairs were reduced in UGI-expressing clones by at least six-fold (58.7 X 10^{-6} versus 9.01 X 10^{-6}; *P* = 0.015) (Fig. 2.5.B.), similarly to that reported for UNG−/− mice (Rada et al., 2002b). To determine the effect of UGI expression on the mutation frequency at A:T base pairs, I measured the reversion frequency of the TAA codon using an ELISPOT assay for IgM secretion (Fig. 2.4.C.). The IgM reversion frequency in UGI-transfected clones was reduced ~three-fold relative to that of the empty vector control (*P* = 0.007), indicating that at least ~2/3 of the mutations at A:T base pairs depended on UNG, possibly in concert with the Msh2/Exo1 pathway (see below). As expected, UNG inhibition did not perturb MMR activity, as measured by a microsatellite instability assay (Fig. 2.4.D.) (Vo et al., 2005).
Figure 2.4. Expression of UNG inhibitor UGI in R67 cells reduces both A:T mutations and C transversion mutations.

(A) The level of UNG activity in control and UGI-expressing Ramos clones was determined using a substrate cleavage assay. A substrate containing a single U:G mismatch was incubated alone (lane 1), with uracil DNA glycosylase (lane 2), or with 3 X serial dilutions of nuclear extracts from two control (lanes 3 to 8) and two UGI-expressing (lanes 9 to 14) R67 clones. TRN, transition; TRV, transversion; MUT, mutation. (B) Mutation characteristics in the V region of control (left) and UGI-expressing (right) R67 clones are shown. Statistical analyses comparing mutation types were performed using the t-test (the asterisk indicates \( P = 0.0150 \) for G:C transition mutations and \( P = 0.0492 \) for G:C transversion mutations). TRN, transition; TRV, transversion; MUT, mutation. (C) The IgM reversion frequency was determined using the ELISPOT assay for IgM secretion. Individual subclones are represented as light gray circles; black circles represent clones in which no IgM reversion events were observed among \( \sim 5 \times 10^{-5} \) cells. Statistical analyses comparing median reversion frequencies (indicated by numbers at bottom of the panel) were performed using the Mann-Whitney test (***, \( P = 0.0007 \)). (D) MMR activity for Msh2+/− and Msh2−/− pre-B cell lines and R67 clones was quantified using a microsatellite instability assay based on the GFP reversion frequency. Median reversion frequencies are indicated by numbers at bottom of the panel. The GFP open reading frame is interrupted with an out-of-frame dinucleotide microsatellite-like sequence (Vo et al., 2005), and microsatellite instability can lead to GFP expression. Statistical analyses were performed using the Mann-Whitney test (**, \( P = 0.003 \) for results with Msh2−/− pre-B cells compared to those with Msh2−/− pre-B cells).
Figure 2.4.
To determine whether UNG activity is required for linked C transversion and A mutations, I measured whether UGI expression reduced the fraction of IgM-positive clones in which the reversion mutation was associated with a C transversion. As shown in Fig. 2.2.B. and 2.2.C., ~31% (16/51) of IgM-positive Ramos 67 clones had a linked transversion, compared to ~17% (3/18) for the UGI-expressing cells. Combined with the reduced mutation frequency at A:T base pairs in UGI-expressing Ramos cells (Fig. 2.4.C.), these data indicate that ~2/3 of A:T mutations in Ramos cells are produced by replication across an UNG-generated abasic site on the bottom strand. Although A:T mutations occurred when UNG was inhibited (Fig. 2.2.C. and 2.4.C.), the residual A:T mutations were not associated with C transversions. These data suggest that a mechanism that is independent of an abasic site also generates A:T mutations during SHM.
2.5. Discussion

SHM is initiated by the AID-mediated deamination of cytidines. Proteins involved in the MMR pathway have been co-opted by the SHM process to extend mutagenesis from this initial G:C base pair to A:T base pairs. To gain insight into this process, I used a hypermutating B cell line, Ramos, in which I selected for mutations at a specific A:T base pair and then analyzed the other, correlated mutations. I have interpreted these results based on the model in Fig. 2.1.A. This model postulates that the creation of an abasic site, the result of cytidine deamination followed by uridine excision, recruits an error-prone polymerase that introduces mutations at A:T base pairs. As predicted by the model, I observed that C transversions, a proxy measurement for the replication of an abasic site, were strongly associated with A:T mutations in Ramos cells (Fig. 2.2.).

It is evident from these data and from data for UNG−/− mice (Rada et al., 2002b) that not all mutations at A:T base pairs arise from this pathway, as A:T mutations occurred even when UNG was deficient (Fig. 2.2.A.), and the residual A:T mutations were not strongly correlated with the occurrence of C transversions (Fig. 2.2.B.). A possible explanation is that these mutations result from an MMR-dependent but UNG-independent mechanism. These data argue that in the absence of UNG, A:T mutations are made by a qualitatively different mechanism, perhaps still MMR-dependent but not involving an abasic site.

The analysis of C transition mutations suggested that both top and bottom strands were mutated to a similar extent (see Fig. 2.4.), indicating that AID targets both strands equally. However, C transversions occurred predominantly in the bottom strand (Fig. 2.2.), suggesting that the MMR system preferentially excises the top strand, leading to DNA polymerase η-induced A mutations in the top strand. One potential explanation for the
preferential excision of the top strand by the MMR pathway is that the lagging strand is synthesized off the bottom strand during DNA replication. This would place Okazaki fragments on the top strand in the new daughter cell. Since the MMR system repairs mismatches more efficiently on the lagging strand than on the leading strand (Pavlov et al., 2003), this would lead to the preferential excision of the top strand (Pavlov et al., 2003).

The disproportionate position of Okazaki fragments in the top strand could occur if the Ig genes in B cells were replicated by an origin of replication that is located 3’ of the V region, such as those found in the 3’ enhancer region (Zhou et al., 2002) and the intronic enhancer (Ariizumi et al., 1993).

The model illustrated in Fig. 2.1.A., in which translesional synthesis begins at the abasic site, predicts that all C transversion mutations on the bottom strand are upstream of the mutated A:T base pair. However, I found that bottom strand C transversions occurred both upstream and downstream of the mutated A:T base pair in Ramos cells (Fig. 2.2.). One explanation has been suggested by Ohm-Laursen and Barington, who observed an inverse correlation between the mutation rate at A:T base pairs and the distance to the nearest 3’ WRC motif (Ohm-Laursen and Barington, 2007). The authors proposed that this correlation was due to a 3’-to-5’ exonuclease or endonuclease, such as MRE11, which has been shown to increase SHM when ectopically overexpressed in Ramos cells (Yabuki et al., 2005). The 3’-to-5’ exo/endonuclease would be recruited after replication stalls at the abasic site, thus extending the gap (~100 nucleotides) to include an upstream A:T base pair, followed by error-prone gap filling. As a result, A mutations are induced both upstream and downstream of the abasic site, resulting in the accumulation of C transversion mutations on the bottom strand located 3’ and 5’ of the A mutation.
These data reveal that in sequences containing A:T mutations with associated C transversion mutations, a C-to-G mutation predominates. This indicates that a deoxycytidine is frequently inserted opposite the abasic site. I propose that Rev1, a known deoxycytidyl transferase, is involved at this step. Studies of yeast show that Rev1 preferentially inserts C opposite abasic sites in a gapped duplex substrate (Gibbs and Lawrence, 1995), which mimics an MMR-induced excision tract containing an abasic site in the template strand. Moreover, C-to-G transversion mutations were significantly reduced in mutated Ig genes in Rev1-deficient mice (Jansen et al., 2006). Although Rev1 efficiently inserts C opposite abasic sites and other lesions, it does not readily extend from them (Haracska et al., 2001). Following the initial insertion event, Rev1 is replaced by a second translesional polymerase that can extend from a C-abasic site mispair, and in hypermutating B cells this is likely polymerase η. The reason for the utilization of polymerase η is not fully understood but may be related to the finding that Msh2-Msh6 proteins associate physically with, and stimulate the catalytic activity of, polymerase η (Wilson et al., 2005).

Figure 2.2. shows that most mutations at G:C base pairs were located within 100 nucleotides on either side of the mutated A:T base pair. These data suggest that the Exo1 degradation tract is ~200 nucleotides in length. Similar results were found in a study that examined the length of MMR-dependent ssDNA in yeast and mammalian cells (Mojas et al., 2007). Using electron microscopy, the authors demonstrated that the length of Exo1 tracts peaked at ~200 nucleotides. Furthermore, Unniraman and Schatz (Unniraman and Schatz, 2007) showed using a transgenic mouse model that mutations at A:T base pairs accumulated up to ~30 nucleotides away from a G:C-rich tract, suggesting that the Exo1 tract is ~60 nucleotides in length. Based on data generated in this study as well as by Mojas
et al. and Unniraman and Schatz, I conclude that the excision tract produced by Exo1 during MMR is ~200 nucleotides (Mojas et al., 2007; Unniraman and Schatz, 2007).

This analysis also can be used to estimate the error rate of the translesional polymerase η in vivo. Ten of the 51 V genes that I sequenced from the IgM-positive Ramos revertants contained A:T mutations in addition to the A:T mutation that restored IgM production; 8 of these 10 mutations had A in the top strand. Considering that transversions at G:C occurred mostly with C on the bottom strand, these eight A:T mutations probably represent misincorporations opposite T in the bottom strand. These eight A mutations were generated in the course of repairing 51 excision tracts (i.e., 51 revertant clones analyzed) that is estimated to be ~200 nucleotides long in a sequence of 100 bp on either side of the TAA nonsense codon, which contains 24% T's. This calculation thus implies that the error rate of polymerase η opposite template T is ~3.3 X 10^{-3} (i.e., 8/[51 X 200 X 0.24]). This error rate is ~10-fold lower than the in vitro error rate of human polymerase η (3.5 X 10^{-2}) (Rogozin et al., 2001), which may be due to the repair of polymerase η-generated errors in this in vivo system.

Like other seeming paradoxes, the discovery that somatic hypermutation co-opt machinery that otherwise prevents mutations has proved very instructive. The lesson from the immune system is that a G:U mismatch, which AID creates frequently by deaminating cytidine, engenders further mutations. This process of mutagenesis has implications for non-B cells, because all cells must cope with the spontaneous deamination of cytidine. Although the spontaneous rate of deamination is low, the target size (all of the C's in the genome) is large, so the mutational load is potentially high. If MMR and UNG were to act on these spontaneous G:U mismatches in the same manner as that in Ig genes, the system would extend mutagenesis beyond the G:U lesion, clearly a deleterious situation for non-B
cells. One can anticipate from this apparent paradox that some method of avoiding the hypermutation problem has evolved. The solution to this puzzle will prove very interesting to elucidate.
Chapter 3

A:T Mutations in Murine Immunoglobulin Sequences Are Stimulated by the Concerted Action of UNG and Msh2

Chapter 3 was published as:

3.1. Abstract

Somatic hypermutation of the immunoglobulin locus requires the participation of activation-induced cytidine deaminase, uracil-$N$-glycosylase (UNG) and the base excision repair (BER) pathway, and the mismatch repair (MMR) pathway. Deficiency in MMR proteins reduces the A:T mutation load in the V region as does deficiency in the translesional polymerase η and elimination of the ubiquitination sites on PCNA, suggesting that translesional synthesis is important for generating A:T mutations. In this study, I examined the association between A:T and G:C mutations in the V regions of hypermutating murine B cells. I found that A:T mutations are closely associated with G:C transversion mutations on the opposite DNA strand and that this relationship is dependent on the presence of UNG. I further observed that MMR-deficiency results in the selective reduction of G:C transversion mutations on the transcribed DNA strand, indicating that a portion of these mutations is contributed by MMR. Together, these data support a mechanism whereby the MMR and BER pathways interact to produce abasic sites in MMR/Exo1-degradation tracts which stimulate translesion synthesis and the production of mutations at A:T base pairs.
3.2. Introduction

The production of high affinity antibodies by B cells during an immune response occurs via iterative rounds of targeted mutagenesis at the immunoglobulin (Ig) locus via somatic hypermutation (SHM) and the selection of B cells expressing higher affinity BCRs in the germinal centre environment. Mutations are introduced by the enzyme activation-induced cytidine deaminase (AID), which deaminates cytidines (C), thus converting them to uridines (U) to generate a G:U lesion. As discussed in Chapter 2, the presence of U is not tolerated in DNA, thus it is removed using one of two possible mechanisms (Peled et al., 2008). First, U is a substrate for uracil-\textit{N}-glycosylase (UNG), a component of the base excision repair (BER) pathway, which removes U from single stranded (ss)- and double stranded (ds)-DNA. Excision of U generates an abasic site which can be repaired by BER to re-produce the original G:C base pair. Alternatively, the G:abasic site lesion undergoes error-prone repair to generate transition and transversion mutations at G:C base pairs as well as a small number of mutations at A:T base pairs. Details about the error-prone BER remain to be elucidated. Second, the G:U lesion is a non-canonical base pair that is recognized as a mismatch and processed by the mismatch repair (MMR) pathway, which produces mutations at A:T base pairs (Bardwell et al., 2004; Cascalho et al., 1998; Ehrenstein and Neuberger, 1999; Kim et al., 1999; Martin et al., 2003; Phung et al., 1998; Rada et al., 2004; Rada et al., 2002b; Shen et al., 2006; Wiesendanger et al., 2000; Wilson et al., 2005).

