Novel insights into the biochemical mechanisms of RAG1 and RAG2 using mutational approaches

Daniel Thomas Thwaites

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy

The University of Leeds

Faculty of Biological Sciences

School of Molecular and Cellular Biology

July 2019

Intellectual Property and Publication Statements

The candidate confirms that the work submitted is his/her/their own, except where work which has formed part of jointly authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

Chapter 3 contains content adapted from a jointly authored publication: Thwaites, D.T., Carter, C., Lawless, D., Savic, S. and Boyes, J.M. 2019. A novel *RAG1* mutation reveals a critical in vivo role for HMGB1/2 during V(D)J recombination. *Blood*. **133**,p. 820. DTT, SS, and JMB conceived the project, designed the experiments, and wrote the paper. DTT performed the experiments for Figures 1, 2, 4 and 5; and produced Figures 3 and 6. CC, and DL provided intellectual input.

Acknowledgements

I would like to thank my supervisor, Dr Joan Boyes, for the thorough training and support she has provided throughout this project. In particular, I would like to thank Joan for always reminding me to step back, look at the bigger picture and ensure I am asking relevant and important scientific questions. I am also confident that after four years she is well trained in translating my ‘Yorkshire’ language to the Queen’s English! I would also like to thank my co-supervisor, Dr Ian Wood, for his helpful input during my PhD.

I would like to thank current and past members of the Boyes lab; Dr James Scott, Dr Alastair Smith, Zeqian Gao and Mathew Burke, for their support and ideas and making the lab a pleasant place to work. Notably, James and Alastair who welcomed me into the lab and helped me greatly during the early stages. I also thank Professor Peter Stockley and other members of the Stockley group for allowing me to use many pieces of their equipment. In particular, I would like to thank Dr Emma Wroblewski, Dr Nikesh Patel and Dr Dan Maskell for their training and help with the AKTA. I am also very grateful to Dr Sinisa Savic and members of his group for a very successful collaboration.

On a weirdly personal level I would like to thank my trusty Citroen C1 for getting me to and from the lab for the last four years. I’ve put it through 35,000 miles, five punctured tyres and one breakdown, but it still keeps going!

Saving the best till last, I would like to give a huge thank you to my family. To my mum, thank you for always being there for me and being supportive of whatever I do (and for all the free haircuts!). I am in debt to my dad, as always, for the ‘knowledge’ and for keeping my car running at its best. Finally, I would like to thank my sister, Laura, for ensuring I keep a work-life balance and ferrying me to and from Leeds for the last few months when fuel costs became too much. I look forward to Florida in May!

Abstract

V(D)J recombination generates a diverse repertoire of antigen receptors by somatically recombining different gene segments. The reaction is catalysed by the proteins RAG1 and RAG2 which bind and cleave recombination signal sequences flanking each gene segment. However, V(D)J recombination is a daring approach since it involves the breaking and rejoining of the genome, posing a threat to genome stability. Indeed, mistakes do occur, highlighted by 35-40% of lymphoid cancers bearing hallmarks of errors during antigen receptor production. Additionally, mutation of the RAG proteins can cause immunodeficiency, the severity of which reflects the importance of the altered residue(s) during recombination.

Here, I study the RAG proteins in two different ways: using naturally occurring or artificial mutations. Firstly, I analysed two *RAG1* mutations identified in an immunodeficient patient. Both mutations lead to a severe decrease in recombination *in vivo*, and analysis of the reaction *in vitro* revealed that one mutation disrupts the recruitment of the accessory protein HMGB1 to the RAG complex. Together with studies of recombination in HMGB1 deficient cells, this revealed a critical *in vivo* role for HMGB1/2 during V(D)J recombination. Next, I used artificially generated *RAG2* mutations in a bespoke *in vivo* assay to map the region of the RAG2 C-terminus that inhibits an aberrant reaction, namely reintegration, where recombination by-products are reinserted into the genome. This identified four residues in the RAG2 acidic hinge which are vital to suppress this dangerous reaction. Finally, I generated novel chromatin substrates, containing the H3K4me3 modification, to test a model for how the RAG proteins maintain genome integrity. This led to the unexpected discovery that the RAG2 C-terminus inhibits binding and cleavage of nucleosome substrates which lack the H3K4me3 modification. Together, these studies provide novel insights into the function of RAG proteins.

Contents

List of Figures and Tables

Figures

Tables

Abbreviations

AID	Activation induced cytidine deaminase
ALL	Acute lymphoblastic leukaemia
APC	Antigen presenting cell
APE1	Apurinic/apyrimidinic endonuclease 1
ATM	Ataxia telangiectasia mutated
BCR	B cell receptor
BER	Base excision repair
bp	Base pair
BSA	Bovine serum albumin
CAT	Chloramphenicol acetyltransferase
CDK	Cyclin dependent kinase
cDNA	Complementary DNA
CDR	Complementarity determining region
CE	Coding end
CID	Combined immunodeficiency
CJ	Coding joint
CLP	Common lymphoid progenitor
CRISPR	Clustered regularly interspaced short palindromic repeats
cRSS	Cryptic RSS
CSR	Class switch recombination
CTCF	CCCTC-binding factor
CTD	Carboxy-terminal domain
DDBD	Dimerisation and DNA binding domain
DMEM	Dulbecco’s modified Eagle’s medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DNA-PKcs	DNA-dependent protein kinase catalytic subunit
DSB	Double strand break
DTT	Dithiothreitol

EDTA	Ethylenediaminetetraacetic acid
EM	Electron microscopy
ESC	Excised signal circle
Fab	Fragment antigen binding
FACS	Fluorescent activated cell sorting
Fc	Fragment crystallisation
FCS	Foetal calf serum
FISH	Fluorescent *in situ* hybridisation
FITC	Fluorescein isothiocyanate
GFP	Green fluorescent protein
HAT	Histone acetyltransferase
HJ	Hybrid joint
H3K4me3	Histone H3 lysine 4 trimethylation
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HMGB1/2	High mobility group box protein B1/2
HR	Homologous recombination
HSC	Haematopoietic stem cell
IDD	Intrinsically disordered domain
Ig	Immunoglobulin
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ITAM	Immunoreceptor tyrosine-based activation motif
kDa	Kilodalton
KOAc	Potassium Acetate
KRAB	Krüppel associated box
MBP	Maltose binding protein
Mbr	Major breakpoint
MHC	Major histocompatibility complex
MMR	Mismatch repair
NaOAc	Sodium acetate
NBD	Nonamer binding domain

NHEJ	Non-homologous end joining machinery
NH_4OAc	Ammonium acetate
NK	Natural killer
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCC	Post-cleavage complex
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PEI	Polyethyleneimine
PHD	Plant homeodomain
PID	Primary immunodeficiency
PMSF	Phenylmethylsulfonyl fluoride
PreR	Pre-RNase H
qPCR	Quantitative PCR
RAG	Recombination activating gene
RING	Really interesting new gene
RNA	Ribonucleic acid
RNH	RNase H
RSS	Recombination signal sequence
SCID	Severe combined immunodeficiency
SDS	Sodium dodecyl sulphate
SE	Signal end
SEC	Signal end complex
sgRNA	single guide RNA
SHM	Somatic hypermutation
SJ	Signal joint
TBS	Tris buffered saline
TCR	T cell receptor
Tdt	Terminal deoxynucleotidyl transferase
TEV	Tobacco etch virus

TIR	Terminal inverted repeat
TLS	Translation start site
TSS	Transcription start site
UNG	Uracil DNA glycosylase
YY1	Ying Yang1

Chapter 1 - Introduction

A) The adaptive immune system

1.1 Innate and adaptive immunity

Multicellular organisms are continually exposed to a plethora of pathogens but due to our immune system we rarely become ill. There are two main arms of the immune system which work in tandem to defend against these infectious agents and clear them should infection arise; these are known as the innate and adaptive immune systems.

The innate immune system is a rapid first line of defence against infections and some form of this immunity can be found in all plants and animals. The innate immune system is made up of cells of the myeloid lineage such as macrophages which engulf and degrade invading pathogens in a process known as phagocytosis. Macrophages also help to coordinate the immune response through secretion of signalling molecules in order to recruit other immune cells to the site of infection. The innate immune response is the initial response to an unknown pathogen which happens within the first five to seven days of an infection (Murphy and Weaver, 2017). However, this offers little protection against pathogens which are capable of rapidly evolving in order to evade the non-specific nature of the innate immune system. To this end, vertebrates have evolved an adaptive immune system which is capable of mounting a specific defence against rapidly changing pathogens. Some phagocytic cells of the innate immune system, namely but not exclusively dendritic cells, are known as antigen presenting cells (APCs). APCs, through display of antigens derived from ingested pathogens on their cell surface, act as a bridge between the innate and adaptive immune systems.

The adaptive immune system is made up of the antigen-specific lymphocytes known as B lymphocytes and T lymphocytes (referred hereafter as B and T cells) which express highly specific antigen receptors on their cell surface that recognise a single antigen. Both B and T cells must recognise their cognate antigen in order to become activated.

Upon activation, B cells help to neutralise infection through the secretion of antibodies whereas T cells either directly kill infected cells or assist other cells of the immune system. Disfunction of the adaptive immune system can be devastating as infants born without a correctly functioning adaptive immune system have a low chance of survival unless they are protected from invading pathogens (Alberts *et al.*, 2002).

1.2 B lymphocytes and humoral immunity

B cells are responsible for the humoral immune response during which their surface antigen receptor, known as the B-cell receptor (BCR), binds to an antigen and leads to their proliferation and differentiation into plasma cells. These cells are responsible for the secretion of antibodies, which are essentially a secreted form of the BCR and therefore recognise the same antigen. Both the BCR and the secreted antibodies are a class of molecules known as immunoglobulins (Ig).

Antibodies can combat pathogens in a number of different ways. The most crude method is simply binding of the antibody to viruses and toxins, leading to their neutralisation and thus stopping interaction with their target. Another mechanism is required to successfully clear pathogens such as bacteria where binding alone is not sufficient to stop them replicating. In this regard, an antibody bound to the surface of a bacterium can facilitate its ingestion by a phagocytic cell, which is especially important for bacteria that have evolved an outer coat which evades direct detection by a phagocyte. Finally, in some cases the bound antibodies can activate the complement pathway which leads to the formation of a protein complex on the surface of the bacterium. Complement coating can kill the bacterium directly by disrupting the cell surface membrane and causing cell lysis or it can enhance ingestion by a phagocyte (Murphy and Weaver, 2017).

1.3 Immunoglobulin structure

Immunoglobulins are Y-shaped proteins which consist of two identical heavy chains (IgH) and two identical light chains (IgL), linked by disulphide bridges. The

immunoglobulin can contain either kappa (κ) or lambda (λ) light chains, but in a given immunoglobulin these are always of the same isotype. Together, the heavy and light chains form the variable regions, for antigen binding, and the constant region, which specifies the effector function of the immunoglobulin (Figure 1.1). Heavy chains contain three constant domains whereas the light chain contains just one.

The BCR and the antibody have an almost identical structure except for a small C-terminal portion of the heavy chain. In the BCR this region is hydrophobic which anchors the molecule into the cell membrane, whereas in the antibody this region is hydrophilic to facilitate secretion (Murphy and Weaver, 2017).

Each immunoglobulin contains two variable regions which are made up of three hyper variable loops known as the complementary determining regions (CDRs). The three CDRs of the heavy chain pair with those of the light chain to form the highly specific antigen-binding site, capable of distinguishing between proteins that differ by just a single amino acid. The sequence of the CDRs is highly diverse between developing B cells and thus enables binding to a vast number of antigens.

Once folded, the immunoglobulin consists of three equally sized globular domains: two identical Fab fragments (Fragment antigen binding), containing the antigen binding site, and a single Fc fragment (Fragment crystallisation), consisting of two constant regions of the heavy chain. These fragments were originally observed via limited proteolysis of an antibody with the protease papain, cleaving the molecule at the N-terminus of the disulphide bond to yield the two Fab fragments and the single Fc fragment (Figure 1.1) (Murphy and Weaver, 2017).

Similar to the light chain, the heavy chain also contains several different isotypes: μ, δ, γ, α and ε. However, unlike the different light chain isotypes, the different heavy chain isotypes each have specific functions. This is determined by the C-terminal region of the heavy chain and leads to the five different classes of immunoglobulins: IgM, IgD,

IgG, IgA and IgE. IgG is the most abundant immunoglobulin and can be divided into a number of distinct subclasses, in humans these are IgG_1, IgG_2, IgG_3 and IgG_4. Naive B cells only express membrane bound immunoglobulins of the IgM and IgD classes, but upon activation during an immune response the heavy chain constant region can be swapped with the other isotypes in a process known as class switch recombination (Section 1.22). The class to which the antibody switches to is directed by different cytokines released by CD4+ T cells in response to a specific type of infection. For example, as part of the immune response at mucosal surfaces, class switching to IgA allows transport of the immunoglobulin into mucosal secretions by association with the polymeric Ig receptor (pIgR), expressed by cells of mucosal epithelium (Woof and Russell, 2011). Therefore, class switching of antibodies tailors the immune response to most efficiently combat a specific pathogen.

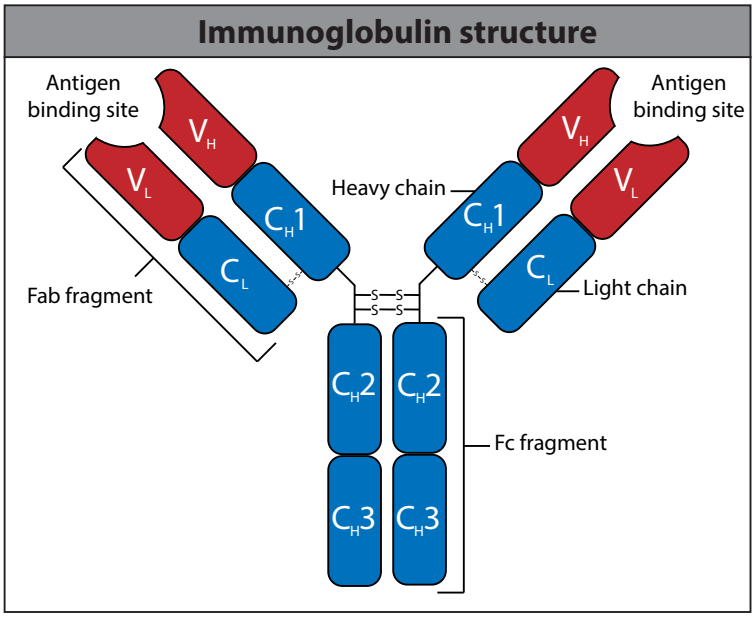


Figure 1.1 - Structure of an immunoglobulin molecule

An immunoglobulin contains two identical heavy chains and two identical light chains, joined by disulphide bonds (-S-S-). The Fab (fragment antigen binding) fragment contains the antigen binding site which is formed by both heavy and light chain variable regions (V_L and V_H). The single light chain constant region (C_L) and the first of the three heavy chain constant regions (C_H1) are also found in the Fab fragment. The remaining two heavy chain constant regions (C_H2 and C_H3) are located in the Fc (fragment crystallisation) fragment.

1.4 T lymphocytes and cellular immunity

T cells are responsible for the cellular immune response and, unlike B cells, they do not directly recognise the native antigen. Instead T cells recognise antigens that have been processed and displayed on the cells surface of APCs or infected cells via a cell surface glycoprotein known as the major histocompatibility complex (MHC). T cells are therefore important in recognising cells which have already been infected by a pathogen, such as a virus. There are two different classes of MHC molecules; MHC class I which is expressed by most cells of the body and displays antigens derived from the cytosol, and MHC class II which is expressed mainly by APCs and displays antigens derived from proteins degraded in intracellular endosomal vesicles.

T cells have a number of different classes which have varying responses to antigen exposure and this class is determined by the expression of the cell surface markers CD4 or CD8. CD8+ cells are cytotoxic T cells which directly kill infected cells via recognition of antigens displayed on MHC class I. On the other hand, CD4+ cells are known as T helper cells which do not directly kill cells but instead direct the activity of other cells of the immune system through secretion of cytokines after recognition of antigens displayed by MHC class II (Murphy and Weaver, 2017).

T helper cells can be divided into several subtypes, with the main types being T_H1, T_H2, T_H17 and T_{FH}. T_H1 cells, through the release of interferon-γ (IFN-γ), help to further activate bacterially infected macrophages which enhances their microbicidal activity to eliminate the intracellular bacteria. Infections by parasites, such as helminths, are controlled by T_H2 cells which secrete cytokines to aid class switching in B cells in order to produce IgE antibodies whose primary role is to fight parasitic infections. T_H17 cells secrete the cytokine IL-17 which, in response to infection by bacteria and fungi, induces chemokine production in local fibroblasts and epithelial cells and leads to the recruitment of neutrophils to the site of infection (Murphy and Weaver, 2017). The final subset of T helper cells are T_{FH} cells which provide help to B cells by assisting isotype switching and affinity maturation (Crotty, 2014).

Regulatory T cells (T_{reg}) are another subset of CD4+ cells whose function, unlike T helper cells, is to repress the immune response via limiting the activation of other CD4+ cells (Yamane and Paul, 2013). A primary function of this subset is maintaining peripheral tolerance and therefore preventing autoimmune diseases (Vignali *et al.*, 2008).

Once an infection has been cleared by the immune system, the majority of activated B and T cells die apart from a small number which remain as memory B and T cells. These memory cells can rapidly differentiate into effector lymphocytes upon secondary exposure to the same antigen during re-infection, reducing the response time of the immune system and giving long-lasting immunity. This forms the basis of vaccination (Murphy and Weaver, 2017).

1.5 T cell receptor structure

A T cell receptor (TCR) consists of two polypeptide chains, linked by a disulphide bond and resemble the Fab fragment of an antibody (Figure 1.2). Two differing forms of TCR exist; the predominant α:β TCR made up of α and β chains, and the less common γ:δ TCR, made up of γ and δ chains. Each polypeptide chain is made up of a variable region, both containing three hypervariable loops that together form the antigen binding site, and a constant region, followed by a transmembrane domain and short cytoplasmic tail. TCRs differ from immunoglobulins in a number of ways: they are always membrane bound, only have one antigen binding site and can only recognise small peptide antigens presented by MHC proteins rather than any form of organic antigen. Furthermore, TCR:antigen interactions usually have much lower affinity and specificity than immunoglobulin:antigen interactions.

The TCR, like the BCR, has no intrinsic signalling capacity and therefore requires association with a complex of additional accessory transmembrane proteins: the CD3 complex consisting of three polypeptide chains (CD3γ, CD3δ and CD3ε), and a

homodimer of ζ chains. Immunoreceptor tyrosine-based activation motifs (ITAMs) found within this complex become phosphorylated upon TCR ligand binding, leading to initiation of intracellular signalling by SH2 domain containing proteins and ultimately T cell activation (Brownlie and Zamoyska, 2013). A similar mechanism of signal transduction mediated by ITAM containing transmembrane proteins also occurs in B lymphocytes when the membrane bound BCR binds its cognate antigen.

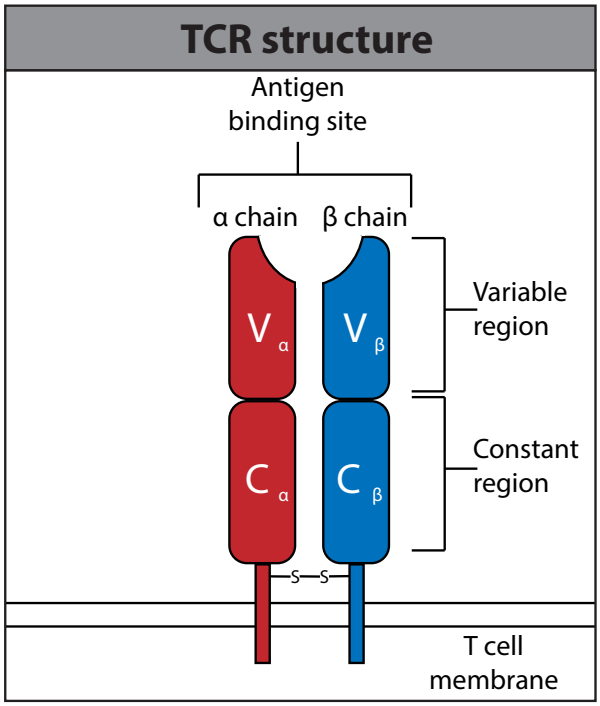


Figure 1.2 - Structure of a T cell receptor

A T cell receptor (TCR) contains two polypeptide chains (α and β or γ and δ), joined by disulphide bonds (-S-S-). Each chain contains a single variable region (V_α and V_β), which both together form the antigen binding site, and a single constant region (C_α and C_β). A short C-terminal cytoplasmic tail anchors the TCR to the cell membrane. The diagram depicts a TCR consisting of α and β chains.

B) V(D)J recombination

1.6 Generating a diverse antigen receptor repertoire

An effective adaptive immune system must be capable of producing a vast repertoire of unique antigen receptors in order to combat the potentially limitless number of

pathogens to which an individual can be exposed. Indeed, the number of immunoglobulin and TCR specificities found in the human body can exceed 10^{11} (Murphy and Weaver, 2017). However, as the entire human genome only contains around 20,000 protein coding genes (Pertea *et al.*, 2018), a paradox is how the vast antigen receptor repertoire can be generated. This diversity is achieved, in part, by the mixing and matching of antigen receptor gene segments within the developing B and T lymphocytes. This process is known as V(D)J recombination.

The antigen receptor loci consist of many copies of variable (V), diversity (D) and joining (J) gene segments (Table 1.1). The immunoglobulin light chain (Igκ and Igλ) and the TCR α and γ loci contain just V and J gene segments whereas the immunoglobulin heavy chain (IgH) and TCR β and δ loci contain all three types of gene segment types (Gellert, 2002). One of each of the V, D (if present) and J gene segments are stochastically joined to form the variable exon of the antigen receptor (Figure 1.3). Given that humans have approximately 320 different light chain combinations (200 Igκ and 120 Igλ) and approximately 11,000 different heavy chain combinations, together these provide approximately 3.5 million different potential antibody specificities (Janeway, 2001). However, due to differences in gene segment usage frequency and the incompatibility of some heavy and light chain combinations, in reality, this number is likely to be lower. Nevertheless, combinatorial diversity contributes significantly to the overall diversity of the antigen receptor repertoire. The remaining diversity is provided by imprecise gene segment joining (Section 1.7) and somatic hypermutation (Section 1.21).

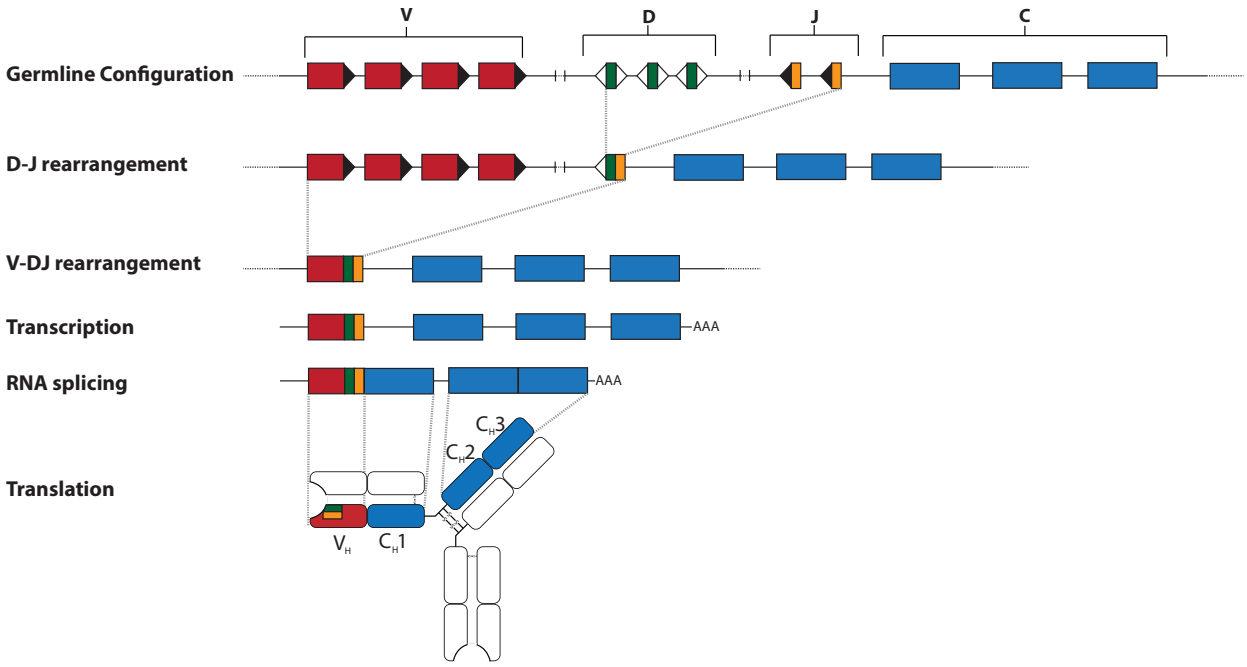


Figure 1.3 - Rearrangement of gene segments at a heavy chain locus

Schematic depicting gene segment rearrangements occurring at a simplified heavy chain locus. Variable (V) and joining (J) gene segments are flanked by 23-RSSs (black triangles), whereas diversity (D) gene segments are flanked by 12-RSSs (white triangles). D-J recombination occurs first, followed by V-DJ recombination to form the variable exon. The resulting transcript is then spliced in order to join the variable exon to the three constant (C) exons. Rearrangements at the light chain loci occur in a similar manner but lack D gene segments. For kappa light chains, splicing occurs to a single constant region whereas for lambda light chains splicing occurs to one of four different constant regions. Following translation, the heavy chain pairs with a light chain and, along with another pair of identical heavy and light chains, forms the antibody molecule.

	Immunoglobulin		T cell receptor	
	Igκ + Igλ	IgH	TCR α	TCR β
Variable (V)	~70	~40	~70	52
Diversity (D)	0	23	0	2
Joining (J)	5 (κ) 4 (λ)	6	61	13

Table 1.1 - The numbers of human immunoglobulin and T cell receptor gene segments

Displayed are the numbers of human immunoglobulin light chain (Igκ + Igλ), heavy chain (IgH) and α: β T cell receptor gene segments (Murphy and Weaver, 2007).

Gene segments to be recombined are flanked by recombination signal sequences (RSSs) which contain highly conserved 7 base pair (bp) heptamer (CACAGTG) and 9 bp nonamer (ACAAAAACC) sequences, separated by either a 12±1 bp or 23±1 bp spacer sequence, giving a 12- or 23-RSS respectively (Figure 1.4). The spacer sequences, other than their length, are much more variable than the sequence of the heptamer and nonamer, although evidence suggests there may be some sequence conservation at particular positions (Ramsden *et al.*, 1994). Efficient recombination only occurs between two gene segments whose RSSs contain spacers of different lengths; known as the 12/23 rule (Tonegawa, 1983). Crucially, each gene segment type is flanked by either a 12- or 23-RSS to ensure only functional rearrangements occur. For example, at a light chain locus, all V segments are associated with an RSS of one spacer length and J segments the other, therefore enforcing functional recombination between V and J segments.

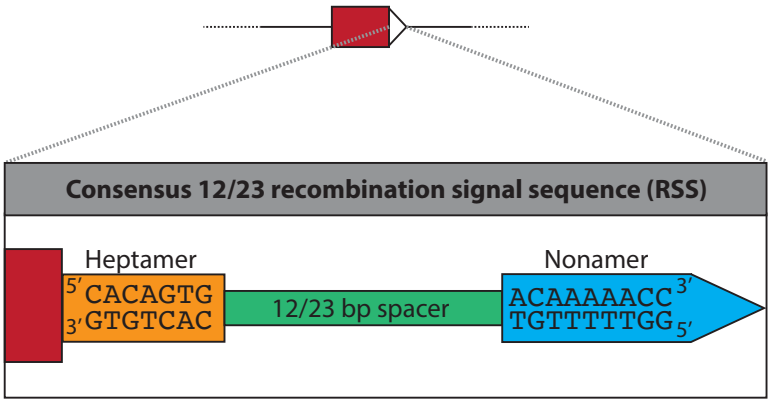


Figure 1.4 - Recombination signal sequence

Each RSS contains a conserved 7 bp heptamer (5' CACAGTG 3') and 9 bp nonamer (5' ACAAAAACC 3') sequence, separated by a less conserved 12 or 23 bp spacer sequence. Efficient recombination only occurs between two gene segments which are flanked by RSSs of differing spacer lengths, in accordance with the 12/23 rule (Tonegawa, 1983).

1.7 Overview of V(D)J recombination

V(D)J recombination is initiated by the lymphoid-specific RAG (recombination activating gene) endonuclease complex which is comprised of the proteins RAG1 and RAG2

(Oettinger *et al.*,1990). The reaction can be divided into three clear stages: binding and synapsis of a 12- and 23-RSS by the RAG complex, cleavage of the DNA at the heptamer-gene segment boundary and finally joining of the two gene segments (Figure 1.5).

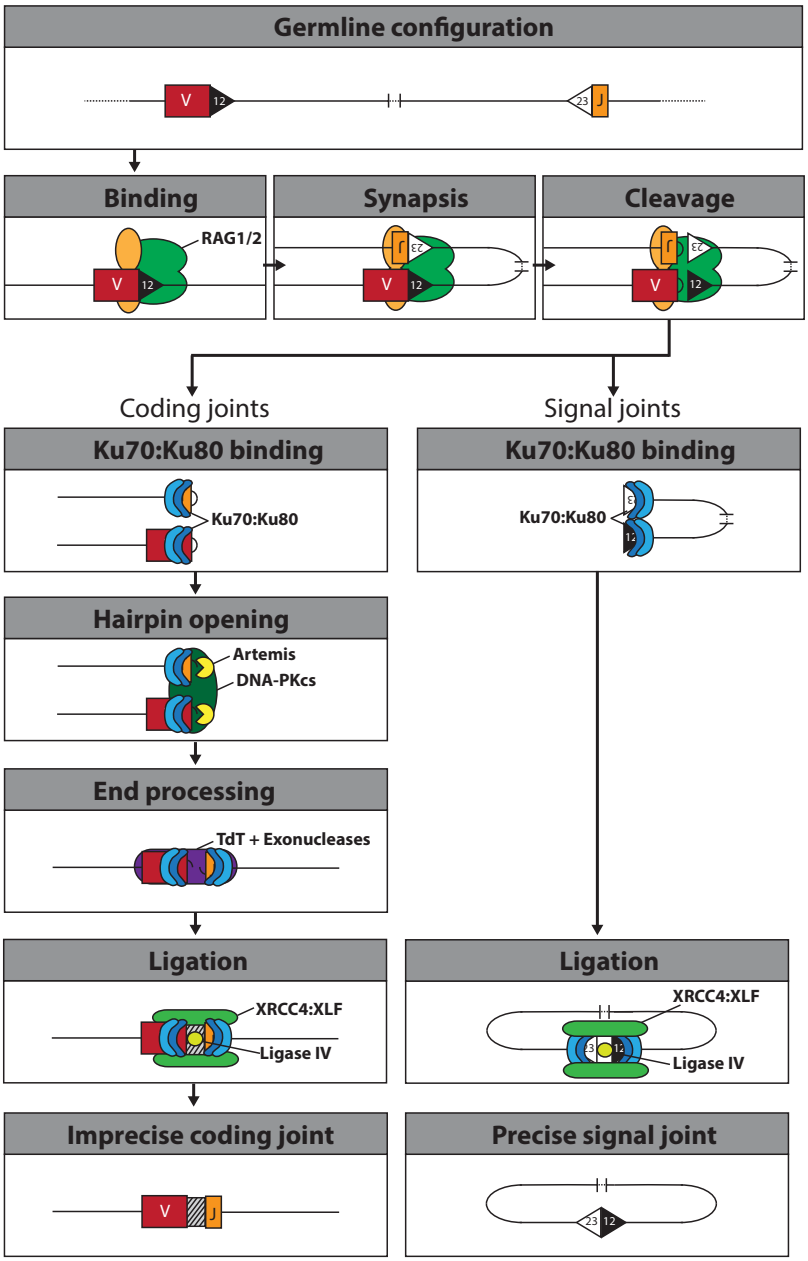


Figure 1.5 - Mechanism of V(D)J recombination

Diagram represents a simplified overview of V(D)J recombination. RAG (green and yellow ovals) binds an RSS (black or white triangles) in the genome and synapses with a second RSS of differing spacer length. RAGs then nick the DNA 5' of the heptamer and catalyse coupled cleavage via a direct transesterification reaction, resulting in a pair of hairpin sealed coding and and 5' phosphorylated blunt signal ends. Coding and signal ends are repaired in slightly

different ways. For the signal ends, the Ku70:Ku80 complex binds to the broken ends and recruits ligaseIV and XRCC4:XLF. The ends are then ligated without any further processing to leave a precise signal joint. The coding ends are also bound by Ku70:Ku80 which then recruits DNA-PKcs and Artemis. The endonuclease activity of Artemis opens the hairpins and the ends are then processed by TdT and exonucleases. Ligation occurs as with signal ends and results in an imprecise coding joint.

To initiate V(D)J recombination the RAG complex binds to the RSSs that flank the V, D and J gene segments and brings them into close proximity via the formation of a synaptic complex. Next, a nick is introduced at the 5' boundary of the heptamer of both RSSs, leading to the formation of a free 3' hydroxyl group (3'OH) on one strand. RAGs then catalyse the coupled cleavage of the RSSs through a direct transesterification reaction where the free 3'OH attacks the opposite DNA strand (Schatz and Swanson, 2011). The result is hairpin sealed coding ends abutting the gene segments and 5' phosphorylated blunt signal ends at the RSSs. The latter remain associated with RAGs in a post-cleavage complex (PCC) (Agrawal and Schatz, 1997; Hiom and Gellert, 1998; Jones and Gellert, 2001). RAG-mediated cleavage is discussed in more detail in Section 1.10. Subsequent opening, processing and joining of the hairpin coding ends by the non-homologous end joining (NHEJ) machinery generates the variable exon and the blunt signal ends are also joined to form the signal joint (Helmink and Sleckman, 2012).

Due to their differing DNA end configurations the coding and signal ends are resolved in slightly different ways. The coding ends are firstly bound by the Ku70:Ku80 protein complex which subsequently recruits the DNA-dependent protein kinase catalytic subunit (DNA-PKcs):Artemis complex (Swanson, 2004). The endonuclease activity of the DNA-PKcs:Artemis complex nicks the hairpin at a random point, which results in opening of the hairpin. Should the point of opening occur away from the apex of the hairpin, a 3' single stranded overhang is produced which is then filled in to produce palindromic (P) nucleotides (Chang and Lieber, 2016). Terminal deoxynucleotidyl transferase (TdT) can also be recruited to process the DNA ends through the addition

of non-templated (N) nucleotides (Repasky *et al.*, 2004) (Figure 1.6). Next, the two ends are paired, non-complementary nucleotides removed by an exonuclease and gaps filled in by DNA polymerases (DNA Pol μ or DNA Pol λ). Finally, the two ends are ligated by the DNA ligase IV:XRCC4 complex, forming the imprecise coding joint (Lieber, 2010). The addition of P and N nucleotides adds to the junctional diversity of the resulting variable exon and therefore the overall diversity of the antigen receptor repertoire. However, this comes at a cost as two out of three events leads to an out-of-frame join.

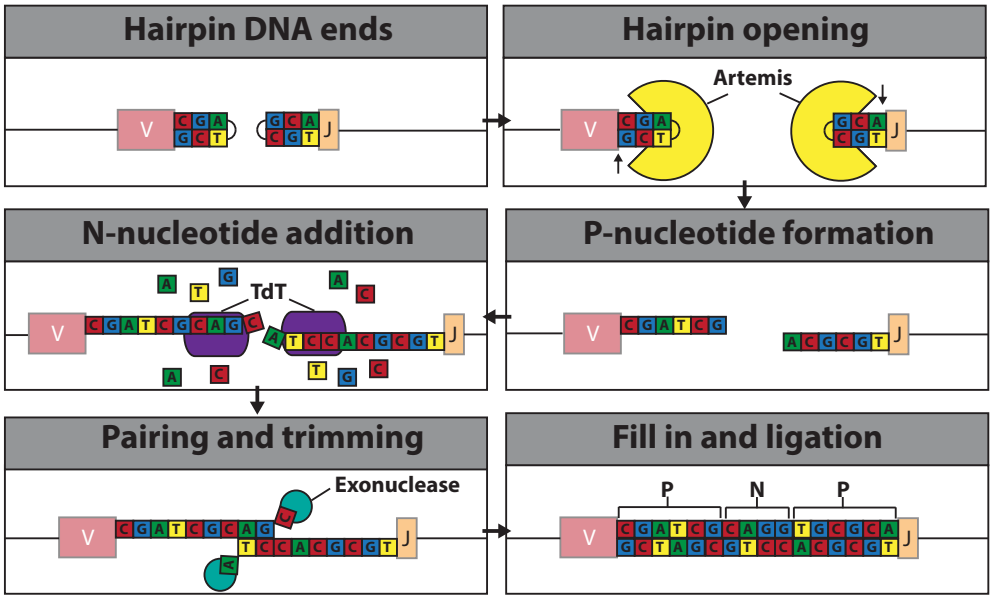


Figure 1.6 - Generation of junctional diversity

Further diversity is introduced at the junction between two gene segments by the palindromic (P) and non-templated (N) nucleotide addition. Should Artemis open the hairpin at a site away from the apex of the hairpin then a single stranded overhang is formed, this gives rise to P nucleotides. TdT can then randomly add additional N nucleotides. The ends are then paired and any non-Watson-Crick base paired are removed. Finally the ends are filled in to form the processed coding joint.

The signal ends are joined in a similar way but, due to their blunt nature, they require much less processing. The DNA ends are firstly bound by the Ku70:Ku80 protein complex which then facilitates their joining by the DNA ligase IV:XRCC4 complex, producing a precise signal joint. Depending upon the orientation of the RSSs, the signal joint is either retained in the genome or removed as an excised signal circle

(ESC). Convergent orientation of RSSs leads to deletional recombination and the formation of an ESC, whereas co-oriented RSSs leads to inversional recombination and the retention of the signal joint in the genome (Gauss and Lieber, 1992) (Figure 1.7).

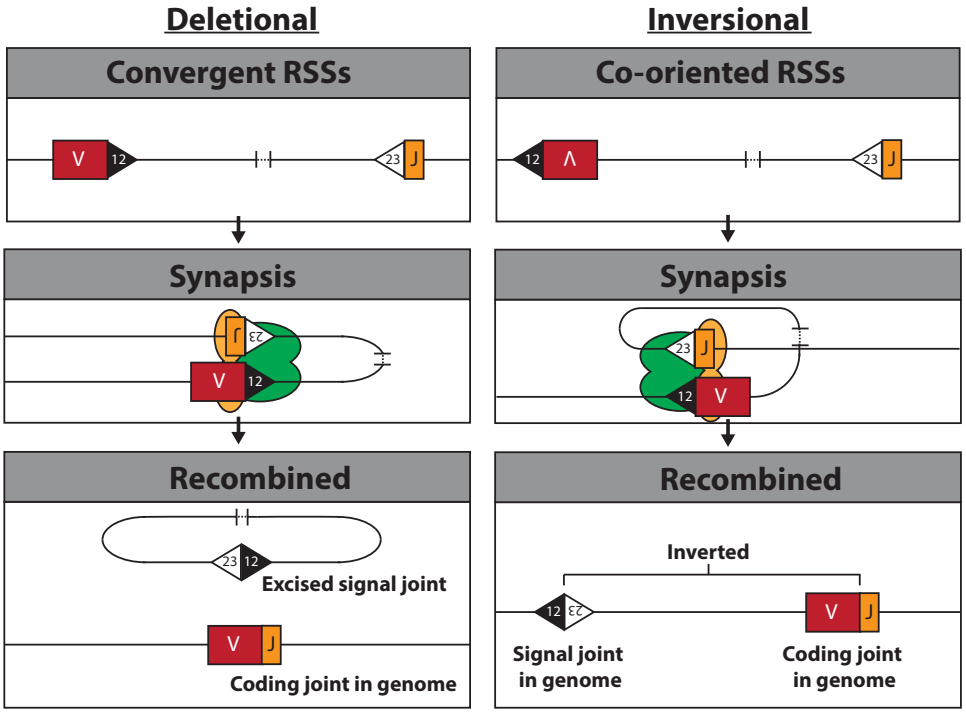


Figure 1.7 - Deletional and inversional V(D)J recombination

V(D)J recombination can occur via a deletional or inversional process which lead to a different outcome for the signal joint. Which method of recombination occurs depends upon the orientation of the RSSs. If RSSs are positioned in a convergent orientation, then deletional recombination occurs and the signal joint is removed from the genome as an excised signal circle (left panel). Should the RSSs be co-oriented then inversional recombination occurs and the signal joint is retained in the genome (right panel).

1.8 The RAG proteins

i) Discovery of the lymphoid specific proteins RAG1 and RAG2

RAG1 was initially identified to initiate low efficiency V(D)J recombination by transfection of *Rag1* cDNA into fibroblasts (Schatz *et al.*, 1989). Shortly after the co-factor RAG2 was identified which caused a thousand fold increase in recombination activity when it was co-transfected with RAG1 (Oettinger *et al.*, 1990). Both proteins were then shown to be vital for the generation of an effective immune system as mice

deficient in either RAG1 or RAG2 have no mature B or T lymphocytes, due to their inability to perform V(D)J recombination. Subsequent *in vitro* studies using purified RAG1 and RAG2 proteins identified that only these two proteins are required for the cleavage reaction during V(D)J recombination (McBlane *et al.*, 1995). The *in vitro* cleavage activity of RAGs was also shown to be greatly stimulated by the addition of the high mobility group box (HMGB) proteins 1 or 2 (van Gent *et al.*, 1997), but its involvement in the reaction *in vivo* was contested until very recently (Aidinis *et al.*, 1999; Calogero *et al.*, 1999; Thwaites *et al.*, 2019) (Section 1.9). The role of HMGB1/2 during V(D)J recombination *in vivo* is discussed further in Chapter 3.

ii) Origin and evolution of RAG1 and RAG2

Since the discovery of RAG1, similarities between V(D)J recombination and cut-and-paste transposition have been drawn (Sakano *et al.*, 1979). The RAG1 and RAG2 proteins are highly conserved across all jawed vertebrates, as is their genomic organisation with both genes located adjacent to each other, just 8 kb apart and are divergently transcribed from single exons (Schatz *et al.*, 1992). The fact that the RAG complex can catalyse a transposition reaction *in vitro* (Agrawal *et al.*, 1998), DNA cleavage by the RAG complex resembles that of transposases, and that the RSSs also resemble terminal inverted repeats (TIR) that are the target sites of transposases (van Gent *et al.*, 1996), gives merit to the hypothesis that the taming of an ancient transposon was the origin of V(D)J recombination (Jones and Gellert, 2004; Carmona and Schatz, 2017).

Extensive work has been done to identify a primordial *Rag1-Rag2* transposon (Thompson, 1995; Fugmann, 2010). A number of candidate transposable elements (TEs) have been proposed but the most favoured is those of the *Transib* family, many of which share sequence similarity with RAG1, including the catalytic residues (Kapitonov and Jurka, 2005). Additionally, TIRs of the *Transib* family contain a heptamer sequence with a 5’ CAC, a region of the RSS which is critical for RAG cleavage (Carmona and Schatz, 2017), and some members, such as Hztransib, even

share similar cleavage reaction intermediates as V(D)J recombination and can catalyse recombination of extra-chromosomal templates when co-transfected with vertebrate RAG2 (Hencken *et al.*, 2012). However, TEs from this family can only account for a *RAG1*-like gene, suggesting an independent origin for *RAG2* (Kapitonov and Jurka, 2005). A potential precursor to RAG2, SpRAG2L, was identified in the purple sea urchin along with a RAG1-like protein SpRAG1L (Fugmann *et al.*, 2006). Like Hztransib, SpRAG1L is able to catalyse V(D)J recombination when paired with vertebrate RAG2. However, although SpRAG2L shares a similar protein fold and C-terminal elements as vertebrate RAG2, it is unable to stimulate V(D)J recombination by vertebrate RAG1, Hztransib or SpRAG1L (Fugmann *et al.*, 2006; Wilson *et al.*, 2008). Due to this, the origin of RAG2 still remains unclear.

The lack of TIRs flanking the *SpRAG1L-SpRAG2L* locus also provided no support to the identification of an assumed *Rag1-Rag2* transposon. However, the existence of an ancient RAG transposon was finally confirmed with the discovery of *ProtoRAG*. *ProtoRAG* is a cut-and-paste transposon identified in lancelets that contains *RAG1-like* and *RAG2-like* genes flanked by TIRs resembling a RSS heptamer (Huang *et al.*, 2016). The current dogma is that acquisition of a *Transib* family TE gave rise to RAG1, followed by further acquisition of RAG2 from an unknown source, giving rise to the ancestral *Rag1-Rag2* transposon flanked by TIRs which resemble an RSS (Carmona and Schatz, 2017).

iii) RAG1

RAG1 is the larger of the two recombinase proteins and provides the catalytic activity of the recombinase complex (Kim *et al.*, 1999). Most of the biochemical studies have been performed using murine RAG1 and RAG2 proteins, which have >90% sequence similarity to the human protein. Extensive mutagenesis of RAG1 and the recent solution of its high resolution structure, both in its apo form and in complex with an RSS, has revealed a wealth of knowledge about its major domains and their function (Notarangelo *et* al., 2016; Rodgers, 2017). RAG1 is comprised of 1040 amino acids

which can be truncated to core RAG1 (cRAG1, aa 384-1008) which is sufficient for V(D)J recombination and substantially easier to purify and therefore has been used is most RAG studies (Schatz and Swanson, 2011) (Figure 1.8). However, cRAG1 knock-in mice have slightly impaired V(D)J recombination and decreased mature lymphocytes, highlighting the importance of the highly conserved non-core domain (Dudley *et al.*, 2003).

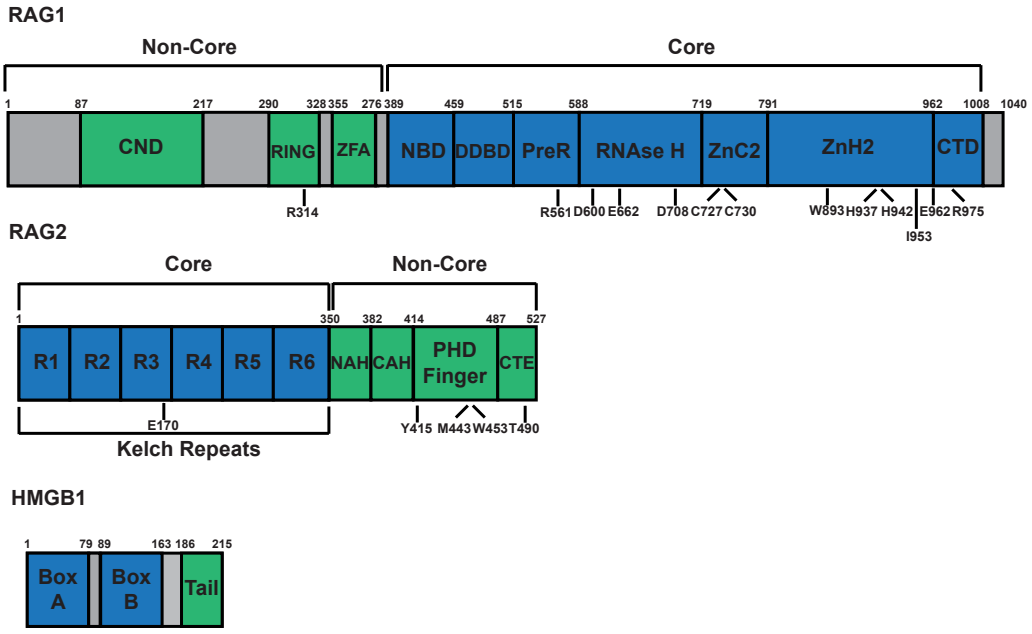


Figure 1.8 - Linear diagram of RAG1, RAG2 and HMGB1 proteins

A schematic showing the main domains of the proteins RAG1, RAG2 and HMGB1. Amino acid boundary numbers are listed above each domain and any important residues are denoted below the domain within which they fall. Top: RAG1 non-core domains are shown in green whereas core domains are shown in blue, non-defined regions are grey. CND, central non-core domain; RING, really interesting new gene, ZFA, zinc finger A; NBD, nonamer binding domain; DDBD, dimerisation and DNA binding domain; PreR, Pre-RNase H; CTD, C-terminal domain. Middle: RAG2 core domains are shown in blue and non-core domains in green. R1-R6 represent the six kelch motif repeats which make up the beta-propeller structure of core RAG2. NAH, N-terminal acidic hinge; CAH, C-terminal acidic hinge; CTE, C-terminal extension. Bottom: HMGB1 Box A and Box B are shown in blue and the acidic C-terminal tail is show in green. The grey boxes represent linkers between each domain.

The majority of non-core RAG1 is located in the N-terminus of the protein (aa 1-383) and although poorly understood, the domain architecture of this region has been well defined (Arbuckle *et al.*, 2011). Early work on non-core RAG1 revealed a region of basic amino acids in the N-terminus that when mutated leads to a substantial decrease in recombination activity (McMahan *et al.*, 1997). The exact function of this basic region is still unclear but it has been reported to be important for zinc binding and association with nuclear import proteins SRP1 and KPNA1 (Rodgers *et al.*, 1996; Cortes *et al.*, 1994; Simkus *et al.*, 2009a). Despite interaction with these proteins, this non-core region of RAG1 is not necessary for nuclear import (Simkus *et al.*, 2009a). Instead, a nuclear localisation sequence found in the core of RAG1 fulfils this role through binding KPNA2 (Spanopoulou *et al.*, 1995). Although not fully mapped, there is evidence for a region of non-core RAG1 that is involved in association with the repair proteins Ku70/ Ku80 (Raval *et al.*, 2008). The interaction with these repair proteins is not absolutely required to perform the joining step of V(D)J recombination but instead may help co-ordinate repair and could explain why removal of the RAG1 non-core domain leads to increased accumulation of cleaved signal ends (Steen *et al.*, 1999).

The more C-terminal part of non-core RAG1 (aa 265-383) is comprised of a C_3HC_4 RING (really interesting new gene) finger domain with E3 ubiquitin ligase activity (Yurchenko *et al.*, 2003). The RING finger domain highlights the important regulatory role of the RAG1 N-terminus due to strong correlation between RAG1 ubiquitin ligase activity and overall V(D)J recombination activity (Simkus *et* al., 2007; Simkus *et al.*, 2009b). Mutation of a key cysteine residue (C325) leads to loss of RAG1 mediated ubiquitin ligase activity and blocks the autoubiquitylation of lysine 233 in RAG1 and also ubiquitylation of histone H3 (Jones and Gellert, 2003; Grazini *et al.*, 2010; Deng *et al.*, 2015). Interestingly, V(D)J recombination activity in $RAG1^{C325Y}$ mice is severely impaired yet recombination activity on episomal substrates in transiently transfected fibroblasts is largely unaffected (Deng *et al.*, 2015). Given that episomal substrates consist of a mix of chromatinised and non-chromatinised DNA (Vaughan *et al.*, 2006), the lack of ubiquitin ligase activity caused by the C325Y mutation may impair V(D)J

recombination on chromosomal substrates. Indeed, the N-terminal 218 amino acids of RAG1 preferentially interact with non-ubiquitylated histone H3, immobilising RAG1 and blocking cleavage (Grazini *et al.*, 2010; Deng *et al.*, 2015). Mono-ubiquitinylation of histone H3 by the RAG1 RING finger releases RAG1 from the nucleosome and facilitates cleavage (Deng *et al.*, 2015).

In contrast to non-core RAG1, the domains, function and structure of core RAG1 is fairly well understood. Core RAG1 can be divided into seven main structural modules (Kim *et al.*, 2015) (Figure 1.8). The nonamer binding domain (NBD) and the dimerisation and DNA binding domain (DDBD) are two of these structural modules that together form the main RSS binding and RAG1 dimerisation sites of the RAG complex (Rodgers, 2017). The NBD binds and synapses both nonamer sequences of the 12/23-RSSs, forming a tightly interwoven dimer (Yin *et al.*, 2009). A pair of Gly-Gly-Arg-Pro-Arg (GGRPR) motifs, an AT-hook found in many DNA binding proteins (Aravind and Landsman, 1998), in the NBD dimer are responsible for its targeting and binding to the well conserved AT-rich nonamer sequence (Yin *et al.*, 2009). The DDBD is attached to the NBD via a flexible linker and is positioned towards the heptamer end of the RSS, which it contacts (Ru *et al.*, 2015). The remaining five modules (PreR; Pre-RNase H, RNH; RNase H, ZnC2, ZnH2 and CTD; carboxy-terminal domain) make up the active catalytic site for DNA cleavage, which contains the conserved DDE motif (D600, D708 and E962) (Kim *et al.*, 2015; Landree *et al.*, 1999). The PreR acts as a linker between the DDBD and the RNH, where two of the catalytic amino acids of the DDE motif reside (D600 and D708) (Kim *et al.*, 2015; Rodgers, 2017). The final DDE residue, E962, is located after two important zinc-binding modules: ZnC2 and ZnH2. These regions contain a pair of cysteine (C727 and C730) and histidine (H937 and H942) residues which together form a single zinc-binding site, located close to the DDE active site, and are essential for DNA cleavage activity and dimerisation of RAG1 (Kim *et al.*, 2015; Gwyn *et al.*, 2009; Rodgers *et al.*, 1996). The seventh and final module, the CTD, folds back on itself and associates with the DDBD, providing additional residues to the active

site and assisting RAG1 dimerisation (Kim *et al.*, 2015; Rodgers, 2017; Huye *et al.*, 2002).

iv) RAG2

RAG2 provides an essential regulatory role to the overall RAG recombinase by directing and modulating binding and cleavage activities of RAG1 (De *et al.*, 2004; Lu *et al.*, 2015). The full length protein consists of 527 amino acids and, like RAG1, can be truncated to an active core form (cRAG2, aa 1-382) (Cuomo and Oettinger, 1994; Sadofsky *et al.*, 1994) (Figure 1.8). The entirety of the non-core region is located in the C-terminus of the protein (aa 383-527).

Core RAG2 is folded into a six-bladed β-propeller like structure, with each blade formed by a Kelch-like motif (Callebaut and Mornon 1998; Kim *et al.*, 2015). β-propellers are a well conserved protein structure and are involved in mediating protein-protein interactions via the flat surface provided on both sides of the 'doughnut' shaped core (Chen *et al.*, 2011). Indeed, one face of the RAG2 β-propeller has extended loops which are vital for interacting with the RAG1 PreR, RNH and ZnC2 domains (Kim *et al.*, 2015; Aidinis *et al.*, 2000). Although RAG2 has no inherent DNA binding activities in the absence of RAG1, it contacts RAG1 near the active site, assisting in its formation and directs DNA cleavage (Swanson and Desiderio, 1999; Kim *et al.*, 2015).

Like the non-core RAG1 regions, non-core RAG2 provides an important regulatory role. Full-length RAG2 is easier to purify than full-length RAG1 and consequently the RAG2 non-core regions are better understood. Structural studies of the RAG proteins included full-length RAG2 but due to the flexible nature of the RAG2 C-terminus, the RAG2 non-core region is missing from all currently published structures (Kim *et al.*, 2015; Ru *et al.*, 2015; Kim *et al.*, 2018; Ru *et al.*, 2018a).

The RAG2 C-terminus (aa 383-527, Figure 1.8) is composed of a highly conserved but intrinsically disordered acidic hinge (aa 383-413), a plant homeodomain (PHD) finger

(aa 414-487) (Callebaut and Mornon, 1998; Elkin *et al.*, 2005) and a C-terminal extension (aa 488-527).

The RAG2 PHD finger binds to tri-methylated lysine 4 on histone H3 (H3K4me3), an epigenetic mark associated with active chromatin, which is found throughout the genome and at actively recombining loci (Matthews *et al.*, 2007; Ramon-Maiques *et al.*, 2007; Liu *et al.*, 2007a; Morshead *et al.*, 2003; Schatz and Ji, 2011) (Section 1.16). Sterile transcription at the antigen receptor loci is required for RSS accessibility (Sections 1.14/17); this also deposits H3K4me3 in these regions, aiding recruitment of RAG2 via its PHD finger (Schatz and Ji, 2011). RAG2 binding to H3K4me3 not only facilitates recruitment of RAG1 but also enhances the binding and catalytic activity of the complex (Lu *et al.*, 2015; Bettridge *et al.*, 2017; Ward *et al.*, 2018). As mentioned previously, the RAG2 C-terminus is missing from published high resolution structures however, the PHD finger structure has been resolved in isolation (Elkin *et al.*, 2005; Ramon-Maiques *et al.*, 2007; Matthews *et al.*, 2007). Three aromatic residues (Y415, M443 and W453) within the PHD finger fold into an aromatic channel which forms the binding site for H3K4me3. Mutation of any of these key residues leads to a loss of H3K4me3 binding and a severe decrease in recombination activity (Matthews *et al.*, 2007). The aromatic channel differs from the conventional aromatic cage of other histone mark readers in that it is open from the top and not closed on all four sides (Matthew *et al.*, 2007; Ruthenburg *et al.*, 2007). The open design of the aromatic channel allows occupancy of a self proline residue from an N-terminal peptide chain in the absence of H3K4me3 (Ramon-Maiques *et al.*, 2007). Binding of a self peptide may form part of the auto inhibition imposed on the RAG complex by the RAG2 C-terminus.

An acidic hinge region is found in the intervening region between the RAG2 core and the PHD finger (aa 383-413). This region allosterically imposes inhibition on RAG1 which decreases its RSS binding and catalytic cleavage activity (Lu *et al.*, 2015; Bettridge *et al.*, 2017). Upon PHD finger binding to H3K4me3 a number of conformational changes are induced in both RAG1 and RAG2 which relieve the

autoinhibition and promote the catalytic activity of the RAG complex (Bettridge *et al.*, 2017). The autoinhibitory domain was mapped to a region within residues 388-405 of the acidic hinge and mutation of this region to alanine leads to an increase in RAG binding, cleavage and overall V(D)J recombination activity and also rescues activity of PHD finger mutants (Bettridge *et al.*, 2017; Lu *et al.*, 2015). A number of amino acids in this region (Y402, N403, D406 and E407) have been shown to directly interact with the core histones, which is required for efficient recombination at endogenous loci (West *et al.*, 2005). This interaction with the core histone is independent of the PHD finger and its exact function is unclear.

Historically, most studies defined core RAG2 as requiring amino acids 1-382 (Cuomo and Oettinger, 1994; Sadofsky *et al.*, 1994) but more recently it has been discovered that proteins further truncated to residue 350 retain recombination activity (Coussens *et al.*, 2013). This coincides with the end of the core β-propeller and also marks the end of the portion of RAG2 that could be resolved in published structures because of inherent flexibility after this point (Callebaut and Mornon, 1998; Kim *et al.*, 2015; Ru *et al.*, 2015). Due to the disordered and acidic nature of this region, it is now categorised as an extension to the C-terminal acidic hinge described above. However, this region has a distinct role in directing the joining phase of V(D)J recombination. A conserved patch of acidic amino acids (aa 370-383) are required to direct repair by the classical NHEJ (cNHEJ) machinery rather than the more error prone alternative NHEJ (aNHEJ) pathway, safeguarding the integrity of the genome (Coussens *et al.*, 2013; Gigi *et al.*, 2014).

The final region of non-core RAG2 is the C-terminal extension (aa 488-527), which is located after the PHD finger. The main function of the C-terminal extension is the regulation of RAG2 protein degradation throughout the cell cycle (Lin and Desiderio, 1993). RAG2 is marked for degradation by cyclin A/Cdk2 phosphorylation of a threonine at position 490 which lies within a cyclin-dependent kinase (Cdk) target site. This phosphorylation event is further dependent on a basic region of adjacent amino

acids (aa 499-508) (Li *et al.*, 1996). Phosphorylated T490 is then recognised by Skp2-SCF which ubiqutinylates RAG2, leading to its degradation via the ubiquitin-proteasome pathway (Jiang *et al.*, 2005). Mutation of T490 to alanine or deficiency of Skp2 overcomes the destruction of RAG2 at the G1-to-S phase transition and allows RAG2 expression outside of the G0/G1 phases (Li *et al.*, 1996; Jiang *et al.*, 2005). Restricting RAG2 expression and therefore V(D)J recombination to the G0/G1 phases of the cell cycle is crucial to direct correct repair of V(D)J recombination intermediates by the NHEJ pathway, which is most active at these cell cycle stages (Fukushima *et al.*, 2001; Takata *et al.*, 1998; Jiang *et al.*, 2005) (Section 1.13).

1.9 High mobility group box (HMGB) proteins

High mobility group (HMG) proteins are a group of highly abundant proteins which have a plethora of functions. These include DNA binding in the nucleus and mitochondria, signalling regulators in the cytoplasm and even acting as an extracellular inflammatory cytokine (Malarkey & Churchill, 2012). Additionally, HMGB proteins also have an essential role in V(D)J recombination. HMG proteins were first discovered over forty years ago as chromatin-associated proteins which have fast migration through polyacrylamide gels, hence the name ‘high mobility’ (Goodwin *et al.*, 1973). There are three main superfamilies of HMG proteins: HMGA (HMG-AT-hook), HMGN (HMG-nucleosome binding) and HMGB (HMG-box) proteins (Bustin, 2001), with the latter being the most abundant.

HMGB1/2 are by far the most well studied of the HMG-box proteins. HMGB1 is ubiquitously expressed in all cell types and is the most abundant non-histone protein, with approximately one HMGB1 molecule for every ten nucleosomes (Thomas and Travers, 2001). In contrast, HMGB2 is expressed during embryonic development but in adults is confined to testicular and lymphoid cells (Ronfani *et al.*, 2001). Both proteins contain two HMG-boxes (A and B) which are highly basic motifs of ~75 amino acids that fold into a characteristic ‘L’ shape made up of three α-helices (Štros, 2009) (Figure 1.8). Binding of a HMG box to the minor grove of the DNA is facilitated by hydrophobic

and electrostatic interaction between the inside face of the 'L' shaped box and the DNA. This then allows the partial intercalation of bulky hydrophobic residues between two base pairs (King and Weiss, 1993). Widening of the minor grove is facilitated by a kink introduced into the DNA via this partial intercalation, thus inducing a bend towards to major grove. Box B has two regions of intercalating residues and so induces a bend angle of approximately 90°, whereas Box A has just one region thereby inducing a lesser bend angle of 60° (Jamieson *et al.*, 1999). The C-terminal tail of HMGB1/2 is made up of a long stretch (20-30 amino acids) of negatively charged aspartic acid and glutamic acid residues. This acidic tail can modulate DNA interactions of the HMG boxes by interaction with the concave DNA-binding face of the boxes, thereby lowering the affinity with DNA (Stott *et al.*, 2014).

Although RAG1 and RAG2 are sufficient for DNA cleavage *in vitro* (McBlane *et al.*, 1995), addition of HMGB1/2 both stimulates cutting up to 100 fold (van Gent *et al.*, 1997) and decreases the K_d of the RAG complex for RSS binding (Lovely *et al.*, 2015). Binding is particularly enhanced to the 23-RSS (van Gent *et al.*, 1997). HMGB1/2 bends the longer 23 bp spacer in order to position the heptamer and nonamer a similar distance apart as in the 12-RSS (Ru *et al.*, 2015; Kim *et al.*, 2018). However, bending still occurs at the 12-RSS and is required for correct positioning in the complex, facilitating efficient coupled cleavage (Kim *et al.*, 2018; Zagelbaum *et al.*, 2016). Structural studies of the RAG complex revealed a 150° bend in the 23-RSS spacer, stabilised by binding of both boxes of HMGB1, whereas the 12-RSS has only a 60° bend stabilised by binding of Box A alone (Kim *et al.*, 2018). Both bend angles are crucial for cleavage and are thought to facilitate the binding of the RAG NBD and RNH domains to the nonamer and heptamer of both 12- and 23-RSSs, despite their different spacer lengths, and to correctly position both 'CAC's of the heptamer within the active site (Kim *et al.*, 2018; Ru *et al.*, 2018b).

Despite this substantial *in vitro* evidence, the role of HMGB1/2 during V(D)J recombination *in vivo* remained controversial (Thwaites *et al.*, 2019; Meek, 2019).

HMGB1 was shown to stimulate cleavage of a 12-RSS packaged on a nucleosome, suggesting involvement of HMGB1/2 in RSS cleavage *in vivo* (Kwon *et al.*, 1998). Additionally, over-expression of HMGB1/2 enhanced RAG:RSS binding *in vivo*, accompanied by up to a 4-fold increase in recombination of an extra-chromosomal substrate (Aidinis *et al.*, 1999). However, the observation that HMGB1 knockout mice produce normal levels of immunoglobulins and a complete T cell receptor repertoire cast serious doubt upon the involvement of HMGB1/2 *in vivo* (Calogero *et al.*, 1999).

1.10 RAG:RSS complex formation and the molecular basis of V(D)J recombination

In vitro the RAG proteins can form two major complexes with different stoichiometries. Both complexes contain two molecules of RAG1 but differ in the number of RAG2 molecules: SC1 contains one RAG2 protein whereas SC2 contains two RAG2 proteins (Swanson, 2002). HMGB1/2 is recruited to the complex after RAG binds to the RSS (Little *et al.*, 2013), leading to the formation of HSC1 or HSC2. It is currently puzzling how HMGB1/2 is recruited to the 23-RSS as the linker between Box A and Box B is threaded underneath the 23 bp spacer, suggesting that the RAG:RSS complex may have to partially dissociate to allow HMGB1/2 to bind (Kim *et al.*, 2018). The active RAG complex is most likely a heterotetramer consisting of two RAG1:RAG2 dimers bound to a 12/23-RSS pair plus two molecules of HMGB1 (Swanson *et al.*, 2004; Ru *et al.*, 2018b) (Figure 1.9a).

The recombination reaction begins by RAG binding to either a 12- or 23-RSS to form either a 12- or 23-SC (signal complex) respectively (Figure 1.9B). In the absence of a partner RSS, single stranded DNA nicking 5’ of the heptamer CAC can occur to leave a 3’OH on the coding segment end and a 5’ phosphate on the RSS end (Eastman and Schatz, 1997). Subsequent capture of a partner RSS leads to the formation of the synaptic paired complex (PC) within which both the 12- and 23-RSS are nicked (Swanson, 2002; Ru *et al.*, 2015). Next, the 3’OH of the coding end is directed to attack the opposite strand in a direct transesterification reaction to generate a pair of hairpin

sealed coding ends and a pair of blunt signal ends, which are held in the post-cleavage complex (PCC) prior to joining by the NHEJ pathway (Schatz and Swanson, 2011).

In recent years a number of high-resolution structures of the RAG:RSS complex have been solved by cryo-EM and X-ray crystallography, giving a more in-depth picture into the molecular mechanism of V(D)J recombination (Yin *et al.*, 2009; Kim *et al.*, 2015; Ru *et al.*, 2015; Kim *et al.*, 2018; Ru *et al.*, 2018a). The RAG proteins form a 'Y' shaped structure (Figure 1.9A) within which both molecules of RAG1 interact with the heptamer and nonamer the RSSs and represent the stalk of the 'Y', whereas RAG2 is found associated with the two coding ends in the branches of the 'Y'. Each RSS is cooperatively recognised by both RAG1-RAG2 dimers. For example, the 12-RSS coding segment is bound by the first RAG1-RAG2 dimer and the same RAG1 molecule recognises the beginning part of the 12-RSS heptamer (CAC) via its RNH domain. The more distal parts of the 12-RSS heptamer (3') and adjacent spacer nucleotides are recognised by the DDBD and CTD of the second RAG1 molecule, and the 12-RSS nonamer is bound mainly by the NDB of the second RAG1 molecule. However, the NBD dimer is a tightly interwoven structure where both RAG1 NBDs make contact with both nonamers (Yin *et al.*, 2009). It is quite remarkable that despite the differences in length of the 12- and 23-RSSs, they are contacted by the same residues of the RAG proteins. This is due to the malleability of the RSS which is facilitated by the large bends introduced by HMGB1/2 (Kim *et al.*, 2018). The mechanism behind the 12/23 rule was also revealed by these high-resolution structures. Binding of the flexible NBD dimer to one RSS leads to tilting of the entire dimer in a way that provides a complementary binding platform only for an RSS with the complementary spacer length (Ru *et al.*, 2015).

More recently, insights into the two catalytic steps, nicking and hairpin formation, have also been uncovered by these high-resolution structures (Kim *et al.*, 2018; Ru *et al.*, 2018a). Both steps are mediated by DNA distortions of the RSS within the active site, allowing the same active site to be used for both nicking and hairpinning, despite few

conformational changes of the site occurring between the two steps (Ru *et al.*, 2018b). To facilitate nicking, melting of the heptamer occurs which leads to the unwinding of the second and third positions of the heptamer (Ru *et al.*, 2018a). This positions the scissile phosphate in the active site and allows RAG to precisely nick the DNA 5' of the boundary between the coding segment and heptamer (Santagata *et al.*, 1999; Ru *et al.*, 2018a). This observation also explains why the 5' CAC sequence of the heptamer is so highly conserved, due to its weak base pairing and corresponding propensity for strand melting (Ramsden *et al.*, 1994; Ru *et al.*, 2018b). Further DNA distortions then occur to facilitate hairpin formation within the same active site. Base flipping of both nucleotides either side of the nick and a subsequent 180° rotation of the coding end, positions the scissile phosphate and attacking 3' OH of the coding end into the active site (Bischerour *et al.*, 2009; Ru *et al.*, 2015; Kim *et al.*, 2018). Hairpin formation is then catalysed by the DDE motif in the active site which coordinates two metal ions (Mg^{2+}) to allow nucleophilic attack of the scissile phosphate by the 3' OH on the opposite DNA strand, leading to hairpin sealed coding ends (Kim *et al.*, 1999).

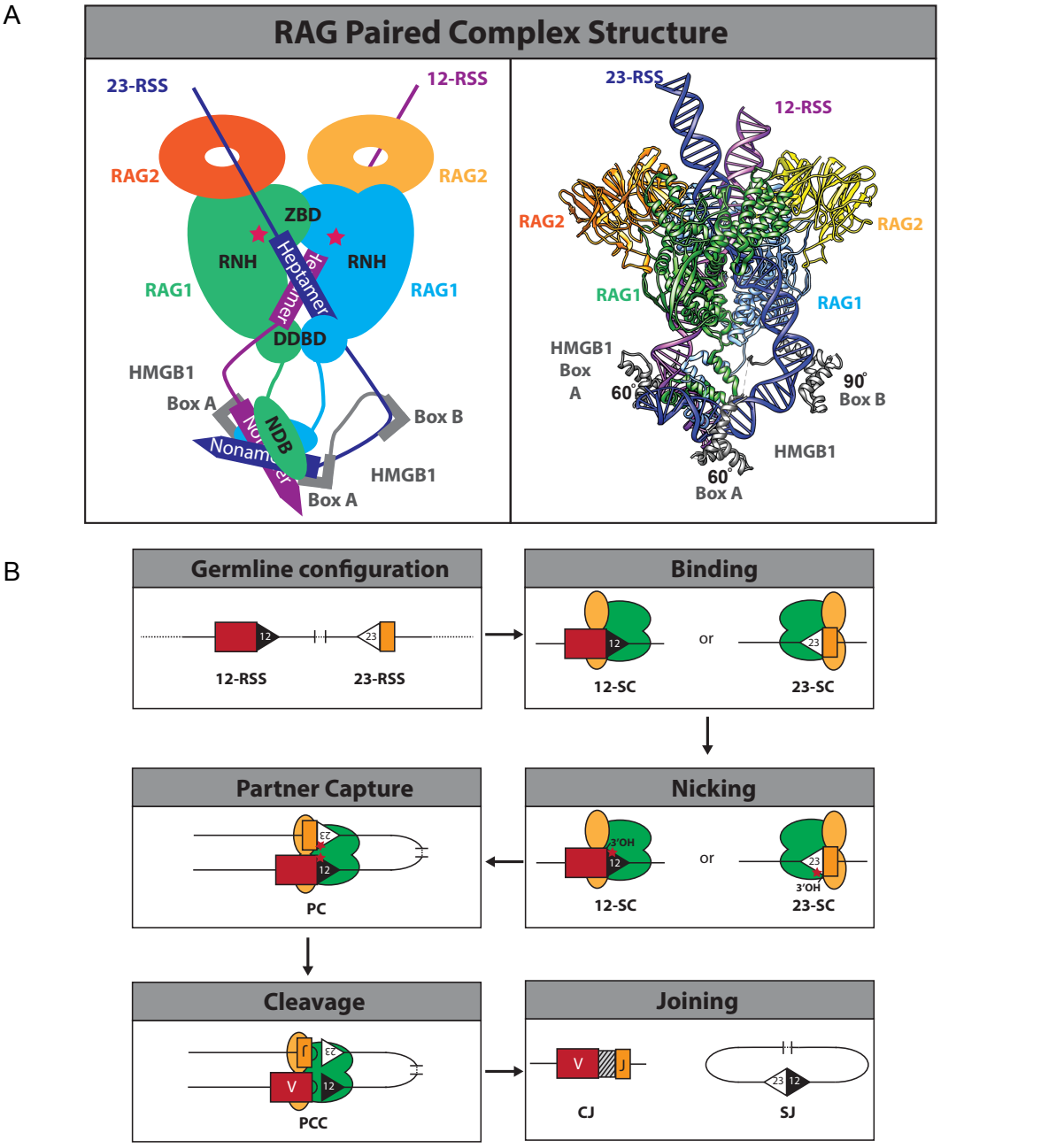


Figure 1.9 - RAG complex structure and molecular mechanism of V(D)J recombination

(A) Left: Simplified diagram of the RAG heterotetramer in complex with a 12- and 23-RSS. Each molecule of RAG1 is shown in green or blue and each molecule of RAG2 is shown in orange or yellow. The 12-RSS is shown in purple and the 23-RSS is blue. HMGB1 is represented in grey, both boxes can be seen at the 23-RSS whereas only Box A was visible in the high resolution structures at the 12-RSS. The active site of the RAG proteins is represented by red stars. Right: Cryo-EM structural model of the RAG complex (PDB 6CG0; Kim *et al.*, 2018). Colours of the proteins and DNA match those of the simplified diagram (Left). Bend angles induced in the 12 and 23 bp spacer are also shown. The structure of HMGB1 Box A from the 23-RSS was

superimposed over a density in the 12-RSS spacer that likely represents Box A that was not fully resolved in the original structure. (B) Diagram showing the different steps during V(D)J recombination, including the different intermediate RAG complexes which form during the reaction. Firstly a RAG heterotetramer (green and yellow ovals) bind to a single RSS to form a 12- or 23-SC (signal complex) and nick the DNA 5' of the heptamer to release a free 3' hydroxyl group. A partner RSS of the different spacer length is then captured to form the paired complex (PC) within which both RSSs are nicked. The free 3'OH then attacks the phosphate on the opposite strand to leave a pair of hairpin sealed coding ends and blunt signal ends. The ends are held by the RAG complex in a post-cleavage complex (PCC) before they are joined by the NHEJ machinery to produce a precise signal joint and imprecise coding joint.

C) Regulation of V(D)J recombination

1.11 Restraining the recombinase

Although V(D)J recombination allows the production of an efficient adaptive immune system, it is a daring approach in the pursuit of diversity. The reaction is inherently dangerous as it requires the generation of several double stranded DNA breaks (DSBs) in the genome, an event that cells normally aim to prevent. Therefore, V(D)J recombination must be properly regulated so that dangerous outcomes such as chromosomal translocations, that can contribute to cancer, can be avoided. To this end, various mechanisms are in place that govern V(D)J recombination throughout B and T cell development.