The MMR pathway has evolved to correct DNA error produced during DNA replication (Modrich and Lahue, 1996). However, in centroblast B cells, the MMR pathway produces mutations at A:T base pairs as evidenced by the reduction of these mutations.
when the MMR pathway is absent or defective (Bardwell et al., 2004; Cascalho et al., 1998; Ehrenstein and Neuberger, 1999; Kim et al., 1999; Martin et al., 2003; Phung et al., 1998; Rada et al., 2004; Shen et al., 2006; Wiesendanger et al., 2000; Wilson et al., 2005). The A:T mutation signature in hypermutated V regions resembles that of the translesional polymerase η (Rogozin et al., 2001), and indeed (Delbos et al., 2007; Delbos et al., 2005; Faili et al., 2004; Yavuz et al., 2002; Zeng et al., 2004; Zeng et al., 2001), it has been demonstrated that polymerase η is essential to the hypermutation process. Although the reason for the recruitment of polymerase η to the site of mismatch repair tracts during SHM is not known, the translesion synthesis pathway sheds some light on this mechanism. During DNA replication, lesions that block the advance of the replication machinery induce ubiquitination of the DNA processivity clamp PCNA and the recruitment of translesional polymerases which facilitate bypass of the lesion. Langerak et al. have demonstrated that mice harboring the PCNAK164R/K164R mutation are unable to ubiquitinate PCNA at the critical residue and lack mutations at A:T base pairs in the V region (Arakawa et al., 2006; Langerak et al., 2007; Roa et al., 2008). Together, these data suggest that the general translesion synthesis pathway is induced at the hypermutating V region. To further elucidate this mechanism, I analyzed Ig sequences from hypermutating wildtype, UNG−/− and MMR-deficient murine B cells. I have shown that the DNA repair mechanisms used to process the G:U lesion are not mutually exclusive, but rather that interactions between the BER and MMR pathways enable the recruitment of error-prone polymerases and contribute to extensive mutagenesis at the V region.
3.3. **Materials and Methods**

3.3.1. **Amplification and analysis of murine Ig sequences.**

All murine V region sequence data were generated from sorted PNA$^{hi}$ B220$^{+}$ germinal center B cells isolated from Peyer’s patches or spleen as described previously (Martin et al., 2003). Mutations were analyzed in the intronic JH2-JH4 region or the intronic VHJ558-JH4 rearrangement flanking region (wild-type and UNG$^{-/-}$ mice). Wild-type mouse data were obtained from previously published works (Bardwell et al., 2003; Martin et al., 2003; Rada et al., 2002b). UNG$^{-/-}$ sequence data were generously provided by J. Di Noia, C. Rada, and M. Neuberger. Additional UNG$^{-/-}$ sequence data were generated from genomic DNA kindly provided by H. Ming and U. Storb (Shen et al., 2006). The VHJ558-JH4 flanking region was amplified from genomic DNA as previously described (Rada et al., 2002b). To compensate for the unequal distribution of nucleotides in the sequenced region (the C:G ratio was ~0.7:1, and the A:T ratio was ~0.8:1), the data shown in Fig. 3.1.C. were normalized for nucleotide content according to the following formula: percent bottom strand C mutations = 100(bottom strand C mutations/number of bottom strand C's)/[(bottom strand C mutations/number of bottom strand C's) + (top strand C mutations/number of top strand C's)]. The same formula was used to normalize for A:T mutations.

3.3.2. **Statistical analysis.**

Murine data were graphed using GraphPad software (Prism), and statistical analyses were performed using the unpaired t-test, Fisher’s exact test, and Mann-Whitney test.
3.4. Results

3.4.1. Murine Ig sequences with A:T mutations are enriched for C transversion mutations.

In Chapter 2, I examined the association of A and C mutations in the V regions of hypermutating Ramos B cell clones. I observed that clones with an A mutation were enriched for a nearby C transversion mutation, that A mutations occurred predominantly on the top (non-transcribed) strand while C transversion mutations occurred predominantly on the bottom (transcribed) strand, and that the association between A mutations and C transversion mutations was lost upon inhibition of UNG. These results led to a model in which MMR preferentially excises the top strand and UNG creates an abasic site in the bottom strand, which stimulates translesion synthesis, Rev1-induced mutations opposite the abasic site and polymerase η-induced A:T mutations during the gap-filling step of MMR.

In this chapter, I investigated whether a similar relationship between MMR and BER operates in the V regions of murine hypermutating B cells. I examined the JH2-JH4 or VHJ558-JH4 intronic Ig sequence from PNAbi Peyer’s patch B cells from nine wild-type mice and five UNG−/− mice (see Fig. 3.1.). A limitation in examining linked A:T and G:C mutations in Ig sequences derived from mice is that they are often heavily mutated, which is likely due to multiple rounds of mutational events during the life cycle of the centroblast. Therefore, I restricted this analysis to sequences containing a maximum of five mutations per sequence (see Appendix 2), thereby minimizing the possibility that mutations arose independently.
Figure 3.1. Characteristics of mutations derived from wild type and UNG−/− mice.

Characteristics of mutations derived from the intronic JH2-JH4 region or the intronic V_{H}J558-JH4 region in wildtype (left) and UNG−/− (right) mice are displayed. Column 1 includes 9 wildtype mice, 182 sequences, 1516 mutations; column 2 includes 5 UNG−/− mice, 173 sequences, 1418 mutations. Statistical analyses comparing mutation types in wildtype and UNG−/− mice were performed using the t-test (** = P<0.0001 for G:C transition mutations; *** = P<0.0001 for G:C transversion mutations; ** = P=0.0046 for A:T mutations, unrestricted dataset). B. Mutation frequency at A:T basepairs relative to mutation frequency at G:C basepairs (set at 1) in the 9 wildtype and 5 UNG−/− mice examined above. (* = P= 0.0332; t-test).
Figure 3.1.
As shown in Fig. 3.2.A, ~57 and ~85% of Ig sequences that contained at least one A:T mutation and had up to five total mutations also harbored a G:C mutation in wild-type and UNG−/− mice, respectively. To examine the relationship between C transversion mutations and A:T mutations in the wild-type-restricted data set, I measured the number of sequences with C transition or C transversion mutations with respect to the presence or absence of A:T mutations. For wild-type mice, in sequences with no A:T mutations, only 21% (5/24) of those sequences harbored a C transversion mutation (Fig. 3.2.B). However, among sequences that contain A:T mutations, 65% (20/31) also contained C transversion mutations (P = 0.0023) (Fig. 3.2.B). Thus, sequences with A:T mutations are enriched for C transversion mutations in murine B cells, thereby corroborating the Ramos cell data. However, Ig sequences from UNG−/− mice showed no difference in C transversion mutations whether or not a sequence harbored an A:T mutation (Fig. 3.2.B), highlighting, just as with the UGI-Ramos data (Fig. 2.4.B.), that a mechanism independent of abasic sites produces A:T mutations during SHM.

3.4.2. Adenine mutations on the top strand are associated with cytosine transversion mutations on the bottom strand in murine B cells

I next examined whether specific mutations displayed a strand bias exclusively in sequences with A:T mutations. To eliminate bias due to the unequal distribution of nucleotides in the sequenced region, all values were normalized for nucleotide content (see Materials and Methods). As shown in Fig. 3.3.A, only 28% of A:T mutations had the A on the bottom strand, indicating that A mutations occurred preferentially in the top strand (P < 0.0001), consistently with previous data for mice (Mayorov et al., 2005) and for Ramos
Figure. 3.2. Sequences with A:T mutations are enriched for G:C transversion mutations.

(A) Percentages of sequences with associated A:T and G:C mutations from nine wild-type mice and five UNG−/− mice. All analyzed sequences contained up to five mutations per sequence with a minimum of one mutation at an A:T base pair. (B) Comparison of G:C transition and G:C transversion mutations in sequences with and without A:T mutations in wild-type (WT) and UNG−/− mice (for the wild-type mice, \( n = 31 \) without A:T mutations and \( n = 24 \) with A:T mutations; for UNG−/− mice, \( n = 20 \) without A:T mutations and \( n = 41 \) with A:T mutations). Sequences with G:C transversion mutations that also contained G:C transition mutations were scored as sequences with G:C transversion mutations. Statistical analyses were performed using Fisher’s exact test (**, \( P = 0.0023 \) for sequences with and without A mutations from wild-type mice for G:C transversion mutations).
Figure 3.2.
Figure 3.3. Transversion mutations at C on the bottom strand are associated with mutations at A on the top strand in wild-type mice.

(A) Percentages of mutations located on the bottom strand were determined for wild-type and UNG⁻/⁻ mice in unrestricted (UnRes; columns 1 to 3 and 7 to 9) and restricted (Res; columns 4 to 6 and 10 to 12) data sets. Sequences that did not contain A:T mutations were excluded from this analysis. Mutations were normalized for nucleotide content in the sequenced region (see Materials and Methods). Statistical analyses were conducted using the t-test (*, \( P = 0.02 \) for the comparison of C transversion mutations in the restricted data set between wild-type and UNG⁻/⁻ mice). Restricted dataset includes sequences containing 1 to 5 mutations per sequence. Unrestricted dataset includes all sequences with 1 or more mutations per sequence. (B) Location of C and A mutations in the JH2-JH4 and VHJ558-JH4 intronic regions of wild-type mice. Data shown are from sequences containing A mutations in the top strand (top) or the bottom strand (middle) or containing no A mutations (bottom). Top strand mutations and bottom strand mutations are depicted above and below the midline, respectively. Statistical analyses were performed using the t-test (*, \( P = 0.0249 \) for bottom strand transition mutations [top panel]; ***, \( P < 0.0001 \) for bottom strand transversion mutations [top panel]). (C) Strand bias of A mutations in sequences containing a single C mutation in the top strand (top) or bottom strand (bottom). A mutations that are linked to a C transition mutation are represented by a square, and those linked to a C transversion mutation are represented by a circle. Statistical analyses for the strand bias of A mutations were performed using the t-test (***, \( P < 0.0001 \)), and those for the location of A mutations relative to C transition and transversion mutations were performed using Fisher's exact test (\( P = 0.0015 \)).
Figure 3.3.
Importantly, 72% of C transversion mutations occurred on the bottom strand \( (P = 0.00392) \) (Fig. 3.3.A.) in wild-type mice, displaying the same strand preference as that in Ramos cells (Fig. 2.2.C.), and most of these mutations depended on UNG (Fig. 3.3.A.). Moreover, in wild-type mice, in sequences containing A mutations in the top strand, 100% of transversion mutations and 65% of transition mutations at C were located on the bottom strand \( (P < 0.0001 \text{ and } P = 0.0249, \text{ respectively}) \) (Fig. 3.3.B., top). In contrast, sequences containing A mutations on the bottom strand showed no strand preference for C mutations \( (P = 0.6261; P = 0.2195 \text{ for C transversion and C transition mutations, respectively}) \) (Fig. 3.3.B., middle). There also was no strand preference in sequences with no mutations at A:T base pairs \( (P = 0.5796; P = 0.5734 \text{ for C transversion and C transition mutations, respectively}) \) (Fig. 3.3.B., bottom).

### 3.4.3. Bottom strand C transversion mutations are preferentially located upstream of A mutations

To extend this analysis, I determined whether A mutations display a strand bias in sequences harboring a single C mutation. Sequences with a top strand C mutation do not display a significant strand bias with A mutations \( (P = 0.1114) \) (Fig. 3.3.C., top). In contrast, sequences with a bottom strand C mutation display a bias of A mutations to the top strand \( (P < 0.0001) \) (Fig. 3.3.C., bottom), which is consistent with data shown in Fig. 3.3.B. and with Ramos data (Fig. 2.2.C.). Notably, bottom strand C transversion mutations are preferentially located upstream of A mutations \( (P = 0.0015) \) (Fig. 3.2.E., bottom), which is consistent with the model shown in Fig. 2.1.A. However, as noted above, this relationship was not observed in Ramos cells (Fig. 2.2.C.). Collectively, these data show that A mutations
on the top strand are associated with C transversion mutations on the bottom strand in murine B cells, supporting the notion that the excision of the top strand by Exo1 and the removal of U on the bottom strand by UNG generates A mutations on the top strand.

### 3.4.4. Deficiency of MMR leads to decreased C transversion mutations.

The model (Fig. 2.1.A.) envisages that the AID-generated G:U mismatch recruits the MMR system, which sometimes excises the G-containing strand exposing the U to UNG, leading to frequent transversion mutations. Because UNG displays greater activity on U in ssDNA than in dsDNA (Krokan and Wittwer, 1981), deficiency in MMR should protect U from UNG and lead to an increase in transition mutations at C on the bottom strand. As such, I examined Ig sequences from wild-type and MMR-compromised (i.e., Msh2<sup>−/−</sup>, Msh2<sup>G674A</sup>, Msh6<sup>−/−</sup>, and Exo1<sup>−/−</sup>) mice from previously published data (Bardwell et al., 2004; Li et al., 2006; Martin et al., 2003) (see Fig. 3.4.).