Firstly, the RAG proteins are only expressed in cells of the lymphoid lineage which restricts RAG activity to B and T cells. Secondly, recombination within B and T cells only occurs at the correct loci for that cell type, *Ig* loci in B cells and *Tcr* loci in T cells. Additionally, specific *Ig* and *Tcr* loci are only rearranged at distinct stages during B and T cell development, such as at the immunoglobulin heavy chain locus in pro-B cells or at the immunoglobulin light chain in pre-B cells. Rearrangements at these loci also occur in a defined order, for example at the immunoglobulin heavy chain locus D to J rearrangement occurs first, followed by V to DJ rearrangement. Finally, allelic exclusion

restricts rearrangements to just one allele. The following sections describe how these restrictions are imposed.

1.12 B and T cell development

Hematopoietic stem cells (HSCs) of the bone marrow are capable of generating all cell types of the blood. Initially, HSCs generate progenitors of the lymphoid and myeloid lineages. The common lymphoid progenitor (CLP) is the precursor to B, T and natural killer (NK) cells (Murphy and Weaver, 2017). A complex network of transcription factors provide the necessary signalling to drive the differentiation of CLPs to cells of the B or T cell lineages (Glimcher and Singh, 1999). Further advancement through the different cell stages is driven by a feedback mechanism between productive antigen receptor expression and developmental progression (Bassing *et al.*, 2002).

Expression of the transcription factor Pax5, the guardian of B cell identity, leads to commitment to the B cell lineage and gives rise to pro-B cells (Nutt *et al.*, 1999) (Figure 1.10). V(D)J recombination initiates in intermediate pro-B cells at the IgH locus, first D to J gene segment rearrangement occurs, followed by V to DJ rearrangement in late pro-B cells. Productive rearrangement forms the μ heavy chain which associates with surrogate light chains, λ5 and V-pre-B, to form the pre-B receptor (pre-BCR) (Bassing *et al.*, 2002). Expression of the pre-BCR on the cell surface of pro-B cells signals their expansion and differentiation into the large pre-B cell stage and cessation of heavy chain rearrangements (Geier and Schlissel, 2006). Further progression into the small pre-B cell stage is accompanied by light chain recombination of V and J gene segments at either the kappa or lambda loci (Gorman and Alt, 1998). Following productive light chain rearrangement, the product is expressed and pairs with the μ heavy chain to form the BCR which is then displayed on the surface of an immature B cell. Binding of the BCR to its cognate antigen and stimulation by a helper T-cell leads to activation and progression to the mature B cell stage, accompanied by somatic hypermutation and class switch recombination.

T cell development occurs in a similar manner to that of B cells. Like Pax5 expression in B cells, commitment to the T cell lineage is dependent upon expression of the transcription factor Notch1 (Pui *et al.*, 1999). The CD4-/CD8- double negative stage in TCRαβ cells mirrors the pro-B stage, with D to J rearrangements occurring before V to DJ at the TCRβ locus. Productive rearrangement leads to the expression of the β-chain along with a surrogate α chain (Tα), which forms the pre-T cell receptor (pre-TCR). Pre-TCR expression leads to proliferation and differentiation to CD4+/CD8+ double positive T cells (pre-T cells). Rearrangement of the TCRα locus then occurs at the pre-T cell stage leading to expression of the α-chain which pairs with the β-chain to form the T cell receptor. Unlike B cells, T cells do not undergo class switch recombination and were also thought, until recently, to not perform somatic hypermutation. Evidence of somatic hypermutation was recently discovered at the TCRα locus in shark T cells, at similar frequency to that of B cells (Ott *et al.*, 2018).

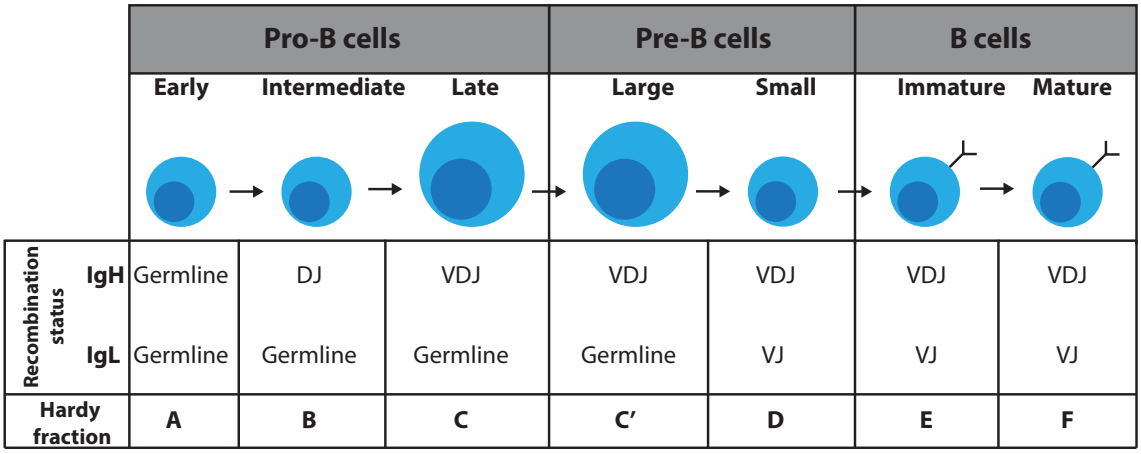


Figure 1.10 - B cell development

The recombination status of gene segment at the heavy chain (IgH) or light chain (IgL) loci are shown for each stage in B cell development. Fractions for each cell stage originally defined by Hardy are also shown (Hardy *et al.*, 1991).

1.13 Regulation of RAG expression

Expression of RAG proteins is regulated in a lineage and stage specific manner, consistent with the requirements for cell-specific and stage-specific V(D)J recombination. This is achieved by the use of a number of lymphoid specific and

ubiquitous transcriptions factors in conjunction with a highly conserved enhancer element, Erag (Fuller and Storb, 1997; Lauring and Schlissel, 1999; Hsu *et al.*, 2003). During B and T cell development there are two main waves of RAG expression which coincide with products of V(D)J recombination (Wilson *et al.*, 1994; Grawunder *et al.*, 1995). The first wave of RAG expression occurs in pro-B and pro-T cells and facilitates recombination at the heavy chain and TCR loci respectively. Expression of the pre-BCR and pre-TCR is followed by a sharp decrease in RAG expression levels, followed by rapid proliferation and entry into the pre-B and pre-T cell stages (Grawunder *et al.*, 1995). The second wave of RAG expression occurs upon exit from the cell cycle and facilitates recombination at the light chain or TCRα loci (Wilson *et al.*, 1994). Further rearrangements are then prevented by RAG silencing after BCR expression in the B cell or engagement of the TCR with a MHC molecule during positive selection (Brändle *et al.*, 1992).

RAG proteins can become re-expressed in the latter stages of B and T cell development during central tolerance. Central tolerance eliminates self reactive lymphocytes by either clonal deletion, cellular inactivation or receptor editing. Upregulation of RAG proteins during receptor editing allows further rearrangement of the light chain or TCRα loci in order to generate a functional antigen receptor with no auto-reactivity (Nemazee, 2006).

As described previously (Section 1.8 iv), V(D)J recombination is also restricted to G0/ G1 phases of the cell cycle via degradation of RAG2 outside of these phases, to ensure that RAG mediated breaks are repaired by the NHEJ pathway, rather than homologous recombination (HR), which is primarily active during these cell stages. This also prevents recombination occurring during DNA replication and cell division which could lead to aberrantly repaired breaks or segregation of broken chromosomes.

1.14 RSS chromatin accessibility

The eukaryotic genome is packaged into a DNA-protein complex known as chromatin. Chromatin is comprised of repeating units of nucleosomes which consist of 146 bp of DNA wrapped around a histone octamer. The histone octamer is a tight globular protein complex made up of a histone H3/H4 tetramer and two histone H2A/H2B dimers (Luger *et al.*, 1997). Histone H1 binds to DNA at the nucleosome entry/exit points and to linker DNA between nucleosomes to confer partial protection of linker DNA (Allan *et al.*, 1980). Long histone tails protrude from the globular nucleosome core and can become modified by a number post-translational modifications, such as acetylation and methylation, which forms the basis of the epigenetic code. These epigenetic marks confer a functional effect on the surrounding chromatin, governing if a particular part of the genome is open and accessible or closed and inaccessible (Jenuwein and Allis, 2001).

RAG proteins recognise and recombine the same elements, a 12- or 23-RSS, found at all antigen receptor loci, yet recombination in B and T cells is confined to either immunoglobulin or TCR loci respectively. Additionally, within a given lineage, all gene segments have these same RSSs but loci rearrange in an ordered manner e.g. heavy chain followed by light chain rearrangement. This phenomenon can be explained by differential accessibility to the RSSs, caused by open or closed chromatin states. This is dubbed the ‘accessibility hypothesis’ which proposes that RSSs of different loci and gene segments are only accessible for recombination in a particular cell type or stage in development (Yancopoulos and Alt, 1985). Although this was a widely accepted hypothesis at the time, it took many years before direct evidence of chromatin accessibility during V(D)J recombination was directly measured (Stanhope-Baker *et* al., 1996; Krangel, 2015). In this study, nuclei were extracted from B and T cells at differing stages of development and subjected to *in vitro* RAG cleavage. Sites of cleavage were measured and correlated strongly with sites in loci which are recombined at the respective stage of development or cell type, the first direct evidence that RSS

accessibility in chromatin directs lineage-specific and temporal order of gene segment recombination (Stanhope-Baker *et al.*, 1996).

RSSs are preferentially packaged into nucleosomes via positioning which is mediated by the nonamer sequence itself (Baumann *et al.*, 2003). This renders the RSS inaccessible for RAG binding and therefore inhibits V(D)J recombination (Kwon *et al.*, 1998; Golding *et al.*, 1999; McBlane and Boyes. 2000). Therefore, some form of remodelling must occur to allow an RSS to become accessible to the RAG proteins and initiate V(D)J recombination (Bevington and Boyes, 2013; Pulivarthy *et al.*, 2016).

1.15 Histone acetylation

Histones are highly basic proteins and therefore are attracted to negatively charged DNA via electrostatic attractions, helping to form a highly condensed chromatin structure. Acetylation of lysine residues found within the N-terminal tails of histones leads to neutralisation of the basic charge and facilitates chromatin decompaction (Hizume *et al.*, 2011). Indeed, acetylation of a single lysine residue at position 16 in histone H4 (H4K16Ac) has been shown to abolish higher order chromatin structure (Shogren-Knaak *et al.*, 2006). Hyperacetylated regions of the genome are associated with high levels of transcription and increased accessibility, due to their more open conformation (Krajewski and Becker, 1998; Clayton *et al.*, 2006).

Acetylation of histones H3 and H4 is tightly associated with the regulation of V(D)J recombination (Roth and Roth, 2000; Huang and Muegge, 2001; Espinoza and Feeney, 2005). One early observation was that hyperacetylation of the TCRα/δ loci correlated strongly with increased accessibility and V(D)J recombination (McMurry and Krangel, 2000). Histone acetylation has also been shown to play a role in the stage specific regulation of V(D)J recombination. In early pro-B cells, rearrangement of D and J gene segments at the heavy chain locus occurs first and is associated with hyperacetylation. In contrast, the V segments are only acetylated later, once D to J joining has occurred, enforcing the stepwise rearrangement of gene segments at the

immunoglobulin heavy chain locus (Chowdhury and Sen, 2001). The importance of histone acetylation during V(D)J recombination is likely due to its role in altering higher order chromatin structure (McBlane and Boyes, 2000). This is highlighted by the fact that histone acetylation is not sufficient to permit RAG cleavage of a RSS when it is packaged into a nucleosome *in vitro*, but is sufficient to allow V(D)J recombination *in vivo* (Golding *et al.*, 1999; McBlane and Boyes, 2000). The changes in the higher order chromatin structure to a more open conformation due to histone acetylation also aid the recruitment of remodelling complexes to their nucleosome targets, facilitating RAG accessibility to the RSS and therefore initiating V(D)J recombination (Nightingale *et al.*, 2007).

1.16 Histone methylation

The addition of methyl groups to lysine residues is another way in which histone tails can become modified. Lysines can have three methylation states; mono-, di-, and tri-methylation. Methylation of different lysine residues is associated with either transcriptional activation or repression. For example, tri-methylation of lysine 9 or lysine 27 of histone H3 (H3K9me3, H3K27me3) are repressive signals whereas tri-methylation of lysine 4 of histone H3 (H3K4me3) is associated with activation. This latter mark, H3K4me3, is essential for V(D)J recombination. Peaks of H3K4me3 are found at the ends of actively recombining regions of the immunoglobulin heavy chain and TCRβ loci (Morshead *et al.*, 2003), and acquisition of H3K4me3 at antigen receptor loci correlates with initiation of recombination at these loci during different stages of development (Perkins *et al.*, 2004; Goldmit *et al.*, 2004; Fitzsimmons *et al.*, 2007; Xu and Feeney, 2009).

The discovery of a PHD finger in RAG2 provided the link between the recombinase and H3K4me3 (Elkin *et al.*, 2005; Matthews *et al.*, 2007; Ramon-Maiques *et al.*, 2007). RAG2 binding to this mark aids in recruitment of RAG2 to RSSs in actively transcribed regions of the genome (Ji *et al.*, 2010; Schatz and Ji, 2011), and induces a number of conformational changes in the RAG complex which overcomes autoinhibition and leads

to an increase in RAG binding and cleavage activities (Lu *et al.*, 2015; Bettridge *et al.*, 2017) (Section 1.8 iv).

1.17 Non-coding transcription

The accessibility hypothesis described above (Section 1.14) is facilitated by non-coding transcription through the antigen receptor loci (Yancopoulos and Alt, 1985). Initiation of V(D)J recombination coincides with an up regulation of non-coding transcription at the recombining loci and is critical for rearrangements to occur at a given locus (Schlissel and Baltimore, 1989; Düber *et* al., 2003; Abarrategui and Krangel, 2006). Notably, the Krangel laboratory observed a severe drop in V and J rearrangements at the *TCRα* locus when transcription was blocked downstream of the T early α promoter (Abarrategui and Krangel, 2006). Additionally, lower levels of H3K4me2, H3K4me3 and H3K36me3 were observed at J segments downstream of the transcriptional terminator. Together these studies suggest that non-coding transcription is required to open the chromatin structure of downstream regions, which contain the RSSs, via the deposition of activating histone modifications (Abarrategui and Krangel, 2006; Abarrategui and Krangel, 2007).

During transcription RNA polymerase II recruits Set1, a histone methyltransferase, which deposits H3K4me3 at these transcribed regions (Ng *et al.*, 2003). Therefore non-coding transcription is also implicated in the recruitment and activation of the recombinase via binding of the RAG2 PHD finger to H3K4me3 (Abarrategui and Krangel, 2009; Matthew *et al.*, 2007; Ramon-Maiques *et al.*, 2007; Lu *et al.*, 2015; Bettridge *et al.*, 2017) (Section 1.8 iv).

RNA polymerase II also facilitates the deposition of acetylation marks in transcribed regions via the recruitment of histone acetyltransferases (HATs) (Wittschieben *et al.*, 1999). This aids in the formation of a more open chromatin architecture at these transcribed regions and therefore has implications in activating antigen receptor loci for recombination to occur (McBlane and Boyes, 2000) (Section 1.15).

However, H3K4me3 and histone acetylation are insufficient to allow initiation of V(D)J recombination (Bevington and Boyes, 2013). Although deposition of activating marks within recombining loci aids in the opening of the higher order chromatin structure, it does not explain how individual nucleosomes are remodelled to allow RAG access to an RSS. This could be explained by the increased recruitment of remodelling complexes to open regions (Nightingale *et al.*, 2007), but an inherent mechanism during transcription could also facilitate accessibility to an RSS occluded by a nucleosome. During transcription transient eviction of a H2A/H2B dimer from the nucleosome occurs (Kireeva *et al.*, 2002). This releases 35-40 bp of nucleosomal DNA and has been shown to be enough to allow RAG binding and cleavage to an RSS, both *in vitro* and *in vivo*, that would otherwise be blocked by the intact nucleosome (Bevington and Boyes, 2013). This may play an important part in maintaining genomic stability by only transiently allowing RAG access to RSSs during transcription, stopping genome wide RAG activity which could lead to the generation of many aberrant DNA breaks.

1.18 Allelic exclusion

As each lymphocyte genome contains two alleles of each antigen receptor loci, one question is how are productive rearrangements restricted to just one allele? (Sonoda *et al.*, 1997). The second allele either remains in the germline configuration or undergoes non-productive rearrangement, as two thirds of recombination events are out-of-frame. A number of models have been proposed to explain how this allelic exclusion is imposed; the asynchronous, stochastic, and feedback inhibition models (Vettermann and Schlissel, 2010).

Asynchronous recombination models rely on chromatin accessibility being the cause of mono-allelic rearrangements. There are two models for asynchronous recombination: the probabilistic model and the instructive model (Vettermann and Schlissel, 2010). The probabilistic model proposes that, due to limitations in chromatin accessibility, the low

efficiency of recombination means that the chance of bi-allelic rearrangement occurring would be approximately 0.25% (Perry *et al.*, 1980; Liang *et al.*, 2004; Schlimgen *et al.*, 2008; Vettermann and Schlissel, 2010). In the instructive model the asynchronous replication of each allele during early embryogenesis epigenetically marks the early replicating allele to become more accessible and therefore the allele upon which recombination can occur (Mostoslavsky *et al.*, 2001).

The stochastic model proposes that the high frequency of out-of-frame rearrangements means there is a low probability of a productive rearrangement occurring on both alleles (Coleclough *et al.*, 1981). However, only ~1% of B cells actually have rearrangements on both alleles (Casellas *et* al., 2007; Barreto and Cumano, 2000), but the stochastic model would allow for ~20% (Wabl and Steinberg, 1992).

In the feedback inhibition model, the products or intermediates of V(D)J recombination leads to inhibition of further recombination. Expression of a BCR/TCR leads to a cascade of signalling pathways which stops rearrangements occurring on the second allele. This is achieved by down regulation of RAG expression (Grawunder *et al.*, 1995) and reducing accessibility to the antigen receptor loci.

The RAG proteins themselves have recently been directly implicated in imposing mono-allelic recombination (Hewitt *et al.*, 2009; Hewitt *et al.*, 2017). Firstly, both homologous antigen receptor loci are thought to pair, in a RAG1 dependent manner, and cleavage on one allele induces ATM-dependent positioning of the second allele to pericentric heterochromatin (Hewitt *et al.*, 2009). This compartment of the nucleus is composed of tightly packed and inactive chromatin, and movement of the second allele to this region stops further rearrangements due to inaccessibility. Secondly, mutation of a conserved ATM-kinase phosphorylation site in RAG2 (S365) leads to bi-allelic cleavage, suggesting a feedback control mechanism involving RAG2 phosphorylation by ATM to enforce mono-allelic recombination (Hewitt *et al.*, 2017).

1.19 Long-range interactions

The genome is often visualised and represented as a linear moiety, in reality it is organised into a complex network of folds and loops which allows parts of the genome which are far apart in the linear sequence to be in close proximity (Oudelaar *et al.*, 2017). This nuclear organisation is dynamic and determines or reflects the function of different cell types and stages. However, it is unclear whether a given nuclear organisation is a direct consequence of nuclear processes occurring in that cell, or is folded deliberately to promote these processes (Oudelaar *et al.*, 2017).

Gene segments within the antigen receptor loci can be separated by up to 3 Mb of intervening sequence. This spacial separation of gene segments and regulatory elements contributes to the cell-specific and stage-specific regulation of V(D)J recombination. Indeed, use of fluorescent in situ hybridisation (FISH) to examine locus configuration has shown that immunoglobulin loci in T cells are in an extended spacial configuration whereas in pro- and pre-B cells these loci become contracted to facilitate recombination (Kosak *et* al., 2002; Roldan *et al.*, 2004; Sayegh *et al.*, 2005; Skok *et al.*, 2007). Three major transcription factors have been implicated in mediating these long-range interactions; Pax5, YY1 (Ying Yang 1), and CTCF (CCCTC-binding factor).

Pax5 is required for immunoglobulin heavy chain locus contraction in order to facilitate V to DJ recombination in the late pro-B cell stage (Fuxa *et al.*, 2004). YY1 is also important for immunoglobulin heavy chain locus contraction (Liu *et al.*, 2007b; Medvedovic *et al.*, 2013) but is potentially important in mediating stage specific contractions as it has also been implicated in contraction of the kappa light chain locus (Liu *et al.*, 2007b; Pan *et al.*, 2013). CTCF, along with cohesin, form large chromatin loops in order to mediate long range interactions (Tolhuis *et al.*, 2002, Parelho *et al.*, 2008). Loops formed by CTCF and cohesin at the *IgH* (Ebert *et al.*, 2013; Guo *et al.*, 2011), *IgK* (Ribeiro de Almeida *et al.*, 2015) and *TCRα* (Seitan *et al.*, 2011) have all been implicated in the regulation of V(D)J recombination.

D) Generation of additional antigen receptor diversity during B cell maturation

1.20 Further antibody diversification

V(D)J recombination in isolation is not capable of generating the vast antibody repertoire of the humoral immune system. Upon binding of a naive B cell to its cognate antigen, immunoglobulin genes can be further modified in order to fine tune antibody:antigen affinity or change the function of the antibody in processes known as somatic hypermutation (SHM) and class switch recombination (CSR), respectively. Both processes primarily occur in the germinal centers of lymph nodes and rely on activity of the enzyme activation-induced deaminase (AID) (Muramatsu *et al.*, 2000; Cattoretti *et al.*, 2006; Methot and Di Noia, 2017). The following sections describe both of these processes, their role in antibody diversification, and the molecular mechanism behind them.

1.21 Somatic hypermutation

SHM introduces point mutations into the variable regions of the immunoglobulin genes in order to alter antibody binding. The frequency of these point mutations is estimated to be one mutation per cell division (Rajewsky *et al.*, 1987). Mutations which confer an advantage in antigen recognition are selected and undergo further rounds of SHM and selection to progressively increase the affinity of the antibody. Therefore, B cells that have undergone terminal differentiation and committed to memory often have ~10-20 mutations clustered at the complementary determining regions (CDRs) (Tiller *et al.*, 2007). Broadly neutralising antibodies, such as those that can recognise HIV, can have more than 100 mutations found in both the CDRs and the scaffold of the antibody (Klein *et al*, 2013). However, such antibodies are rare but have great potential in recognising prominent pathogenic viruses that have high variability. Therefore, there is emerging interest in understanding and harnessing the process behind their generation (Burton and Hangartner, 2016; Methot and Di Noia, 2017).

SHM is initiated by AID which is a cytosine deaminase belonging to the APOBEC family. This catalyses the conversion of cytosine to uracil, introducing a U:G mismatch (Bransteitter *et* al., 2003; Conticello *et al.*, 2005) (Figure 1.11). AID primarily deaminates cytosines in single stranded DNA, such as 'DNA bubbles' which form during transcription (Dickerson *et al.*, 2003; Larijani *et al.*, 2007; Tchernaenko *et al.*, 2008). Indeed, AID activity and SHM at the immunoglobulin variable regions is closely coupled to transcription and can even be induced aberrantly at the constant regions by cloning a promoter upstream of these regions (Ramiro *et al.*, 2003; Peters and Storb, 1996). Additionally, the observation that AID in B cells associates with RNA polymerase II at actively transcribing immunoglobulin variable regions further strengthens the idea that transcription is required for SHM (Nambu *et al.*, 2003; Willmann *et al.*, 2012). The consensus WRC motif sequence (where W = A or T and R = A or G) is the preferred site for AID deamination and is found abundantly within SHM hotspots of the complementary determining regions (CDRs) DNA, which codes for the region of the antibody which binds the antigen (Pham *et* al., 2003; Di Noia and Neuberger 2007; Yeap *et al.*, 2015).

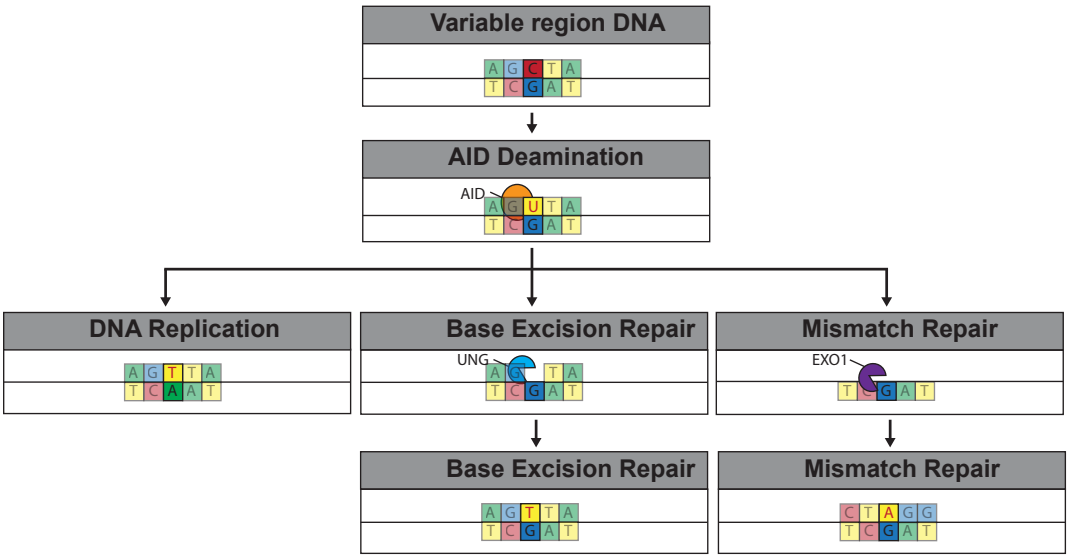


Figure 1.11 - Somatic hypermutation

Somatic hypermutation is initiated by AID activity on single stranded DNA bubbles found within the variable region DNA. This leads to the deamination of cytosine to uracil and therefore leaves a U:G mismatch. Repair of the lesion can occur by DNA replication, base excision repair (BER) or mismatch repair (MMR). If DNA replication encounters a U then it is simply recognised at T

and therefore leads to a C to T mutation. During BER the uracil is removed by uracil DNA glycosylase (UNG) to leave an abasic site which is repaired by error prone polymerases. During MMR exonuclease 1 (EXO1) can remove patches of nucleotides surrounding the U:G mismatch and lead to a greater number of mutated nucleotides.

SHM usurps a number of different ubiquitous cellular processes to resolve the U:G mismatch that was introduced by AID, including; base excision repair (BER), mismatch repair (MMR), or DNA replication (Figure 1.11). These processes normally function in a way that ensures genome fidelity, but during SHM they are hijacked to become mutagenic.

During BER the uracil base is removed from the DNA by uracil-DNA glycosylase (UNG) to leave an abasic site (Krokan and Bjørås, 2013). The abasic site is then cleaved by apurinic/apyrimidinic exonuclease 1 (APE1) to produce a single stranded nick in the DNA. The U:G mismatch can be faithfully repaired (C:G) via recruitment of the high fidelity DNA polymerase β and DNA ligase 3, facilitated by the activation and recruitment of PARP1 and XRCC4 to the nick. However, during SHM the U:G mismatch is erroneously repaired by addition of any of the four nucleotides via translesional DNA synthesis, performed by error prone polymerases such as REV1 or Polη (Methot and Di Noia, 2017).

During normal MMR the MSH2-MSH6 (MutSα) or MSH2-MSH3 (MutSβ) heterodimers recognise the mismatch which leads to the recruitment of MLH1 and PSM2. Following the recruitment of further elements, a single stranded nick is introduced near to the mismatch and exonuclease 1 (EXO1) excises the mismatch and an undefined patch of nucleotides surrounding the mismatch. The DNA is then faithfully repaired by high fidelity DNA polymerases (Polδ and Polε) and DNA ligase I. However, like in BER, error prone polymerases are recruited to repair the mismatch, facilitating mutation of the U:G mismatch and surrounding nucleotides (Peled *et al.*, 2008).

Finally, repair of the U:G mismatch by DNA replication always leads to a mutation. When the polymerase encounters the uracil during replication, it interprets it as a thymidine and therefore incorporates an adenine, resulting in a C:G to T:A mutation (Methot and Di Noia, 2017).

1.22 Class switch recombination

A naive B cell can express either IgM or IgD immunoglobulins via the alternative splicing of μ or δ heavy chain constant region exons. Upon B cell activation, this constant region can be changed to the other isotypes (IgG, IgA or IgE) in a process known as class switch recombination (CSR) (Figure 1.12).

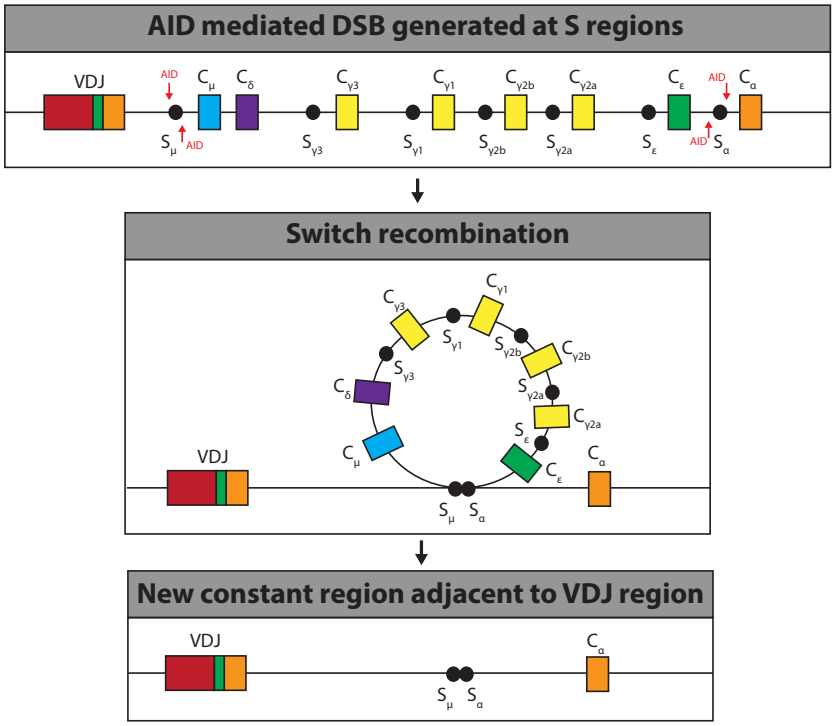


Figure 1.12 - Class switch recombination

AID activity at switch regions (S, black circles) found upstream of heavy chain constant regions initiates class switch recombination. This leads to the formation of two double strand breaks which are repaired, removing the intervening DNA. Consequently, a new constant region becomes joined to the variable region.

Like SHM, CSR is initiated by the enzyme AID whose activity relies on non-coding transcription through switch (S) regions found upstream of all heavy chain constant

region exons, expect IgD (Figure 1.12) (Xu *et al.*, 2012). Switching to a particular isotype is directed by T cell signalling which activates non-coding transcription through the S region of the required constant region exon (Stavnezer *et al.*, 2008). The S regions are highly enriched in 5'-AGCT-3' repeats which are the targets of the cytosine deaminase activity of AID (Xu *et al.*, 2010). Transcription-mediated single stranded DNA bubbles are generated in the S regions and this facilitates the deamination of cytosine to uracil by AID (Nambu *et al.*, 2003). UNG removes the uracil from the U:G mismatch and the resulting abasic site is cleaved by APE1 to leave a single stranded nick in the DNA. Should nicks form on opposing strands in close proximity, this leads to the formation of a DNA double stranded break (Stavnezer *et al.*, 2008). The donor and acceptor DSBs are then joined by the NHEJ machinery in an intrachromosomal deletional recombination reaction which excises the intervening DNA (Figure 1.12). Consequently, this joins the variable exon of the immunoglobulin gene to the new constant region exon.

E) RAGs and disease

1.23 RAGs and immunodeficiency - severity, prevalence and clinical outcome.

Due to their critical role in V(D)J recombination, mutation of the RAG proteins in humans can have serious pathogenic outcomes. Mutations which completely or partially diminish RAG activity during V(D)J recombination have a negative effect on the efficacy of the immune system and therefore can lead to immunodeficiency (Notarangelo *et al.*, 2016).

One major form of immunodeficiency is severe combined immunodeficiency (SCID), a group of disorders in which there are serious abnormalities in the production and function of T and B cells (Fischer *et al.*, 2015). Due to the lack of an effective immune system, patients present very early in life and fatally succumb to infections unless their immune system is successfully rescued by haemotopoetic stem cell transplantation (Pai *et al.*, 2014; Fischer *et al.*, 2015). T and B cell double negative SCID (T^-B^- SCID)

can be divided into two major forms. One form is associated with an increased cellular sensitivity to radiation and is often caused by a DNA repair defect (Nicolas *et al.*, 1998). The other form does not have this same sensitivity to radiation and is mainly caused by inactivating RAG mutations (Schwarz *et al.*, 1996). SCID is a relatively rare condition with an incidence of 1 case in every 58,000 live births (Kwan *et al.*, 2014), with 12% of these cases being caused by RAG deficiency (Kumánovics *et al.*, 2017). This highlights the critical role of the RAG proteins during the production of an effective immune system.

Unlike T^-B^- SCID, patients with Omenn syndrome have hypomorphic RAG mutations which retain some, but substantially reduced, levels of recombination activity (Villa *et al.*, 1998). However, Omenn syndrome is still a severe form of immunodeficiency which presents early in life and can also be fatal if untreated. As these patients retain low levels of RAG activity they do have some circulating lymphocytes. However, T cells are heavily skewed towards CD4+ T helper 2 cells which causes extremely high serum levels of IgE antibodies (de Saint-Basile *et al.*, 1991; Elnour *et al.*, 2007). Around 40% of Omenn patients have hypomorphic RAG mutations (Kwan *et al.*, 2014).

More recently RAG deficiencies have been identified in patients with a milder immunodeficiency phenotype, who present later in life. In one study, three individuals, from different families, between the ages of 2-10 were diagnosed with a type of primary immunodeficiency (PID) known as combined immunodeficiency (CID) (Schuetz *et al.*, 2008). This is a milder form of immunodeficiency where patients can produce a limited repertoire of TCR and antibody specificities but often present with granulomas and have severe complications after a viral infection. All of these CID patients were compound heterozygous for hypomorphic mutations in RAG1 or RAG2, which retained between 3-30% residual recombination activity (Schuetz *et al.*, 2008).

Analysis of whole genome sequencing data from ~60,000 individuals has estimated that the incidence of disease causing hypomorphic RAG mutations is 1:181,000

(Kumánovics *et al.*, 2017). This is twice the rate of inactivating RAG mutations observed in SCID and Omenn syndrome patients combined (1:336,000) (Kwan *et al.*, 2014; Kumánovics *et al.*, 2017). Therefore, hypomorphic RAG mutations are likely to contribute significantly to cases of PID where the cause is undiagnosed (Kumánovics *et al.*, 2017). Indeed, the increased popularity of whole genome sequencing has lead to the identification of hypomorphic RAG mutations in PID patients which survived to adulthood (Lawless *et al.*, 2018).

Together these studies suggest an emerging role for RAG deficiencies as one of the leading causes of early- and late-onset immunodeficiency. Therefore, there is a large amount of interest in the study of RAG mutations identified in these patients. This has revealed a wealth of knowledge about the function of the RAG proteins, that may one day aid in the treatment of immunodeficiency.

1.24 Mutations in key RAG1/2 regions lead to immunodeficiency

Approximately 150 RAG1 and 57 RAG2 mutations have been identified in immunodeficient patients (Notarangelo *et al.*, 2016). There is a strong correlation between the impact of these RAG1 and RAG2 mutations on recombination activity and the severity of the clinical outcome (Lee *et* al., 2014; Tirosh *et al.*, 2018). Within RAG1 the majority of these mutations (~120) fall within the crucial core region and are therefore associated with severe forms of immunodeficiency such as SCID and Omenn syndrome. Although most of these mutations are found in the zinc binding regions of core RAG1 (ZnH2 and ZnC2), when the number of mutations are normalised to the size of each domain the highest rate of mutation is found within the NBD, followed closely by the CTD (Notarangelo *et al.*, 2016). Mutations associated with SCID and Omenn syndrome which cause a severe impairment to RAG activity can be divided into four distinct categories (Kim *et al.*, 2015).

The first category are mutations which destabilise the overall tertiary structure of the RAG1/RAG2 complex. These mutations are often found within, or surrounding, the zinc

binding domains of RAG1. Mutations such as C727E, a key residue in zinc ion co-ordination, have been identified which can perturb the structure of the ZnC2 and ZnH2 domains, both of which are essential structural modules involved in formation of the RAG active site (Kim *et al.*, 2015; Gwyn *et al.*, 2009; Rodgers *et al.*, 1996) (Figure 1.8; Section 1.8 iii). Hydrophobic residues buried in the adjacent CTD have also been mutated in some cases to hydrophilic residues (W893R and I953R), likely leading to destabilisation of the CTD which would affect the structure of the zinc binding regions (Kim *et al.*, 2015) (Figure 1.8).

Mutation of solvent exposed polar residues form the second category of RAG mutations. These are generally mutations of positively charged residues in the NBD, DDBD or CTD that are important for DNA binding (Huye *et al.*, 2002). For example, twelve SCID/Omenn syndrome mutations found within the NBD are of polar residues which are important for the structural integrity and sequence specificity of the NBD (Yin *et al.*, 2009).

The third category of mutations are those which are found surrounding the active site. These mutations can either perturb the structural integrity of the active site or alter its DNA binding properties (Kim *et al.*, 2015).

The final category are those mutations which interfere with RAG1/RAG2 dimerisation. Most of the disease causing mutations found in core RAG2 are clustered at the interface between RAG1 and RAG2 (Kim *et al.*, 2015). Residues such as E170 in RAG2 and R561 in RAG1 form salt bridges to aid in dimerisation. Both of these residues have been observed in immunodeficient patients (Ko *et al.*, 2004; Dorna *et al.*, 2019) (Figure 1.8).

Due to the importance of the RAG1 core domains, there are very few hypomorphic mutations located in these regions, but those that do exist are thought to effect the overall quality of V(D)J recombination. One example is the mutation R975Q which

causes hypersensitivity to coding segment sequences and therefore restricts the repertoire of coding segments which can be involved in recombination (Wong *et al.*, 2008) (Figure 1.8). Mutations of the non-core domain can be less severe and have been identified in CID patients, such as R314W which is a key residues of the non-core RING domain (Schuetz *et al.*, 2008) (Figure 1.8).

In contrast to RAG1, disease mutation rate in RAG2 is highest in the non-core region. Of the 13 reported SCID/Omenn mutations, 12 are found in the PHD finger (Notarangelo *et al.*, 2016). These mutations destabilise the structure of the RAG2 PHD finger or, in the case of W453R, directly disrupt the H3K4me3 binding pocket (Mathews *et al.*, 2007; Couëdel *et al.*, 2010) (Figure 1.8). These observations clearly demonstrate the crucial role of the RAG2 PHD finger, and therefore the non-core domain, during V(D)J recombination.

1.25 Errors during antigen receptor production lead to cancer

V(D)J recombination is an inherently dangerous process as it is mediated by the formation of double stranded DNA breaks in the genome. Mistakes during the resolution of the DSBs generated during V(D)J recombination can lead to chromosomal translocations and genomic instability, both of which can contribute to development of cancer. Indeed, 35-45% of leukaemia and lymphomas bear hallmarks of errors during V(D)J and class switch recombination (Lieber, 1993). Moreover, a number of aberrant recombination events involving the excised by-product of V(D)J recombination have also been identified as having potential oncogenic effects. The following sections discuss how mistakes during V(D)J recombination can lead to cancer.

1.26 Chromosomal translocations

Recurrent translocations found in lymphoid tumours are thought to cause malignancy in a number of different ways (Nussenzweig and Nussenzweig, 2010). One mechanism is by the activation of proto-oncogenes or inactivation of tumour suppressor genes by translocation of one of these genes to a strong regulatory element found in the

immunoglobulin loci (Adams *et al.*, 1985; Vanasse *et al.*, 1999). A classical example is the t(14;18) translocation which is found in nearly all follicular lymphomas (Raghavan *et al.*, 2001). This translocation causes juxtaposition of a strong immunoglobulin heavy chain enhancer on chromosome 14 with the BCL2 promoter on chromosome 18, leading to constitutive overexpression of BCL2. BCL2 is an anti-apoptotic protein which acts to inhibit the release of cytochrome c from the mitochondria (Kridel *et al.*, 2012), therefore overexpression of BCL2 promotes cellular survival and aids in malignant transformation.

A further mechanism by which chromosomal translocations promote lymphoid malignancy is by the production of a fusion protein with oncogenic potential. The BCR-ABL kinase fusion protein is one example which is a hallmark of chronic myeloid leukaemia but also seen in some cases of acute lymphoblastic leukaemia (ALL). This fusion protein is formed by the t(9;22)(q34;q11) translocation, known as the Philadelphia chromosome, and promotes transformation by aberrantly activating the RAS-MAPK, JAK-STAT, and PI3K-AKT pathways (Teitell and Pandolfi, 2009). Another example is the t(12;21) translocation which forms the ETV6-RUNX1 (also known as TEL-AML1) fusion protein. This translocation is present in 22% of all childhood ALL cases, making it the most common genetic aberration in these patients (Shurtleff *et al.*, 1995). ETV6-RUNX1 expression cases a stall in B-cell development at the pro-B cell stage, consequently prolonging the stage where RAG proteins are highly expressed which leads to accumulations of further translocations (Pappaemmanuil *et al.*, 2014).

Chromosomal translocations require the formation of two DSBs in the genome. For translocations which involve the juxtaposition of a proto-oncogene next to a strong immunoglobulin enhancer element, the break at the immunoglobulin loci is often the result of RAG or AID activity. The source of the other break can be caused by either lymphoid specific processes such as mistargetted RAG and AID activity, or by non-lymphoid specific processes such as replication fork collapse or exposure to ionising radiation (Tsai and Lieber, 2010). Fusion protein translocations, such as BCR-ABL and

ETV6-RUNX1, do not show any obvious lymphoid specific causes. The cause of these breaks are not well defined and are likely to form by ubiquitous processes as seen in other non-lymphoid cancers.

1.27 Mistargeted RAG activity

i) Substrate selection errors

Chromosomal translocations can be caused by aberrant V(D)J recombination between a genuine RSS in the antigen receptor loci and any of the estimated 10 million RSS-like sequences, known as a cryptic RSS (cRSS), found throughout the genome (Lewis *et al.*, 1997; Roth, 2003) (Figure 1.13). cRSS are defined as any sequence of 12 or 23 basepairs in length which can be recombined by the RAG complex. Indeed, cRSS sequences are readily observed at translocation breakpoints from B and T cell malignancies (Tycko and Sklar, 1990). One study of known cRSSs revealed that only the first three nucleotides of the heptamer (CAC) were conserved across all sequences analysed and each cRSS typically contained only 9 of the 16 canonical heptamer and nonamer nucleotides (Lewis *et al.*, 1997).

However, a number of observations indicate that substrate-selection errors may not be a major cause of translocations in lymphocytes (Roth, 2003). Firstly, cRSSs identified at translocation breakpoint often contain heptamer and nonamer sequences which differ greatly from the consensus sequence. Studies of RSS sequence have revealed that mutating even just a single critical nucleotide in the heptamer sequence is enough to decrease DSB formation at these sequences by at least 100-fold (Steen *et al.*, 1997). This likely explains why some cRSSs are used more efficiently by RAG proteins, such as the LMO2, TAL1 and TAL2 cryptic RSS, whereas others are used much less efficiently such as cRSSs found at translocation breakpoints involving the BCL1 and BCL2 loci (Raghavan *et al.*, 2001; Marculescu *et al.*, 2002). Instead, the latter translocations likely occur due to a different mechanism and contain sequences resembling RSSs by chance. The second argument against the involvement of cRSSs during chromosomal translocations is that most breakpoints do not occur at the

heptamer-coding sequence boundary, this is inconsistent with RAG cleavage. Finally, many breakpoints also contain short direct repeats, a hallmark of cleavage events involving the formation of short single stranded overhangs, this is also inconsistent with RAG cleavage (Küppers and Dalla-Favera, 2001; Roth, 2003).

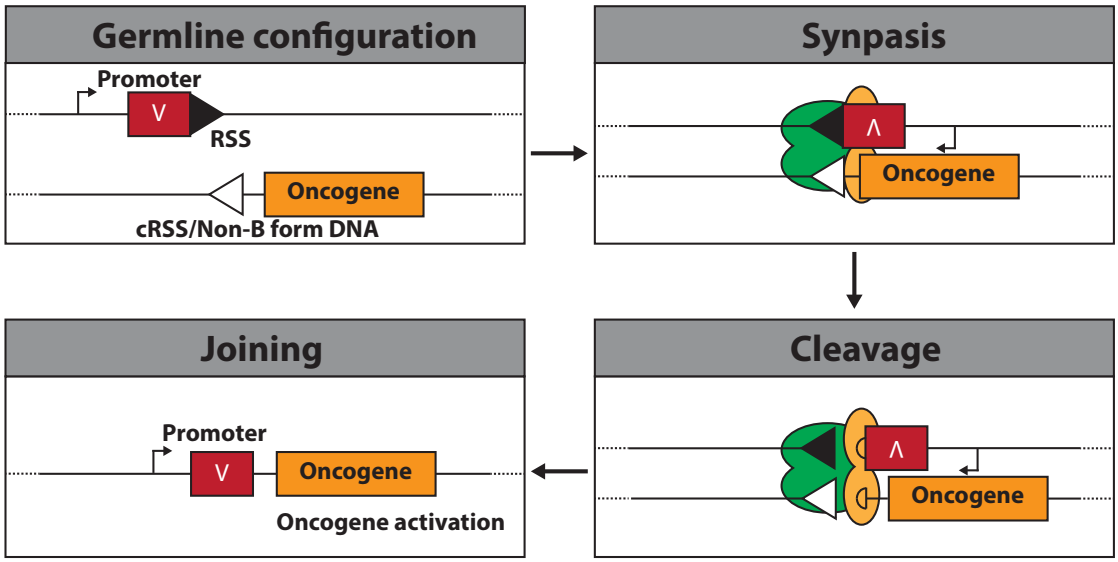


Figure 1.13 - Substrate selection errors

Strong regulatory elements, such as a promoter (black arrow), from the antigen receptor loci can be brought into range of a proto-oncogene by aberrant recombination of a genuine RSS (black triangles) with a cryptic RSS (cRSS) or non-B form DNA (white triangles).

ii) Structure-specific RAG binding

The t(14;18) translocation is a prevalent genetic aberration of follicular lymphoma yet, as mentioned above, there is no cRSS at the BCL2 major breakpoint (Mbr) that can be used efficiently by RAGs. However, the involvement of RAG in this translocation cannot be ruled out. The BCL2 Mbr reacts with bisulphite, a marker for single stranded DNA, which suggests it contains a region of non-B form DNA (Raghavan *et al.*, 2004a). RAG proteins are able to efficiently bind and cleave non-B form DNA of the BCL2 Mbr, *in vitro* and *in vivo* (Raghavan *et al.*, 2004b; Raghavan *et al.*, 2005), and consequently the BCL2-IGH translocation is readily observed in transfected lymphoid cells (Nambiar and Raghavan, 2012). The discovery that DNA melting of the heptamer in the RAG active site initiates cleavage (Ru *et* al., 2018a) may explain the ability of RAGs to act on non-

B form DNA and cause chromosomal translocations. Together this suggests the t(14;18) translocation may be caused by aberrant recombination between an RSS at the IgH locus and non-B form DNA at the BCL2 Mbr.

iii) End Donation

The above two models of chromosomal translocations occur by normal, although misguided, V(D)J recombination between two regions of the genome synapsed by the RAG proteins. However, in some translocations the non-immunoglobulin parter does not contain any hallmarks of RAG activity. Notably, approximately 50% of RAG mediated chromosomal aberrations only involve one RSSs (Papaemmanuil *et al.*, 2014). Another model, known as end donation, proposes that RAGs cause a DSB at a genuine RSS which then becomes joined to a second DSB that has formed by an entirely different process (Bakhshi *et al.* 1987; Lewis, 1994) (Figure 1.14). These secondary breaks could be formed via ionising radiation or collapse of replication forks. However, the ends generated by the RAG complex are thought to be heavily protected in the post-cleavage complex which also helps enforce their correct resolution (Deriano *et al.*, 2011; Coussens *et al.*, 2013). Therefore the mechanism by which an unrelated DSB can become joined to a RAG-mediated DSB is unclear. One study suggests that certain non-consensus heptamer sequences can destabilise the RAG post-cleavage complex, releasing the broken ends and facilitating end donation (Arnal *et al.*, 2010), but further investigation is required.

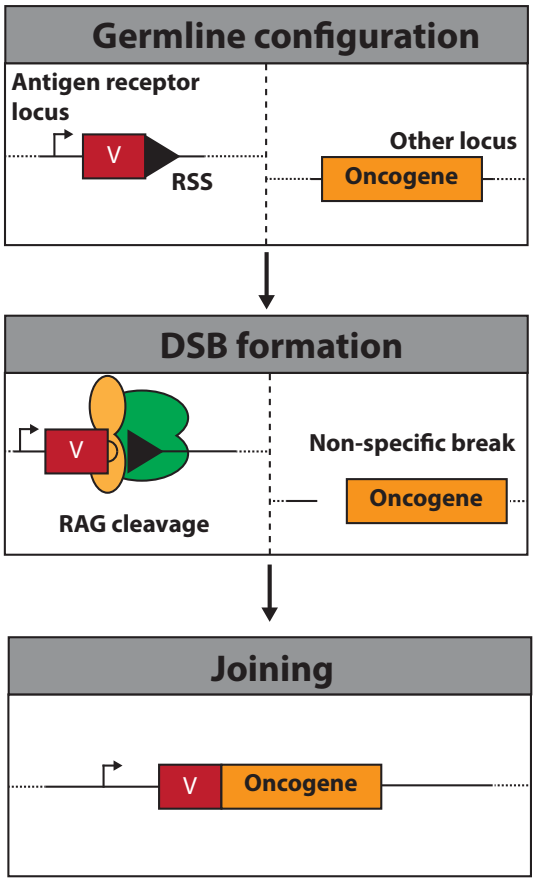


Figure 1.14 - End donation

The end donation model proposes that a DSB generated by RAG cleavage becomes joined to an unrelated DSB, facilitating chromosomal translocations in lymphocytes.

1.28 Dangers of the ESC

The DNA excised from the genome during V(D)J recombination, the excised signal circle (ESC), was thought to be an inert by-product of the reaction. More recently evidence has been emerging that suggests the ESC can pose a significant threat to genomic stability.

After cleavage, the RAG complex remains stably bound to the signal ends, which are then joined to form the ESC only when RAGs are down-regulated (Ramsden and Gellert, 1995). The ESC can also become rebound by RAGs after formation due to the presence of two RSSs in a head-to-head configuration (Kirkham *et al.*, 2019). Both of these situations allow the ESC to continue to be an active element which can be used

in a number of different aberrant reactions. The following sections describe three such aberrant reactions.

i) Transposition

Since RAG1 evolved from an ancient transposase (Section 1.8 ii) it is not surprising that it can catalyse an aberrant transposition-like reaction involving the cleaved RSS signal ends. During RAG-mediated transposition, the free 3'-OH group, found on the signal end, attacks the phosphodiester backbone of the genome (Figure 1.15). This allows the random RAG-mediated insertion of the ESC into any part of the genome via an NHEJ independent mechanism (Agrawal *et al.*, 1998; Hiom *et al.*, 1998). This could have potentially dangerous outcomes as it could juxtapose strong regulatory elements from the antigen receptor loci next to proto-oncogenes or tumour suppressors, similar to the way some chromosomal translocations contribute to oncogenesis (Section 1.26). Transposition could occur directly after coding end release from the PCC when the RAGs are still associated with the newly formed signal end, or it could occur after they are joined by RAGs opening the ESC via a nick-nick mechanism (Neiditch *et al.*, 2002).

Although the RAG proteins, especially their truncated core regions, can catalyse transposition with high efficiency *in vitro* (Agrawal *et al.*, 1998; Hiom *et al.*, 1998), there have been very few identified incidents of transposition *in vivo*. One event was observed in a human T cell line where an excised signal fragment from the *TCR*α locus became integrated into an intron of the *HPRT* gene (Messier *et al.*, 2003). However, there are no known cases of leukaemia or lymphoma where RAG-mediated transposition is implicated in its cause. This is likely due to the fact that *in vivo* full-length RAG2 is used and the RAG2 C-terminus has been shown to greatly inhibit transposition (Elkin *et al.*, 2003).

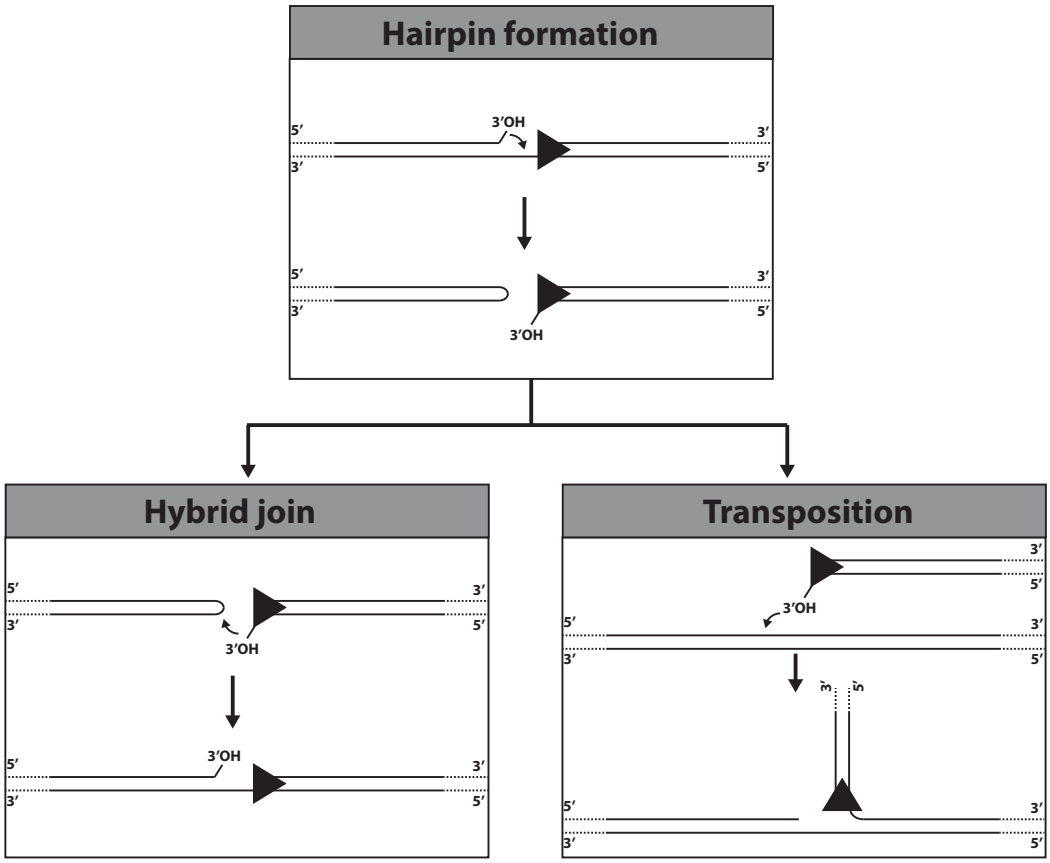


Figure 1.15 - Hairpin formation, hybrid joining and transposition

RAGs mediate hairpin formation via nicking of the DNA 5' of the heptamer to generate a free 3' hydroxyl (3'OH) group which can attack the opposing strand. This leaves a hairpin sealed coding end and a blunt signal end which also has a free 3'OH. The signal end 3'OH can attack the hairpin coding end, leading to the aberrant joining of hairpin and blunt ends which is known as a hybrid join. Transposition occurs by the same mechanism but the signal end 3'OH attacks the phosphodiester backbone of any region of the genome, leading to insertion of the signal end.

ii) Reintegration

Whereas transposition can lead to the insertion of the ESC sequence into any region of the genome, reintegration has target specificity. During the reintegration reaction the RAG:ESC complex can synapse with a genuine RSS or one of the 10 million cryptic RSSs in the genome and undergo a *trans* V(D)J recombination reaction (Vanura *et al.* 2007; Curry *et al.*, 2007) (Figure 1.16). Here, one of the signal ends from the ESC becomes joined to the signal end of the RSS, forming a precise signal joint, whilst the

other signal end joins to the coding end to form a pseudohybrid joint which bears hallmarks of processing similar to a normal coding join (Vanura *et al.*, 2007). The outcome is reintegration of the ESC into the genome at the target RSS.

Reintegration was first proposed based on observations from transfection experiments in mouse fibroblasts which revealed that RAGs could efficiently recombine an ESC into a target RSS (Vanura *et al.*, 2007). This study also demonstrated that reintegration was not restricted to genuine RSSs as ESC insertions were readily observed at cRSSs from the *LMO2* and *TAL2* oncogenes (Vanura *et al.*, 2007). Subsequent analysis of signal end integrations in developing thymocytes revealed that nearly half of the 48 characterised insertions were targeted to cryptic RSSs and showed no hallmarks of transposition (Curry *et al.*, 2007). Additionally, reintegration events were seven-fold more frequent in mice expressing core RAG, suggesting a role for the RAG2 C-terminus in inhibiting the reintegration reaction (Curry *et al.*, 2007).

Reintegration is a potentially dangerous reaction in that, similar to transposition, it could introduce strong regulatory elements from antigen receptor loci into regions of the genome close to proto-oncogenes and tumour suppressors. In contrast to transposition, reintegration is much more readily observed *in vivo* (Vanura *et al.*, 2007; Curry *et al.*, 2007), but there are still no reported cases of reintegration leading to leukaemia. However, many leukaemia and lymphoma cases display oncogene activation, such as *LMO2*, when no cytogenetic abnormalities are detected (Ferrando *et al.*, 2002). This leads to the possibility that reintegration is a potential driver of malignancy in such cases and could go undetected due to the relatively small size of the ESC which would not be observed by traditional cytogenetic techniques. It is estimated that ESC reintegration could occur in 1 in 10,000 to 1 in 1,000,000 lymphocytes, which equates to around 5000 reintegration events per human per day (Vanura *et al.*, 2007). Together this suggests that ESC reintegration is a significant threat to genomic instability.

iii) Cut-and-run

Whilst studying the reintegration reaction, the Boyes laboratory recently discovered an unexpected reaction, termed cut-and-run, with potentially devastating consequences for genomic stability. During the cut-and-run reaction RAGs can synapse the ESC with an RSS or cRSS in the genome, but in contrast to reintegration, cleavage only occurs at the target RSS and the ESC remains intact (Figure 1.16). This asymmetric cleavage is dependant on a separate RAG complex being bound to each of the head-to-head RSSs found on the ESC. Indeed, mutation of one of the RSSs of the ESC leads to restoration of symmetric cleavage (Kirkham *et al.*, 2019). Blocked cleavage of the ESC may be caused by steric clashes or due to both RAG complexes attempting to position one of the two head-to-head RSSs of the ESC within their active sites simultaneously. Thus, one factor that may determine whether reintegration or cut-and-run occurs is the number of RAG complexes bound to the ESC, one or two respectively.

After cleavage, the subsequent RAG signal end complex (SEC) containing an intact ESC and cleaved RSS is much less stable than a normal SEC containing a cleaved 12- and 23-RSS pair, which leads to the release of the cleaved RSS from the complex (Kirkham et al., 2019). This leaves RAGs bound to the ESC to facilitate the 'run' part of the reaction in which the RAG:ESC complex could potentially cut another RSS or cRSS in the genome. Indeed, transfection of RAG expression vectors and an extra-chromosomal ESC into fibroblast cells lead to increased γH2AX foci, a marker of DSBs, compared with RAG expression alone or in conjunction with an extra-chromosomal RSS (Kirkham et al., 2019).

These observations lead the authors to investigate whether the ESC can contribute to genomic instability. To this end, the human B cell line, REH, was transduced with an ESC or a single RSS and the number of translocations associated with RSSs were scored. ESC transduced cells had increased translocations involving broken RSSs compared with the control or 12- and 23-RSS transduced cells (Kirkham *et al.*, 2019). Analysis of breakpoints from these cells revealed breaks which map to within a 100 bp

window of 28 different ALL patient breakpoints. Additionally, breaks were increased in 9 of the 13 RSS enriched genes which are commonly mutated in ETV6/RUNX1-positive ALL (Kirkham *et al.*, 2019; Papaemmanuil *et al.*, 2014). Together these data suggest that cut-and-run is a potent cause of genomic instability and a likely cause of genetic aberrations, at least in ALL patients.

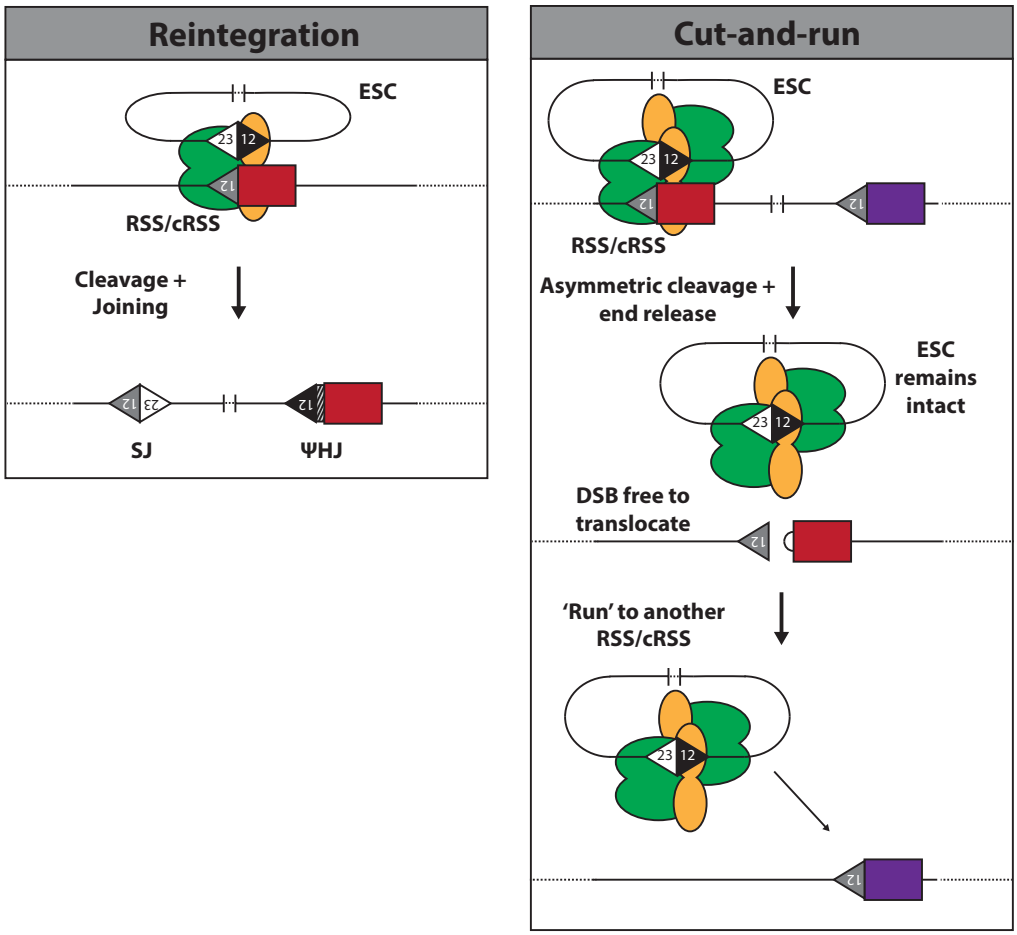


Figure 1.16 - Reintegration and Cut-and-run

Diagram of the reintegration and cut-and-run reactions. Left: Reintegration occurs via a trans-V(D)J recombination reaction between the ESC and a RSS or cRSS in the genome. This leads to a formation of a precise signal joint, containing the target RSS and the partner RSS from the ESC, and a pseudo hybrid join where the coding segment from the target RSS becomes joined to the other RSS of the ESC. Right: When each of the RSSs of the ESC is bound by a separate RAG complex cut-and-run occurs. One RAG complex synapses the ESC with an RSS or cRSS in the genome and leads to asymmetric cleavage and release of the target RSS. The released broken RSS is then capable of translocating with another break in the genome. As the ESC remains intact and bound by RAGs it can continue to 'run' to other RSSs in the genome and trigger further DSBs.

1.29 Protection of the genome by the RAG2 C-terminus

Despite the many mechanisms discussed above whereby RAG activity can lead to genomic instability, the RAG proteins inherently protect against these dangerous outcomes. More specifically the RAG2 C-terminus plays an important role in maintaining genome integrity during V(D)J recombination and inhibiting lymphomagenesis. This vital role of the RAG2 C-terminus was strongly emphasised by the observation of rapid thymic lymphoma formation, bearing complex genetic aberrations of the antigen receptor loci, in p53 deficient mice expressing core RAG2. The same outcome was not observed in p53 deficient mice alone or those also expressing core RAG1 (Deriano *et al.*, 2011).

One major mechanism whereby the RAG2 C-terminus protects genome integrity is by enforcing repair of RAG-mediated breaks by the classical NHEJ (cNHEJ) pathway. Firstly, the RAG2 C-terminus contains a degradation signal which restrains RAG expression, and therefore V(D)J recombination, to the cell stage which favours repair by NHEJ rather than HR (Section 1.8 iv). Secondly, a region of the acidic hinge within the RAG2 C-terminus inhibits repair by the error-prone alternative NHEJ (aNHEJ) pathway, ensuring repair by cNHEJ (Coussens *et al.*, 2013; Gigi *et al.*, 2014) (Section 1.8 iv). Other than processing that occurs on the hairpin coding end, repair by the cNHEJ machinery is highly accurate and efficient (Lieber, 2010), whereas repair by aNHEJ can lead to large deletions or short duplications and is implicated in chromosomal translocations (Deriano and Roth, 2013; Brandt and Roth, 2009). It is therefore beneficial for the RAG complex to inherently direct the repair of breaks it makes via the less erroneous cNHEJ pathway.

There is strong evidence that the RAG complex directs the joining phase of V(D)J recombination. Firstly, mutations which lead to an unstable post-cleavage complex (PCC), which holds the broken DNA ends before NHEJ repair, results in incorrect repair (Lee *et al.*, 2004). Secondly, RAG mutations exist that have normal cleavage activity

but are deficient in coding and signal join formation (Schultz *et al.*, 2001; Huye *et al.*, 2002). Thirdly, the RAG proteins co-purify with the Ku70/Ku80 complex (Raval *et al.*, 2008) and the RAG2 C-terminus has functional redundancy with XLF, a DNA repair protein which forms long filaments to tethering broken DNA ends during repair (Lescale *et al.*, 2016). Finally, coding joins form much more readily during *in vitro* cleavage reactions, supplemented with NHEJ proteins, when the RAG proteins remain associated with the ends (Leu *et al.*, 1997; Ramsden *et al.*, 1997).

Repair of a standard break in the genome by the NHEJ machinery normally involves the joining of just two breaks, however repair of breaks during V(D)J recombination is novel as it involves four breaks which must be correctly resolved. Loss of the RAG2 C-terminus leads to an increase in aberrant V(D)J recombination reactions such as transposition, reintegration and the formation of a hybrid join (Figure 1.15) (Elkin *et al.*, 2003; Vanura *et al.*, 2007; Sekiguchi *et al.*, 2001). Transposition is very rare *in vivo*, likely due to suppression by the RAG2 C-terminus (Elkin *et al.*, 2003), but the insertion event in the only reported natural transposition occurred at a site which likely contains cruciforms structures terminating in a hairpin (Roth, 2003). Therefore, all of these reactions share one common feature, the incorrect joining of a hairpin sealed coding end to a blunt signal end. This suggests a mechanism where the RAG2 C-terminus directs end repair to ensure correct hairpin-to-hairpin and blunt-to-blunt joining. How the RAG2 C-terminus governs this vital process is currently unknown.

F) Aims

Mutation of RAG proteins has the potential to cause immunodeficiency and lymphoid cancers. Whole genome sequencing of patients with SCID, Omenn syndrome, and PID (Kwan *et al.*, 2014; Kumánovics *et al.*, 2017; Lawless *et al.*, 2018) suggests an emerging role for RAG mutations as one of the leading causes of early- and late-onset immunodeficiencies (Sections 1.24 and 1.25). In collaboration with Dr Sanisa Savic, I was provided with two novel *RAG1* mutations identified in an adult patient with PID. The initial aims of this project were:

1. Develop a quantitative *in vivo* V(D)J recombination assay.
2. Determine the recombination activity of the *RAG1* mutations *in vivo*.
3. Analyse cleavage and binding activity of the *RAG1* mutation *in vitro*.

During this initial analysis, I discovered that RAG1 proteins carrying the R401W mutation are unable to catalyse V(D)J recombination and appear to have a deficiency in binding to the accessory protein HMGB1. Since the role of HMGB1 during V(D)J recombination *in vivo* was controversial, I set out to determine if the R401W mutation indeed affects HMGB1 binding and if this is the cause of decreased V(D)J recombination. Therefore, the aims of this project became:

1. Determine if the mutation R401W causes a deficiency in HMGB1 binding.
2. Investigate the mechanism behind the deficiency in HMGB1 binding.
3. Assess if HMGB1 is required for efficient V(D)J recombination *in vivo.*

The results from this project are presented in Chapter 3.

Aberrant V(D)J recombination reactions have the potential to cause genome instability and to aid lymphomagenesis. However, the RAG2 C-terminus has been demonstrated to suppress these aberrant reactions and to maintain genome stability (Roth, 2003;

Deriano *et al.*, 2011; Sections 1.27 - 29). I set out to identify which regions of the RAG2 C-terminus are important in suppressing one of these aberrant reactions, namely reintegration (Section 1.29). The aims of this project were:

1. Identify the region of the RAG2 C-terminus involved in suppressing reintegration.
2. Investigate the mechanism of suppression by this region.
3. Determine the involvement of this region in recombination and other aberrant reactions.

These aims are addressed in Chapters 4 and 5. In Chapter 4, I describe the development of an *in vivo* reintegration assay and the analysis of RAG2 C-terminal mutants in this assay. This led to the identification of a region in the RAG2 acidic hinge which is vital to suppress reintegration. I then analysed the effect of mutating this region on recombination and hybrid joint formation and explored the nature of the suppressive activity further. This suppression likely occurs during the joining step but high-resolution structures of the RAG post-cleavage complex and the RAG2 C-terminus are currently missing. In Chapter 5, I set out to develop an *in vitro* system with the aim of using this in structural studies to solve the aforementioned structures. This would reveal a wealth of knowledge regarding to joining step of V(D)J recombination and help in understanding the nature of the suppressive activity of the RAG2 acidic hinge.

Chapter 2 - Materials and Methods

A) Common Buffers

Alkaline lysis buffer I

50 mM	Glucose
25 mM	Tris-HCl (pH 8.0)
10 mM	EDTA (pH 8.0)

Alkaline lysis buffer II

0.2 M	NaOH
1% (w/v)	SDS

Alkaline lysis buffer III

3 M	KOAc
5 M	Glacial acetic acid

DNA sample loading buffer (6x)

15% (w/v)	Ficoll®-400
20 mM	Tris-HCl (pH 8.0)
60 mM	EDTA (pH 8.0)
0.48% (w/v)	SDS
0.03% (w/v)	Xylene Cyanol
0.03% (w/v)	Bromophenol Blue

Protein sample loading buffer (6x)

6% (w/v)	SDS
50% (v/v)	Glycerol
60 mM	Tris-HCl (pH 6.8)
640 mM	β-Mercaptoethanol
0.06% (w/v)	Bromophenol Blue

TE (Tris-EDTA) buffer

10 mM	Tris-HCl (pH 8.0)
1 mM	EDTA (pH 8.0)

TAE (Tris-acetate-EDTA) buffer

40 mM	Tris base
20 mM	Glacial acetic acid
1 mM	EDTA (pH 8.0)

TBE (Tris-borate-EDTA) buffer

90 mM	Tris base
90 mM	Boric acid
2 mM	EDTA (pH 8.0)

TBS (Tris buffered saline) buffer

50 mM	Tris-HCl (pH 7.4)
150 mM	NaCl
2 mM	KCl

TGS (Tris-glycine-SDS) buffer

25 mM	Tris base
192 mM	Glycine
0.1% (w/v)	SDS

PBS (Phosphate buffered saline) buffer

137 mM	NaCl
2.7 mM	KCl
4.3 mM	Na_2HPO_4
1.47 mM	KH_2PO_4

Western blot semi-dry transfer buffer

48 mM	Tris base
39 mM	Glycine
20% (v/v)	Methanol
0.04% (w/v)	SDS

B) Media

Dulbecco's Modified Eagle Medium (DMEM)

DMEM medium (Sigma-Aldrich, D6546) supplemented with:

10%	Foetal Calf Serum (PAA Laboratories)
2 mM	L-Glutamine
50 μg/ml	Streptomycin
50 U/ml	Penicillin

Lysogeny broth (LB)

1% (w/v)	Bacto-tryptone
0.5% (w/v)	Yeast Extract
0.5% (w/v)	NaCl

LB-Agar

1% (w/v)	Bacto-tryptone
0.5% (w/v)	Yeast Extract
0.5% (w/v)	NaCl
1.5% (w/v)	Agar

Super optimal broth (SOB)

2% (w/v)	Bacto-tryptone
0.5% (w/v)	Yeast Extract
0.5% (w/v)	NaCl
0.019% (w/v)	KCl

Terrific broth

1.2% (w/v)	Bacto-tryptone
2.4% (w/v)	Yeast Extract
0.4% (v/v)	Glycerol
170 mM	KH_2HPO_4
270 mM	K_2HPO_4

The salts were added following autoclaving

2x YT broth

1.6% (w/v) Bacto-tryptone
1% (w/v) Yeast Extract
0.5% (w/v) NaCl
Adjust pH to 7.0 with NaOH

C) DNA and RNA based methods

2.1 Restriction digestion of DNA

Restriction digestions were performed using enzymes from New England Biolabs (NEB). Reactions were performed as specified by the manufacturer's instructions, using the appropriate buffer for each enzyme. For a diagnostic digestion, 0.5 µg of DNA was used in a total volume of 10 µl and a small excess of restriction enzymes were added and incubated for 1 hour at 37°C. For larger amounts of DNA (>1 µg), such as the amount needed in a vector or insert digest for cloning, the reaction volume was scaled up to 200 µl and the incubation was performed overnight at 37°C.

2.2 Agarose gel electrophoresis

DNA was separated according to size by electrophoresis through an agarose gel containing 1X TAE buffer and 5 µg/ml ethidium bromide. The concentration of agarose (0.8 - 2%) used was dependent upon the size of the DNA being analysed. Prior to loading, DNA sample loading buffer was added. Electrophoresis was performed using the Bio-Rad Sub-Cell® GT system. Gels were electrophoresed in 1x TAE buffer at 80 V for a small gel (7 x 10 cm) or 120 V for a large gel (15 x 10 cm), until sufficient separation was achieved. DNA was visualised in the gel by exposure to UV and imaged using a Bio-Rad gel documentation system.

2.3 Isolation of DNA fragments from agarose gels

DNA fragments were separated by agarose gel electrophoresis (Section 2.2) and visualised under UV light. Fragments were excised using a clean scalpel blade,

working quickly to avoid over exposure of the DNA to damaging UV light. The resulting gel slice was placed in a dialysis membrane (Thermofisher, BioDesignDialysis Tubing, 8000 MWCO) containing 1x TAE buffer, clipped at both ends. DNA fragments were eluted from the gel slice into the surrounding buffer by electrophoresis at 80 V for one hour. The polarity of the current was reversed for ~30s to recover any DNA stuck to the surface of the dialysis membrane. The buffer containing the eluted DNA was subjected to two rounds of phenol:chloroform extraction (Section 2.4) and ethanol precipitated (Section 2.5). DNA pellets were resuspended in an appropriate amount of TE buffer.

2.4 Phenol:chloroform extraction

A equal volume of a 1:1 (v/v) phenol and chloroform solution was added to the DNA sample, vortexed for 30 seconds and centrifuged at 16,000 *g* for 2 minutes. The upper phase containing the DNA was transferred to a new tube and the lower phase containing proteins and other organic contaminants was discarded. This was repeated until there was no longer a visible white interface between the two phases, after which the DNA was ethanol precipitated (Section 2.5).