Figure 3.5. displays the percentage of C mutations on the top or bottom strand in wild-type and MMR-deficient mice. While transition and transversion mutations at C were similar between wild-type and MMR-deficient mice on the top strand, I observed a two-fold decrease in the fraction of C mutations that were C-to-G transversion mutations ($P = 0.014$) on the bottom strand in MMR-deficient mice ($P = 0.003$) and a corresponding ~1.3-fold increase in C-to-T transition mutations in MMR-deficient mice ($P = 0.003$). These data indicate that ~50% of C transversion mutations on the bottom strand are produced indirectly by MMR proteins and suggest that in the absence of MMR, a larger proportion of AID-generated U on the bottom strand is instead being replicated, leading to an increase in C transition mutations. Since AID is slightly more active on the top than the bottom strand
Figure 3.4. Ig sequence data for wild type and MMR⁻/⁻ mice

Ig sequence data for combined (A) and individual (B) WT and MMR⁻/⁻ mice that was used to generate data shown in Figure 3.5.
Figure 3.4.

A

<table>
<thead>
<tr>
<th></th>
<th>WT</th>
<th></th>
<th>MMR-/--</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>To:</td>
<td>From:</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>C</td>
<td>G</td>
<td>T</td>
</tr>
<tr>
<td>A</td>
<td>-</td>
<td>9</td>
<td>33</td>
<td>31</td>
</tr>
<tr>
<td>C</td>
<td>11</td>
<td>-</td>
<td>5</td>
<td>33</td>
</tr>
<tr>
<td>G</td>
<td>30</td>
<td>11</td>
<td>-</td>
<td>10</td>
</tr>
<tr>
<td>T</td>
<td>11</td>
<td>24</td>
<td>10</td>
<td>-</td>
</tr>
</tbody>
</table>

B

<table>
<thead>
<tr>
<th>Msh2G674A +/-</th>
<th>A</th>
<th>C</th>
<th>G</th>
<th>T</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Msh2G674A +/-</td>
<td>A</td>
<td>1</td>
<td>9</td>
<td>7</td>
<td>17</td>
</tr>
<tr>
<td>Msh2G674A +/-</td>
<td>C</td>
<td>0</td>
<td>1</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>Msh2G674A +/-</td>
<td>G</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>11</td>
</tr>
<tr>
<td>Msh2G674A +/-</td>
<td>T</td>
<td>2</td>
<td>8</td>
<td>3</td>
<td>13</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Msh2G674A +/-</th>
<th>A</th>
<th>C</th>
<th>G</th>
<th>T</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Msh2G674A +/-</td>
<td>A</td>
<td>3</td>
<td>6</td>
<td>7</td>
<td>16</td>
</tr>
<tr>
<td>Msh2G674A +/-</td>
<td>C</td>
<td>0</td>
<td>1</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>Msh2G674A +/-</td>
<td>G</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>Msh2G674A +/-</td>
<td>T</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>EXO1 +/-</th>
<th>A</th>
<th>C</th>
<th>G</th>
<th>T</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXO1 +/-</td>
<td>A</td>
<td>4</td>
<td>14</td>
<td>10</td>
<td>28</td>
</tr>
<tr>
<td>EXO1 +/-</td>
<td>C</td>
<td>7</td>
<td>3</td>
<td>14</td>
<td>24</td>
</tr>
<tr>
<td>EXO1 +/-</td>
<td>G</td>
<td>13</td>
<td>4</td>
<td>6</td>
<td>23</td>
</tr>
<tr>
<td>EXO1 +/-</td>
<td>T</td>
<td>5</td>
<td>10</td>
<td>2</td>
<td>17</td>
</tr>
</tbody>
</table>
Figure 3.5. Bottom, but not top, strand C transversion mutations are reduced in MMR-deficient mice.

The percentages of C-to-T, C-to-G, and C-to-A mutations on the top strand (left) and on the bottom strand (right) in wild-type and MMR-deficient mouse is displayed. Data were collected from the JH2-JH4 region of wild-type (two Msh2G674A*/*, one Msh2G674A*/*, and one Exo1*/) and MMRdeficient (one Msh2−/−, one Exo1−/−, one Msh6−/−, and three Msh2G674A*/* mouse from previous publications (see Fig. 3.4.). Statistical analyses were performed using the *t*-test (the asterisk indicates *P* = 0.0250 for C to T mutations on the bottom strand and *P* = 0.0121 for C to G mutations on the bottom strand).
Figure 3.5.

![Graph showing mutations at dC (%) for C>T, C>G, and C>A in wildtype (n = 4) and MMR-/- (n = 6) with error bars indicating variability. The graph is divided into top strand and bottom strand comparisons.](image)
and the UNG pathway has not been shown to display a strand bias, this result indicates that
the MMR and BER pathways cooperate to produce C transversion mutations on the bottom
strand and A mutations on the top strand, as inferred from the analysis of Ramos cells.
3.5. Discussion

The SHM process is initiated by AID-generated G:U lesions in the V region and the processing of this lesion by DNA repair mechanisms achieves the full spectrum of mutations at G:C and A:T base pairs is achieved. In Chapter 2, I conducted an analysis in Ramos cells to gain insight into the use of the BER and MMR pathways in hypermutating B cells. Our data revealed a model in which C deamination followed by U excision results in the creation of an abasic site, which stimulates the recruitment of an error-prone polymerase that introduces mutations at A:T base pairs. Here I sought to determine whether the same model operates in hypermutating murine B cells. These data reveal that in sequences containing A mutations with associated C transversion mutations, a C-to-G mutation predominates, indicating that a C is frequently inserted opposite the abasic site (Fig. 3.2.B.). Additionally, this analysis revealed a strand polarity in that A mutations on the top strand were associated exclusively with C transversions in the bottom strand (Figure 3.3.B.). I also demonstrated that A mutations occurred both 5' and 3' of the C transversion, indicating that additional mechanisms induce translesion synthesis upstream of the abasic site. These data substantiate the model presented in the Ramos study and suggest that an abasic site lesion produced by the excision of U by UNG in an Exo1/MMR-generated ssDNA tract stimulates translesion synthesis and the recruitment of error-prone polymerases such as the deoxycytidyl transferase Rev1 and the A:T mutator polymerase η (Gibbs and Lawrence, 1995; Haracska et al., 2001; Wilson et al., 2005).

I also examined the effect of MMR deficiency on the frequency of top and bottom strand C transition and transversion mutations. The data reveal that absence of MMR results in the selective reduction of C transversion mutations and a corresponding increase
in C transition mutations on the bottom strand and suggest that MMR contributes to ~50% of bottom strand C transversion mutations (Fig. 3.5.). However, data presented here and by other laboratories (Rada et al., 2002b) clearly illustrate that that not all mutations at A:T base pairs arise from this pathway, as A:T mutations occurred even when UNG was deficient. It is possible that an MMR-dependent, but UNG- and/or abasic site-independent mechanism activates translesion synthesis to give rise to A:T mutations.

The data presented here corroborate earlier findings that reveal an interaction between the BER and MMR pathways as a critical step for extending AID-generated G:U lesions to a broad spectrum of mutations. However, these findings also underscore that MMR-induced mutagenesis is not yet fully understood. Given that hypermutating B cells are equipped to transform DNA repair mechanisms into mutagenesis pathways, it is possible that similar mechanisms could be active in other cell types, and thus it will be important to further elucidate this process.
Chapter 4

shRNA-Mediated Silencing of MRE11 Does Not Affect SHM in Ramos Cells
4.1. Abstract

Somatic hypermutation (SHM) generates mutations at G:C and A:T base pairs. While both mutation types require the action of AID, mutations at A:T base pairs are additionally dependent on the MMR pathway and DNA polymerase η. Recent reports suggest that the Mre11-Nbs1-Rad50 complex is involved in SHM and MMR. These findings prompted me to examine the participation of this complex in the generation of A:T mutations in the Ig V region. I reduced the expression level of Mre11 in the hypermutating Ramos cell line by RNAi and measured various parameters. Although I observed the selective advantage of Mre11\textsuperscript{high} cells over Mre11\textsuperscript{low} clones in culture, I did not observe significant differences between control and Mre11\textsuperscript{low} clones in survival following treatment with ionizing radiation, MMR status, and A:T mutation frequency. These results indicate that successful knock down of Mre11 in Ramos clones is technically challenging due to its requirement in vital cellular processes and that Mre11 levels I obtained were insufficiently reduced to discern any effect on MMR or SHM function.
4.2. Introduction

The Mre11-Rad50-Nbs1 (MRN) complex is involved in numerous cellular pathways, including DNA repair, DNA replication, telomere maintenance, and cell cycle signaling. MRN is a key player in the cellular response to double strand breaks (DSB) as it is the first sensor of this type of DNA damage. It is instrumental in the recruitment and activation of the ataxia mutated (ATM) kinase, which initiates a downstream protein cascade that enables cell cycle arrest and the ligation of the broken DNA ends. Null mutations in any of the MRN components result in early embryonic lethality in mice while humans carrying hypomorphic mutations in Mre11 or Nbs1 are affected by the ataxia telangiectasia-like mutated (ATLD) and Nijmegen syndromes, respectively, both of which are characterized by immunodeficiency, radiation sensitivity, cell cycle checkpoint defects, chromosomal instability, and increased cancer susceptibility. Mre11 is a multifunctional enzyme with DNA unwinding, annealing and endo-/exonuclease activities. It forms the core of the MRN complex and interacts with DNA ends, Rad50, Nbs1, as well as with itself. Rad50 serves as a bridge that connects two ends of broken DNA molecules, while Nbs1 is the regulatory subunit that is involved in ATM activation. MRN defects result in failure to activate cell cycle checkpoint and DNA repair factors.

Prior to the discovery of the MutLα endonuclease and at the time of this study, the multi-functional MRN complex was proposed as the exonuclease involved in 3’-directed MMR. Mre11 was shown to associate with Mlh1 in yeast (Her et al., 2002) and Mre11 knock-down in HeLa cells resulted in increased microsatellite instability (MSI), a hallmark of MMR deficiency, and inhibition of 3’-, but not 5’-directed repair of an MMR substrate (Vo et al., 2005). The importance of the MRN complex in antibody diversification processes has
been demonstrated *in vivo* and *in vitro*. The conditional knock-out of Nbs1 in murine B cells diminished the level of CSR, while patients with hypomorphic mutations in Mre11 or Nbs1 have reduced numbers of isotype switched-peripheral blood B cells (Gregorek et al., 2002). Nbs1 overexpression in Ramos cells resulted in a modest increase in SHM and was dependent on the Mre11-interaction domain (Yabuki et al., 2005). Furthermore, Nbs1-overexpressing cells showed an increased proportion of C transversion mutations. To identify a mechanistic role for MRN in SHM, Larson *et al.* examined the enzymatic function of Mre11/Rad50 *in vitro* and found that the protein harbors AP lyase activity on ssDNA and is enriched in V region DNA of hypermutating B cells (Larson et al., 2005). The authors suggested that Mre11-induced cleavage of UNG-generated abasic sites produces a DNA break that cannot be directly extended by a DNA polymerase, but requires further processing, which necessitates the recruitment of translesional DNA polymerases and results in error-prone repair of the G:U lesion (Bailly and Verly, 1987; Kim and Linn, 1988). These data prompted me to investigate whether Mre11 is involved in SHM, focusing especially on its potential involvement in the MMR arm of the process and the generation of mutations at A:T base pairs in Ramos cells.
4.3. Materials and Methods

4.3.1. Cell culture

Ramos 67 and Ramos 80 cells were maintained as previously described (Zhang et al., 2001), and Abelson pre-B cell lines 15-63 (Msh2−/−) and 8-58 (Msh2−/−) were maintained in RPMI medium (Invitrogen) with 10% bovine calf serum (HyClone), penicillin (100 U/ml), and streptomycin (0.1 mg/ml; Sigma).

4.3.2. Plasmids and transfections.

To measure MMR activity, the microsatellite-like plasmid pCA-OF (Vo et al., 2005) was kindly provided by C. Her. To knock down Mre11, the pGeneClip system was used as per manufacturer’s directions (Promega). Briefly, 2 constructs, each targeting a unique region of Mre11, were ligated into the pGeneClip vector and simultaneously transfected into Ramos cells. Control clones were transfected with a pGeneClip vector containing a scrambled sequence. Plasmids were linearized with PvuI (pGeneClip-hMre11 and pGeneClip-Con) and BglII (pCA-OF). For cell transfections, ~4 X 10⁶ log-phase cells were mixed with 10 μg of linearized plasmid DNA in 4-mm cuvettes and electroporated (Gene Pulser Xcell) at 250 V and 950 F for Ramos cells and at 450 V, 950 F, and 150 Ω for pre-B cells. Cells were diluted in appropriate media and plated in 96-well plates. Transfected cells were incubated at 37°C for 24 h, stable clones were selected with puromycin (0.8 μg/ml; pGeneClip-hMre11, pGeneClip-SCR) and blasticidin (2.5 μg/ml for Ramos and 25 μg/ml for pre-B cells; pCA-OF).
4.3.3. **Western blot analysis.**

Ramos cell lysates were prepared by cell lysis in 1% SDS/T.E. buffer and transferred to nitrocellulose membranes. Mre11 and β-actin were detected using mouse anti-human Mre11 (GeneTex), goat anti-mouse IgG-hHRP, rabbit anti-mouse β-actin (Jackson), and goat anti-rabbit (SB) primary and secondary antibodies (Abcam), respectively. Protein levels were quantified by densitometry analysis using ImageQuant software, version 5.0 (Molecular Dynamics).