2.5 Ethanol precipitation of DNA

DNA was precipitated by the addition of 0.1 volumes of 3M sodium acetate pH 5.2 and two volumes of 100% ethanol. If the amount of DNA was expected to be 20 µg or lower, 1 µl of 20 mg/ml glycogen was added to act as a carrier. The tube was inverted several times to mix, incubated on dry ice for 5 minutes and centrifuged at 20,000 *g* for 10 minutes at 4°C. The supernatant was removed and the DNA was de-salted in 1 ml 70% ethanol, vortexed and centrifuged again at 20,000 *g* for 3 minutes at 4°C. Finally, the supernatant was removed and the pellets were air dried before resuspension in TE.

2.6 PCR with *Taq* polymerase

Routine PCR or the first round of nested PCR was performed using *Taq* DNA polymerase (M0267S, NEB). Reactions were typically 25 µl and contained 1X ThermoPol buffer (20 mM Tris-HCl pH 8.8, 10 mM $(NH_4)_2SO_4$, 10 mM KCl, 2 mM

$MgSO_4$, 0.1% Triton X-100), 200 µM dNTPs, 0.2 µM of each primer, 0.625 units *Taq* DNA polymerase and an appropriate amount of template DNA

Cycling conditions were typically as follows:

95°C 30 seconds

95°C 15 seconds

T_m* 15 seconds

68°C 1 minute/kb

68°C 5 minutes

30 cycles

**Calculated using the NEB Tm Calculator for Taq DNA polymerase with ThermoPol buffer*

2.7 PCR with Q5 polymerase

PCR to generate fragments for cloning or site-directed mutagenesis (Section 2.20) was performed using Q5® High-Fidelity DNA polymerase (M0491S, NEB). Reactions were typically 25 µl and contained 1x Q5 Reaction Buffer, 200 µM dNTPs, 0.5 µM of each primer, 0.5 units Q5 DNA polymerase and 1-10 ng of plasmid DNA as template. For difficult to amplify templates, such as lentiviral vector backbones, 1x Q5 Enhancer was added.

Cycling conditions were typically as follows:

98°C 30 seconds

98°C 10 seconds

T_m* 20 seconds

72°C 30 seconds/kb

72°C 3 minutes

25-35 cycles

**Calculated using the NEB Tm Calculator for Q5 DNA polymerase*

2.8 Ligation of DNA

DNA was ligated using T4 DNA Ligase (M0202S, NEB). Reactions were in a final volume of 10 µl and contained 1X T4 Ligase Buffer (50 mM Tris pH 7.5, 10 mM $MgCl_2$,

1 mM ATP, 10 mM DTT), 200 units of enzyme and typically 100 ng of vector with a three fold molar excess of insert. Sticky end ligations were incubated at room temperature for a minimum of 20 minutes whereas blunt end ligations were incubated at 16°C overnight. Vectors generated by restriction digestion were dephosphorylated with calf intestinal phosphatase (CIP, M0290S, NEB) to prevent intramolecular ligation. Inserts generated from annealed oligonucleotides or by PCR were phosphorylated prior to ligation using T4 polynucleotide kinase (T4 PNK, M0201S, NEB). Phosphorylation reactions were typically in a volume of 25 µl and contained 1X T4 Ligase Buffer, 5 units of T4 PNK and up to 300 pmol of 5' end termini and were incubated at 37°C for 30 minutes.

2.9 Preparation of chemically competent *E.coli*

Chemically competent DH5α or Endura™ (Lucigen) *E.coli* were prepared by treatment with $CaCl_2$ (Sambrook *et al.*, 1989). A single colony from a freshly streaked plate of *E.coli* was used to inoculate 50 ml LB in a 250 ml flask and incubated overnight at 37°C with shaking (200 rpm). Following this, 4 ml was transferred to 400 ml of LB in a 2L flask and incubated at 37°C with shaking (200 rpm) until an OD_{600} of approximately 0.375 was reached. The culture was cooled on ice for 5-10 minutes in eight 50 ml aliquots and cells were harvested by centrifugation at 1600 *g* for 7 minutes at 4°C. Each pellet was gently resuspended in 10 ml of ice cold $CaCl_2$ solution (60 mM $CaCl_2$, 15% glycerol, 10 mM PIPES pH 7.0) and centrifuged at 1100 *g* for 5 minutes at 4°C. Pellets were again resuspended in 10 ml $CaCl_2$ solution but were incubated on ice for 30 minutes for centrifugation. Finally, each cell pellet was resuspended in 2 ml ice cold $CaCl_2$ solution and 200 µl aliquots were prepared and immediately snap frozen on dry ice and stored at -80°C until use.

2.10 Transformation of *E.coli*

Plasmids containing repetitive sequence, such as those used for lentivirus production, were transformed into chemically competent Endura™ (Lucigen) *E.coli* whereas all

other plasmids were transformed into chemically competent DH5α *E.coli*. After thawing an aliquot of competent cells on ice, an appropriate amount of DNA was added to 50 µl of cells and incubated for 30 minutes on ice. The cells were heat shocked for 30 seconds in a 42°C water bath and returned to ice for two minutes. Following this, 300 µl SOB was added and the cells were incubated at 37°C for one hour. For site directed mutagenesis or transformation of a ligation reaction, 150 µl of cells was spread on an LB agar plate containing the appropriate antibiotics, whereas for transformation of an existing plasmid approximately 10 µl of cells was used. Plates were incubated at 37°C overnight.

2.11 Small scale purification of plasmid DNA from *E.coli* (miniprep)

A single bacterial colony was picked from an LB agar plate, used to inoculate 2 ml of LB and incubated for 16 hours at 37°C with shaking (200 rpm). The culture was transferred to a 1.5 ml centrifuge tube and the remaining 0.5 ml of culture was stored at 4°C for subsequent use in large scale plasmid purification. The cultures were centrifuged at 12,000 *g* for 30 seconds, the supernatant was discarded and the bacterial pellet was fully resuspended in 100 µl of alkaline lysis solution 1 (50 mM glucose, 25 mM Tris.HCl pH 8.0, 10 mM EDTA pH 8.0) by vortexing. Cells were lysed by the addition of 200 µl of alkaline lysis solution 2 (0.2 M NaOH, 1% w/v SDS), gently mixed by inverting the tube and incubated on ice for 5 minutes. Next, 150 µl of alkaline lysis solution 3 (3M KOAc, 11.5% v/v glacial acetic acid) was added and the tube was mixed vigorously for 10 seconds and stored on ice. After 5 minutes, the samples were centrifuged at 12,000 *g* for 5 minutes and the supernatant transferred to a fresh tube. The samples were subjected to one round of phenol:chloroform extraction (Section 2.4) and ethanol precipitated (Section 2.5) without the addition of extra salt since there is sufficient salt in the alkaline lysis solutions. The DNA pellet was de-salted by the addition of 1 ml 70% ethanol and centrifuged at 12,000 *g* for 3 minutes. Finally, the pellet was resuspended in 50 µl TE and 2 µl of RNase A (10 mg/ml) was added.

Typically, samples were analysed by restriction digestion (Section 2.1) to identify the correct clones.

2.12 Large scale purification of plasmid DNA from *E.coli* (midiprep)

The remaining culture from Section 2.12 was used to inoculate 100 ml of LB in a 250 ml flask. Cultures were incubated for 16 hours at 37°C with shaking (200 rpm) and cells were harvested by centrifugation at 3795 *g* for 20 minutes. Plasmid DNA was purified using the Qiagen Plasmid Midi Kit (Qiagen, #12943) as per the manufacturer's instructions. Purified DNA was resuspended in 200 µl TE and the concentration was measured using a DS-11 spectrophotometer (DeNovix) and adjusted to 1 mg/ml.

2.13 Genomic DNA extraction from mammalian cells

Approximately 1 x 10^6 cells were transferred to a 1.5 ml micro centrifuge tube, pelleted by centrifugation at 800 *g* for 3 minutes and washed twice in 1 ml PBS. Cells were lysed by resuspending the cell pellet in 500 µl of cell lysis buffer (200 mM NaCl, 10 mM Tris-HCl pH 7.4, 2 mM EDTA, 0.2% w/v SDS, 200 µg/ml Proteinase K) and incubated overnight at 55°C. Following this, 500 µl of isopropanol was added directly to the lysate and immediately mixed by inversion. A pipette tip was used to transfer the pellet of genomic DNA to a new micro-centrifuge tube containing 1 ml 70% ethanol which was centrifuged at 20,000 *g* for 5 minutes at 4°C. The supernatant was discarded and the pellet was completely air dried before being resuspended in 100 µl TE. Tubes were placed at 37℃ for 2 hours with agitation to aid resuspension of genomic DNA.

2.14 Extraction of total RNA from mammalian cells

RNA was extracted from mammalian cells using TRI Reagent (T9424, Sigma). Approximately 1 x 10^6 cells were transferred to a 1.5 ml micro centrifuge tube, pelleted by centrifugation at 800 *g* for 3 minutes and washed twice in 1 ml PBS. Cells were lysed by the addition of 500 µl TRI Reagent and incubated at room temperature for 5 minutes. Following this, 100 µl of chloroform was added and the sample was vortexed for 30 seconds. After incubation at room temperature for 5 minutes, samples were

centrifuged at 12,000 *g* for 15 minutes at 4°C. The upper aqueous phase, which contains RNA, was transferred to a new 1.5 ml micro-centrifuge tube and 250 µl of isopropanol was added, tubes were mixed by inversion and incubated at room temperature for 10 minutes. The precipitated RNA was pelleted by centrifugation at 12,000 *g* for 10 minutes at 4°C and the pellet was washed with 1 ml of 70% ethanol. The RNA pellet was completely air dried and resuspended in 50 µl RNase free water. Contaminating genomic DNA was removed by treatment with DNase I (18047019, ThermoFisher). Reactions were performed using 2 Kuniz unites of DNase I, in DNase I reaction buffer (10 mM Tris-HCl pH 7.5, 2.5 mM $MgCl_2$, 0.5 mM $CaCl_2$) and were incubated at 37°C for 20 minutes. DNase I was removed by phenol:chloroform extraction (Section 2.4), and the RNA ethanol precipitated (Section 2.5) and resuspended in 50 µl TE. The concentration of RNA was determined by absorbance at 260 nm using a DS-11 Spectrophotometer (DeNovix) and 1 µg was used to synthesis cDNA.

2.15 Synthesis of cDNA from RNA

M-MLV Reverse Transcriptase (28025013, Invitrogen) was used to synthesise cDNA from total RNA extracted from mammalian cells (Section 2.14). For first strand synthesis, a reaction containing 1 µg RNA, 0.5 µg oligo dT primer, 1 µl 10 mM dNTPs and water to 12 µl was heated to 65°C for 5 minutes. Following this, the reaction was immediately placed on ice and 4 µl First-Strand buffer (250 mM Tris-HCl pH 8.3, 375 mM KCl, 15 mM $MgCl_2$), 2 µl 100 mM DTT and 1 µl RNasin (RNase Inhibitor, Promega) were added. The reaction was pre-heated to 37°C for 2 minutes before adding 1 µl of M-MLV Reverse Transcriptase, and then incubated at 37°C for a further 50 minutes. The reaction was stopped by heating to 70°C for 15 minutes and the cDNA was phenol:chloroform extracted (Section 2.4), ethanol precipitated (Section 2.5) and resuspended in 50 µl TE.

D) Common protein based methods

2.16 Preparation of whole cell protein extracts

Typically, mammalian cells from a confluent well of a 6-well plate were used to make protein extracts. The media was aspirated and the cells were washed twice which ice cold PBS. A cell scraper was used to remove the cells from the well and then transferred to a 1.5 ml micro-centrifuge tube in 1 ml ice cold PBS. Cells were harvested by centrifugation at 800 *g* for 5 mins at 4°C and resuspended in 50 µl of a 3:1 mix of RIPA (150 mM NaCl, 1% v/v IGEPAL CA-630, 0.5% w/v NaDOC, 0.1% w/v SDS, 50 mM Tris pH 8.0) and lysis buffer (5% w/v SDS, 150 mM Tris pH 6.8, 30% glycerol) supplemented with 1 x Pierce™ Protease Inhibitor (A32953, ThermoFisher). Lysates were immediately boiled for 5 minutes and then cleared by centrifugation at 20,000 *g* for 10 minutes at 4°C. Samples were either used immediately for SDS-PAGE (Section 2.17) or stored at -80°C.

2.17 SDS-PAGE of proteins

SDS-PAGE gels were cast using a Mini-PROTEAN casting system (Biorad) using gel solutions prepared according to Laemmli (1970). Separating gels were typically 10% (37.5:1 acrylamide:bis-acrylamide), with a stacking gel of 4% (37.5:1 acrylamide:bis-acrylamide). Before loading on the gel, protein samples were mixed with protein sample loading buffer to 1x final concentration and boiled for 2 minutes. Gels were submerged in 1x TGS running buffer and electrophoresed at 125 - 180V in a Mini-PROTEAN Tetra Cell gel tank (Biorad) until the dye front was at the bottom of the gel. Gels were either stained with coomassie blue (Section 2.18) or transferred to PVDF membrane for western blotting (Section 2.19).

2.18 Coomassie blue staining

SDS-PAGE gels were stained with coomassie blue mainly for analysis of purified proteins. Gels were submerged in 100 ml of staining solution (40% v/v ethanol, 10% v/v glacial acetic acid, 0.1% w/v coomassie blue R-250) in a sealable plastic container and microwaved until the solution boiled. After agitation of the gels for 15 minutes, the stain was removed and the gels were washed twice in ~100 ml deionised water. To destain, gels were submerged in 100 ml of destaining solution (40% v/v ethanol, 10% v/v glacial acetic acid), microwaved until the solution boiled and agitated for 5-10 minutes. The destaining step was repeated, typically two to three times, until stained protein bands were clearly visible on the gel. Stained gels were imaged using a Syngene G:Box ChemiXRG Imaging System.

2.19 Western blotting

Western blotting was used to detect expression of proteins in wild type or transfected mammalian cells. SDS-PAGE gels and blotting papers (Whatman 3MM) were soaked in western blot semi-dry transfer buffer and proteins were transferred to PVDF membrane (Immobilon P 0.45 µM, Merck) using a TransBlot Turbo transfer system (Biorad) running at 25 V for 40 minutes. Membranes were blocked in a blocking buffer, containing either 5% w/v non-fat milk power, 3% BSA or 20% horse serum in TBS with 0.05% v/v Tween-20 (TBST), for 30 minutes at room temperature. All primary antibody hybridisations were overnight at 4°C, whereas secondary or tertiary antibody hybridisations were for 1 hour at room temperature. After each hybridisation, membranes were washed for one hour in TBST, replacing the buffer every 5 minutes with fresh TBST. Membranes were developed by incubation with SignalFire™ ECL reagent (6883P3, CST) for 3 minutes at room temperature and imaged using a Sysgene G:Box ChemiXRG Imaging System.

Blots were stripped if a second protein was to be analysed. Membranes were incubated in 100 ml of stripping buffer (2% w/v SDS, 62.5 mM Tris pH 6.8, 115 mM β-

mercaptoethanol) at 55°C for 10 minutes and then washed for 1 hour with TBST. Following this, membranes were blocked and hybridised with antibodies as described above.

E) *In vivo* RAG assays

2.20 Site-directed mutagenesis

Mutations were introduced into mouse RAG1 or RAG2 cDNA by inverse PCR using NEB Q5® DNA Polymerase. Mutagenic primers were designed using the NEBaseChanger® tool. Reactions were performed as described in Section 2.7 using 5 ng of plasmid DNA as template and 28 cycles of PCR.

Following PCR, 10 µl of each reaction was analysed by agarose gel electrophoresis (Section 2.2) to confirm the presence of amplified plasmid. Next, 0.5 µl of DpnI (R0176S, NEB) was added to 15 µl of PCR product and incubated at 37°C for a minimum of one hour to remove template DNA. Next, 2 µl of DpnI treated PCR product was added to a 10 µl reaction containing 1x T4 Ligase buffer, 0.5 µl T4 DNA Ligase and 0.5 µl T4 PNK to enable ligation of the linear, dephosphorylated PCR product. Reactions were incubated at room temperature for at least 30 minutes before being transformed into chemically competent *E.coli* (Section 2.10). Plasmid DNA from minipreps (Section 2.11) of colonies was analysed by restriction digestion (Section 2.1) and the presence of the intended mutation was confirmed by Sanger sequencing (Eurofins Genomics).

2.21 Generation of RAG1 mutants

The plasmid pCS2MT-RAG1 (Figure 2.1) was used to express RAG1 in mammalian cells. This expression vector contains the cDNA sequence of full-length mouse RAG1 with an N-terminal 6x Myc Tag. Mutations were introduced by Q5 site directed mutagenesis (Section 2.20), using the mutagenic primers listed in Table 2.1.

#	Name	Sequence
1	R401W F	GTCACTGACG**TGG**AGGGCGCAGAAAC
2	R401W R	AGGAGATGCTGGCGAGGC
3	R504Q F	GCATGCTCTTC**A**GAATGCCGAGA
4	R504Q R	AAAGGTTGAAAAATCTGCCTC
5	R401L F	GTCACTGACG**CT**AAGGGCGCAGAAAC
6	R401L R	AGGAGATGCTGGCGAGGC
7	R401K F	TCACTGACGA**A**AAGGGCGCAG
8	R401K R	CAGGAGATGCTGGCGAGGC

Table 2.1 - List of primers used to introduce mutations into pCS2MT-RAG1

Mutagenic primers were designed using the NEBaseChanger® tool. The mutated nucleotides are shown in bold.

2.22 Generation of RAG2 C-terminal mutants

The plasmid pEFXC-RAG2 (Figure 2.2) was used to express RAG2 in mammalian cells. This expression vector contains the cDNA sequence of full-length mouse RAG2 with an N-terminal FLAG Tag. Mutations were introduced by Q5 site directed mutagenesis (Section 2.20), using the mutagenic primers listed in Table 2.2. To generate truncations and internal deletions, mutagenic primers were designed to flank the region to be deleted (Table 2.2, Primers 1-7). For the alanine patch mutants, mutagenic primers were designed to delete the desired 3-4 codons and replace them with alanine codons by adding these sequences to the 5’ end of the primers (Table 2.2, Primers 8-17).

#	Name	Sequence
1	487Δ/413Δ/403Δ/393Δ F	TAATTTAGCAAAAGCCCC
2	487Δ R	TGCTCTTGCTATCTGTAC
3	413Δ R	GGTTACAGACTCGTCATC
4	403Δ R	ATTGTAGGTGTCAAATTCATC

#	Name	Sequence
5	393Δ R	AAAACTGGTTGCTTCAGC
6	383-413Δ F	GGCTACTGGATAACATGTTG
7	383-413Δ R	CTCTGAGTCTTCAAAGGGA
8	383-386A F	gctgccAGTGCTGAAGCAACCAGT
9	383-386A R	agctgcCTCTGAGTCTTCAAAGGGAG
10	387-390A F	gcagcaACCAGTTTTGATGGTGAC
11	387-390A R	agcagcGAAACAAAATTCCTCTGAGTC
12	391-393A F	tgctGATGGTGACGATGAATTTG
13	391-393A R	gcagcTGCTTCAGCACTGAAACA
14	383-390A F	tgccGCAACCAGTTTTGATGGTGACG
15	383-390A R	gctgcGGCAGCAGCTGCCTCTGA
16	387-394A F	gctgccGAATTTGACACCTACAATG
17	387-394A R	agctgcTGCTGCAGCAGCGAAACA
18	W453A F	GGATGGGCAC**GC**GGTACATGCC
19	W453A R	CCATGAGAACAATAGATCATG

Table 2.2 - List of primers used to introduce mutations into pEFXC-RAG2

Mutagenic primers were designed using the NEBaseChanger® tool. Mutated nucleotides are shown in bold and alanine codon sequences are shown in lower case on the 5' end of the primer.

2.23 Cell culture of mammalian cells

NIH3T3, COS7 and HEK293T cells were maintained in 15 - 20 ml complete DMEM in a 75 cm^2 flask and incubated at 37°C in 5% CO_2. At ~90% confluency, approximately every two days, cells were passaged by splitting 1:10. To this end, the DMEM was aspirated and the cells were washed twice with 10 ml PBS, dispensed on the top of the flask to avoid disrupting the cells. Next, 1 ml of Trypsin-EDTA (T3924, Sigma Aldrich) was added and the flask was incubated at room temperature for ~1 minute until the cells detached. To quench the Trypsin-EDTA, 10 ml of complete DMEM was added and cells were pelleted by centrifugation at 300 *g* for 3 minutes. Finally, cells were

resuspended in 10 ml complete DMEM and 1 ml of cells was added to a new 75cm^2, flask containing 15-20 ml complete DMEM.

2.24 Transfection of mammalian cells with polyethylenimine (PEI)

The day before transfection, cells were seeded at 1 - 2 x 10^5 cells per well of a 6-well plate in 2 ml complete DMEM or at 1 - 3.5 x10^6 cells per 10 cm^2 dish in 9 ml complete DMEM. Three hours prior to transfection, the media was replaced with fresh complete DMEM without antibiotic. Transfections were performed using PEI at a 3:1 (DNA:PEI) ratio. Firstly, DNA and PEI were diluted in Opti-MEM (Gibco, 31985062) in separate tubes and mixed well. Next, the diluted PEI was added to the diluted DNA and the sample was immediately vortexed for 10 seconds. Transfection mixes were incubated at room temperature for 15 minutes before being gently added dropwise to each respective well or dish. Transfected cells were typically incubated (Section 2.23) for a further 48 hours before being harvested for analysis or protein purification.

2.25 Hirt extraction of plasmid DNA from transfected mammalian cells

Plasmid DNA was recovered from transfected mammalian cells using a modification of the plasmid DNA extraction method described by Hirt (1967). The culture media was aspirated and cells were washed twice with 2 ml of room temperature PBS. Cells were lysed directly on the plate by the addition of 400 µl Hirt I solution (0.6% SDS, 10 mM Tris-HCl pH 8.0, 1 mM EDTA) to each well, followed by gentle agitation of the plate and incubation at room temperature for 10 minutes. Next, 100 µl Hirt II solution (5 M NaCl, 10 mM Tris-HCl pH 8.0, 1 mM EDTA) was added to each well and mixed. Lysates were transferred to a 1.5 ml micro-centrifuge tube and incubated at 4°C for 16 hours. Samples were centrifuged at 20,000 *g* for 40 minutes at 4°C and the supernatant was transferred to a fresh tube. The DNA was extracted with phenol:chloroform (Section 2.4) until no visible white interface remained, followed by one final extraction with just chloroform to remove residual phenol. To precipitate the plasmid DNA, 200 µl NH_4OAc and 900 µl ethanol was added and the samples were incubated on dry ice for 10 minutes, followed by centrifugation at 20,000 *g* for 10 minutes at 4°C. The supernatant

was removed and the pellets were washed with 1 ml 70% ethanol to remove any precipitated salt. Finally, pellets were air dried and resuspended in 50 µl TE.

2.26 *in vivo* extrachromosomal V(D)J recombination assay

Recombination of the substrate plasmid pJH299 (Figure 2.3) was determined using NIH3T3 cells that were also transfected with RAG expression vectors. NIH3T3 cells were seeded at 2 x 10^5 cells per well in a 6-well plate and transfected with PEI (Section 2.24). For transfection, 4.5 µg PEI was diluted in 100 µl Opti-MEM and added to 1 µg of the recombination substrate pJH299, 160 ng pCS2MT-RAG1, and 320 ng pEFXC-RAG2, also diluted in 100 µl Opti-MEM. After 48 hours plasmid DNA was isolated by Hirt extraction (Section 2.25).

Recombination activity was determined using a nested qPCR approach. First round amplifications were performed using *Taq* DNA polymerase (Section 2.6), with 0.2 µM of each primer DR55 and 1233 (Table 2.3, 1 + 2) and 5 µl Hirt extracted DNA. Thermocycling comprised 19 cycles of 15 seconds at 95°C, 15 seconds at 55°C, and 30 seconds at 68°C. Following the first round of PCR, samples were used immediately for qPCR or stored at 4°C.

qPCR reactions were in a final volume of 10 µl and contained 1x Luna Universal qPCR Master Mix (M3003S, NEB), 0.4 µM of both SJ-F and SJ-R primers (Table 2.3, 3 + 4), 0.1 µM JH299 SJ probe (Table 2.3, 5), and 1 µl first round PCR product. qPCR was performed in a RotorGene 6000 cycler with thermocycling comprised of 40 cycles of 5 seconds at 95°C, and 30 seconds at 60°C. The fluorescent signal from each reaction was acquired at the end of every cycle using the green fluorescence channel and the C_t was determined using the analysis function in the RotorGene 6000 software (version 1.7).

Recombination was normalised to the total amount of JH299 substrate plasmid recovered. This was determined by a second qPCR reaction using primers which

amplify a region of pJH299 which is unaltered by recombination. qPCR reactions were in a final volume of 10 µl and contained 1x SensiFAST SYBR Master Mix (QP100023, Bioline), 0.4 µM of both CAT-F and CAT-R primers (Table 2.3, 6 + 7), and 1 µl of 1:100 dilution of the first round PCR product. Thermocycling comprised of 40 cycles of 5 seconds at 95°C, 10 seconds 63°C and 10 seconds 72°C.

2.27 *in vivo* extrachromosomal reintegration assay

An existing plasmid, pcDNA3.1-12-RSS (Figure 2.4; Kirkham *et al.*, 2019), which contains a consensus 12-RSS was used as the acceptor substrate for the reintegration assay. The ESC donor substrate was generated by digesting an existing pcDNA3.1-ESC (Kirkham *et al.*, 2019) vector with NheI and HindIII to destroy a primer binding site which was also present in pcDNA3.1-12RSS. Complementary oligonucleotides were designed with a random 20 bp sequence with 5' NheI and 3' HindIII overhangs (Table 2.3, 27 + 28). The oligonucleotides were annealed and phosphorylated, and ligated into the digested pcDNA3.1-ESC vector as outlined previously (Section 2.8). This generated pcDNA3.1-ESC-Reint (Figure 2.5).

COS7 cells were seeded at 2 x 10^5 cells per well in a 6-well plate the day before transfection. Transfection were performed using PEI (Section 2.24) and contained 500 ng each of pcDNA3.1-12RSS and pcDNA3.1-ESC-Reint, 500 ng pCS2MT-RAG1 and up to 500 ng pEFXC-RAG2. After 48 hours, plasmid DNA was recovered by Hirt extraction (Section 2.25).

The level of reintegration was determined by probe based qPCR. A typical qPCR reaction contained Luna Universal Probe qPCR Master Mix (M3004, NEB), 1 µl Hirt extracted DNA (around 100 ng), 4 pmol Reint F and Reint R primers (Table 2.3, 29 + 30) and 10 pmol of Reint probe (Table 2.3, 31), in a total reaction volume of 10 µl. qPCR was carried out using a Rotor Gene 6000 cycler and 40 cycles of 5 seconds at 95°C and 30 seconds at 60°C. Reintegration activity was normalised to the total amount of recovered 12-RSS substrate using primers Neo-F and Neo-R (Table 2.3, 32

+ 33) which amplify a region of pcDNA3.1-12RSS which remains unaltered by reintegration. Reactions contained 1x SensiFAST SYBR Master Mix (QP100023, Bioline), 1 ul 1:500 dilution Hirt extracted DNA and 4 pmol of each primer. qPCR reactions were carried out as above but with 40 cycles of 5 seconds at 95°C, 10 seconds at 60°C and 10 seconds at 72°C.

2.28 Analysis of hybrid joint formation

The ability of RAG2 mutants to catalyse the formation of hybrid joints was analysed using the pJH299 plasmid and a semi-quantitative PCR approach. Samples generated for the analysis of recombination activity (Section 2.26) were used as template in the hybrid joint PCR. During hybrid joint formation on pJH299, the 23-RSS becomes aberrantly joined to the 12-RSS coding segment and the intervening DNA, containing the 12-RSS, is excised. Reactions were 10 µl and were performed using Q5 DNA polymerase (Section 2.7), using primers DR99 and DR100 (Table 2.3, 34 + 35) and ~300 ng Hirt extracted DNA as template. The amount of template DNA used in each reaction was normalised before the hybrid join PCR by qPCR using primers CAT-F and CAT-R (Table 2.3, 6 + 7), as described in Section 2.26. Cycling comprised 32 cycles of 98°C for 5 seconds, 67°C for 10 seconds and 72°C for 10 seconds. Reactions were analysed by 2% agarose gel electrophoresis (Section 2.2) and stained using ethidium bromide. The primers amplify a 478 bp region on pJH299 before hybrid joint formation; if a hybrid joint has formed the PCR product becomes 192 bp. The levels of hybrid joint formation were quantified from the gels by 2D densitometry using ImageJ.

2.29 Analysis of RAG1 protein expression in transfected NIH3T3 cells

Protein expression levels of the different Myc-tagged RAG1 mutants in transfected NIH3T3 cell lysates was analysed and compared with that of wild type RAG1. Transfections were performed in a similar way to the extrachromosomal V(D)J recombination assay (Section 2.16) except that they were scaled up 5-fold to aid the detection of RAG1. 0.5 µg pTkβ-galactosidase (Clontech) was co-transfected and

western blotted as a loading control. After 48 hours, protein extracts were made (Section 2.16) and 20 µl of each sample was separated by SDS-PAGE (Section 2.17).

Western blotting was performed as described in Section 2.19, using a three-layer approach to increase sensitivity. Membranes were blocked with 5% w/v non-fat milk powder in TBST, which was also used to dilute antibodies. Primary hybridisation was with a 1:2000 dilution of goat anti-c-Myc antibody (9E10, Abcam), secondary hybridisation was with a 1:10,000 dilution of a rabbit anti-goat antibody (5210-0170, Sera Care), and tertiary hybridisation was with a 1:20,000 dilution of HRP-conjugated anti-rabbit antibody (A120-101P, Bethyl). Membranes re-probed for β-galactosidase using a 1:5000 dilution of a mouse anti-β-galactosidase antibody (Z3781, Promega).

2.30 Analysis of RAG2 protein expression in transfected COS7 cells

COS7 cells were transfected as described in Section 2.27 but also with 100 ng pTkβ-galactosidase (Clontech) for use as a transfection and loading control. Forty eight hours post transfection, whole cell protein extracts were made (Section 2.16) and separated by 10% SDS-PAGE (Section 2.17). Proteins were transferred to PVDF membrane and western blotted as described in Section 2.19. Membranes were blocked in 20% horse serum (16050122, ThermoFisher) in TBST and blots were incubated with a 1:2000 dilution of anti-FLAG M2 antibody (Sigma, F3165). β-galactosidase was western blotted as described in Section 2.29.

F) *In vitro* RAG assays

2.31 Gel purification of DNA oligonucleotides

Synthetic DNA oligonucleotides for *in vitro* RAG binding and cleavage assays were supplied by Sigma-Aldrich, purified by desalting. Oligonucleotides were further purified by separation on denaturing PAGE gels, filled by extraction using the “crush-and-soak” method (Ellington and Pollard, 2001). Briefly, lyophilised oligonucleotides were

resuspended in 1x formamide loading buffer (50% v/v formamide, 1 mM EDTA, 0.01% w/v Bromophenol blue, 0.01% w/v Xylene cyanol), boiled for 3 minutes and separated by size on a 12% (19:1) polyacrylamide gel containing 7 M urea and 1x TBE that was run at 500 V for 1 hour. After electrophoresis, one lane was excised and stained with ethidium bromide for 10 minutes and the location of the oligonucleotide was visualised using UV light and measured with a ruler. This measurement was used to locate the oligonucleotide on the unstained gel which was excised using a sterile scalpel blade. The gel slice was finely diced and transferred to a 1.5 ml micro-centrifuge tube containing 1 ml elution buffer (10 mM Tris-HCl pH 8.0, 100 mM NaOAc, 0.5% SDS, 1 mM EDTA). Tubes were boiled for 3 minutes, freeze-thawed 3 times and placed on a rotating wheel overnight at room temperature. The next day, gel pieces were removed by centrifugation at 20,000 *g* for 2 minutes and the supernatant was passed through a 0.2 µM syringe filter to remove residual polyacrylamide. The volume of oligonucleotide solution was reduced to 400 µl by repeated butanol extraction, followed by phenol:chlorofrom extraction (Section 2.4) and ethanol precipitation (Section 2.5).

2.32 Annealing of oligonucleotides

Unlabelled 12-RSS or 23-RSS oligonucleotides were generated by annealing gel-purified complementary single stranded oligonucleotides (DAR39/DAR40 for the 12-RSS, DG61/DG62 for the 23-RSS; Table 2.3, 36-39). Annealing reactions contained 100 µm of each strand, 100 mM NaCl, and 0.5x TE in a total volume of 100 ul. Reactions were boiled for 3 minutes and left to cool gradually overnight to allow annealing of the strands. Annealed 12-RSS and 23-RSS oligonucleotides were stored at 4°C until use.

2.33 Preparation of radio-labelled RSSs

Radioactively labelled RSSs were used in RAG binding assays since they have higher sensitivity compared with fluorescently labelled RSSs. The “top” strand of the 12-RSS (DAR39; Table 2.3, 36) or 23-RSS (DG61; Table 2.3, 38) was 5’-end labelled with [γ-^{32}P]-ATP using T4 polynucleotide kinase. Reactions were in a final volume of 15 µl

and contained 10 pmol oligonucleotide, 10 units T4 PNK, 1x T4 PNK buffer (70 mM Tris-HCl pH 7.6, 10 mM $MgCl_2$, 10 mM DTT), and 2.5 µl [γ-^{32}P]-ATP and were incubated at 37°C for 30 minutes. The volume was increased to 50 µl with TE and unincorporated [γ-^{32}P]-ATP was removed using a MicroSpin™ G-25 column (GE Healthcare) for 2 min at 700 *g*. The purified radio-labelled oligonucleotide was annealed to the complementary unlabelled reverse strand as described in Section 2.32.

2.34 Preparation of fluorescently labelled RSSs

Fluorescently labelled 12-RSS and 23-RSSs were generated by PCR with Herculase II Fusion DNA polymerase (600675, Agilent). PCR reactions contained either 1 ng of annealed 12-RSS or 23-RSS oligonucleotides as template, 25 pmol of each primer, Alexa Fluor 488-labelled RSS-F and unlabelled RSS-R (Table 2.3, 25 + 40), 250 µM dNTPs and Herculase II Fusion DNA polymerase. PCR comprised 35 cycles of 10 seconds at 95°C, 15 seconds at 60°C and 15 seconds 72°C. Typically five 50 µl reactions were performed which would generate enough fluorescently labelled 12-RSS or 23-RSS for >100 standard RAG cleavage reactions. Reactions were separated on a 15% (19:1) polyacrylamide gel containing 1x TBE, by running at 170 V for 2 hours. A blue LED torch (UltraFire H-B3 470 nm) was used to locate the fluorescent RSS in the gel which was excised with a sterile surgical scalpel blade. DNA was recovered from the gel by electroelution. Gel slices were placed in dialysis tubing with 1 ml 1x TBE and electrophoresed at 80 V for 2 hours. The polarity for reversed for 1 minute and the DNA was recovered from the buffer by ethanol precipitation (Section 2.5). The pellet was resuspended in 10 mM Tris-HCl pH 7.5 and stored at 4°C.

2.35 Purification of RAG proteins from HEK293T cells

HEK293T cells were co-transfected with pEF-McR1 (wild type or mutant; Figure 2.6) and pEF-McR2 (Figure 2.7) using PEI (Section 2.24). These expression vectors contain the cDNA sequences for mouse core RAG1 (aa 384-1008) or mouse core RAG2 (aa 1-382) fused to an N-terminal maltose binding protein (MBP) amylose affinity tag. The day before transfection, 10 cm^2 dishes (typically 15 per purification) were

seeded with 2 x 10^6 HEK293T cells/dish in complete DMEM and the media was replaced with 9 ml DMEM without antibiotics 3 hours prior to transfection. For transfection, 30 µg PEI was diluted in 500 µl serum free DMEM and this was added to a mix of 5 µg pEF-McR1 and 5 µg pEF-McR2, also diluted in 500 µl serum free DMEM. The PEI:DNA mix was added to the cells after incubation at room temperature for 15 minutes.

After 48 hours, the media was aspirated and cells from each 10 cm^2 dish were harvested by scraping into 10 ml of ice cold PBS. Cells were pooled in 50 ml falcon tubes and pelleted by centrifugation at 800 *g* for 3 minutes at 4°C. Pellets were washed twice by resuspension in 50 ml ice cold PBS, followed by centrifugation. Pellets were either used immediately for purification or snap frozen on dry ice and stored at -80°C until use. All steps during purification were performed on ice or at 4°C. Cells were resuspended in 10 ml of buffer A (0.5 M NaCl, 10 mM NaPi pH 7.2, 0.25% v/v Tween-20, 1 mM DTT) supplemented with 1 mM benzamidine-HCl, 1 mM PMSF, and 1 x Pierce™ Protease Inhibitor (Thermo, A32953), and lysed by 20 strokes of a Dounce homogeniser using the tight pestle. The lysates were clarified by centrifugation at 11,000 g (Beckman SW55Ti rotor, 30,000 rpm) for 40 minutes at 4°C.

The supernatant was batch bound for 30 minutes to 1 ml of amylose resin (NEB, E8021), pre-equilibrated in buffer A, and applied to a 10 ml gravity flow column. The resin was washed with 10 column volumes (CVs) of buffer A, followed by 10 CVs of buffer B (0.5 M NaCl, 10 mM NaPi pH 7.2, 1 mM DTT). MBP-RAG proteins were eluted with 10 CVs of buffer C (0.5 M NaCl, 10 mM NaPi pH 7.2, 1 mM DTT, 10 mM maltose), collecting 1 ml fractions in 1.5 ml siliconised micro centrifuge tubes. Fractions were analysed by SDS-PAGE and those containing MBP-RAG proteins were pooled and dialysed twice for 3 hours against dialysis buffer (25 mM Tris-HCl pH 8.0, 150 mM KCl,

10% v/v glycerol, 1 mM DTT). Dialysed MBP-RAG proteins were aliquoted, snap frozen on dry ice and stored at -80°C until use.

Purification of full-length RAG2 plus core RAG1 was performed in a similar manner. HEK293T were transfected as described above with pEF-McR1 and pEF-MfR2 (Figure 2.8). Cells were resuspended in 1.5 ml ice cold buffer RSB (10 mM Tris-HCl pH 7.5, 10 mM NaCl, 0.5% v/v IGEPAL CA-630, 5 mM $MgCl_2$, 1 mM PMSF) and allowed to swell on ice for 5 minutes. Next, 2.2 ml buffer LSB (20 mM Tris-HCl pH 7.5, 1 M NaCl, 0.2% v/v IGEPAL CA-630, 0.2 mM $MgCl_2$, 1 mM PMSF) was added and lysates were rocked for one hour at 4°C. The remainder of the purification was performed as described above for core RAG proteins.

2.36 Purification of HMGB1 from bacterial cells

The full-length rat HMGB1 expression plasmid, pETM11-HMGB1 (Figure 2.9) was a kind gift from Professor Marco Bianchi (Vita-Salute San Raffaele University, Milan, Italy). This expression vector contains the cDNA sequence of full-length rat HMGB1 with a 6xHis tag fused to the N-terminus, separated by a Tobacco Etch Virus (TEV) cleavage site to allows removal of the 6xHis tag from HMGB1 using TEV protease. BL21(DE3) cells were transformed with pETM11-HMGB1 as described in Section 2.10, plated on LB agar plates with 50 µg/ml kanamycin and incubated at 37°C overnight. A single colony was used to inoculate 10 ml of LB media with 50 µg/ml kanamycin (LB/kanamycin) and this was incubated at 37°C with shaking (200 rpm) overnight. The next day, the 10 ml culture was transferred to 1L of (LB/kanamycin) in a 2 litre flask and was incubated at 37 with shaking (200 rpm) until an OD_{600} of 0.6 - 0.8 was reached. Expression of 6xHis-HMGB1 was induced by the addition of 0.5 mM IPTG and the culture was incubated at 25°C with shaking (200 rpm) overnight.

Cells were harvested by centrifugation at 4000 *g* for 20 minutes at 4°C and the pellet was washed once with ice cold PBS. The pellet was resuspended in 20 ml lysis buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 10 mM imidazole, 2 mM β-mercaptoethanol, 0.2% v/v IGEPAL CA-630) supplemented with 1 mM benzamidine-HCl, 1 mM PMSF, and 1 x Pierce™ Protease Inhibitor (Thermo, A32953). Cells were lysed by 10 cycles of sonication on ice which comprised 10 seconds on and 10 seconds off, at 10 microns. The lysate was cleared by centrifugation at 16,000 *g* for 30 minutes at 4°C and filtered through a 0.2 µM syringe filter.

The cleared lysate was batch bound to 5 ml Ni-NTA resin (30210, Qiagen) on a rotating wheel for 30 minutes and then applied to a gravity flow chromatography column. The flow through was collected and batch bound again to 2 ml Ni-NTA resin which was applied to a second column. Columns were washed with 10 CVs lysis buffer containing protease inhibitors, 5 CVs of lysis buffer without protease inhibitors, 5 CVs wash buffer 1 (20 mM Tris-HCl pH 8.0, 1 M NaCl, 10 mM imidazole, 2 mM β-mercaptoethanol), and 5 CVs wash buffer 2 (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 30 mM imidazole, 2 mM β-mercaptoethanol). 6xHis-HMGB1 was eluted with elution buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 300 mM imidazole, 2 mM β-mercaptoethanol) and fractions were analysed by SDS-PAGE (Section 2.17) and coomassie staining (Section 2.18),; those containing 6xHis-HMGB1 were pooled and the concentration determined by Bradford assay.

To remove the 6xHis tag, TEV protease was added at a ratio 1:50 (TEV:6xHis-HMGB1) and the mixture was dialysed overnight against dialysis buffer (20 mM Tris-HCl pH 8.0, 2 mM β-mercaptoethanol, 20 mM NaCl). This step was omitted for the purified 6xHis-HMGB1 used in supershift experiments which required the 6xHis tag. Uncleaved 6xHis-HMGB1 and TEV protease, which also contains a His tag, was removed from the

mix by batch binding to 3 ml Ni-NTA for 30 minutes and the supernatant, containing cleaved HMGB1, was retained.

HMGB1 was further purified by cation exchange chromatography. HMGB1 was loaded onto a 1 ml HiTrap SP XL column (GE Healthcare), pre-equilibrated with buffer A (20 mM Tris-HCl pH 8.0, 80 mM NaCl, 1 mM DTT), and washed with 9 CVs of buffer A. HMGB1 was eluted with a continuous gradient of 80 - 650 mM NaCl (20 mM Tris-HCl pH 8.0, 1 mM DTT) over 25 CVs, collecting 1 ml fractions. Fractions were analysed by SDS-PAGE and those containing HMGB1 were pooled and dialysed twice for 3 hours against storage buffer (25 mM Tris-HCl pH 8.0, 150 mM KCl, 10% v/v glycerol, 2 mM DTT). Aliquots were snap frozen on dry ice and stored at -80°C until use.

2.37 *In vitro* RAG cleavage assay

RAG cleavage assays were performed using fluorescently labelled RSS probes (Section 2.34). Reactions were in a final volume of 10 µl and contained 25 mM MOPS pH 7.0, 50 mM potassium acetate, 1 mM DTT, 100 µg/ml BSA, 1 mM $MgCl_2$ (1 mM $MnCl_2$ for Figure 3.12B), 8 nM fluorescently labelled oligonucleotide substrate, 20 nM unlabelled partner oligonucleotide (except Figure 3.12B), 100 ng (400 nM) HMGB1 and 150 ng of purified wild type or mutant RAG proteins. Where indicated, HMGB1 was omitted. Initial reactions (Figure 3.9) were for 10 minutes in the presence of 20% DMSO, which enhances RAG cleavage. Subsequent cleavage reactions were in the absence of DMSO for 60 minutes to accurately determine the effects of HMGB1 (Bergeron *et al.*, 2006).

Reactions were incubated at 37°C before stop buffer was added to give final concentrations: 50 mM Tris-HCl pH 8.0, 0.1% w/v SDS, 5 mM EDTA, 0.175 mg/ml proteinase K, 1 x DNA loading buffer, followed by incubation at 37°C for 30 minutes. The reaction was loaded onto a 15% polyacrylamide gel in 1x TAE and separated for 2 hours at 170V. Gels were visualised using a FLA-5100 imager (Fujifilm, Tokyo, Japan).

2.38 *In vitro* RAG binding assay

RAG-RSS binding assays were performed using the same conditions as for RAG cleavage (Section 2.37), except the divalent cation was Ca^{2+} and radiolabelled RSSs were used to increase sensitivity. For the 6xHis-HMGB1 super-shift assay, 100 ng of 6xHis-HMGB1 and 0.5 µl of mouse anti-6x-His monoclonal antibody (STJ96906; St John's Laboratory) were used. Binding reactions were loaded onto a 4% (19:1) polyacrylamide gel containing 0.5x TBE, pre-chilled to 4°C in a cold room. Gels were run at 170V for 2 hours (except for gels in Figure 3.13A which were run for 3 hours) and dried on to Whatman 3MM chromatography paper using a Bio-Rad gel drier with an 80°C heated lid for 1 hour. The dried gel was exposed to a phosphor imager screen and visualised using an FLA-5100 imager (Fujifilm).

G) Generation of HMGB1 deficient NIH3T3 cells

2.39 Design and cloning of sgRNAs

Guide RNAs (sgRNAs) to target Cas9 to the murine *Hmgb1* locus for knockout (Section 2.42) or knockdown (Section 2.44) experiments were designed using the MIT CRISPR design tool (http://crispr.mit.edu/) and modified to include complementary overhangs to those produced by digestion with BsmBI. Complementary single stranded oligonucleotides were synthesised by Integrated DNA Technologies and purified by desalting (Table 2.4). Oligonucleotides were simultaneously phosphorylated and annealed in a 10 µl reaction containing 100 pmol of each oligonucleotide, 1x T4 Ligase Buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 1 mM ATP, 10 mM DTT) and 5 units T4 PNK. Reactions were incubated in a thermocycler at 37°C for 30 minutes before being heated to 95°C for 5 minutes and then slowly cooled to 25°C at 5°C/minute. The annealed oligonucleotides were diluted 1:200 in sterile water and 1 µl was used in a ligation reaction (Section 2.8), with 50 ng of either lenti-CRISPR-Puro V2 (Figure 2.10) or lenti-sgRNA-Zeo (Figure 2.11) which had been digested with BsmBI and gel purified.

Ligation reactions were transformed into Endura *E.coli* cells (Section 2.10) and the presence of the sgRNA in the plasmid was confirmed by Sanger sequencing (Eurofins).

2.40 Production of lentiviruses

Lentiviruses were produced by transfection of HEK293T cells with lentiviral packaging plasmids using PEI (Section 2.24). The day before transfection, HEK293T cells were seeded at 3.5 x 10^6 cells per 10 cm^2 dish in 9 ml complete DMEM. Three hours prior to transfection, the medium was replaced with 9 ml fresh complete DMEM without antibiotics. For transfection, 30 µg PEI was diluted in 500 µl serum free DMEM and added to a mix of 4 µg of packaging vector pCMVR8.74, a gift from Didier Trono (Addgene plasmid #22036), 2 µg of the vector that expresses the coat protein, pMD2.G, a gift from Didier Trono (Addgene plasmid #12259) and 4 µg of transfer vector (containing sgRNA sequence, or cDNA of protein to be expressed), also diluted in 500 µl serum free DMEM. Transfection mixes were incubated at room temperature for 15 minutes and added to each respective 10 cm^2 dish of cells. The medium was changed 16 hours post transfection and the supernatant containing lentiviral particles was harvested at 48 hours and 72 hours and used immediately for transduction.

2.41 Lentiviral transduction of NIH3T3 cells

NIH3T3 cells were seeded at 1 x 10^5 cells per well of a 6-well plate. After 24 hours, the medium was replaced with 1 mL of each lentiviral supernatant containing 6 µg/mL polybrene (TR-1003- G, Merck Kenilworth). Plates were centrifuged at 800 *g* for 30 minutes at 32°C. At 24 hours post transduction, the medium was replaced with media containing the appropriate antibiotic for selection, such as: 5 µg/ml puromycin dihydrochloride (sc-108071B, Santa Cruz Biotechnology), 400 mg/mL zeocin (J67140.XF, Alfa Aesar), or 10 mg/mL blasticidin S HCl (sc-495389, Santa Cruz Biotechnology).

2.42 Deletion of *Hmgb1* exon 2 in NIH3T3 cells using CRISPR-Cas9

Two sgRNA sequences which target Cas9 to regions upstream (Table 2.4, 1 + 2) and downstream (Table 2.4, 3 + 4) of mouse *Hmgb1* exon 2 were cloned into lenti-CRISPR-Puro V2 and lenti-sgRNA-Zeo respectively (Section 2.39). Lentiviruses were produced using these transfer vectors (Section 2.40) and were used to transduce NIH3T3 cells (Section 2.41). 24 hours post-transduction, the medium was replaced with complete DMEM containing 5 µg/ml puromycin dihydrochloride and 400 mg/mL zeocin to select cells which had been successfully transduced with both lentiviruses.

Following antibiotic selection, genomic DNA was extracted from wild type NIH3T3 cells or the pool of transduced cells (Section 2.13), and this was used as template for PCR to detect the presence of a deletion of *Hmgb1* exon 2. PCR reactions were performed using Q5 polymerase (Section 2.7) with 0.5 µM of both HMGB1 Del F and R primers (Table 2.3, 8 + 9) and 200 ng of genomic DNA as template. Cycling comprised 30 cycles of 98°C for 10 seconds, 71°C for 20 seconds and 72°C for 30 seconds. The resulting PCR products were separated on a 2% agarose gel (Section 2.2) to check for the presence of a band corresponding to a deletion of *Hmgb1* exon 2 (468 bp).

HMGB1 protein expression in the pool of *Hmgb1* exon 2 knockout cells was determined by western blotting and compared with wild type NIH3T3 cells. Protein sample preparation (Section 2.16) and western blotting (Section 2.19) was performed as described previously, using a 15% SDS-PAGE gel (Section 2.17). Blots were incubated with a 1:1000 dilution of mouse monoclonal anti-HMGB1 antibody (sc-56698, Santa Cruz Biotechnology) in blocking buffer overnight at 4°C. After washing, blots were incubated with a 1:20,000 dilution of a HRP-conjugated anti-mouse antibody in blocking buffer at room temperature for 1 hour.

The pool of *Hmgb1* exon 2 deletion cells was seeded at 1 cell per well in a 96 well plate by flow cytometry (BD, FACSMelody). Cells were maintained in 1 ml complete DMEM, cultured for 3 weeks, and transferred to a well of a 6 well plate and cultured until ~90%

confluent before half of the cells were harvested to prepare genomic DNA (Section 2.13). The copy number of *Hmgb1* was determined by SYBR green qPCR (Section 2.26) using primers which are complementary to a region of *Hmgb1* exon 2 which would be deleted if the knockout was successful (Table 2.3, 10 and 11). *Hmgb1* exon 2 copy number was compared to that of *Hprt*, which was also determined by qPCR using primers HPRT F and R (Table 2.3, 12 and 13).

2.43 Cloning lenti dCas9-KRAB-T2A-BLAST

To generate lenti-dCas9-KRAB-T2A-BLAST, first VP64 was deleted from the plasmid lenti-dCAS-VP64_Blast (Addgene Cambridge, MA, USA; 61425, a gift from Feng Zhang) via Q5 PCR (Section 2.7) using primers VP64 Del F and R primers (Table 2.3, 14 and 15), to generate lenti-dCas9-T2A-BLAST. Cycling comprised 30 cycles of 98°C for 15 seconds, 60°C for 30 seconds and 72°C for 5 minutes 30 seconds.

The cDNA sequence encoding the KRAB domain from human ZNF10 (aa 2-97) was amplified from HEK293T cell cDNA (Section 2.14 and 2.15) using Q5 DNA polymerase. Reactions were the same as above but used primers hZNF10 F and R (Table 2.3, 16 and 17) and 5 ng HEK293T cell cDNA. The PCR product was gel purified (Section 2.3) and ligated into the lenti-dCas9-T2A-BLAST vector (Section 2.9), between the dCas9 and T2A-BLAST sequences to generate lenti-dCas9-KRAB-T2A-BLAST (Figure 2.12), which was verified by sequencing.

2.44 Knockdown of *Hmgb1* in NIH3T3 cells using CRISPRi

Three sgRNA sequences were designed to target dCas9-KRAB to the start of the *Hmgb1* gene were designed using the MIT CRISPR design tool (http://crispr.mit.edu/; Table 2.4, 5-10). These were cloned into lenti-sgRNA-Zeo as described in Section 2.39. Separate lentiviruses for each sgRNA and lenti-dCas9-KRAB-T2A-BLAST were produced as described in Section 2.40 and were used to transduce NIH3T3 cells as described in Section 2.41. At 24 hours post transduction, the media was replaced with media containing 10 mg/mL blasticidin S HCl (sc-495389, Santa Cruz Biotechnology) and 400 mg/mL zeocin (J67140.XF, Alfa Aesar). Knockdown of HMGB1 protein

expression was confirmed by western blotting using an anti-HMGB1 antibody as described in Section 2.42. Cells transduced with the TSS1 sgRNA (Table 2.4, 5 + 6; HMGB1 KD cells) gave the highest knockdown of HMGB1 protein expression and were used the bulk of experiments shown.

The recombination assay was performed using the HMGB1 KD cells as described in Section 2.26. HMGB1 KD cells were maintained in DMEM, supplemented with 10 mg/mL blasticidin S HCl and 400 mg/mL zeocin to avoid silencing of the integrated lentivirus. The media was replaced with complete DMEM containing no antibiotics three hours prior to transfection since excessive cell death was observed if the antibiotics were present during transfection.

2.45 Cloning HMGB1 and HMGB2 expression vectors

To generate the lenti-P2A-Puro lentiviral vector, firstly, a fragment containing the U6 promoter and gRNA scaffold was deleted from lentiCRISPR v2 (Addgene 52961, a gift from Feng Zhang) by digestion with KpnI (R3142S, New England Biolabs) and EcoRI (R3101S, New England Biolabs) and the overhangs removed using T4 DNA polymerase (M0203S, New England Biolabs). Next, the Cas9 sequence was removed via Q5 PCR using primers P2A F and EF1α R (Table 2.3, 18 + 19). Rat HMGB1ΔC cDNA was amplified from pETM11-HMGB1 (Figure 2.3) also using Q5 DNA polymerase, and primers rHMGB1 F and rHMGB1ΔC R (Table 2.3, 20 + 21); this was ligated into the lenti-P2A-Puro vector to generate lenti-HMGB1ΔC-P2A-Puro (Figure 2.13) and its sequence verified.

The HMGB2ΔC-P2A-Puro and HMGB2-P2A-Puro lentiviral vectors were generated in a similar way. First, the mouse HMGB2 cDNA sequence was amplified from NIH3T3 cDNA (Section 2.14 and 2.15) by Q5 PCR with primers mHMGB2 F and mHMGB2ΔC R (HMGB2ΔC; Table 2.3, 22 + 23) or mHMGB2 F and mHMGB2 R (HMGB2; Table 2.3, 22 + 24); these were also ligated into the lenti-P2A-Puro vector to generate lenti-

HMGB2ΔC-P2A-Puro (Figure 2.14) or lenti-HMGB2-P2A-Puro (Figure 2.15) and their sequences verified.

2.46 Analysis of HMGB1 and HMGB2 protein expression

Lentiviral particles were produced as described previously (Section 2.40) and HMGB1 KD cells were transduced with either the HMGB1ΔC-P2A-Puro, HMGB2-P2A-Puro or HMGB2ΔC-P2A-Puro lentivirus (Section 2.41). 24 hours post transduction, the mediaum was replaced with complete DMEM containing 5 µg/ml puromycin dihydrochloride (sc-108071B, Santa Cruz Biotechnology). The presence of HMGB1ΔC-P2A, HMGB2-P2A-Puro or HMGB2ΔC-P2A was confirmed by western blotting (Section 2.19) using either a mouse monoclonal anti-HMGB1 antibody (sc-56698, Santa Cruz Biotechnology) or a rabbit monoclonal anti-HMGB2 antibody (14163T, CST, Danvers, MA, USA).

H) Generation and analysis of nucleosome RSS substrates

2.47 Cloning of pUC19-601-7-12/23-RSS

A gene block containing a Widom 601 sequence upstream of a 23-RSS sequence was synthesised by Integrated DNA Technologies (IDT) and ligated into pUC19 which had been digested with SmaI (R0141S, NEB), to generate pUC19-601-23-RSS. Site-directed mutagenesis (Section 2.20) was used to mutate the 23-RSS to a 12-RSS using primers 601 12-RSS F and R (Table 2.3, 41 + 42) and a 7 bp linker sequence between the 601 sequence and the RSS was added in a similar manner using primers 601 Linker F and R (Table 2.3, 43 + 44). This generated pUC19-601-7-12-RSS and pUC19-601-7-23-RSS (Figure 2.16 and Figure 2.17).

2.48 Large scale production of 601-7-12/23-RSS fragments

Large amounts of unlabelled 601-7-12-RSS or 601-7-23-RSS were generated by large scale plasmid preparation using a method adapted from Dyer *et al.*, 2004. A single colony was used to inoculate 5 ml of terrific broth (TB), supplemented with 100 µg/ml

ampicillin, and incubated at 37°C for 6 hours with shaking (200 rpm). Next, the pre-culture was split between two 2L flasks containing 500 ml TB with 100 µg/ml ampicillin and incubated overnight at 37°C with shaking (200 rpm). Cells were harvested by centrifugation at 4000 *g* for 20 minutes at 4°C and the pellet was resuspended in 60 ml alkaline lysis solution I. Following this, 120 ml of alkaline lysis solution II was added, gently mixed by inversion and incubated on ice for 10 minutes. Next, 90 ml of alkaline lysis solution III was added, mixed by shaking and incubated on ice for 20 minutes. Lysates were cleared by centrifuged at 10,000 *g* for 20 minutes at 4°C and then filtered through Whatman 3MM filter paper. Next, 0.52 volumes of isopropanol was added and incubated at room temperature for 15 minutes before centrifugation at 10,000 *g* for 30 minutes at room temperature. Pellets were washed with 70% ethanol before being resuspended in 2 ml TE. Fifty microlitres of RNase A (10 mg/ml) was then added and the nucleic acids were incubated at 37°C overnight.

The suspension was extracted three times with an equal volume of phenol and once with an equal volume of chloroform. PEG precipitation of DNA was used to separate digested RNA from plasmid DNA by adding, 5 M NaCl and 40% w/v PEG-8000 to the suspension to give a final concentrations of 0.5 M NaCl and 10% w/v PEG-8000. The sample was agitated at 37°C for 5 minutes and incubated on ice for 30 minutes before centrifugation at 3000 *g* for 20 minutes at 4°C. The supernatant, which contains digested RNA, was discarded and the DNA pellet was resuspended in 2 ml TE. PEG was removed by two extractions with an equal volume of chloroform and plasmid DNA was ethanol precipitated (Section 2.5). The pellet was resuspended at a concentration of 1.5 mg/ml in TE. Typically 25 mg of plasmid DNA was generated from 1L of culture.

EcoRV restriction sites flank the 601-7-12-RSS and 601-7-23-RSS fragments to allow their excision from the pUC19 backbone by digestion with EcoRV. Reactions were performed with plasmid DNA at a concentration of 1 mg/ml diluted in 1x CutSmart® Buffer (50 mM KoAc, 20 mM Tris-acetate pH 7.9, 10 mM magnesium acetate, 100 µ/ml BSA; NEB), with 30 units of EcoRV-HF (R3195S, NEB) per nanomole of EcoRV site.

Reactions where incubated for 16 hours at 37°C before being analysed for digestion efficiency by 2% agarose gel electrophoresis (Section 2.2). If digestion was incomplete an additional 50% EcoRV-HF was added and incubated again at 37°C for 16 hours. The enzyme was removed by two rounds phenol:chloroform extraction (Section 2.4) and the DNA was ethanol precipitated (Section 2.5) and resuspended at a concentration of 1 mg/ml in TE.

The excised EcoRV fragment was separated from the pUC19 backbone by PEG precipitation of DNA. This was achieved by the addition of 5 M NaCl and 40% w/v PEG-8000 to the digested DNA to final concentrations of 0.5 M NaCl and 7% w/v PEG. The mix was incubated on ice for 1 hour before centrifugation at 20,000 *g* for 20 minutes at 4°C. A small amount of sample from the resulting supernatant and resuspended pellet was analysed by 2% agarose gel electrophoresis (Section 2.2) to check complete separation of the fragment from DNA backbone. PEG was removed from the supernatant, which contains the excised fragment, by two extractions with an equal volume of chloroform and the DNA was ethanol precipitated. Finally, pellets were washed in 70% ethanol and resuspended in an appropriate amount of TE and the concentration was adjusted to 1 mg/ml. Approximately 2 mg of purified 601-7-12RSS or 601-7-23-RSS per litre of culture were generated using this method.

2.49 Production of fluorescently labelled 601-7-12/23-RSS fragments

Fluorescently labelled 601-7-12-RSS and 601-7-23-RSS fragments were generated by PCR using a forward primer with a 5' AlexaFluor 488 dye molecule (Table 2.3, 25). PCR reactions (50 µl) were performed using an in house produced Pfu-Sso7d polymerase (Wang *et al.*, 2004) and contained 1x HF Pfu Buffer (150 mM Tris-HCl pH 10, 50 mM KCl, 50 mM NH_4OAc, 10 mM $MgCl_2$, 0.5% Triton X-100, 0.5 mg/ml BSA), 200 µM dNTPs, 0.5 µM of each 488 Label F and 601 Label R (Table 2.3, 25 + 26), 1 ng pUC19-601-12-RSS or pUC19-601-7-23-RSS as template, and 1 unit Pfu-Sso7d polymerase. Cycling consisted of 35 cycles of 98°C for 5 seconds, 44°C for 15 seconds and 72°C for 15 seconds. Fluorescent 601-7-12-RSS and 601-7-23-RSS fragments

were purified by PAGE as described in Section 2.34. Typically 1 µg of purified fluorescent fragment could be generated from a single 50 µl PCR.

2.50 Purification of histone proteins from *E.coli*

Xenopus laevis histones were expressed and purified from *E.coli* as described by Dyer *et al.*, 2004. Chemically competent BL21(DE3) cells were transformed with pET-histone expression vectors for each histone; H2A, H2B, H3 or H4. Four preculture tubes containing 4 ml of 2x YT media, containing 100 µg/ml ampicillin and 0.1% glucose, were inoculated with a single colony and incubated at 37°C. After three hours, all four precultures were transferred to 100 ml 2x YT with 100 µg/ml ampicillin and 0.1% glucose and incubated at 37°C until an OD_{600} of 0.4 was reached. The preculture was divided equally between six 2L flasks containing 1L 2x YT media with 100 µg/ml ampicillin and 0.1% glucose and incubated again at 37°C until an OD_{600} of 0.4 was reached. Expression was induced by the addition of IPTG to a final concentration of 0.2 mM and incubated at 37°C for a further 2 hours. Cells were harvested by centrifugation at 4000 *g* for 20 minutes at room temperature, pellets were resuspended in 35 ml of wash buffer (50 mM Tris-HCl pH 7.5, 100 mM NaCl, 1 mM benzamidine, 1 mM β-mercaptoethanol) and flash frozen in liquid nitrogen.

Inclusion bodies containing recombinant histones were prepared from the resuspended cells. The cell suspension was thawed in a 37°C water bath, lysozyme (L6876-1G, Sigma) was added to a final concentration of 1 mg/ml and incubated on ice for 30 minutes. The volume of the cell suspension was adjusted to 150 ml using wash buffer and the solution was sonicated as described previously (Section 2.36). Inclusion bodies were pelleted by centrifugation at 23,000 *g* for 20 minutes at 4°C. Pellets were washed by resuspending in 150 ml TW (wash buffer containing 1% Triton X-100) and centrifuged at 23,000 *g* for 10 minutes at 4°C. The wash step was repeated once with TW and twice with wash buffer. Following this, 1 ml of 100% DMSO was added to the pellet and incubated at room temperature for 30 minutes. The pellet was minced with a spatula and 40 ml of unfolding buffer (7 M guanidinium hydrochloride, 20 mM Tris-HCl

pH 7.5, 10 mM DTT) was slowly added. The solution was gently mixed for one hour at room temperature using a magnetic stir plate, followed by repeatedly passing through a pipette until the pellet was fully resuspended. Insoluble material was removed by centrifugation at 23,000 *g* for 10 minutes.

The supernatant was loaded onto a HiLoad 26/60 Superdex 75 prep grade gel filtration column (GE Healthcare), pre-equilibrated in SAU-1000 buffer (7 M guanidinium hydrochloride, 20 mM NaOAc pH 5.2, 1 M NaCl, 5 mM β-mercaptoethanol, 1 mM EDTA) with a flow rate of 3 ml/minute. The elution profile was monitored at wavelengths of 260 nm and 280 nm, and 5 ml fractions were collected. Peak fractions were analysed by 18% SDS-PAGE (Section 2.17). The A_{260}/A_{280} ratio of histone containing fractions was determined using a DS-11 spectrophotometer (DeNoix) and those containing a high A260/A280 ratio were discarded, due to excessive DNA contamination. Histone fractions were pooled, dialysed against three changes of distilled water containing 2 mM β-mercaptoethanol at 4°C and lyophilised.

Next, 60 - 100 mg of lyophilised histone protein was resuspended in 5 ml SAU200 buffer (7 M Urea, 20 mM sodium acetate pH 5.2, 0.2 M NaCl, 5 mM β-mercaptoethanol, 1 mM EDTA) and centrifuged at 23,000 *g* for 10 minutes to remove insoluble material. Histones were loaded onto an SP Sepharose XK26/20 column (GE healthcare), pre-equilibrated with SAU200. Histone proteins were eluted using a gradient of SAU200 and SAU600 (7 M Urea, 20 mM sodium acetate pH 5.2, 0.6 M NaCl, 5 mM β-mercaptoethanol, 1 mM EDTA), as specified by Luger *et al.*, (1999) for each individual histone. Fractions of 8 ml were collected and analysed by 18% SDS-PAGE. Those containing pure histone protein were dialysed against three changes of distilled water containing 2 mM β-mercaptoethanol at 4°C and lyophilised.

2.51 Site-specific substitution of a H3K4me3 analog

The lysine at position 4 in histone H3 was mutated to cysteine by site directed mutagenesis (Section 2.20) using primers H3 K4C F and R (Table 2.3, 45 + 46). An

existing cysteine at position 110 was mutated to alanine to prevent this becoming alkylated, also by site directed mutagenesis using primers H3 C110A F and R (Table 2.3, 47 + 48), to generate histone H3 C110A K4C. This mutated histone H3 was expressed and purified in *E.coli* as described previously (Section 2.50). To alkylate the cysteine introduced at position 4, 10 mg of lyophilised mutant histone H3 was resuspended in 980 µl alkylation buffer (1 M HEPES pH 7.8, 4 M guanadinium chloride, 10 mM D/L-methionine) and 20 µl freshly prepared 1 M DTT was added (Simon *et al.*, 2007). Histones were reduced for 1 hour at 37°C before 100 mg of solid (2-bromoethyl)trimethylammonium bromide was added and the reaction was heated to 50°C. After 2.5 hours, 10 µl 1 M DTT was added and the reaction was incubated at 50°C for a further 2.5 hours. The reaction was quenched by the addition of 50 µl 14.2 M β-mercaptoethanol, before being dialysed against three changes of distilled water containing 2 mM β-mercaptoethanol at 4°C and lyophilised.

Successful substitution of H3K4$_C$me3 was confirmed by western blotting (Section 2.19) using an anti-H3K4me3 antibody (ab8580, Abcam) at a concentration of 1:1000, diluted in 3% BSA/TBST.

2.52 Histone octamer refolding

Each lyophilised histone was dissolved in unfolding buffer (7 M guanidinium hydrochloride, 20 mM Tris-HCl pH 7.5, 10 mM DTT) to a concentration of 2 mg/ml and incubated at room temperature for 30 minutes. Insoluble material was removed by centrifugation at 20,000 *g* for 10 minutes at 4°C. The concentration of each unfolded histone was determined by measuring the absorbance at 280 nm using a DS-11 spectrophotometer (DeNovix). All four histone proteins were mixed at equimolar ratios and the final protein concentration was adjusted to 1 mg/ml using unfolding buffer. This was dialysed against three changes of 2 L refolding buffer (2 M NaCl, 10 mM Tris-HCl pH 7.5, 1 mM EDTA, 5 mM β-mercaptoethanol) at 4°C. The first two dialysis steps were performed for 4 hours each whereas the final was performed overnight. The following day, precipitated protein was removed by centrifugation at 20,000 *g* for 10 minutes at

4°C and the supernatant was concentrated to 1 ml using an Ultracel 25 ml centrifugation filter (UFC803024, Amicon). A HiLoad 16/60 Superdex 200 prep grade column (GE) was pre-equilibrated with refolding buffer and the concentrated histone octamers were loaded at a flow rate of 1 ml per minute. Fractions of 1 ml were collected and analysed by 18% SDS-PAGE and coomassie blue staining. Histone octamers with equimolar amounts of the four histone proteins were pooled and concentrated to 5 mg/ml as described above. Histone octamers were stored at -20°C in 50% v/v glycerol until use.

2.53 Microscale reconstitution of nucleosome core particles

Microscale reconstitutions were performed prior to larger scale salt gradient deposition reconstitution (Section 2.54) to optimise the Octamer:DNA ratio. 1 µg of fluorescently labelled or unlabelled 601-7-12-RSS or 601-7-23-RSS was diluted in 9 µl 2 M NaCl and a range of octamer amounts was added. The mix was incubated at room temperature for 30 minutes and then 10 µl of 10 mM HEPES-NaOH pH 8.0 was added to give a final concentration of 1 M NaCl. After incubating at room temperature for 1 hour, the following additions of 10 mM HEPES-NaOH pH 8.0 were performed: 5 µl (→ 0.8 M NaCl); 5 µl (→ 0.67 M NaCl); 70 µl (→ 0.2 M NaCl); and 100 µl (→ 0.1 M NaCl); with 1 hour incubation at room temperature after each addition. Each microscale reconstitution was analysed by native PAGE. Gels were 5% (59:1 acrylamide:bis-acrylamide) and contained 20 mM HEPES-NaOH pH 8.0 and 0.1 mM EDTA. Gels were run at 100 V for 5 hours and fluorescently labelled nucleosomes were imaged with a FLA-5100 gel scanner (Fujifilm) using the 473 nm laser. For unlabelled nucleosomes, the gels were stained with dsGreen gel staining solution (10010, Lumiprobe).

2.54 Reconstitution of nucleosomes by salt gradient dilution

Reconstitution reactions were performed as described in Section 2.53, using the optimal Octamer:DNA ratio determined by microscale reconstitution and scaled up to either 10 µg or 100 µg DNA for fluorescent or unlabelled nucleosomes respectively. Reactions were dialysed against 1L of reconstitution buffer 1 (1.2 M NaCl, 10 mM

HEPES-NaOH pH 8.0, 0.5 mM EDTA, 5 mM β-mercaptoethanol) for 2 hours at 4°C. The step dialysis was repeated against four further reconstitution buffers containing decreasing salt concentrations: 1.0 M, 0.8 M, 0.6 M and 0 M NaCl. All dialysis steps were performed at 4°C for 2 hours apart from the last step which was performed overnight. Reconstituted nucleosomes were analysed by native PAGE (Section 2.53) and stored at 4°C until purification (Section 2.55).

2.55 Purification of nucleosomes

Fluorescently labelled and unlabelled nucleosomes were purified using two different methods due to differences in the amount of starting material. Unlabelled nucleosomes were purified by sucrose density gradient centrifugation. A reconstitution reaction containing at least 100 µg of unlabelled DNA was loaded onto a 4 ml continuous 15-40% sucrose gradient (10 mM HEPES-NaOH pH 8.0, 0.5 mM EDTA) in a 5 ml ultra centrifuge tube. The gradient was centrifuged at 110,000 *g* for 18 hours at 4°C in an SW55Ti ultra-centrifuge rotor. Following centrifugation, a glass capillary tube, connected to a peristaltic pump, was carefully inserted into the gradient. Fractions of 200 µl were taken, starting from the base of the gradient, using the peristaltic pump on a very low speed (~0.5 ml per minute). The DNA and protein content of each fraction was determined by measuring the absorbance at 260 and 280 nm with a DS-11 spectrophotometer (DeNovix) and peak fractions were analysed by 6% native PAGE (Section 2.53). Fractions containing purified nucleosomes were pooled and dialysed against three changes of 1 L HE buffer (10 mM HEPES-NaOH pH 8.0, 0.5 mM EDTA) and stored at 4°C until use.

Fluorescently labelled nucleosome reconstitutions could not be purified by sucrose density gradient centrifugation due excessive dilution of nucleosomes. Instead, a preparative gel electrophoresis method was used to purify fluorescently labelled nucleosome (Studitsky *et al.*, 1994). A 5% native polyacrylamide gel (59:1, acrylamide:bis-acrylamide) containing 2.5% glycerol, 20 mM HEPES-NaOH pH 8.0 and 0.1 mM EDTA was pre-electrophoresed at 150 V for 9 hours in HE buffer. The buffer

was replaced with fresh HE and the gel was electrophoresed again for 30 minutes at 150 V. The buffer was replaced again and the voltage lowered to 100 V before an entire 10 µg fluorescent reconstitution reaction, with sucrose added to a final concentration of 10%, was loaded onto the gel. Gels were electrophoresed at 100 V for 8 hours and the fluorescent nucleosome band was located using a blue LED torch (UltraFire H-B3 470 nm) and excised with a sterile scalpel blade. The gel slice was finely diced and placed in 200 µl HE/BSA (200 ug/ml BSA) and left to rotate overnight at 4°C. The following day, tubes containing the diced gel slice were centrifuged at 20,000 *g* for 2 minutes and the supernatant containing eluted nucleosomes were transferred to a 1.5 ml siliconised micro-centrifuge tube. Purified nucleosomes were stored at 4°C until use.

2.56 Mapping of nucleosome positions using micrococcal nuclease

Micrococcal nuclease (MNase) reactions were performed in a final volume of 200 µl and contained 500 ng purified nucleosome and 1x MNase reaction buffer (35 mM NaCl, 10 mM HEPES-NaOH pH 8.0, 1 mM DTT, 2 mM $CaCl_2$). The reaction was pre-warmed to 37°C for 5 minutes before adding 0.6 units MNase (9013-53-0, Worthington). After incubation at 37°C for 2 minutes, the reaction was stopped by the addition of 5 µl 0.5 M EDTA. Reactions were extracted twice with phenol:chloroform (Section 2.4) and then the DNA was ethanol precipitated (Section 2.5) using 20 µg glycogen as carrier.

The DNA pellet was resuspended in 20 µl TE and 4 µl 6x DNA sample loading dye was added. This was loaded onto a 6% polyacrylamide gel (19:1, acrylamide:bis-acrylamide) containing 1x TAE and the gel was run at 170 V until the bromophenol blue was ~80% of the way down the gel. The gel was stained with 2 µg/ml ethidium bromide in 1x TAE for 10 minutes and visualised on a transilluminator. The band corresponding to ~147 bp was excised from the gel using a sterile scalpel and eluted from the gel using the “crush-and-soak” method (Section 2.31).

A 20 µl restriction digest containing 1x CutSmart® Buffer (50 mM KoAc, 20 mM Tris-acetate pH 7.9, 10 mM magnesium acetate, 100 µg/ml BSA; NEB), 100 ng 147 bp extracted DNA fragment and 5 units AccI (R0161S, NEB) was incubated at 37°C overnight. DNA sample loading dye was added to 1x final concentration and the reaction was loaded onto a 10% polyacrylamide gel (19:1, acrylamide:bis-acrylamide) containing 1x TAE and the gel was run at 170 V until the bromophenol blue was ~80% of the way down the gel. The gel was stained using dsGreen gel staining solution (10010, Lumiprobe) and visualised with a FLA-5100 gel imager (Fujifilm) using the 473 nm laser.