4.3.4. **Measurement of mismatch repair activity.**

To quantify MMR activity, pCA-OF-expressing clones were harvested and washed in phosphate-buffered saline (PBS) (Gibco, Invitrogen), and expression of green fluorescent protein (GFP) was determined using a flow cytometer (FACSCalibur; BD) as previously described (Vo et al., 2005). Flow cytometry data were analyzed using FlowJo software.

4.3.5. **ELISPOT assays.**

The ELISPOT assay for IgM secretion was performed as previously described (Martin et al., 2002).

4.3.6. **Clonal survival assays.**

The fraction of surviving clones following treatment with ionizing radiation was determined as previously described for adherent cells with some modifications. Briefly, 125 X 10³ cells were irradiated at 0 to 6 Gy. Cells were plated in 96-well plates at 0.5 to 500 cells/well, depending on irradiation dose used. Surviving clones were counted after ~15 days at 37°C. Control, non-irradiated cells were plated to 0.5 cell/well and the plating
efficiency (PE) was determined by using the following formula: PE = (Number of Clones)/(Total Cells Plated). The surviving fraction (SF) was determined as follows: SF = (Number of Clones)/(PE)(Total Cells Plated). All assays were performed in duplicate.

4.3.7. DNA extraction, PCR, and sequencing in Ramos.

Genomic DNA extractions and V region amplification using Pfu Ultra II polymerase were performed as previously described (Martin et al., 2002).

4.3.8. Statistical analysis.

Ramos data were graphed using GraphPad software (Prism), and statistical analyses were performed using the Mann-Whitney test.
4.4. Results

4.4.1. ShRNA reduces Mre11 expression in Ramos.

The MRN complex has been implicated in SHM and MMR (Her et al., 2002; Larson et al., 2005; Vo et al., 2005; Yabuki et al., 2005). Thus, I wanted to determine whether MRN participates in the generation of MMR-dependent A:T mutations. To this end, I sought to reduce expression levels of Mre11 protein in the hypermutating B cell line Ramos using Mre11-targeted shRNA (shMre11). I used two Ramos variants, Ramos 67 and Ramos 80, each of which harbors a truncating TAA nonsense codon in the immunoglobulin (Ig) locus. A:T mutations at the TAA codon result in reversion of the nonsense codon to a productive amino acid and subsequent IgM expression. Thus, selection of IgM-positive cells enables us to enrich for clones that harbor A:T mutations in their V regions. To assess MMR status, Ramos cells were first transfected with the microsatellite-like pCA-OF vector, which is an indicator of MMR activity. The system consists of the green fluorescent protein (GFP) open reading frame interrupted by a repetitive (CA)$_{25}$ sequence. This microsatellite-like sequence puts the GFP sequence out-of-frame, thereby preventing its expression (Vo et al., 2005). Expansion and contraction of the CA repeats reverts the sequence into frame and permits expression of GFP. Since MMR-deficient cells display increased microsatellite instability (MSI), increased GFP reversion is indicative of a deficiency in MMR.

To reduce Mre11 expression, Ramos pCA-OF cells were transfected with two shMre11 or with scrambled (shCon) constructs. As shown in Figure 4.1., Mre11 protein levels were reduced by up to 95% in shMre11, compared to shCon Ramos clones. Transfection with shMre11 generated a wide range (<10% to 100%) of residual Mre11
Figure 4.1. Mre11 protein level is reduced in Ramos treated with shRNA.

(A) Mre11 protein level of Ramos 67 and Ramos 80 clones shMre11 and shCon clones maintained for 1 month in culture. Ramos 67 clones 9, 12, and 18 and 80 clone 9 (top) and 13 additional Ramos 67 clones (bottom) are displayed. Protein input was normalized using β-actin. ShCon clones were serially diluted 2-fold. (B) Mre11 levels in shMre11 clones relative to shCon clones were quantified by densitometry analysis. (C) Effect of culture time on Mre11 protein level was determined by comparing protein level in 1- and 2-month cultures of Ramos 67 shMre11 clone 9 and Ramos 67 shCon clone 1. Protein input was normalized using Ponceau stain.
Figure 4.1.

A

\[ \text{Ramos 67} \quad \text{Ramos 80} \]

\[
\begin{array}{cccc|cccc}
\text{shCon} & \text{shMre11} & \text{shCon} & \text{shMre11} \\
\hline
\text{s1} & 9 & 12 & 18 & 9 & \hline
80kDa & & & & Mre11 \\
42kDa & & & & \beta\text{-Actin} \\
\end{array}
\]

B

<table>
<thead>
<tr>
<th>Clone</th>
<th>22</th>
<th>23</th>
<th>24</th>
<th>25</th>
<th>26</th>
<th>27</th>
<th>28</th>
<th>29</th>
<th>30</th>
<th>31</th>
<th>32</th>
<th>33</th>
<th>34</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative Mre11 level</td>
<td>0.33</td>
<td>0.16</td>
<td>1.39</td>
<td>0.11</td>
<td>0.86</td>
<td>0.06</td>
<td>0.58</td>
<td>0.16</td>
<td>0.07</td>
<td>0.31</td>
<td>0.39</td>
<td>0.12</td>
<td>0.10</td>
</tr>
</tbody>
</table>

C

\[
\begin{array}{c|ccc}
\text{Input} & \text{shCon} & \text{shMre11} \\
\hline
\text{Culture period (mo)} & 1 & 2 & 1 & 2 \\
80kDa & & & & Mre11 \\
\end{array}
\]
protein levels (see Figure 4.1.A. and 4.1.B.). The variable level of Mre11 levels could be due to failed integration of one of the shMre11 constructs or mutation of one or both constructs, rendering them less effective. Figure 4.1.A. displays Ramos 67 clones 9, 12, and 18 and Ramos 80 clone 9, which were selected for further analysis based on substantially reduced Mre11 levels (Mre11\textsuperscript{low}). Notably, transfection of Ramos 80 with shMre11 constructs yielded two additional clones that expressed extremely low levels of Mre11 upon initial western blot analysis. However, these clones were subsequently lost in culture, possibly caused by Mre11 levels falling below the threshold required for survival.

4.4.2. **Mre11-low Ramos clones do not display increased IR sensitivity.**

Since MRN defects lead to increased sensitivity to ionizing radiation (IR), the survival of Mre11\textsuperscript{low} and control clones following IR treatment was carried out in order to determine whether Mre11 was functionally knocked down. I first determined a working dose of IR by treating Ramos cells with 1 to 30 Gy doses of IR and monitoring cell growth every 24 hours for 5 days. Low IR doses (1 to 4 Gy) had little effect for the first 48 hours, but caused cell death after the 5-day period. In contrast, cells irradiated at high doses (5 to 30 Gy) showed signs of cell death in the early phase post-treatment and no surviving cells were observed after 5 days. Based on these observations, Mre11\textsuperscript{low} and control Ramos clones were irradiated at 1 to 6 Gy doses.

As shown in Figure 4.2., both control and Mre11\textsuperscript{low} Ramos clones displayed reduced clonal growth following irradiation, indicating that the assay had the desired effect. However, the surviving fractions of both control and Mre11\textsuperscript{low} groups were similar at all doses, suggesting that the degree of Mre11 reduction was not sufficient to increase
sensitivity to IR. Thus, it was not possible to conclude from this assay that these Ramos Mre11\textsuperscript{low} clones were functionally compromised in their MRN activity. Nevertheless, it is possible that Mre11 levels were reduced below the threshold required for MMR or SHM activity.

4.4.3. **Mre11\textsuperscript{low} clones restore Mre11 levels.**

When cell lines are carried in culture for prolonged periods of time, clones with a selective advantage such as enhanced survival or proliferation will outcompete others. Since Mre11 is a critical component of numerous vital cellular functions, I explored the possibility that Mre11\textsuperscript{low} clones might be under pressure to restore wild type Mre11 levels. Mre11\textsuperscript{low} clones were carried in culture under selective conditions for two months, at which point protein levels were examined by western blot analysis. Figure 4.1.C. displays the level of Mre11 protein after 1- and 2-month culture periods for control and Mre11\textsuperscript{low} Ramos 67 clones. While Mre11 levels remained constant for control clones across both culture periods, Mre11 protein levels increased to near wild type levels after two months in culture in Mre11\textsuperscript{low} clones. This was likely driven by the reduced survival of Mre11\textsuperscript{low} clones relative to their control counterparts, although it was not observed under IR conditions.

4.4.4. **Mismatch repair is not compromised in Mre11\textsuperscript{low} clones.**

Although Mre11 reduction did not lead to a phenotype in the IR assay, it is possible that Mre11 protein levels were sufficiently low to affect MMR or SHM. Therefore, I examined the MMR status of shMre11 Ramos clones using the GFP-based assay described earlier (Section 4.4.1). To verify that defective MMR results in increased GFP reversion frequency, I first
Figure 4.2. Mre11\textsuperscript{low} clones do not display increased sensitivity to ionizing radiation.

(A) The mean surviving fraction of Mre11\textsuperscript{low} (n=3) and control (n=3) Ramos 67 and Ramos 80 clones following treatment with IR; error bars represent standard error. Each treatment was performed in duplicate. (B) The surviving fraction results of independent Mre11\textsuperscript{low} and control clones are displayed. Each plot represents the mean of two independent experiments and is displayed with standard error.
Figure 4.2.
tested the construct in Msh2−/− (8-58) and Msh2+/- (15-63) pre-B cell clones. Figure 4.3.A. shows that the GFP reversion frequency in Msh2−/− cells is ~ 2 orders of magnitude higher than in Msh2+/- clones (P = 0.0030), indicating that the assay is a reliable method for detecting the MMR status of a cell. However, the GFP reversion frequency in 29 control and 34 Mre11low Ramos clones was similar (P = 0.2145). Moreover, the GFP reversion frequency of all Ramos clones was ~10^-4, similar to that observed in Msh2+/- pre-B cells. This indicated that the reduced levels of Mre11 in the Ramos cells did not affect MMR activity.

4.4.5. A:T mutation frequency is not affected in shMre11 Ramos clones.

I examined the mutation frequency at A:T base pairs in Mre11low Ramos clones by determining the reversion frequency of the TAA nonsense codon in the IgM locus using an ELISPOT assay. Figure 4.3.B. displays the IgM reversion frequency of 3 control and 17 Mre11low Ramos clones. I did not observe a significant difference in IgM reversion frequency between these groups (P = 0.5603), suggesting that the reduced Mre11 protein level does not affect A:T mutations in Ramos clones. Consistent with this finding, I found that the mutation spectrum and mutation frequency of the Ramos 67 V region of 2 control and 2 shMre11 clones did not differ between the two groups (3.43 X 10^-4 and 3.93 X 10^-4 mutations per base pair, in control and Mre11low clones, respectively) (Table 4.1.). Collectively, these data indicate that the reduced Mre11 levels in these Ramos clones did not affect SHM.
Figure 4.3. Mismatch repair activity and A:T mutation frequency are not compromised in Mre11\textsuperscript{low} clones.

(A) MMR activity for Msh2\textsuperscript{+/−} and Msh2\textsuperscript{−/−} pre-B cell lines and R67 clones was quantified using a MSI assay based on the GFP reversion frequency. The GFP open reading frame is interrupted with an out-of-frame dinucleotide microsatellite-like sequence (Vo et al., 2005); MSI at the dinucleotide sequence leads to GFP expression. Individual clones are represented as black symbols. Statistical analyses were performed using the Mann-Whitney test (**, \( P = 0.0030 \) for results with Msh2\textsuperscript{−/−} pre-B cells compared to those with Msh2\textsuperscript{+/−} pre-B cells). (B) The IgM reversion frequency was determined using an ELISPOT assay for IgM secretion. Individual shMre11 (n=17) and shCon (n=3) clones are represented as black circles. Statistical analyses comparing median reversion frequencies were performed using the Mann-Whitney test (\( P = 0.5603 \)).
Figure 4.3.
Table 4.1. Mutation frequency is not decreased in Mre11<sub>low</sub> Ramos clones.

<table>
<thead>
<tr>
<th></th>
<th>shCon</th>
<th>Mre11&lt;sub&gt;low&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent clones</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Nucleotides sequenced</td>
<td>20355</td>
<td>23338</td>
</tr>
<tr>
<td>Total mutations</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Mutation frequency (per bp x 10^{-4})</td>
<td>3.93</td>
<td>3.43</td>
</tr>
<tr>
<td>G:C mutations</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>
4.5. Discussion

The generation of A:T mutations during SHM requires the MMR pathway and DNA polymerase η. In a report by Vo et al. (Vo et al., 2005), it was suggested that the Mre11 component of the MRN complex is the 3’-to-5’ exonuclease that mediates the 3’-directed MMR reaction. Furthermore, the MRN complex was also implicated in the SHM process (Larson et al., 2005; Yabuki et al., 2005). Thus, I sought to determine whether Mre11 contributes to the accumulation of mutations at A:T base pairs in hypermutating Ramos B cells. I reduced Mre11 expression levels in the IgM-negative Ramos 67 and 80 variant clones and successfully identified multiple clones with diminished Mre11 levels (Fig. 4.1.).

Unexpectedly, I did not observe a difference in the survival of Mre11<sub>low</sub> and control clones, suggesting that the residual Mre11 levels in Mre11<sub>low</sub> clones were sufficient to mediate DNA repair following treatment with IR (Fig. 4.2.). However, I did observe that prolonged culture of Mre11<sub>low</sub> clones, even under selective conditions, lead to the upregulation of Mre11 to near wild type levels (Fig. 4.1.C.). Furthermore, initial screens for Mre11<sub>low</sub> clones led to the identification of 2 Ramos 80 clones that expressed Mre11 at extremely low levels, but did not survive the culture period. These data suggest that a very low level of functional Mre11 is sufficient to maintain vital cellular function, but that Ramos cells tolerate low Mre11 levels, but are nevertheless at a disadvantage compared to cells expressing higher levels of the protein.

Using a GFP-based assay to measure MMR activity, I found that Mre11<sub>low</sub> clones retained full MMR function (Fig.4.3.A.) These results contradict those of Vo et al. who found that Mre11 knock down in HeLa cells resulted in ~20-fold increase in GFP-reversion (Vo et al., 2005). Several possibilities may explain this discrepancy. Our clones expressed ~10%
of wild type Mre11 levels, while the expression level in the clone utilized by Her et al. was ~3% of wild type. Thus, it is possible that an effect on MMR function is observed only at extremely reduced Mre11 levels. Moreover, I examined over 30 clones, while Her et al. used only one. As can be seen in Figure 4.3., reversion-based assays often yield a range of data where some clones are located substantially above or below the median value. Therefore, it is difficult to obtain reliable reversion data from only one clone, making it particularly important to screen a larger number of clones. Finally, the difference in GFP reversion between the Msh2+/− and Msh2−/− pre-B cell lines is greater than 2 orders of magnitude, while the difference in the study by Her et al. was only ~1 order of magnitude, suggesting a modest effect on MMR function.

As expected from the MMR data, I did not observe an effect on A:T mutations in the V regions of Mre11	extsuperscript{low} Ramos clones (Figure 4.3.B.). However, this still left open the possibility that Mre11 might be involved in the generating G:C mutations during SHM. However, I did not detect any difference in the mutation frequency or mutation spectrum between control and Mre11	extsuperscript{low} clones. An important caveat is that I observed only 8 substitution mutations, making it difficult to draw conclusions about the mutation spectrum. This is in contrast to the study by Yabuki et al. who observed a modest (~3-fold) increase in SHM in Ramos and DT40 cells in which Nbs1 was overexpressed, accompanied by a bias toward transversion mutations on the non-transcribed strand (Yabuki et al., 2005). Given that the assays used to measure mutation frequency were different, it is possible that small effects on SHM could be observed using different assays.

Collectively, these data indicate that the shRNA strategy used to knock down Mre11 in Ramos cells is effective in reducing Mre11 levels, but it is challenging to obtain viable
clones with extremely low levels of Mre11 (our data and (Reina-San-Martin et al., 2005)).
In the absence of Ramos clones with functional defects in the MRN pathway, it remains
difficult to conclude whether the MRN complex participates in MMR or SHM. This might be
resolved with an inducible Mre11 knock-out in which the ON and OFF status of Mre11
could be temporally controlled. Subsequent to the completion of this study, Kadyrov et al.
revealed that MMR is able to proceed in both 3’ and 5’ directions as a result of the cryptic
endonuclease function of MutLα (Kadyrov et al., 2006). MutLα introduces a nick on either
side of the mismatch specifically in the DNA strand that contains a pre-existing nick – that is, the
nascent strand – thereby providing an entry site for Exo1 on the 5’ side of the mismatch when the
original nick is located on the 3’ side. Although the issue of 3’-directed MMR has now been
resolved, it will be necessary to investigate this matter further before the role of the MRN
complex in SHM is fully elucidated.
Chapter 5

Discussion and Future Directions
Most physiological situations call for the avoidance of mutations as they pose a potentially deleterious threat to the organism. In some cases, however, mutations are desired and necessary in order to create diversity, enable adaptation, and drive evolution. The Ig locus of activated B cells is subject to strong selection pressure and undergoes a rapid evolutionary process in the germinal centre. This is an interesting situation, indeed, as the mutagenesis of the Ig locus is required for the production of a strong immune response and can be viewed as a method used by the organism to keep pace with rapidly evolving pathogens. Moreover, the full spectrum of mutations relies upon the action of DNA repair pathways (MMR and BER) that fulfill a highly uncharacteristic role in their contribution to, rather than elimination of, mutations. A considerable amount of insight has been gained in identifying the factors and pathways involved in the SHM process, but little is known about the triggers or signals that induce these pathways to take on a mutagenic role.

The importance of secondary antibody diversification mechanisms is illustrated by a variety of immunodeficiency syndromes, all characterized by a modified antibody spectrum and increased susceptibility to infection by encapsulated bacteria (Bruton, 1952; Cunningham-Rundles and Bodian, 1999; Hermaszewski and Webster, 1993; Lederman and Winkelstein, 1985; Oksenhendler et al., 2008). Hyper IgM syndrome, due to deficiency in AID, CD40, CD40L (CD154), or UNG, is characterized by lack of SHM and CSR and elevated IgM serum titers with a deficiency in other Ig isotypes (Allen et al., 1993; Callard et al., 1994; Conley et al., 1994; DiSanto et al., 1993; Ferrari et al., 2001; Imai et al., 2003; Korthauer et al., 1993; Lee et al., 2005; Lougaris et al., 2005; Minegishi et al., 2000; Revy et al., 2000). Common variable immunodeficiency (CVID) is caused by other genetic defects, including ICOS, CD19 and TACI, but has similar clinical manifestations (Castigli and Geha, 2007; Castigli et al., 2007; Grimbacher et al., 2003; Kanegane et al., 2007; Pan-
Hammarstrom et al., 2007; Salzer et al., 2004; van Zelm et al., 2006; Zhang et al., 2007).

SHM and CSR processes contribute to autoimmunity diseases as well, through the mutation of the natural antibody repertoire produced by CD5+ B-1 cells (Casali et al., 1987; Duquerroy et al., 2007; Harindranath et al., 1991; Ichiyoshi and Casali, 1995; Kasaian et al., 1994; Li et al., 2000; Mantovani et al., 1993).
5.1. Generation of MMR-dependent mutations at A:T base pairs in the absence of UNG

Approximately 60% of mutations at A:T base pairs in the V region are contributed by the MMR pathway (Martin et al., 2003; Phung et al., 1998; Rada et al., 1998) with the remainder arising from the UNG/BER pathway (Rada et al., 2004). Although previous reports on UNG−/− mice failed to detect any effect on the A:T mutation frequency, our investigations have demonstrated that UNG is instrumental in the generation of a subset of MMR-dependent A:T mutations. Our data led us to develop a model in which the MMR and UNG/BER pathways operate together to process G:U lesions. This processing generates an abasic site located in a ssDNA tract, which activates translesion synthesis by polymerase η that then produces A:T mutations. However, our data also reveal that a portion of mutations at A:T base pairs arise in the absence of UNG, indicating that hypermutating B cells use at least one other mechanism to generate A:T mutations via the MMR pathway.

5.1.1. Non-UNG mediated induction of mutagenic MMR.

Although four uracil DNA glycosylases are found in mammals, UNG is only one with a demonstrated role in SHM. Nevertheless, it is possible that SMUG1, which removes uracil from single- and double-stranded DNA, can substitute for UNG, provided the correct substrate or cell cycle phase. SMUG1 may remove AID-generated uracils to generate abasic sites, the majority of which are repaired faithfully to restore the initial G:C base pair. However, a small number of abasic sites generated by SMUG1 in the context of MMR-degradation tracts may be sufficient to induce translesion synthesis and A:T mutagenesis by DNA polymerase η, without generating G:C mutations. Since a mouse strain doubly
deficient for UNG and SMUG1 would lack both modes of uracil excision, any further effect on the A:T mutation frequency would indicate a role for SMUG1 in the MMR phase of SHM.

Translesion synthesis is activated by the presence of template strand DNA lesions that cannot be bypassed by replicative DNA polymerases, however several reports have suggested that exposed ssDNA and RPA is sufficient for the induction of translesion synthesis (Chang et al., 2006; Davies et al., 2008; Lopes et al., 2006; Northam et al., 2006; Suzuki et al., 2009; Tsuji et al., 2008). It is possible, therefore, that the accumulation of MMR-induced ssDNA gaps may also induce the ubiquitination of PCNA and translesion synthesis during the SHM process. This would provide a potential mechanism for PCNA ubiquitination and translesion synthesis initiation in the absence of abasic site lesions. Biochemical analyses would be required to examine whether the MMR-generated ssDNA gaps induce translesion synthesis in hypermutating B cells.

The work of Langerak et al. (Langerak et al., 2007) indicates that translesion synthesis is instrumental in the generation of mutations at A:T base pairs since mice harboring the PCNA^K164R mutation resemble polymerase η, Msh2−/− and other MMR-deficient strains in their reduced accumulation of A:T mutations in the V region. However, these data also indicate that translesion synthesis induced by ubiquitinated PCNA does not account for all A:T mutations since PCNA^K164R strains of mice possess ~4% A:T mutations, compared to ~60% for wild type mice, with a normal G:C mutation spectra (Langerak et al., 2007). Two other possibilities may account for the continued recruitment of error-prone polymerases to the hypermutating V region in the absence of PCNA-ubiquitin. The first is the induction of translesion synthesis by the alternative DNA clamp, the 9-1-1 complex (Caspari et al., 2000; Thelen et al., 1999; Venclovas and Thelen, 2000), or Rev1. The second
possibility is that the MMR pathway in hypermutating B cells is intrinsically error-prone and utilizes error-prone polymerases rather than the high fidelity DNA polymerase δ.

The 9-1-1 complex, that is, the alternative DNA clamp, consists of Rad9, Rad1, and Hus1 and undergoes post-translational modification in response to DNA damage. It is possible that the 9-1-1 clamp participates in the induction of translesion synthesis in SHM, if not along side of PCNA, then perhaps as an alternative pathway (Sabbioneda et al., 2005). This would explain why the abrogation of the PCNA-ubiquitination results in ~4% mutations at A:T base pairs and has little effect on G:C transversion mutations (Langerak et al., 2007), which also require the action of a translesional polymerase. Conditional knock-outs of this complex will be needed to determine whether it is involved in the generation of mutations during SHM. An alternative explanation is that Rev1, which like ubiquitinated PCNA, possesses binding sites for translesional polymerases and has been postulated to play a structural role in translesion synthesis, may recruit error-prone polymerases to the mutating V region. Although G:C transversion mutations are reduced in Rev1-deficient mice and cell lines, A:T mutations are not perturbed, suggesting that this polymerase does not actively participate in the A:T mutagenesis pathway. Nevertheless, it would be interesting to examine the effect of an enzymatically-inactive, but structurally intact Rev1 protein in SHM.

The possibility that MMR is intrinsically error-prone in hypermutating B cells has recently been suggested by two laboratories (Bhattacharya et al., 2008; Ouchida et al., 2008). They postulate that the MMR pathway is conditioned to recruit translesional polymerases during the synthesis stage of repair and that this leads to an increased basal mutation frequency in activated, but not follicular, B cells. Indeed, Wilson et al. observed
interactions between MutSα and polymerase η and demonstrated that the enzymatic activity of polymerase η is stimulated by MutSα (Wilson et al., 2005). However, an intrinsically mutagenic MMR pathway can potentially induce mutation at non-Ig genes; this would present great dangers to the fate of activated B cells.

5.1.2. A role for MutLα in somatic hypermutation.

In order to initiate the degradation step of MMR, Exo1 requires the presence of a ssDNA break. Under wild type conditions, UNG-mediated uracil excision and APE1-mediated abasic site cleavage create the nick that is then used by Exo1. However, since A:T mutagenesis proceeds in the absence of UNG and possibly in the absence of abasic sites, another mechanism must also produce this nick. If the MMR phase of SHM operates during S phase, then nicks may be produced by Okazaki fragments on the lagging strand and 3’ hydroxyl ends on the leading strand. This situation suggests that, as in conventional MMR, MMR during hypermutation is similarly biased to the lagging strand. This would suggest that A:T mutations occur more frequently on the lagging strand than on the leading strand. This may indeed be the case as A:T mutations are biased to the non-transcribed strand, which corresponds to the lagging strand if a replication fork originates at a 3’ origin of replication.

An alternative possibility is that the MutLα endonuclease is redundant in the presence of UNG, but is required when UNG is absent. This may be the case in the peripheral 5’ and 3’ sections of switch regions where UNG/APE1-mediated processing is insufficient to generate double stranded breaks and the MMR pathway plays a more substantial role. This would explain why Mlh1−/− and Pms2−/− strains of mice do not display
SHM defects, but do show modest effect on CSR. Mice doubly deficient for UNG and Mlh1 or UNG and Pms2 could be used to determine whether MutLα is involved in producing strand breaks used by Exo1, with reduced mutations at A:T base pairs suggesting that the MutLα endonuclease is indeed responsible for producing these nicks.
5.2. Generation of MMR-independent mutations at A:T base pairs

Mutations at A:T base pairs are generated predominantly by the MMR pathway, in concert with polymerase η and ubiquitinated PCNA. The UNG/Ber pathway also contributes to A:T mutagenesis, as seen in the absence of MMR. It is difficult to determine what proportion of A:T mutations are due solely to UNG since UNG−/− mice display a slight reduction in A:T mutations, and these may be additionally dependent on the MMR pathway. Thus, it is possible that UNG-generated A:T mutations arise only in the absence of a functional MMR pathway. Nevertheless, it is interesting to consider how UNG-mediated A:T mutations are generated.