2.57 *In vitro* RAG binding and cleavage of nucleosome substrates

RAG binding and cleavage reactions with nucleosome substrates were performed in a similar way to those with free DNA substrates (Section 2.37 and 2.38), except the reaction buffer contained 5 mM $MgCl_2$ and the reactions were incubated at 37°C for 1 hour. A typical reaction contained 10 ng of purified fluorescently nucleosome substrate with a 5-fold excess of unlabelled nucleosome partner RSS. Reactions were loaded onto a 6% polyacrylamide gel (59:1 acrylamide:bis-acrylamide) containing 10 mM HEPES-NaOH pH 8.0. Gels were electrophoresed in HE buffer (10 mM HEPES-NaOH pH 8.0, 0.1 mM EDTA) at 100 V for 6 hours at 4°C. Imaging was performed on an FLA-5100 gel imager (Fujifilm) using the 473 nm laser.

2.58 Analysis of coding ends retained in RAG post-cleavage complexes

RAG:nucleosome complexes were excised from *in vitro* RAG binding and cleavage gels (Section 2.57) to determine if coding ends are retained more readily in such complexes. A sterile scalpel was used to excise gel slices corresponding to RAG:nucleosome complexes, which were finely diced and placed in 1 ml stop buffer (50 mM Tris-HCl pH 8.0, 0.1% SDS, 15 mM EDTA, 175 µg/ml proteinase K). Microcentrifuge tubes containing the diced gel slices were rotated overnight at room temperature. The following day, the tubes were centrifuged at 20,000 *g* for 2 minutes to remove gel pieces and the supernatant was extracted twice with phenol:chloroform

(Section 2.4) and ethanol precipitated (Section 2.5). The DNA pellet was resuspended in 10 µl TE and loaded onto a 10% polyacrylamide gel (19:1, acrylamide:bis-acrylamide) containing 1x TAE, which was run at 170 V for 3 hours. Gels were stained using dsGreen gel staining solution (10010, Lumiprobe) and imaged as described previously (Section 2.56).

l) Oligonucleotide sequences

Oligonucleotides used during the course of this project are shown in Tables 2.3 and 2.4 and were synthesised by either Sigma-Aldrich or Integrated DNA Technologies (IDT).

#	Name	Sequence
1	DR55	AGAGGGACTGGATTCCAAAGTTCTC
2	1233	CTTTCATTGCCATACG
3	SJ F	CTGCTTGCTGTTCTTGAATGG
4	SJ R	TACAGCCAGACAGTGGAGTA
5	JH299 SJ Probe	/56- FAM/CCAGTCTGT/ZEN/AGCACTGTG CACAGT/ 3IABkFQ/
6	CAT F	CGTATGGCAATGAAAGACGGTGAGC
7	CAT R	CGAAGAAGTTGTCCATATTGGCCACG
8	HMGB1 Del F	TCAACGAGCAGTTAGGCATAC
9	HMGB1 Del R	ATACACGACCCAAGCCATTC
10	HMGB1 Exon 2 F	CCACCCTGAGCAGTTTGAAG
11	HMGB1 Exon 2 R	AAGCCCTCTTACCTTCCACC
12	HPRT F	GGGGGCTATAAGTTCTTTGC
13	HPRT R	TCCAACACTTCGAGAGGTCC
14	VP64 Del F	GAGGGCAGAGGAAGTCTGC
15	VP64 Del R	GTCGCCTCCCAGCTGAGA
16	hZNF10 F	GATGCTAAGTCACTAACTGC
17	hZNF10 R	AACTGATGATTTGATTTCAAATG
18	P2A F	GCAACAAACTTCTCTCTGC
19	EF1α R	CTGTGTTCTGGCGGCAAAC

#	Name	Sequence
20	rHMGB1 F	GCCGCCACCATGGGCAAAGGAGATC CTAA
21	rHMGB1ΔC R	TTTCGCTGCATCAGGTTT
22	mHMGB2 F	GCCGCCACCATGGGCAAGGGTGAC CC
23	mHMGB2ΔC R	CTTTCCTGCTTCACTTTTG
24	mHMGB2 R	TTCTTCATCCTCCTCTTCTTCCTCG
25	488 Label F	AlexaFluor-488/TGATCTGGCCTGTCTTA
26	601 Label R	ATCATCGAATTCGGTTTTTG
27	ESC Reint NheI Top	CTAGCACATAGCATCGATTAAGCAA
28	ESC Reint HindIII Bottom	AGCTTTGCTTAATCGATGCTATGTG
29	Reint F	CTAGCACATAGCATCGATTAAGCAA
30	Reint R	CTGGCTAGCGTTTAAACTTAAGC
31	Reint Probe	/56- FAM/CGAGCTCGG/ZEN/ATCCCTGCAGGGTTT/ 3IABkFQ/
32	Neo F	TGCTCCTGCCGAGAAAGTATC
33	Neo R	TTTCGCTTGGTGGTCGAATG
34	DR99	TCACACAGGAAACAGCTATGACCATG
35	DR100	GGGATATATCAACGGTGGTATATCCAGTG
36	DAR39	GATCTGGCCTGTCTTACACAGTGCTACAGACTGGAA CAAAAACCCTGCAG
37	DAR40	CTGCAGGGTTTTTGTTCCAGTCTGTAGCACTGTGTA AGACAGGCCAGATC
38	DG61	GATCTGGCCTGTCTTACACAGTGGTAGTACTCCACT GTCTGGCTGTACAAAAACCCTGCAG
39	DG62	CTGCAGGGTTTTTGTACAGCCAGACAGTGGAGTACT ACCACTGTGTAAGACAGGCCAGATC
40	RSS Label R	GATCTGCAGGGTTTTTGT
41	601 12-RSS F	ACTGGAACAAAAACCCTGCAGGATG
42	601 12-RSS R	CTGTAGCACTGTGATCGGATGTATATATC
43	601 Linker F	TTACACAGTGCTACAGACTGGA
44	601 Linker R	GACCATCGGATGTATATATCTGACA
45	H3 K4C F	GGGCGGTCTGGCAGGTACGGGCC
46	H3 K4C R	GTAAATCCACCGGAGGGAAG

#	Name	Sequence
47	H3 C110A F	GTGGATGGCGGCCAGGTTGGTG
48	H3 C110A R	GCCAAGAGGGTCACCATC

Table 2.3 - List of oligonucleotides during this project

Oligonucleotides names are sequences are shown. Oligonucleotides were purified by desalting or by HPLC (for fluorescent primers and probes). Stock concentrations of oligonucleotides were typically 100 µM, in distilled water, and were stored at -20°C until use.

#	Name	Sequence
1	HMGB1 Upstream Top	**CACCG**TCTGTGGCTTAACCGCTTCT
2	HMGB1 Upstream Bottom	**AAAC**AGAAGCGGTTAAGCCACAGA**C**
3	HMGB1 Downstream Top	**CACCG**GTAAATTAGGATGAACGGGC
4	HMGB1 Downstream Bottom	**AAAC**GCCCGTTCATCCTAATTTAC**C**
5	HMGB1 TSS1 Top	**CACCG**GTTACAGAGCGGAGAGAGTG
6	HMGB1 TSS1 Bottom	**AAAC**CACTCTCTCCGCTCTGTAAC**C**
7	HMGB1 TSS2 Top	**CACCG**GCGGCGCTGTCTCTATGGAGC
8	HMGB1 TSS2 Bottom	**AAAC**CGGCGCTGTCTCTATGGAGC**C**
9	HMGB1 TLS Top	**CACCG**GGAGATCCTAAAAAGCCGAG
10	HMGB1 TLS Bottom	**AAAC**CTCGGCTTTTTAGGATCTCC**C**

Table 2.4 - List of oligonucleotides used to clone sgRNA sequences

Oligonucleotide names and sequences of those used to clone sgRNA sequences for CRISPR studies are shown. Complementary BsmBI overhangs are shown in bold.

J) Plasmid maps

Maps of plasmids used during this the work presented in this thesis are shown below. Plasmid maps were generated using SnapGene Viewer (Version 4.3.10).

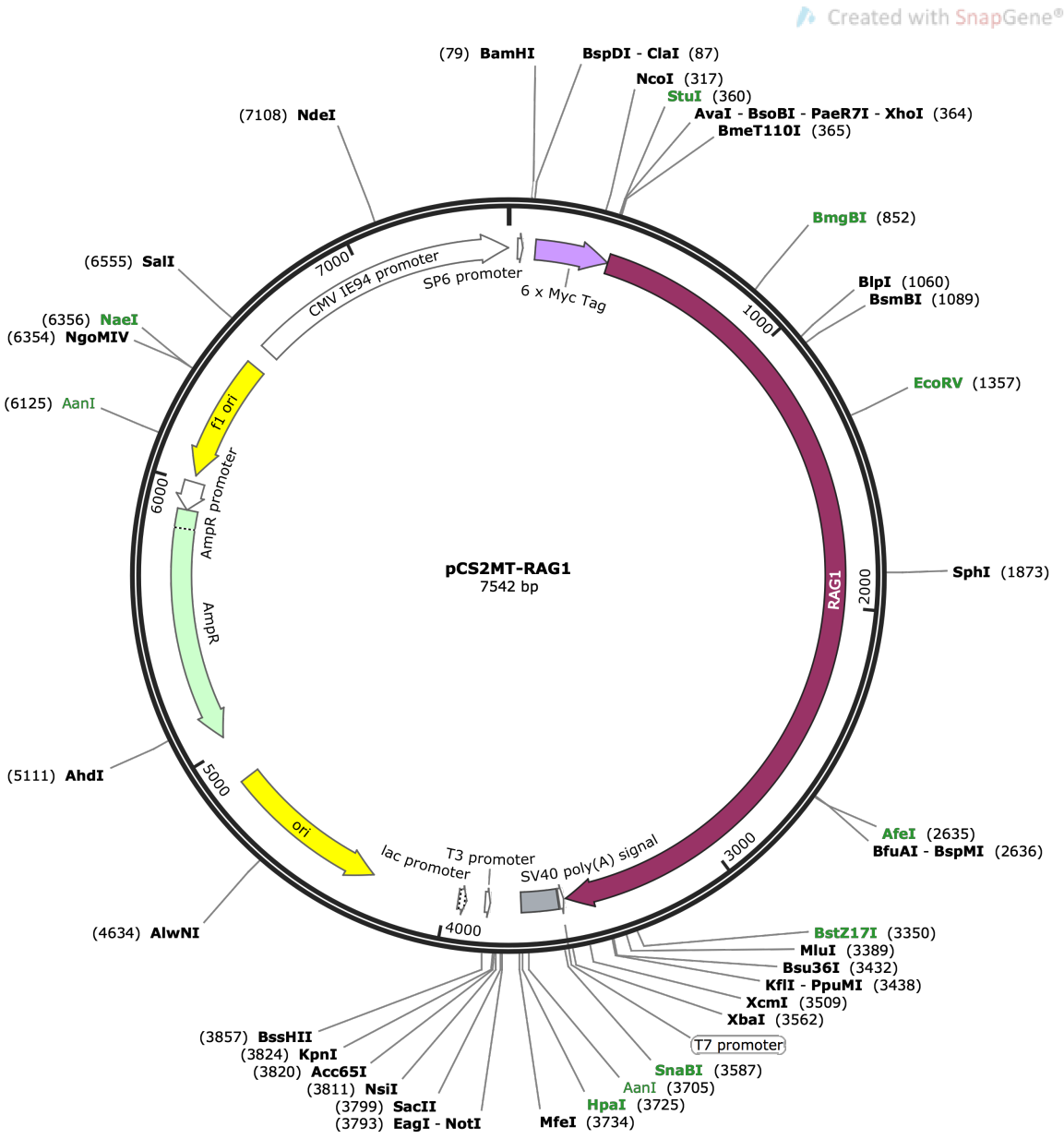


Figure 2.1 - Map of pCS2MT-RAG1

Mouse RAG1 cDNA was cloned into the pCS2MT vector (Rupp *et* al., 1994), in frame with the 6xMyc tag. All unique restriction enzyme sites are shown. AmpR = Ampicillin resistance gene (β-lactamase); CMV = cytomegalovirus; ori = origin of replication; SV40 = Simian virus.

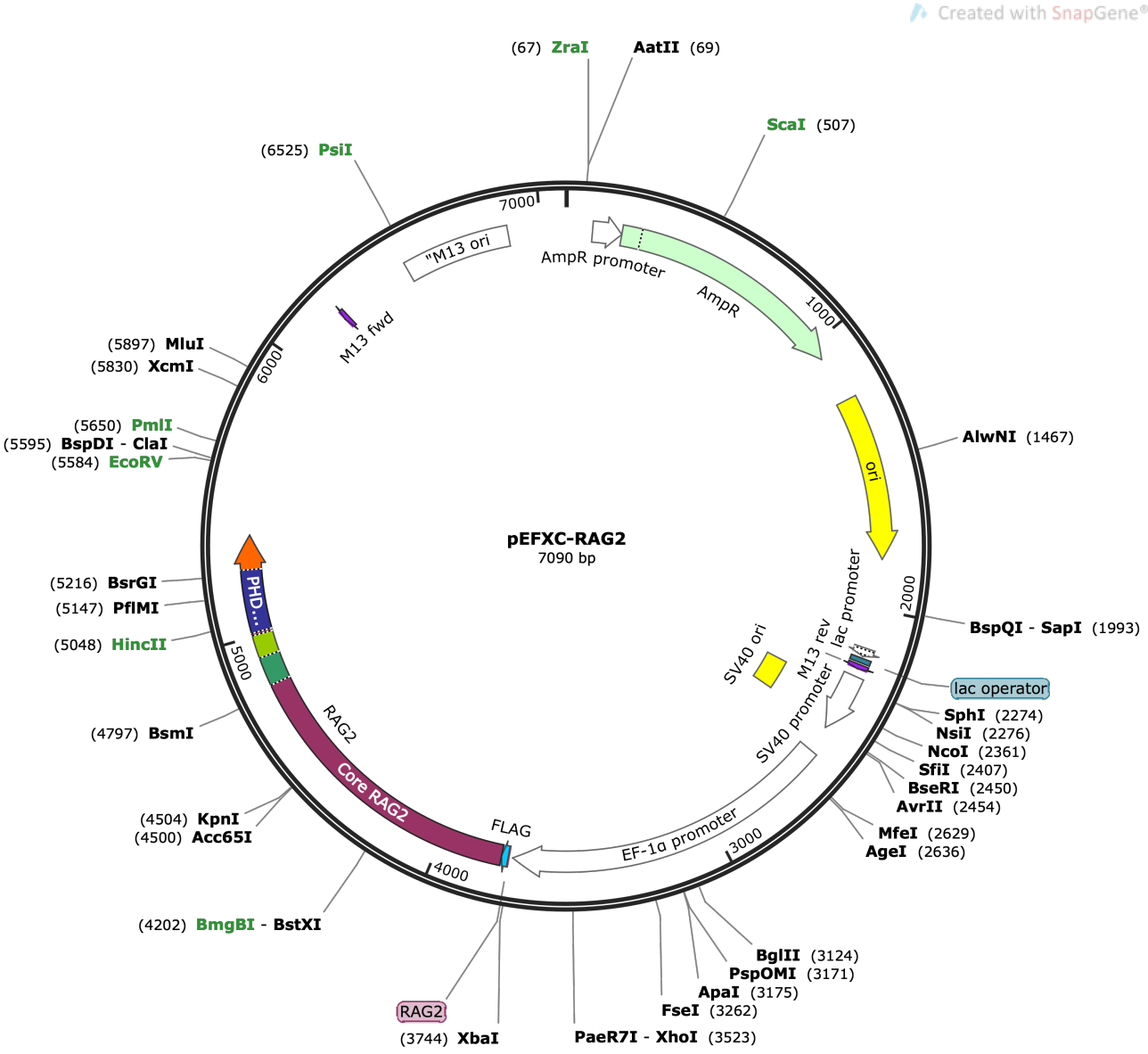


Figure 2.2 - Map of pEFXC-RAG2

Mouse RAG2 cDNA was cloned into the pEF-BOS (Mizushima et al., 1990) digested with XbaI and ClaI. All unique restriction enzyme sites are shown. AmpR = Ampicillin resistance gene (β-lactamase); EF-1α = human elongation factor 1α; ori = origin of replication; SV40 = Simian virus.

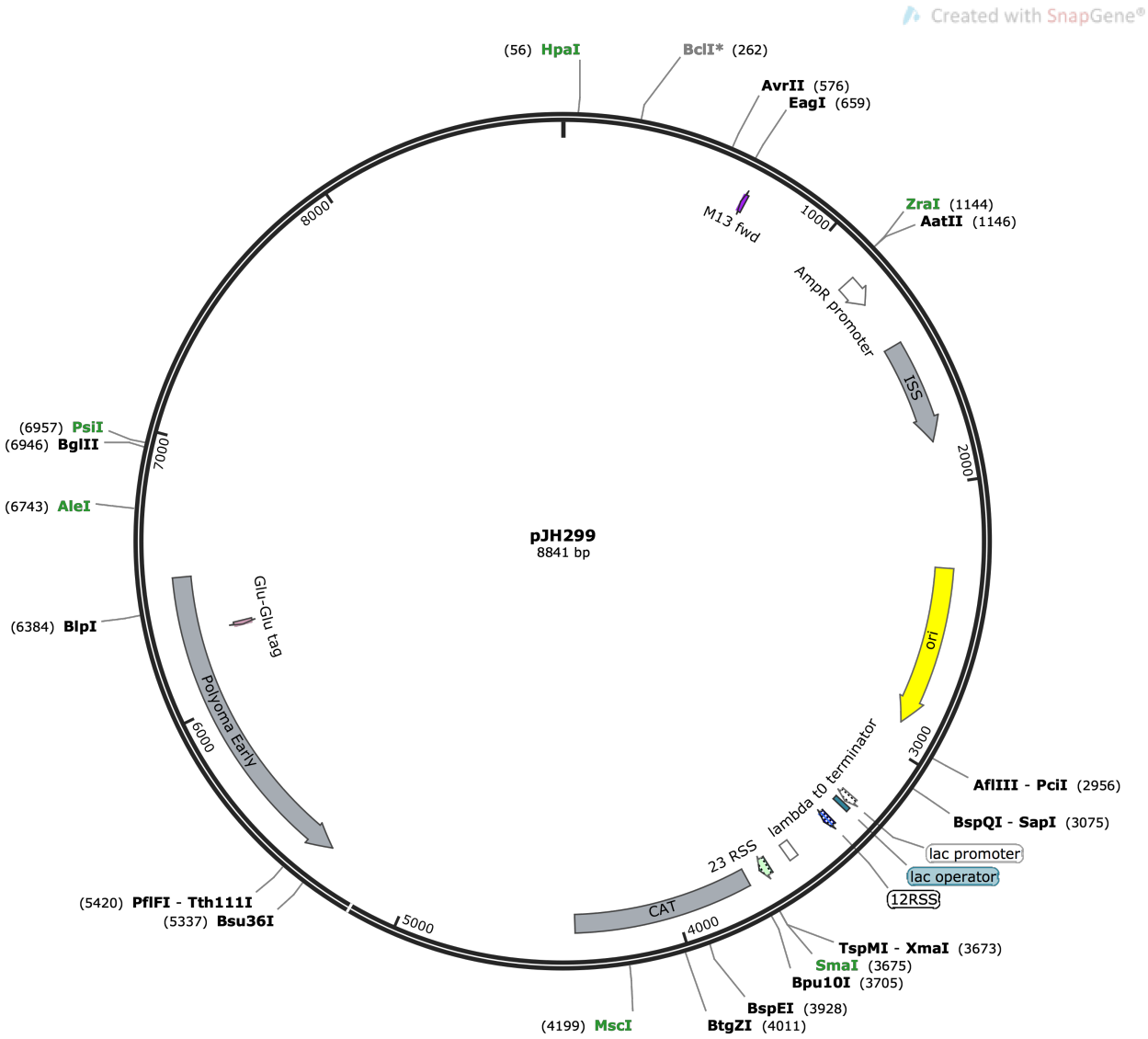


Figure 2.3 - Map of pJH299

Consensus 12-RSS and 23-RSS sequences are positioned in the same orientation, upstream of the CAT gene (Hesse et al., 1989). All unique restriction enzyme sites are shown. AmpR = Ampicillin resistance gene (β-lactamase); ori = origin of replication; CAT = Chloramphenicol resistant gene (Chloramphenicol acetyltransferase).

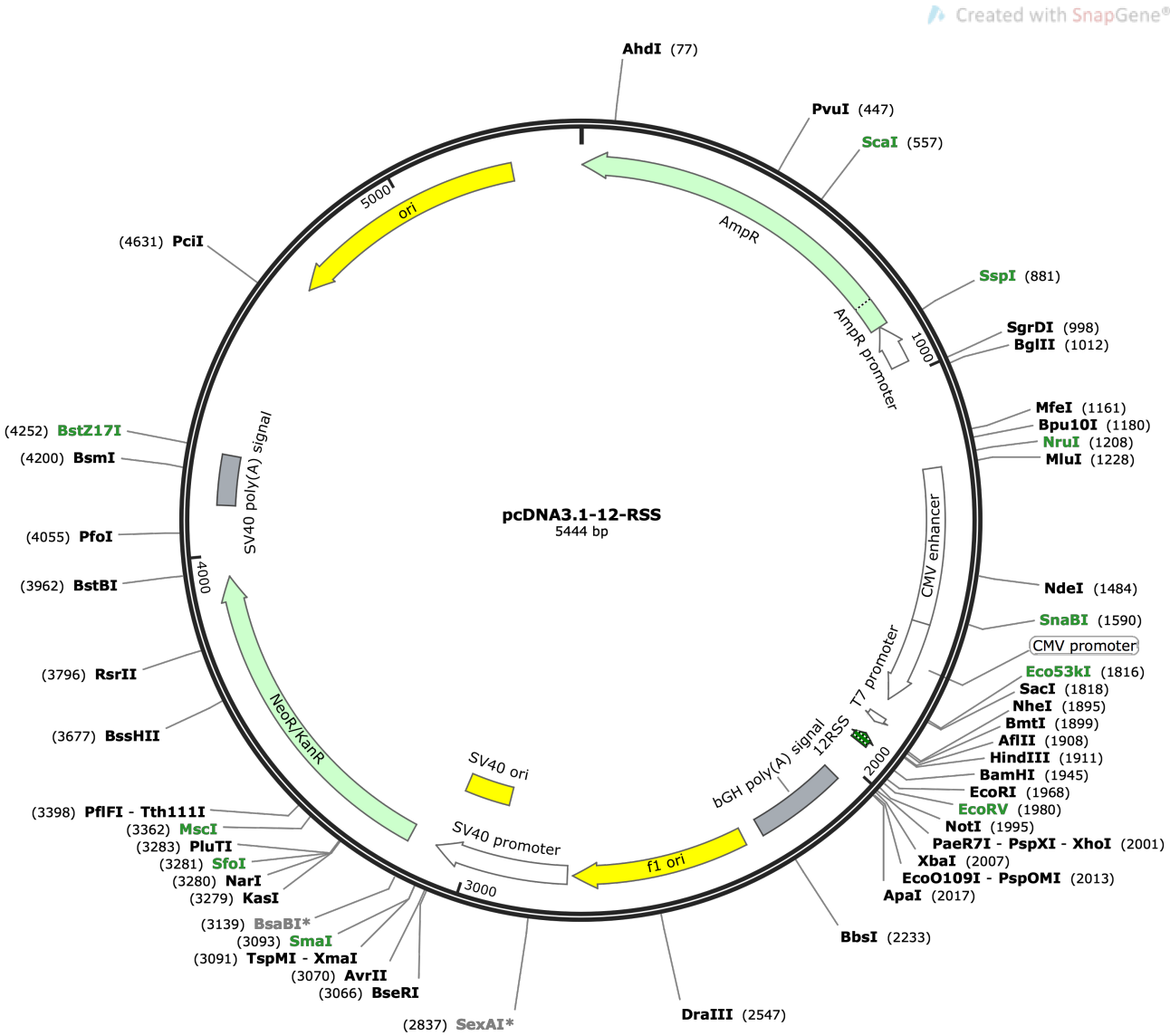


Figure 2.4 - Map of pcDNA3.1-12-RSS

A consensus 12-RSS oligonucleotide was cloned into pcDNA3.1 (Invitrogen) as described in Kirkham *et al.*, (2019). All unique restriction enzyme sites are shown. AmpR = Ampicillin resistance gene (β-lactamase); CMV = cytomegalovirus; ori = origin of replication; SV40 = Simian virus.

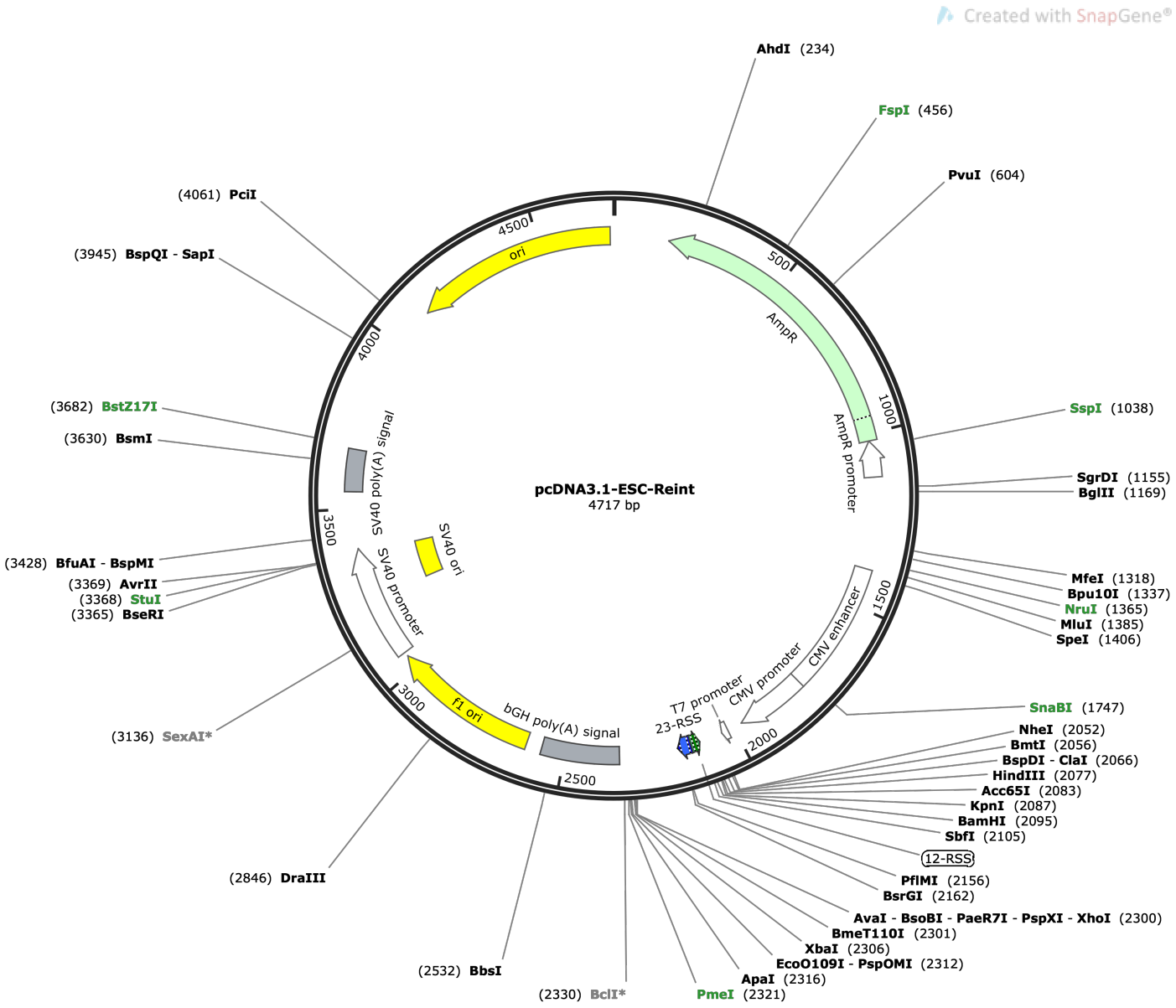


Figure 2.5 - Map of pcDNA3.1-ESC-Reint

pcDNA3.1-ESC was generated by cloning an ESC oligonucleotide into pcDNA3.1 (Invitrogen) as described in Kirkham *et al.*, (2019). A site upstream of the ESC sequence was mutated to include a specific primer binding site and a unique ClaI site to generate pcDNA3.1-ESC-Reint. All unique restriction enzyme sites are shown. AmpR = Ampicillin resistance gene (β-lactamase); CMV = cytomegalovirus; ori = origin of replication; SV40 = Simian virus.

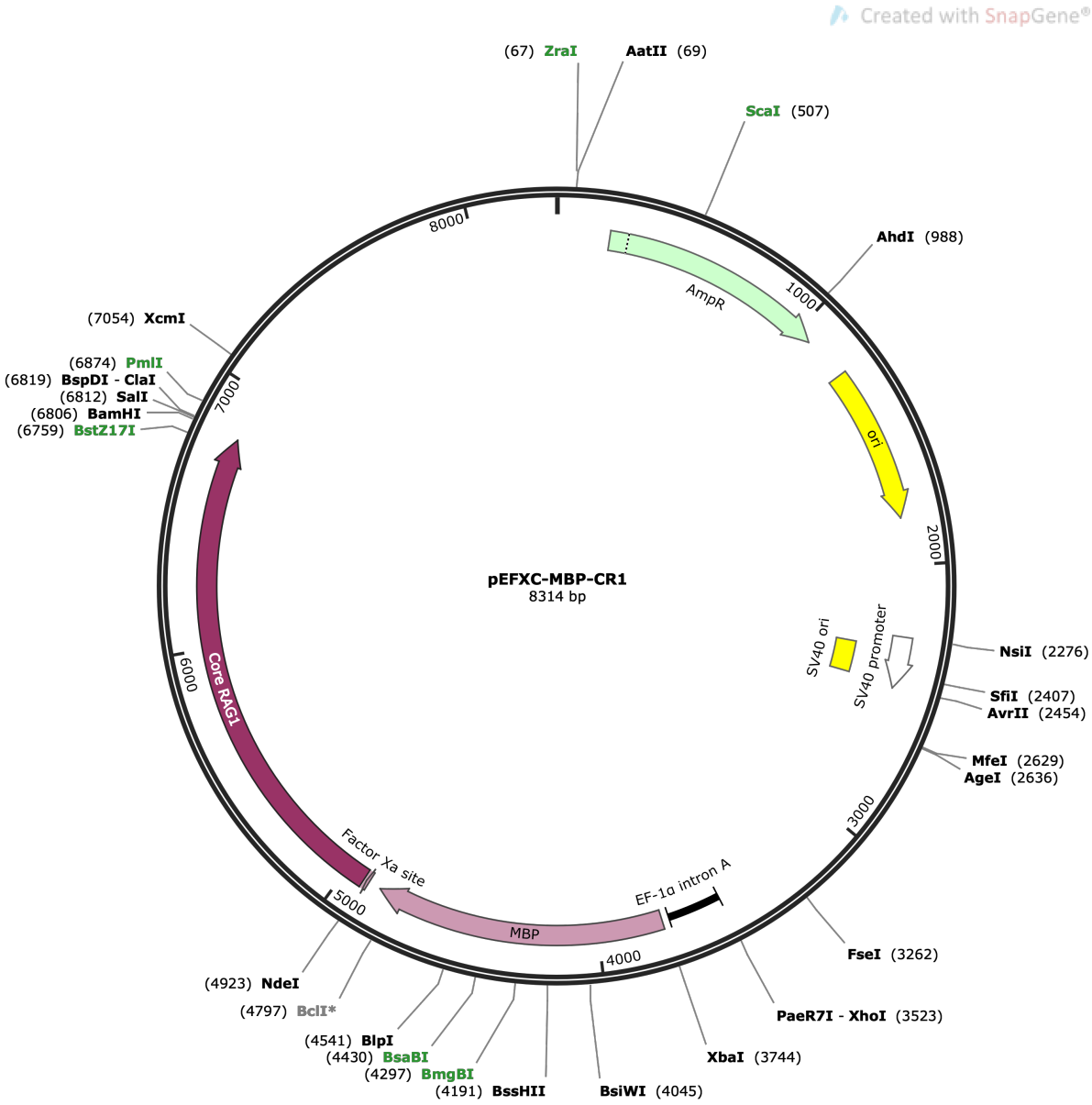


Figure 2.6 - Map of pEF-McR1

Mouse core RAG1 cDNA was cloned into pEF-BOS in frame with a maltose binding protein (MBP) amylose affinity purification tag (Kirkham *et al.*, 2019). All unique restriction enzyme sites are shown. AmpR = Ampicillin resistance gene (β-lactamase); EF-1α = human elongation factor 1α; ori = origin of replication; SV40 = Simian virus.

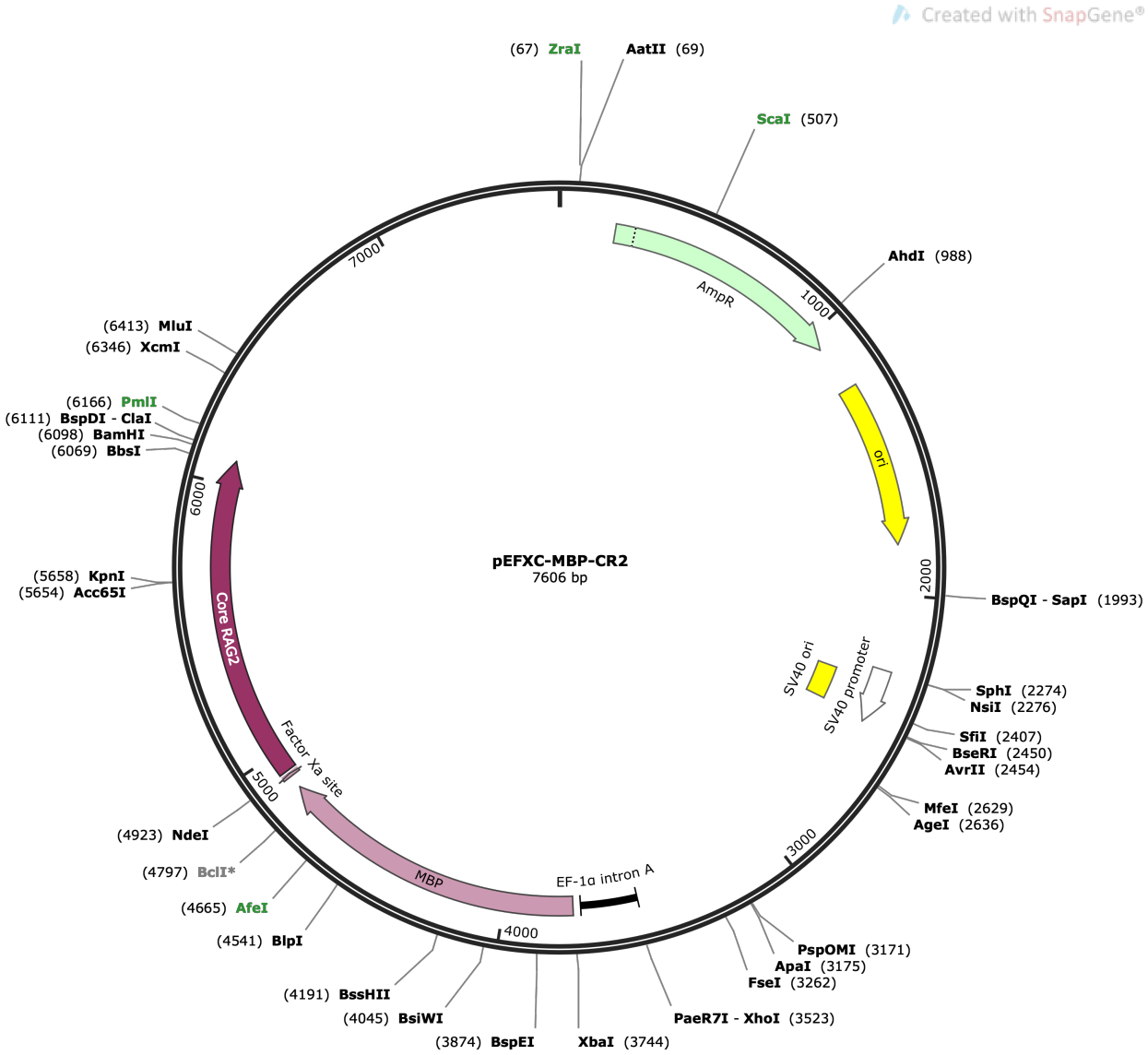


Figure 2.7 - Map of pEF-McR2

Mouse core RAG2 cDNA was cloned into pEF-BOS in frame with a maltose binding protein (MBP) amylose affinity purification tag (Kirkham *et al.*, 2019). All unique restriction enzyme sites are shown. AmpR = Ampicillin resistance gene (β-lactamase); EF-1α = human elongation factor 1α; ori = origin of replication; SV40 = Simian virus.