The BER pathway can proceed via two repair modes: short-patch or long-patch. The former involves removal of the target base by the appropriate glycosylase and production of an abasic site, cleavage of the DNA phosphate backbone by APE1, minor processing and insertion of the correct base by polymerase β, and ligation by DNA ligase I. In contrast, long-patch repair is initiated in the same way, but differs after the cleavage step, with extra processing and degradation of up to ~10 nucleotides by the flap endonuclease Fen1 required to repair the lesion. Although it has not been demonstrated, it is likely that the conversion of AID-generated G:U lesions into transversion mutations at G:C base pairs requires only short patch repair. Nevertheless, insertion of the incorrect base in place of the U requires the action of a translesional polymerase other than polymerase β, which has a relatively low error rate of ~10⁻⁴ (Kunkel, 1985), considerably lower than some of the other translesional polymerases such as polymerase η. An alternative possibility is that following uracil excision, the partially-processed DNA undergoes replication and the abasic
site induces translesion synthesis. This would, however, require that AID-generated G:U lesions are processed during S phase.

In order to produce A:T mutations, long-patch BER accompanied by degradation of the DNA phosphate backbone downstream of the abasic site would be required in order to expose A and T nucleotides in the template strand as substrates for polymerase η. A:T mutations generated by this mechanism would be located within ~10 nucleotides 3’ of the G:U lesion. Since the original lesion may be correctly repaired, its presence would need to be inferred from the AID WRC hot spot motifs. Since the V regions of MMR-deficient mice contain a small number of A:T mutations, an extensive analysis using a large pool of mice is required to conduct this analysis.
5.3. **Translesional polymerases and somatic hypermutation**

The recruitment of DNA polymerase η and its participation in the mutation of A:T base pairs with the exclusion of other translesional polymerases is well established. With the exception of the study that demonstrated that MutSα stimulates polymerase η (Wilson et al., 2005), little is known about the near exclusive use of this polymerase in SHM. The need for polymerase η is evident since it has a high error-rate, one of the highest of all the known error-prone polymerases in eukaryotes, and “specializes” in mutating A:T base pairs (Matsuda et al., 2001). What, however, precludes the action of other polymerases such as polymerase ι, which has an extremely high error rate at ~10⁻¹ mutations per base pair or polymerase κ, which has been shown to replace polymerase η in its absence is not known (Faili et al., 2009; Johnson et al., 2000).

Unrestricted access to the SHM machinery is one explanation for the preferential usage of polymerase η. Unlike polymerase κ, which is localized to the replication factories in only a small proportion of S-phase cells, polymerase η is highly abundant (Kannouche et al., 2001; Ogi et al., 2005). On the other hand, although polymerase ι is highly error-prone opposite template G, it is extremely inefficient at extending from a mismatched primer-template terminus (Johnson et al., 2000). Thus, even if polymerase ι is recruited to the V region, it has little opportunity at extensive DNA synthesis. Polymerase β has been demonstrated to be downregulated in hypermutating B cells, possibly accounting for the use of another polymerase such as Rev1 during the UNG/BER pathway (Poltoratsky et al., 2007). In contrast, polymerase Θ, which is known to efficiently insert nucleotides opposite to and extend from abasic sites, is upregulated in hypermutating B cells (Kawamura et al., 2004; Zan et al., 2005). Furthermore, V regions of polymerase Θ⁻/⁻ mice are reduced for
transversion mutations at G:C and A:T base pairs (Zan et al., 2005). Like polymerase η, polymerase Θ preferentially incorporates the incorrect base opposite template T (Seki et al., 2004). Thus, it is possible that these two polymerases work together to generate mutations during SHM, but with contradicting data, the role of polymerase Θ is still not clear (Martomo et al., 2008; Masuda et al., 2007).
5.4. Conclusions

The identification of AID as the enzyme central to somatic hypermutation, CSR and Ig gene conversion spurred an explosion of knowledge in the field of antibody diversification. The data outlined herein contribute to that field by examining the contribution of MMR to mutagenesis, a role that opposes this DNA repair pathway’s conventional function. An examination of V region mutations in Ramos cells led to a novel mechanism for mutagenic mismatch repair involving bypass of an UNG-generated abasic site. The data demonstrated that the MMR and BER pathways cooperate to induce translesion synthesis and mutagenesis at the V region. These findings were further substantiated by data from murine B cells, which additionally demonstrated that the MMR pathway contributes substantially to transversion mutations at G:C base pairs. Finally, the involvement of the Mre11-Rad50-Nbs1 complex in MMR and the production of mutations at A:T base pairs was examined. Ablating Mre11 levels in Ramos cells while maintaining cell viability proved to be technically challenging. Nevertheless, our data demonstrated that reduced levels of Mre11 do not perturb the mismatch repair or A:T mutation pathways in Ramos cells.
Chapter 6

Appendices
Appendix 1. Primary sequence data of IgM+ Ramos 67 subclones that have associated mutations with the reverted nonsense codon.

31 of 51 clones are shown that have a mutation either at an A:T or G:C basepair (or both) that is associated with the reverted nonsense codon. Because the nonsense codon is within the YYC motif within the framework 3 region, the nonsense codon reverts only to a TAC or a TAT codon, which encodes a tyrosine.
Appendix 1.

ACTGGGGCCCTCCCTCATCCCTTTTACCTATCCATACAAAGGCAACCACCCACATGCAAATCCTCACTTTAGGCAACCACAG

GAAATGACTACACATTTCCTTAAATTCAGGGTCCAGCTCACATGGGAAGTGCTTTCTGAGAGTCATGGACCTCCTGCACA
ATATGAGATGCTGCCCTGATCCAGGCTCAGTGTTCTGTTACACAGGGTTCTGCTCCAGGTGCAGCTACAG

CAGTGGGGCGCAGGACTGTTGAAGCCTTCGGAGACCCTGTCTCTCACCTGCGGTGTTTATGGTGGGTCCTTCAGTGGTTA

GAGAATGATGCTGCCCTGATCCAGGCTCAGTGTTCTGTTACACAGGGTTCTGCTCCAGGTGCAGCTACAG

CAGTGGGGCGCAGGACTGTTGAAGCCTTCGGAGACCCTGTCTCTCACCTGCGGTGTTTATGGTGGGTCCTTCAGTGGTTA

GAGAATGATGCTGCCCTGATCCAGGCTCAGTGTTCTGTTACACAGGGTTCTGCTCCAGGTGCAGCTACAG
GCGTCTGATGGGACTAATCGGAGCCTGGGGCTTGGGGAGCCACATTTGGACGAGATGCCTGAACAAACCAGGGGTCTTAGT
880

GATGGCTGAGGAATGTGTCTCAGGAGCGGTGTCTGTAGGA 920
Appendix 2. Restricted mutation dataset from 9 wildtype and 5 UNG\textendash/- mice.

<table>
<thead>
<tr>
<th>Sequence(^1)</th>
<th>Total Mutations</th>
<th>C(_\text{TRN})</th>
<th>C(_\text{TRV})</th>
<th>A(_\text{MUT})</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mutation</td>
<td>Position</td>
<td>Context(^2)</td>
<td>Mutation</td>
</tr>
<tr>
<td>WT-1-1</td>
<td>5</td>
<td>C&gt;T</td>
<td>379</td>
<td>TATTT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WT-1-2</td>
<td>5</td>
<td>G&gt;A</td>
<td>77</td>
<td>TTTATT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WT-1-3</td>
<td>5</td>
<td>G&gt;A</td>
<td>146</td>
<td>AAATT</td>
</tr>
<tr>
<td></td>
<td>G&gt;T</td>
<td>203</td>
<td>ATTTT</td>
<td>G&gt;A</td>
</tr>
<tr>
<td>WT-1-4</td>
<td>5</td>
<td>G&gt;A</td>
<td>94</td>
<td>GAACA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WT-1-5</td>
<td>5</td>
<td>G&gt;T</td>
<td>113</td>
<td>TGTTA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WT-1-6</td>
<td>4</td>
<td>C&gt;T</td>
<td>72</td>
<td>AATCT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WT-1-7</td>
<td>5</td>
<td>G&gt;A</td>
<td>146</td>
<td>AAATT</td>
</tr>
<tr>
<td></td>
<td>G&gt;A</td>
<td>401</td>
<td>GAACA</td>
<td>G&gt;T</td>
</tr>
<tr>
<td>WT-1-8</td>
<td>4</td>
<td>G&gt;A</td>
<td>146</td>
<td>AAATT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WT-1-9</td>
<td>5</td>
<td>G&gt;C</td>
<td>85</td>
<td>GACCT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WT-1-10</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WT-1-11</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WT-1-12</td>
<td>3</td>
<td>G&gt;C</td>
<td>85</td>
<td>GACCT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WT-1-13</td>
<td>2</td>
<td>C&gt;T</td>
<td>133</td>
<td>AGTTCG</td>
</tr>
<tr>
<td>WT-1-14</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WT-1-15</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WT-1-16</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WT-1-17</td>
<td>1</td>
<td>G&gt;A</td>
<td>94</td>
<td>GAACA</td>
</tr>
<tr>
<td>WT-1-18</td>
<td>1</td>
<td>G&gt;A</td>
<td>135</td>
<td>CCAGC</td>
</tr>
<tr>
<td>WT-1-19</td>
<td>1</td>
<td>G&gt;A</td>
<td>43</td>
<td>AAAAAA</td>
</tr>
<tr>
<td>WT-2-1</td>
<td>4</td>
<td>C&gt;G</td>
<td>124</td>
<td>CCGTG</td>
</tr>
<tr>
<td>--------</td>
<td>---</td>
<td>-----</td>
<td>-----</td>
<td>-------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WT-2-2</td>
<td>5</td>
<td>T&gt;C</td>
<td>59</td>
<td>GGCCT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T&gt;C</td>
<td>303</td>
<td>CCCAT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A&gt;G</td>
<td>354</td>
<td>GAGTA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A&gt;G</td>
<td>358</td>
<td>AAGAG</td>
</tr>
<tr>
<td>WT-2-3</td>
<td>4</td>
<td>G&gt;A</td>
<td>129</td>
<td>AGAGA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C&gt;T</td>
<td>133</td>
<td>ACTCG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A&gt;G</td>
<td>17</td>
<td>GAGCC</td>
</tr>
<tr>
<td>WT-2-4</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WT-2-5</td>
<td>3</td>
<td>G&gt;A</td>
<td>112</td>
<td>TTAGA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G&gt;A</td>
<td>70</td>
<td>TTGAG</td>
</tr>
<tr>
<td>WT-2-6</td>
<td>4</td>
<td>G&gt;A</td>
<td>217</td>
<td>AAAAA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G&gt;A</td>
<td>216</td>
<td>TAGGA</td>
</tr>
<tr>
<td>WT-2-7</td>
<td>3</td>
<td>C&gt;T</td>
<td>221</td>
<td>ATTTG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C&gt;T</td>
<td>379</td>
<td>ATTTT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C&gt;T</td>
<td>388</td>
<td>ACTAC</td>
</tr>
<tr>
<td>WT-2-8</td>
<td>3</td>
<td>C&gt;T</td>
<td>482</td>
<td>CCTTT</td>
</tr>
<tr>
<td>WT-2-9</td>
<td>3</td>
<td>C&gt;T</td>
<td>72</td>
<td>AATCT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G&gt;A</td>
<td>94</td>
<td>GAACA</td>
</tr>
<tr>
<td>WT-2-10</td>
<td>2</td>
<td>G&gt;A</td>
<td>261</td>
<td>TAAGG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G&gt;A</td>
<td>183</td>
<td>TTGGG</td>
</tr>
<tr>
<td>WT-2-11</td>
<td>1</td>
<td>G&gt;A</td>
<td>352</td>
<td>TAAA</td>
</tr>
<tr>
<td>WT-2-12</td>
<td>2</td>
<td>G&gt;A</td>
<td>151</td>
<td>GAAAA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G&gt;A</td>
<td>101</td>
<td>GCGGA</td>
</tr>
<tr>
<td>WT-2-13</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WT-2-14</td>
<td>1</td>
<td>G&gt;A</td>
<td>43</td>
<td>AAAAA</td>
</tr>
<tr>
<td>WT-2-15</td>
<td>1</td>
<td>C&gt;T</td>
<td>179</td>
<td>GCTTT</td>
</tr>
<tr>
<td>WT-2-16</td>
<td>1</td>
<td>G&gt;A</td>
<td>262</td>
<td>AGAGA</td>
</tr>
<tr>
<td>WT-2-17</td>
<td>1</td>
<td>C&gt;T</td>
<td>60</td>
<td>GTTTT</td>
</tr>
<tr>
<td>WT-2-18</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WT-2-19</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WT-2-20</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WT-3-1</td>
<td>1</td>
<td>C&gt;T</td>
<td>482</td>
<td>CCTTT</td>
</tr>
<tr>
<td>WT-3-2</td>
<td>4</td>
<td>G&gt;A</td>
<td>35</td>
<td>CAACT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C&gt;T</td>
<td>51</td>
<td>TATTTG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T&gt;A</td>
<td>52</td>
<td>ACAGC</td>
</tr>
<tr>
<td>WT-3-3</td>
<td>1</td>
<td>G&gt;A</td>
<td>35</td>
<td>CAACT</td>
</tr>
<tr>
<td>WT-3-4</td>
<td>1</td>
<td>G&gt;T</td>
<td>100</td>
<td>AATGA</td>
</tr>
<tr>
<td>WT-3-5</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WT-3-6</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WT-3-7</td>
<td>1</td>
<td>G&gt;A</td>
<td>217</td>
<td>TTACC</td>
</tr>
<tr>
<td>WT-3-8</td>
<td>3</td>
<td>G&gt;C</td>
<td>541</td>
<td>TGCTG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A&gt;T</td>
<td>238</td>
<td>TTGAG</td>
</tr>
<tr>
<td>Gene</td>
<td>No.</td>
<td>Change</td>
<td>Site</td>
<td>Amino Acid Change</td>
</tr>
<tr>
<td>-------</td>
<td>-----</td>
<td>--------</td>
<td>------</td>
<td>-------------------</td>
</tr>
<tr>
<td>WT-3-9</td>
<td>1</td>
<td>A&gt;T</td>
<td>621</td>
<td></td>
</tr>
<tr>
<td>WT-3-10</td>
<td>1</td>
<td>G&gt;A</td>
<td>464</td>
<td></td>
</tr>
<tr>
<td>WT-3-11</td>
<td>1</td>
<td>A&gt;T</td>
<td>168</td>
<td></td>
</tr>
<tr>
<td>WT-4-1</td>
<td>3</td>
<td>G&gt;A</td>
<td>748</td>
<td></td>
</tr>
<tr>
<td>WT-4-2</td>
<td>1</td>
<td>T&gt;C</td>
<td>1452</td>
<td></td>
</tr>
<tr>
<td>WT-4-3</td>
<td>1</td>
<td>A&gt;G</td>
<td>946</td>
<td></td>
</tr>
<tr>
<td>WT-4-4</td>
<td>5</td>
<td>G&gt;A</td>
<td>392</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>C&gt;T</td>
<td>240</td>
<td></td>
</tr>
<tr>
<td>WT-4-5</td>
<td>3</td>
<td>G&gt;C</td>
<td>105</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>A&gt;T</td>
<td>292</td>
<td></td>
</tr>
<tr>
<td>WT-4-6</td>
<td>2</td>
<td>A&gt;T</td>
<td>594</td>
<td></td>
</tr>
<tr>
<td>WT-4-7</td>
<td>1</td>
<td>A&gt;C</td>
<td>569</td>
<td></td>
</tr>
<tr>
<td>WT-5-1</td>
<td>4</td>
<td>T&gt;C</td>
<td>1148</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>A&gt;C</td>
<td>1065</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>A&gt;T</td>
<td>1065</td>
<td></td>
</tr>
<tr>
<td>WT-5-2</td>
<td>1</td>
<td>C&gt;T</td>
<td>413</td>
<td></td>
</tr>
<tr>
<td>WT-5-3</td>
<td>2</td>
<td>C&gt;T</td>
<td>983</td>
<td></td>
</tr>
<tr>
<td>WT-5-4</td>
<td>1</td>
<td>G&gt;A</td>
<td>1201</td>
<td></td>
</tr>
<tr>
<td>WT-6-1</td>
<td>5</td>
<td>A&gt;G</td>
<td>596</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>A&gt;T</td>
<td>447</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>A&gt;G</td>
<td>355</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>A&gt;G</td>
<td>330</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>A&gt;T</td>
<td>325</td>
<td></td>
</tr>
<tr>
<td>WT-6-2</td>
<td>1</td>
<td>C&gt;T</td>
<td>206</td>
<td></td>
</tr>
<tr>
<td>WT-6-3</td>
<td>1</td>
<td>C&gt;A</td>
<td>1154</td>
<td></td>
</tr>
<tr>
<td>WT-6-4</td>
<td>1</td>
<td>A&gt;C</td>
<td>1372</td>
<td></td>
</tr>
<tr>
<td>WT-6-5</td>
<td>1</td>
<td>T&gt;A</td>
<td>1249</td>
<td></td>
</tr>
<tr>
<td>WT-6-6</td>
<td>4</td>
<td>C&gt;T</td>
<td>434</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>C&gt;T</td>
<td>206</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>G&gt;A</td>
<td>195</td>
<td></td>
</tr>
<tr>
<td>WT-6-7</td>
<td>2</td>
<td>A&gt;G</td>
<td>259</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>A&gt;T</td>
<td>213</td>
<td></td>
</tr>
<tr>
<td>WT-6-8</td>
<td>1</td>
<td>T&gt;A</td>
<td>79</td>
<td></td>
</tr>
<tr>
<td>WT-6-9</td>
<td>1</td>
<td>T&gt;A</td>
<td>758</td>
<td></td>
</tr>
<tr>
<td>WT-6-10</td>
<td>4</td>
<td>A&gt;G</td>
<td>596</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>A&gt;C</td>
<td>454</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>A&gt;T</td>
<td>325</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>A&gt;G</td>
<td>141</td>
<td></td>
</tr>
<tr>
<td>WT-6-11</td>
<td>1</td>
<td>T&gt;G</td>
<td>873</td>
<td></td>
</tr>
<tr>
<td>WT-6-12</td>
<td>1</td>
<td>T&gt;A</td>
<td>887</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>WT-7-1</td>
<td>1</td>
<td>G&gt;C</td>
<td>1005</td>
<td>GACTT</td>
</tr>
<tr>
<td>WT-7-2</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WT-7-3</td>
<td>1</td>
<td>G&gt;A</td>
<td>574</td>
<td>CTACC</td>
</tr>
<tr>
<td>WT-7-4</td>
<td>1</td>
<td>T&gt;C</td>
<td>1247</td>
<td>GCGTG</td>
</tr>
<tr>
<td>WT-7-5</td>
<td>5</td>
<td>C&gt;T</td>
<td>1123</td>
<td>ATTTT</td>
</tr>
<tr>
<td>WT-8-1</td>
<td>1</td>
<td>G&gt;A</td>
<td>683</td>
<td>ATAGA</td>
</tr>
<tr>
<td>WT-8-2</td>
<td>3</td>
<td>C&gt;T</td>
<td>555</td>
<td>AGTGG</td>
</tr>
<tr>
<td>WT-8-3</td>
<td>5</td>
<td>G&gt;A</td>
<td>123</td>
<td>GAATT</td>
</tr>
<tr>
<td>WT-9-1</td>
<td>1</td>
<td>C&gt;T</td>
<td>16</td>
<td>CTTCT</td>
</tr>
<tr>
<td>WT-9-2</td>
<td>2</td>
<td>G&gt;A</td>
<td>53</td>
<td>CACT</td>
</tr>
<tr>
<td>WT-9-3</td>
<td>3</td>
<td>T&gt;A</td>
<td>514</td>
<td>GTAGA</td>
</tr>
<tr>
<td>WT-9-4</td>
<td>4</td>
<td>C&gt;T</td>
<td>183</td>
<td>TCTTC</td>
</tr>
<tr>
<td>WT-9-5</td>
<td>4</td>
<td>C&gt;T</td>
<td>497</td>
<td>ATTEC</td>
</tr>
<tr>
<td>WT-9-6</td>
<td>1</td>
<td>C&gt;T</td>
<td>366</td>
<td>TGTGA</td>
</tr>
<tr>
<td>UNG-1-1</td>
<td>5</td>
<td>C&gt;T</td>
<td>298</td>
<td>ATAAAA</td>
</tr>
<tr>
<td>UNG-1-2</td>
<td>4</td>
<td>G&gt;A</td>
<td>353</td>
<td>TAAAA</td>
</tr>
<tr>
<td>UNG-1-3</td>
<td>5</td>
<td>G&gt;A</td>
<td>332</td>
<td>TATT</td>
</tr>
<tr>
<td>UNG-1-4</td>
<td>4</td>
<td>G&gt;A</td>
<td>78</td>
<td>TTATT</td>
</tr>
<tr>
<td>UNG-1-5</td>
<td>3</td>
<td>G&gt;A</td>
<td>95</td>
<td>GAACA</td>
</tr>
</tbody>
</table>

UNG-1-1 | 5 | C>T | 298 | ATAAAA |   |   |
<p>| UNG-1-2 | 4 | G&gt;A | 353 | TAAAA |   |   |
| UNG-1-3 | 5 | G&gt;A | 332 | TATT |   |   |
| UNG-1-4 | 4 | G&gt;A | 78 | TTATT |   |   |
| UNG-1-5 | 3 | G&gt;A | 95 | GAACA |   |   |</p>
<table>
<thead>
<tr>
<th></th>
<th>C&gt;T</th>
<th></th>
<th></th>
<th>A&gt;G</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>UNG-1-6</td>
<td>4</td>
<td>C&gt;T</td>
<td>96</td>
<td>AGTAT</td>
<td>A&gt;G</td>
</tr>
<tr>
<td>UNG-1-7</td>
<td>3</td>
<td>C&gt;T</td>
<td>203</td>
<td>ATTTT</td>
<td>C&gt;A</td>
</tr>
<tr>
<td>UNG-1-8</td>
<td>3</td>
<td>G&gt;A</td>
<td>338</td>
<td>GAATT</td>
<td>T&gt;A</td>
</tr>
<tr>
<td>UNG-1-9</td>
<td>3</td>
<td>G&gt;A</td>
<td>144</td>
<td>GAAAA</td>
<td>T&gt;C</td>
</tr>
<tr>
<td>UNG-1-10</td>
<td>3</td>
<td>G&gt;A</td>
<td>402</td>
<td>CAACA</td>
<td></td>
</tr>
<tr>
<td>UNG-1-11</td>
<td>2</td>
<td>G&gt;A</td>
<td>382</td>
<td>GATTTG</td>
<td>T&gt;C</td>
</tr>
<tr>
<td>UNG-2-1</td>
<td>5</td>
<td>C&gt;T</td>
<td>122</td>
<td>TCTCT</td>
<td>T&gt;G</td>
</tr>
<tr>
<td>UNG-2-2</td>
<td>4</td>
<td>G&gt;A</td>
<td>85</td>
<td>GAATT</td>
<td>A&gt;C</td>
</tr>
<tr>
<td>UNG-2-3</td>
<td>4</td>
<td>C&gt;T</td>
<td>123</td>
<td>TCTCT</td>
<td>A&gt;T</td>
</tr>
<tr>
<td>UNG-2-4</td>
<td>5</td>
<td>C&gt;T</td>
<td>72</td>
<td>AATCT</td>
<td>T&gt;C</td>
</tr>
<tr>
<td>UNG-2-5</td>
<td>4</td>
<td>G&gt;A</td>
<td>99</td>
<td>TTAAG</td>
<td>A&gt;G</td>
</tr>
<tr>
<td>UNG-2-6</td>
<td>4</td>
<td>G&gt;A</td>
<td>319</td>
<td>TAATT</td>
<td>A&gt;G</td>
</tr>
<tr>
<td>UNG-2-7</td>
<td>4</td>
<td>C&gt;T</td>
<td>100</td>
<td>TGTAG</td>
<td></td>
</tr>
<tr>
<td>UNG-2-8</td>
<td>4</td>
<td>G&gt;A</td>
<td>85</td>
<td>GAATT</td>
<td></td>
</tr>
<tr>
<td>UNG-2-9</td>
<td>4</td>
<td>G&gt;A</td>
<td>43</td>
<td>AAAAA</td>
<td>A&gt;C</td>
</tr>
<tr>
<td>UNG-2-10</td>
<td>4</td>
<td>G&gt;A</td>
<td>77</td>
<td>TTATT</td>
<td>A&gt;G</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UNG-2-11</td>
<td>4</td>
<td>G&gt;A</td>
<td>57</td>
<td>CAAGT</td>
<td>A&gt;G</td>
</tr>
<tr>
<td>-----------</td>
<td>---</td>
<td>-----</td>
<td>----</td>
<td>-------</td>
<td>-----</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G&gt;A</td>
<td>290</td>
<td>ATACT</td>
<td>A&gt;G</td>
</tr>
<tr>
<td>UNG-2-12</td>
<td>3</td>
<td>G&gt;A</td>
<td>143</td>
<td>GAAAA</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>G&gt;A</td>
<td>252</td>
<td>ATAAT</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>G&gt;A</td>
<td>360</td>
<td>AAATA</td>
<td></td>
</tr>
<tr>
<td>UNG-2-13</td>
<td>3</td>
<td>G&gt;A</td>
<td>151</td>
<td>GGAAA</td>
<td>C&gt;T</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G&gt;A</td>
<td>365</td>
<td>GAA</td>
<td></td>
</tr>
<tr>
<td>UNG-2-14</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td>A&gt;C</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>T&gt;A</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A&gt;G</td>
</tr>
<tr>
<td>UNG-2-15</td>
<td>3</td>
<td>C&gt;A</td>
<td>95</td>
<td>AGAAT</td>
<td>A&gt;G</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A&gt;G</td>
</tr>
<tr>
<td>UNG-2-16</td>
<td>4</td>
<td>G&gt;A</td>
<td>94</td>
<td>GAACA</td>
<td>C&gt;T</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G&gt;A</td>
<td>331</td>
<td>TAATT</td>
<td>C&gt;T</td>
</tr>
<tr>
<td>UNG-2-17</td>
<td>3</td>
<td>G&gt;A</td>
<td>85</td>
<td>GAATT</td>
<td>A&gt;G</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A&gt;C</td>
</tr>
<tr>
<td>UNG-2-18</td>
<td>4</td>
<td>G&gt;A</td>
<td>85</td>
<td>GAATT</td>
<td>C&gt;T</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>G&gt;A</td>
</tr>
<tr>
<td>UNG-2-19</td>
<td>3</td>
<td>G&gt;A</td>
<td>85</td>
<td>GAATT</td>
<td>C&gt;T</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UNG-2-20</td>
<td>4</td>
<td>C&gt;T</td>
<td>4</td>
<td>TATGG</td>
<td>A&gt;C</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G&gt;A</td>
<td>136</td>
<td>CGACT</td>
<td>A&gt;G</td>
</tr>
<tr>
<td>UNG-2-21</td>
<td>3</td>
<td>G&gt;A</td>
<td>85</td>
<td>GAATT</td>
<td>C&gt;T</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UNG-2-22</td>
<td>2</td>
<td>C&gt;T</td>
<td>133</td>
<td>AGTGG</td>
<td>A&gt;T</td>
</tr>
<tr>
<td>UNG-2-23</td>
<td>2</td>
<td>G&gt;A</td>
<td>151</td>
<td>GGAAA</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>G&gt;A</td>
<td>366</td>
<td>TAATT</td>
<td></td>
</tr>
<tr>
<td>UNG-2-24</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td>A&gt;C</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A&gt;C</td>
</tr>
<tr>
<td>UNG-2-25</td>
<td>2</td>
<td>C&gt;T</td>
<td>367</td>
<td>AGTGG</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>C&gt;T</td>
</tr>
<tr>
<td>UNG-2-26</td>
<td>2</td>
<td>G&gt;A</td>
<td>94</td>
<td>GAACA</td>
<td>T&gt;C</td>
</tr>
<tr>
<td>UNG-2-27</td>
<td>1</td>
<td>C&gt;T</td>
<td>203</td>
<td>ATTTT</td>
<td></td>
</tr>
<tr>
<td>UNG-2-28</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>A&gt;G</td>
</tr>
<tr>
<td>UNG-2-29</td>
<td>1</td>
<td>G&gt;A</td>
<td>57</td>
<td>AGATC</td>
<td></td>
</tr>
<tr>
<td>UNG-2-30</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>A&gt;G</td>
</tr>
<tr>
<td>UNG-2-31</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>T&gt;C</td>
</tr>
<tr>
<td>UNG-2-32</td>
<td>1</td>
<td>C&gt;T</td>
<td>95</td>
<td>AGTAT</td>
<td></td>
</tr>
<tr>
<td>UNG-2-33</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>G&gt;T</td>
</tr>
<tr>
<td>UNG-3-1</td>
<td>4</td>
<td>G&gt;A</td>
<td>404</td>
<td>CAACA</td>
<td>T&gt;G</td>
</tr>
<tr>
<td>--------</td>
<td>---</td>
<td>-----</td>
<td>-----</td>
<td>-------</td>
<td>-----</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C&gt;T</td>
<td>419</td>
<td>ATTTG</td>
<td>T&gt;G</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G&gt;A</td>
<td>322</td>
<td>TAATT</td>
<td></td>
</tr>
<tr>
<td>UNG-3-1</td>
<td>3</td>
<td>G&gt;A</td>
<td>228</td>
<td>GTATG</td>
<td>A&gt;G</td>
</tr>
<tr>
<td>UNG-3-1</td>
<td>2</td>
<td>G&gt;A</td>
<td>404</td>
<td>CAACA</td>
<td>A&gt;G</td>
</tr>
<tr>
<td>UNG-4-1</td>
<td>1</td>
<td>T&gt;C</td>
<td>460</td>
<td>TTGGT</td>
<td></td>
</tr>
<tr>
<td>UNG-4-2</td>
<td>3</td>
<td>C&gt;T</td>
<td>619</td>
<td>AGTTT</td>
<td></td>
</tr>
<tr>
<td>UNG-4-3</td>
<td>2</td>
<td>G&gt;A</td>
<td>97</td>
<td>GAACA</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>C&gt;T</td>
<td>277</td>
<td>GGTTC</td>
<td></td>
</tr>
<tr>
<td>UNG-4-4</td>
<td>5</td>
<td>G&gt;A</td>
<td>356</td>
<td>TAAAA</td>
<td>A&gt;C</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T&gt;A</td>
<td>119</td>
<td>TTAGT</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>A&gt;G</td>
<td>321</td>
<td>GTGGT</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>A&gt;G</td>
<td>330</td>
<td>GTGGT</td>
<td></td>
</tr>
<tr>
<td>UNG-4-5</td>
<td>1</td>
<td>G&gt;A</td>
<td>139</td>
<td>CGALT</td>
<td></td>
</tr>
<tr>
<td>UNG-4-6</td>
<td>3</td>
<td>T&gt;A</td>
<td>381</td>
<td>CTACA</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>T&gt;G</td>
<td>417</td>
<td>AGGAT</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>T&gt;G</td>
<td>508</td>
<td>TGGTG</td>
<td></td>
</tr>
<tr>
<td>UNG-4-7</td>
<td>5</td>
<td>G&gt;A</td>
<td>80</td>
<td>TTATT</td>
<td>A&gt;G</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G&gt;A</td>
<td>88</td>
<td>GAATT</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>G&gt;A</td>
<td>426</td>
<td>ATACA</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>G&gt;A</td>
<td>614</td>
<td>CACTC</td>
<td></td>
</tr>
<tr>
<td>UNG-4-8</td>
<td>5</td>
<td>C&gt;T</td>
<td>21</td>
<td>AATCT</td>
<td>T&gt;C</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C&gt;T</td>
<td>22</td>
<td>ACTTC</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>G&gt;A</td>
<td>60</td>
<td>CAAGT</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>C&gt;T</td>
<td>112</td>
<td>ATTTT</td>
<td></td>
</tr>
<tr>
<td>UNG-4-9</td>
<td>3</td>
<td>G&gt;T</td>
<td>753</td>
<td>ATTTT</td>
<td>T&gt;G</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G&gt;A</td>
<td>790</td>
<td>TTATT</td>
<td></td>
</tr>
<tr>
<td>UNG-4-10</td>
<td>2</td>
<td>A&gt;T</td>
<td>219</td>
<td>TATGAG</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>A&gt;G</td>
<td>222</td>
<td>GAGTC</td>
<td></td>
</tr>
<tr>
<td>UNG-4-11</td>
<td>5</td>
<td>G&gt;A</td>
<td>97</td>
<td>GAACA</td>
<td>A&gt;C</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G&gt;A</td>
<td>226</td>
<td>CTATG</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>G&gt;A</td>
<td>252</td>
<td>TGAAT</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>G&gt;A</td>
<td>335</td>
<td>TATT</td>
<td></td>
</tr>
<tr>
<td>UNG-4-12</td>
<td>5</td>
<td>C&gt;A</td>
<td>162</td>
<td>AAATG</td>
<td>T&gt;A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A&gt;C</td>
<td>110</td>
<td>TACTC</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>A&gt;G</td>
<td>212</td>
<td>GAGAA</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>A&gt;C</td>
<td>473</td>
<td>GGCAT</td>
<td></td>
</tr>
<tr>
<td>UNG-4-13</td>
<td>1</td>
<td>C&gt;T</td>
<td>572</td>
<td>TGTTG</td>
<td></td>
</tr>
<tr>
<td>UNG-4-14</td>
<td>2</td>
<td>A&gt;T</td>
<td>219</td>
<td>TATGAG</td>
<td></td>
</tr>
<tr>
<td>UNG-4-15</td>
<td>3</td>
<td>C&gt;T</td>
<td>629</td>
<td>CATAG</td>
<td>A&gt;G</td>
</tr>
<tr>
<td>----------</td>
<td>---</td>
<td>-----</td>
<td>-----</td>
<td>-------</td>
<td>-----</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G&gt;A</td>
<td>46</td>
<td>AAAAA</td>
<td></td>
</tr>
<tr>
<td>UNG-4-16</td>
<td>3</td>
<td>G&gt;A</td>
<td>51</td>
<td>GAAGC</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>G&gt;A</td>
<td>52</td>
<td>CTATC</td>
<td></td>
</tr>
<tr>
<td>UNG-4-17</td>
<td>2</td>
<td>C&gt;T</td>
<td>100</td>
<td>TGTAG</td>
<td>A&gt;G</td>
</tr>
<tr>
<td>UNG-4-18</td>
<td>1</td>
<td>G&gt;A</td>
<td>97</td>
<td>GAAAC</td>
<td></td>
</tr>
<tr>
<td>UNG-4-19</td>
<td>3</td>
<td>C&gt;T</td>
<td>206</td>
<td>ATTTT</td>
<td>T&gt;G</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A&gt;C</td>
<td>53</td>
<td>TGGCT</td>
<td></td>
</tr>
<tr>
<td>UNG-4-20</td>
<td>2</td>
<td>C&gt;T</td>
<td>10</td>
<td>G&gt;T</td>
<td></td>
</tr>
<tr>
<td>UNG-4-21</td>
<td>1</td>
<td>C&gt;T</td>
<td>71</td>
<td>TTGAG</td>
<td></td>
</tr>
<tr>
<td>UNG-5-1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UNG-5-2</td>
<td>5</td>
<td>C&gt;T</td>
<td>17</td>
<td>CATCG</td>
<td>A&gt;G</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G&gt;A</td>
<td>81</td>
<td>TTAAT</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>G&gt;A</td>
<td>173</td>
<td>GGAAT</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>G&gt;A</td>
<td>238</td>
<td>GTATT</td>
<td></td>
</tr>
<tr>
<td>UNG-5-3</td>
<td>2</td>
<td>G&gt;A</td>
<td>81</td>
<td>TTAAT</td>
<td></td>
</tr>
<tr>
<td>UNG-5-4</td>
<td>3</td>
<td>G&gt;A</td>
<td>103</td>
<td>TTACA</td>
<td>A&gt;T</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A&gt;T</td>
</tr>
<tr>
<td>UNG-5-5</td>
<td>4</td>
<td>C&gt;T</td>
<td>385</td>
<td>TATTT</td>
<td>T&gt;A</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>T&gt;A</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>T&gt;C</td>
</tr>
<tr>
<td>UNG-5-6</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>A&gt;T</td>
</tr>
<tr>
<td>UNG-5-7</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>A&gt;T</td>
</tr>
<tr>
<td>UNG-5-8</td>
<td>3</td>
<td>G&gt;A</td>
<td>427</td>
<td>ATACA</td>
<td>A&gt;G</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A&gt;T</td>
</tr>
<tr>
<td>UNG-5-9</td>
<td>4</td>
<td>G&gt;A</td>
<td>238</td>
<td>GTATT</td>
<td>A&gt;T</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>G&gt;A</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A&gt;G</td>
</tr>
<tr>
<td>UNG-5-10</td>
<td>5</td>
<td>C&gt;T</td>
<td>7</td>
<td>TATGG</td>
<td>C&gt;G</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>T&gt;A</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>C&gt;T</td>
</tr>
<tr>
<td>UNG-5-11</td>
<td>1</td>
<td>C&gt;T</td>
<td>7</td>
<td>TATGG</td>
<td></td>
</tr>
<tr>
<td>UNG-5-12</td>
<td>1</td>
<td>C&gt;T</td>
<td>9</td>
<td>AGTAT</td>
<td></td>
</tr>
<tr>
<td>UNG-5-13</td>
<td>5</td>
<td>C&gt;T</td>
<td>741</td>
<td>TATTC</td>
<td>T&gt;A</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>T&gt;C</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>T&gt;A</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>T&gt;C</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>T&gt;A</td>
</tr>
<tr>
<td>UNG-5-14</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>A&gt;G</td>
</tr>
<tr>
<td>UNG-5-15</td>
<td>5</td>
<td>G&gt;A</td>
<td>103</td>
<td>TTACA</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>C&gt;T</td>
<td>104</td>
<td>TGTAG</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>G&gt;A</td>
<td>155</td>
<td>GGAAA</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>G&gt;A</td>
<td>366</td>
<td>AAATA</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>G&gt;A</td>
<td>452</td>
<td>GGAAC</td>
<td></td>
</tr>
<tr>
<td>UNG-5-16</td>
<td>5</td>
<td>C&gt;T</td>
<td>7</td>
<td>TATTG</td>
<td>C&gt;G</td>
</tr>
<tr>
<td>---------</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-------</td>
<td>-----</td>
</tr>
<tr>
<td></td>
<td>C&gt;T</td>
<td>39</td>
<td>TCTTC</td>
<td></td>
<td>A&gt;T</td>
</tr>
</tbody>
</table>

1 Sequences that have at least one A:T mutation and up to a total of 5 mutations are shown for 9 wildtype (WT) mice and 5 UNG⁻/⁻ (UNG) mice. Unrestricted data is shown in Figure 3.3.

2 Central nucleotide (bold) is the mutated nucleotide. Underlined sequences are WRC/RGY hotspot motifs for G:C mutations or WA hotspot motifs for A:T mutations. TRN = Transition; TRV = Transversion; MUT = Mutation
Chapter 7

References


agammaglobulinemia with normal or elevated levels of IgM (HIM). Comparison of X-linked, autosomal recessive, and non-X-linked forms of the disease, and obligate carriers. J Immunol 153, 3295-3306.


MacDuff, D. A., Neuberger, M. S., and Harris, R. S. (2006). MDM2 can interact with the C-terminus of AID but it is inessential for antibody diversification in DT40 B cells. Mol Immunol 43, 1099-1108.