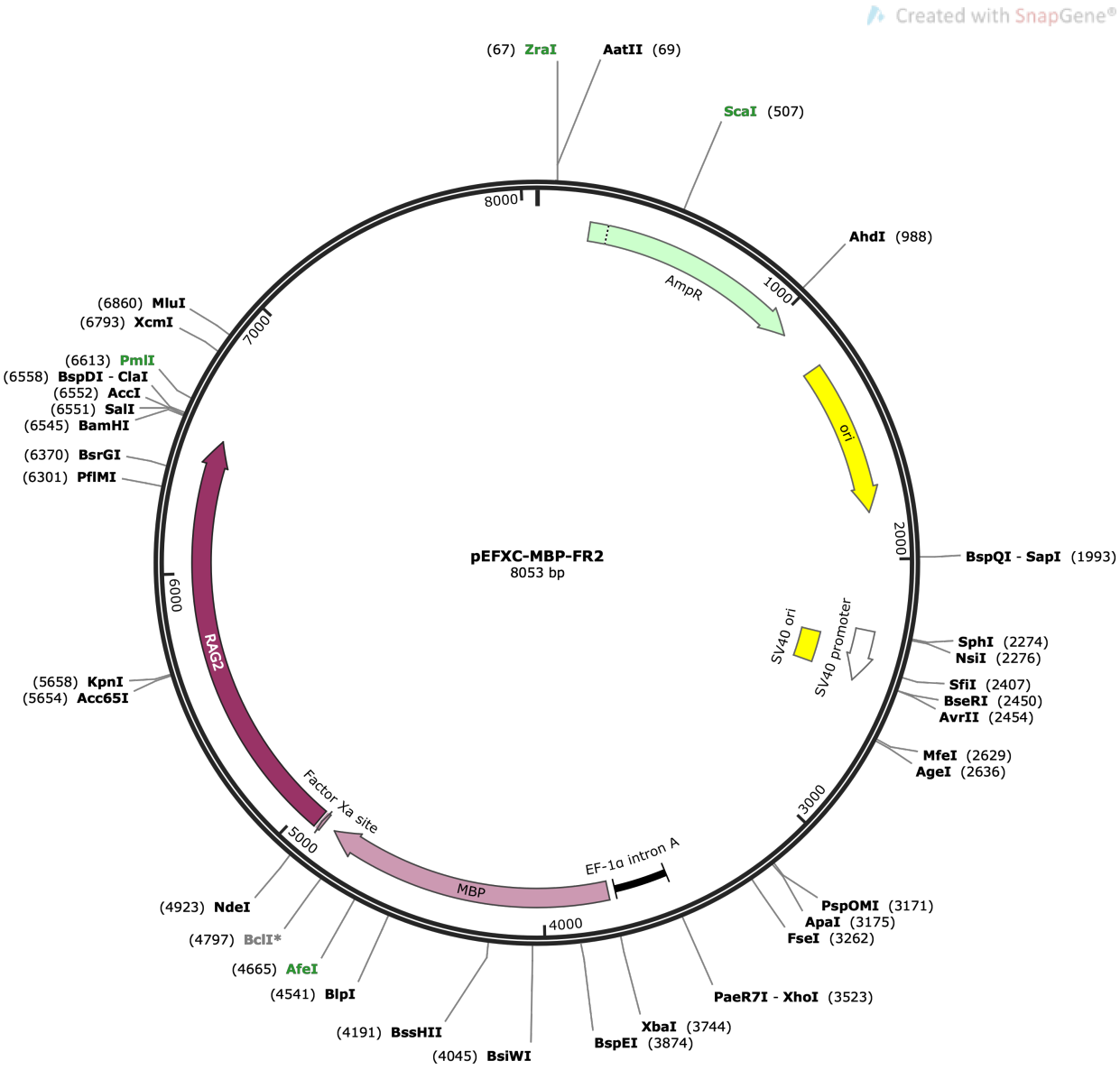


Figure 2.8 - Map of pEF-MfR2

Mouse RAG2 cDNA was cloned into pEF-BOS in frame with a maltose binding protein (MBP) amylose affinity purification tag (Kirkham *et al.*, 2019). All unique restriction enzyme sites are shown. AmpR = Ampicillin resistance gene (β-lactamase); EF-1α = human elongation factor 1α; ori = origin of replication; SV40 = Simian virus.

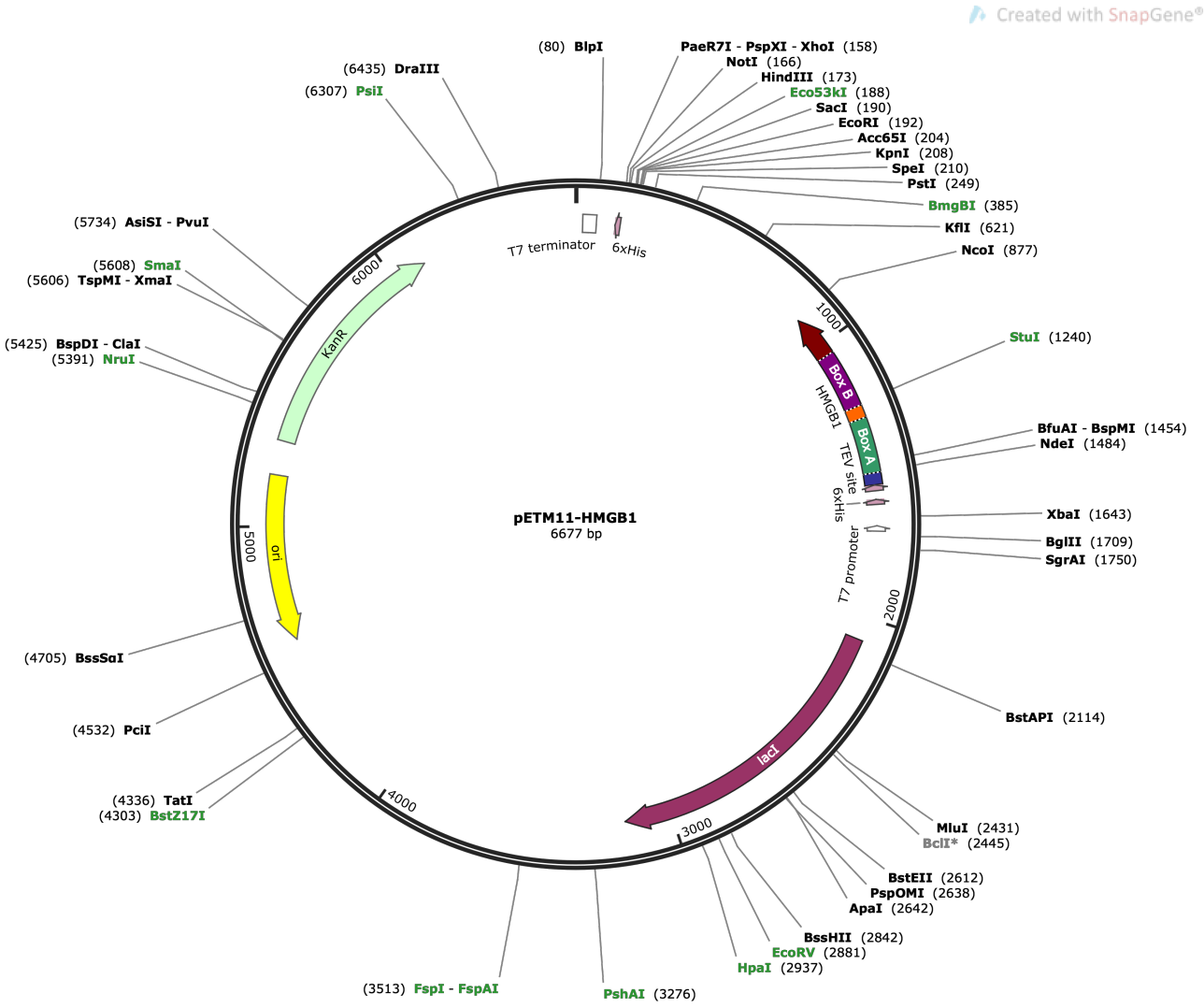


Figure 2.9 - Map of pETM11-HMGB1

pETM11-HMGB1 was a kind gift from Professor Marco Bianchi. Rat HMGB1 cDNA is cloned into the pETM11 vector, in frame with a 6xHis tag and TEV cleavage site. All unique restriction enzyme sites are shown. KanR = Kanamysin resistance gene; ori = origin of replication.

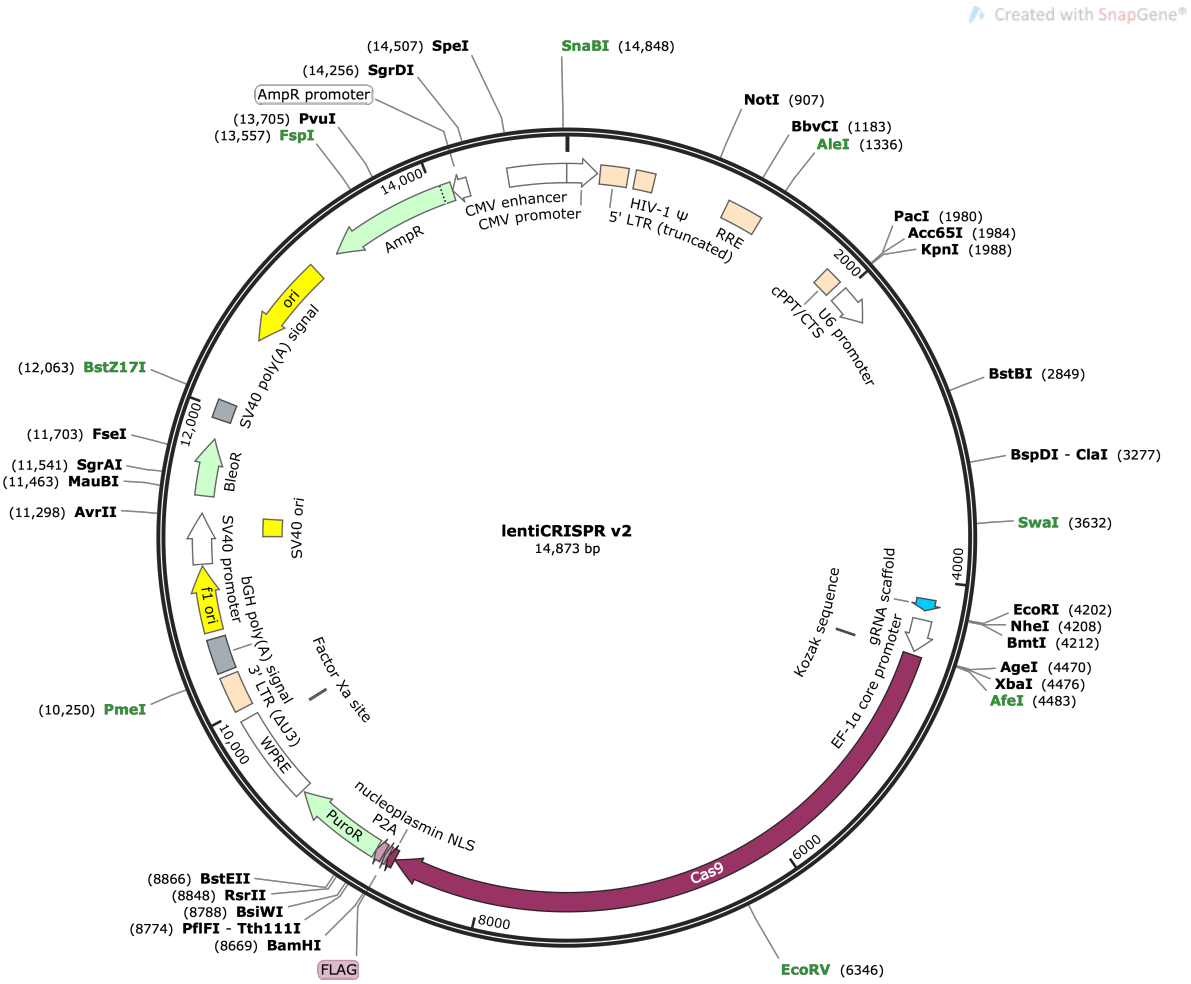


Figure 2.10 - Map of lenti-CRISPR V2

lenti-CRISPR V2 was a gift from Feng Zhang (Addgene plasmid # 52961 ; http://n2t.net/addgene:52961 ; RRID:Addgene_52961). A sgRNA scaffold , under the control of the human U6 promoter, is positioned upstream of the Cas9 cDNA sequence. All unique restriction enzyme sites are shown. AmpR = Ampicillin resistance gene (β-lactamase); CMV = cytomegalovirus; EF-1α = human elongation factor 1α; ori = origin of replication; LTR = Long terminal repeat; RRE = Rev response element; PuroR = Puromycin resistance gene (puromycin N-acetyltransferase); P2A = porcine teschovirus-1 2A self cleaving peptide sequence; SV40 = Simian virus.

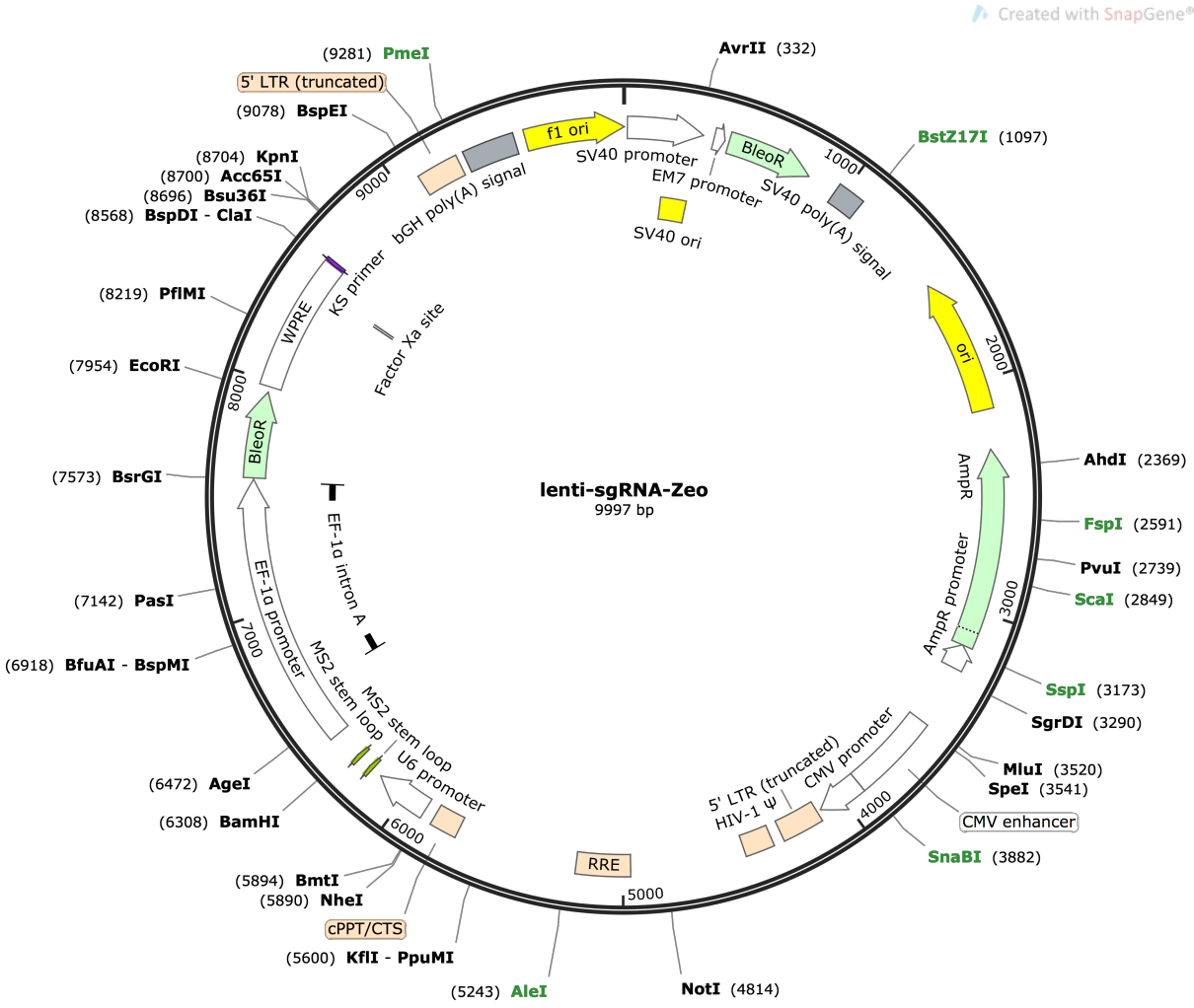


Figure 2.11 - Map of lenti-sgRNA-Zeo

lenti-sgRNA-Zeo backbone was a gift from Feng Zhang (Addgene plasmid # 61427 ; http://n2t.net/addgene:61427 ; RRID:Addgene_61427). A sgRNA scaffold with MS2 stem loops is positioned downstream of the human U6 promoter. The zeocin resistance gene (BleoR) is positioned downstream of the human EF-1α promoter. All unique restriction enzyme sites are shown. AmpR = Ampicillin resistance gene (β-lactamase); CMV = cytomegalovirus; EF-1α = human elongation factor 1α; ori = origin of replication; LTR = Long terminal repeat; RRE = Rev response element; BleoR = Zeocin resistance gene; SV40 = Simian virus; U6 = Human U6 promoter.

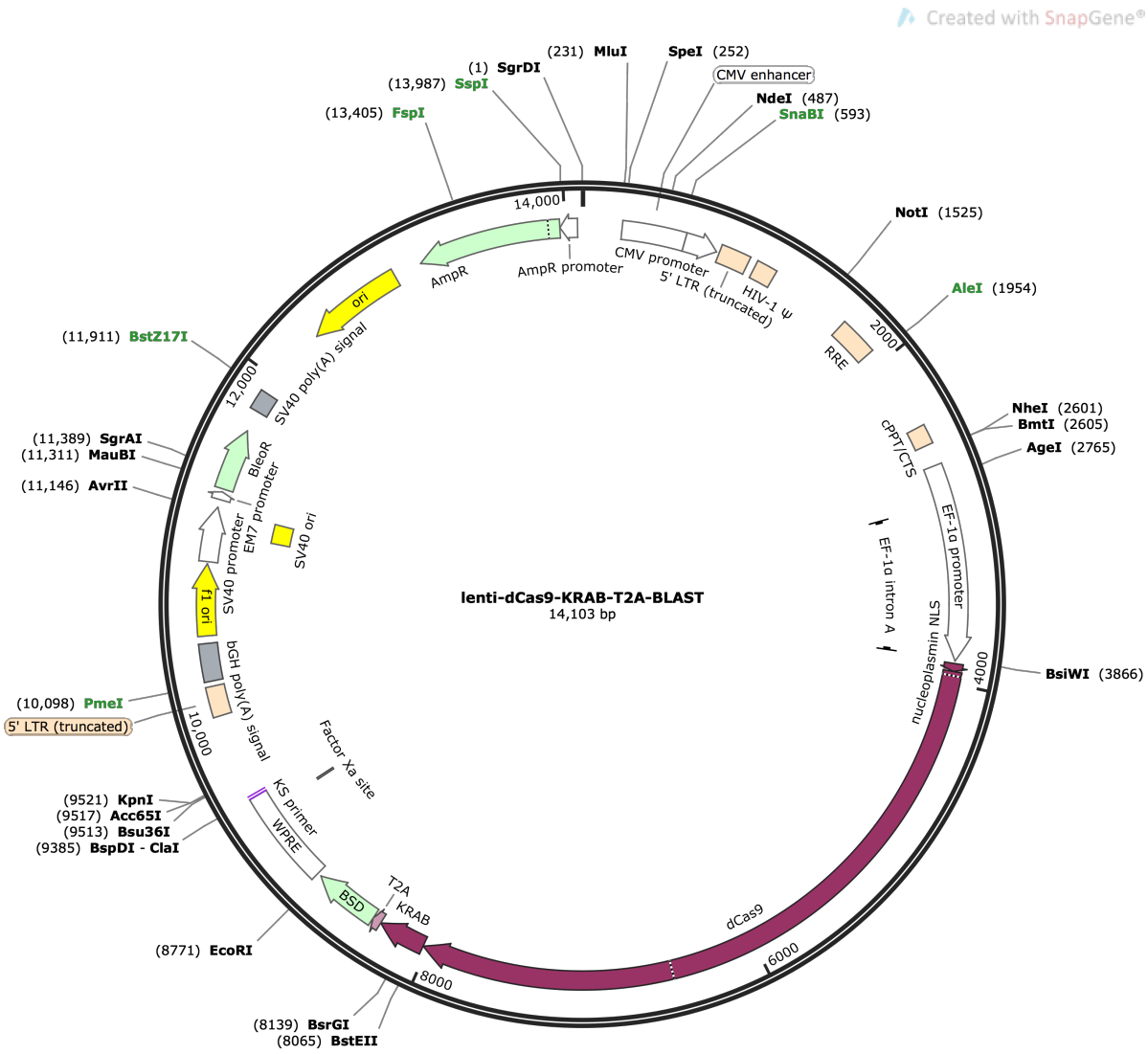


Figure 2.12 - Map of lenti-dCas9-KRAB-T2A-BLAST

The VP64 sequence from lenti dCAS-VP64-Blast, a gift from Feng Zhang (Addgene plasmid # 61425 ; http://n2t.net/addgene:61425 ; RRID:Addgene_61425), was replaced with the cDNA sequence of the KRAB domain from human ZNF10 (aa 2-97). All unique restriction enzyme sites are shown. AmpR = Ampicillin resistance gene (β-lactamase); CMV = cytomegalovirus; EF-1α = human elongation factor 1α; ori = origin of replication; LTR = Long terminal repeat; RRE = Rev response element; BSD = Blasticidin resistance gene; SV40 = Simian virus; T2A = thosea asigna virus 2A self cleaving peptide sequence.

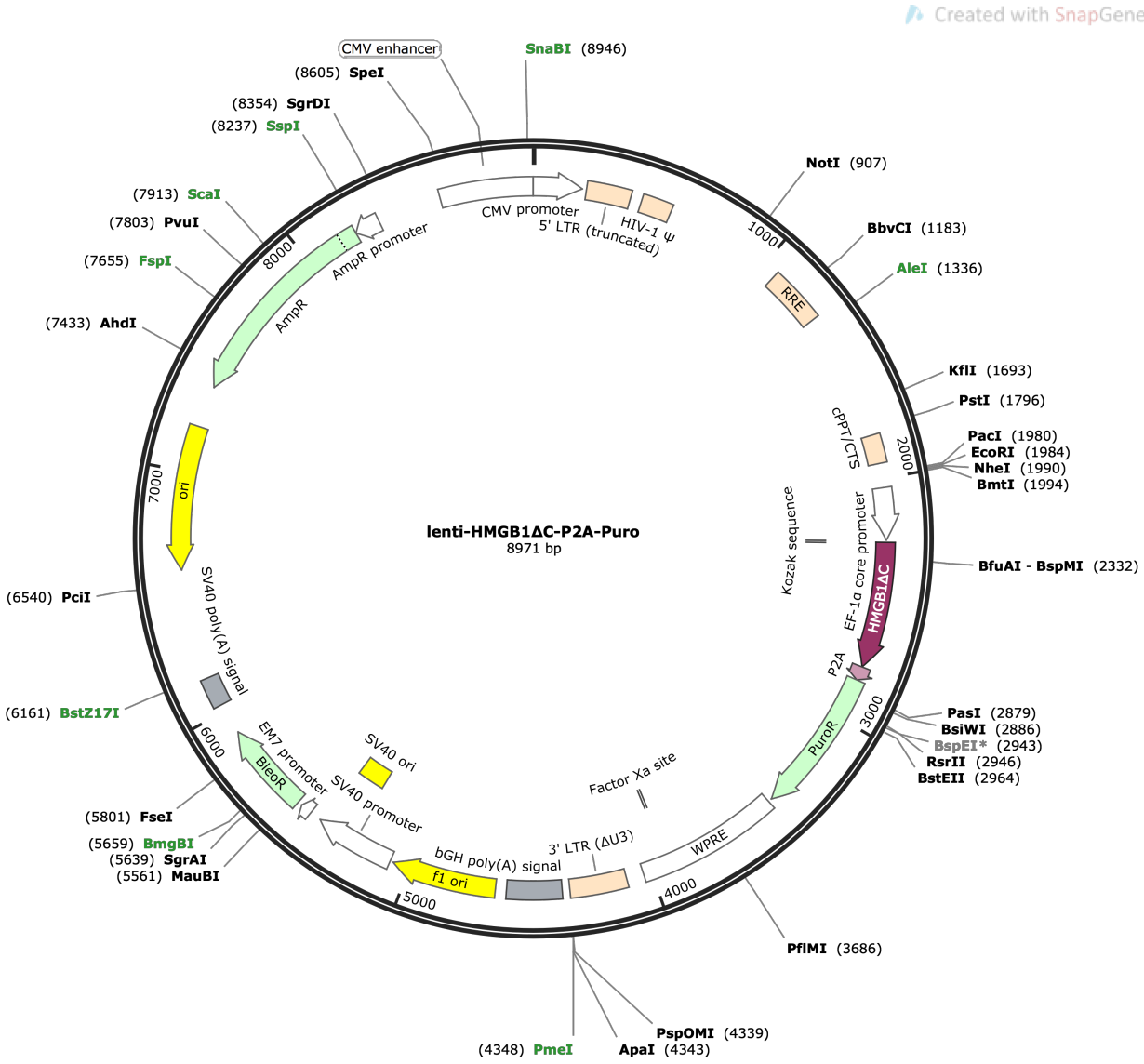


Figure 2.13 - Map of lenti-HMGB1ΔC-P2A-Puro

The Cas9 and gRNA scaffold sequences were removed from lenti-CRISPR V2 (Section 2.45) and the rat HMGB1ΔC cDNA sequence was cloned in frame with the P2A and puromycin resistance gene. All unique restriction enzyme sites are shown. AmpR = Ampicillin resistance gene (β-lactamase); CMV = cytomegalovirus; EF-1α = human elongation factor 1α; ori = origin of replication; LTR = Long terminal repeat; RRE = Rev response element; PuroR = Puromycin resistance gene (puromycin N-acetyltransferase); P2A = porcine teschovirus-1 2A self cleaving peptide sequence; SV40 = Simian virus.

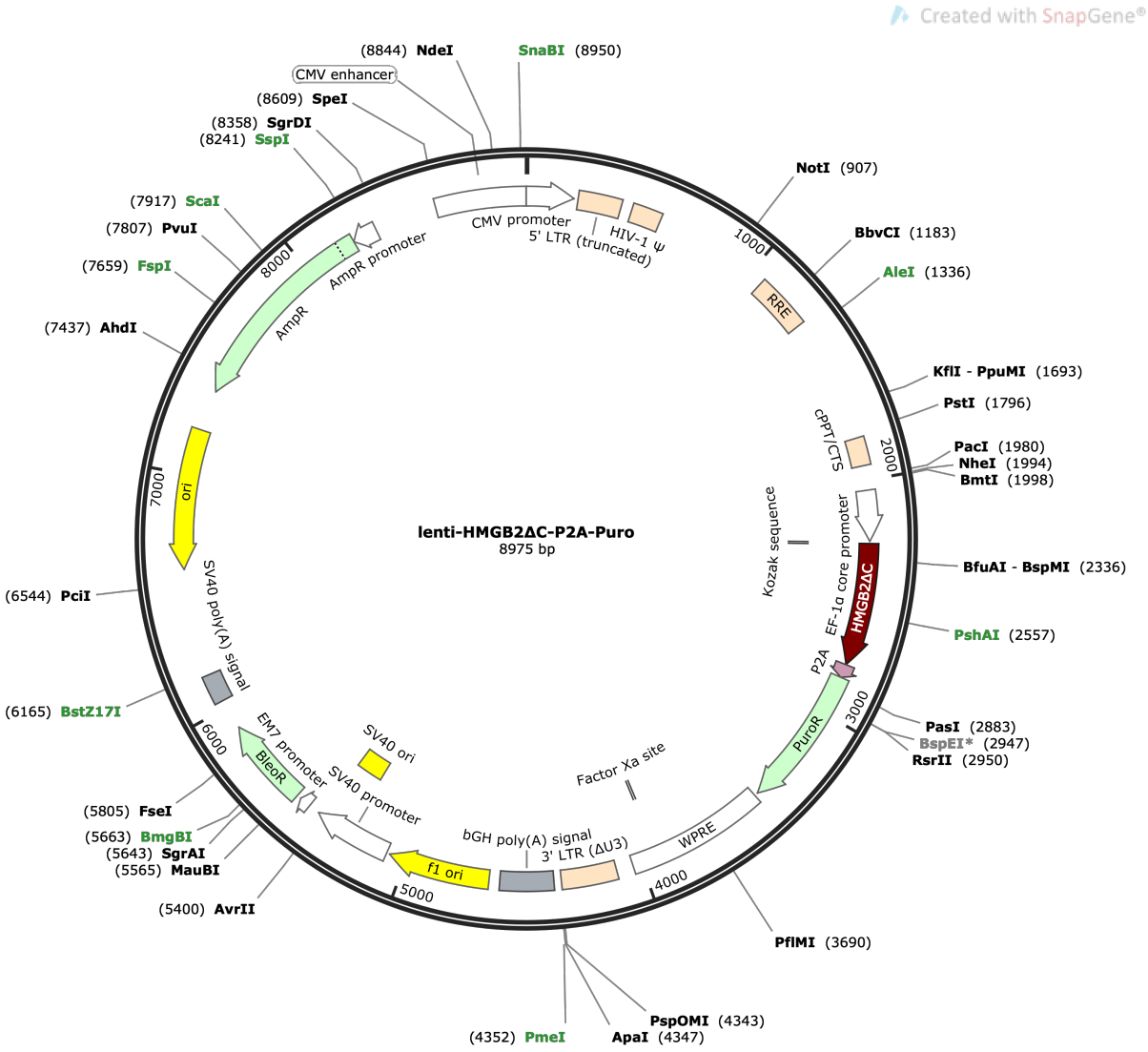


Figure 2.14 - Map of lenti-HMGB2ΔC-P2A-Puro

The Cas9 and gRNA scaffold sequences were removed from lenti-CRISPR V2 (Section 2.45) and the rat HMGB2ΔC cDNA sequence was cloned in frame with the P2A and puromycin resistance gene. All unique restriction enzyme sites are shown. AmpR = Ampicillin resistance gene (β-lactamase); CMV = cytomegalovirus; EF-1α = human elongation factor 1α; ori = origin of replication; LTR = Long terminal repeat; RRE = Rev response element; PuroR = Puromycin resistance gene (puromycin N-acetyltransferase); P2A = porcine teschovirus-1 2A self cleaving peptide sequence; SV40 = Simian virus.

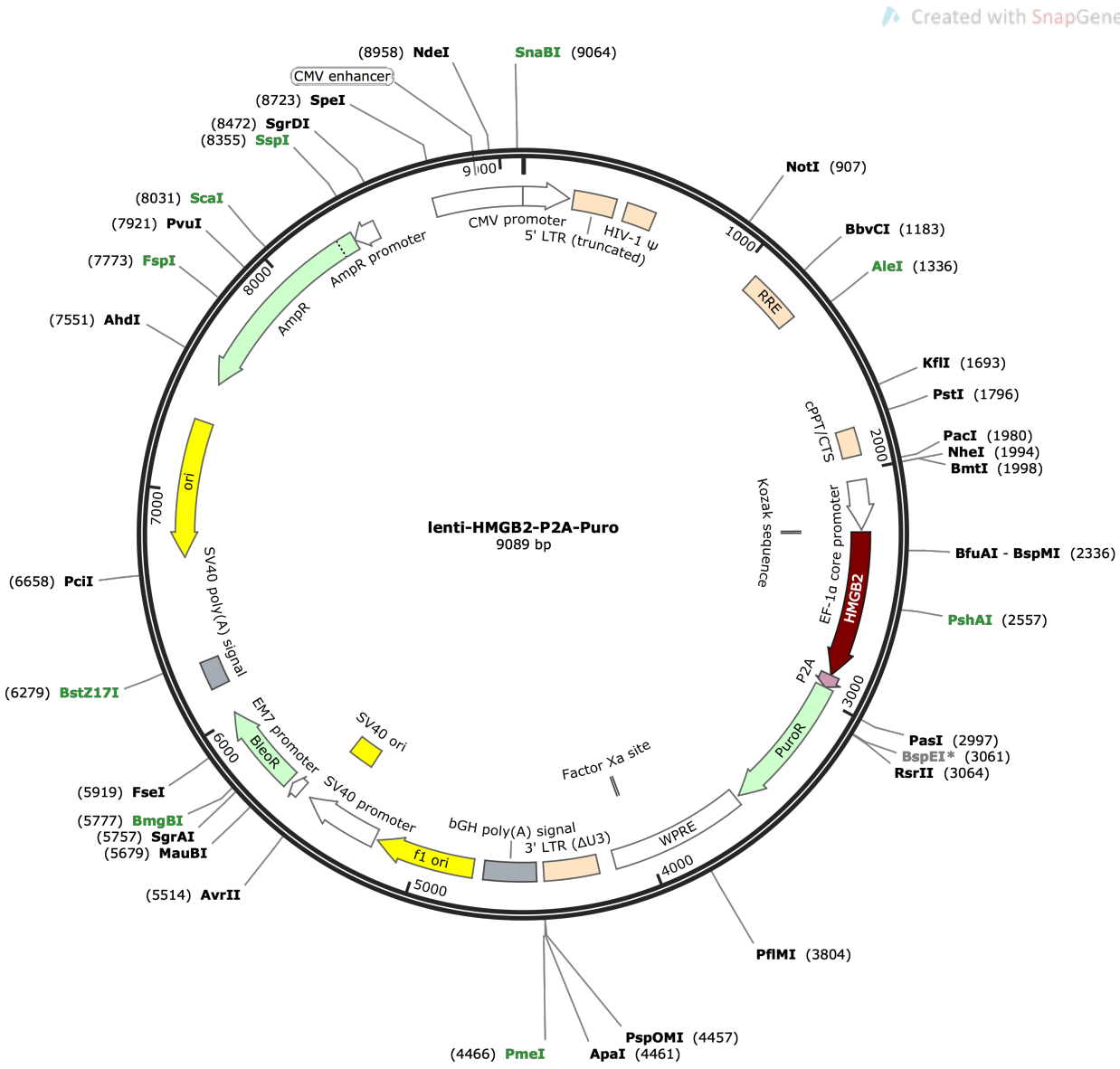


Figure 2.15 - Map of lenti-HMGB2-P2A-Puro

The Cas9 and gRNA scaffold sequences were removed from lenti-CRISPR V2 (Section 2.45) and the rat HMGB2 cDNA sequence was cloned in frame with the P2A and puromycin resistance gene. All unique restriction enzyme sites are shown. AmpR = Ampicillin resistance gene (β-lactamase); CMV = cytomegalovirus; EF-1α = human elongation factor 1α; ori = origin of replication; LTR = Long terminal repeat; RRE = Rev response element; PuroR = Puromycin resistance gene (puromycin N-acetyltransferase); P2A = porcine teschovirus-1 2A self cleaving peptide sequence; SV40 = Simian virus.

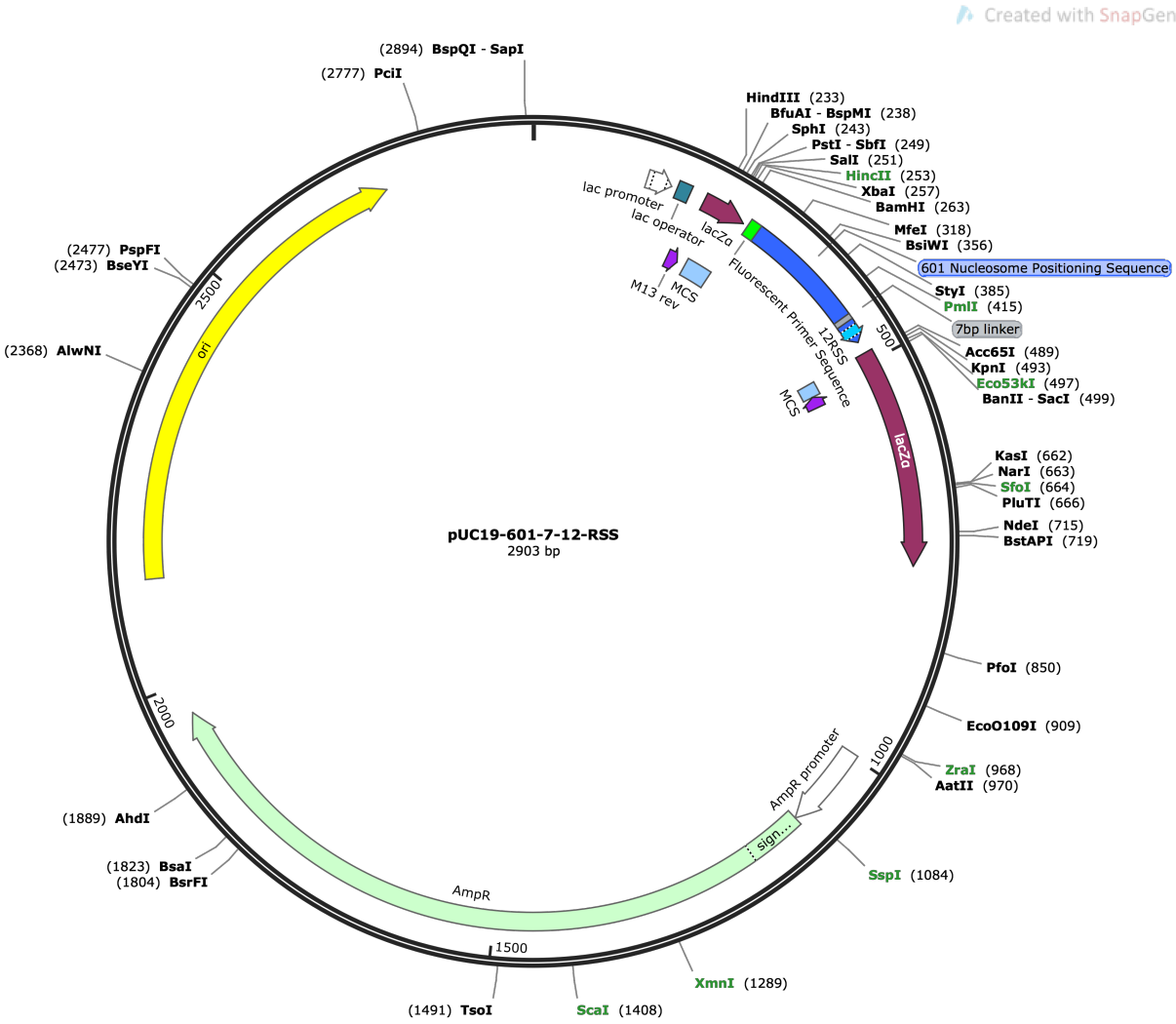


Figure 2.16 - Map of pUC19-601-7-12-RSS

A gene block containing a Widom 601 sequence adjacent to a 23-RSS was cloned into pUC19 digested with SmaI. Site-directed mutagenesis was used to mutate the 23-RSS to a 12-RSS and to introduce a 7 bp linker sequence between the 601 sequence and RSS. All unique restriction enzyme sites are shown. AmpR = Ampicillin resistance gene (β-lactamase); ori = origin of replication; MCS = multiple cloning site.

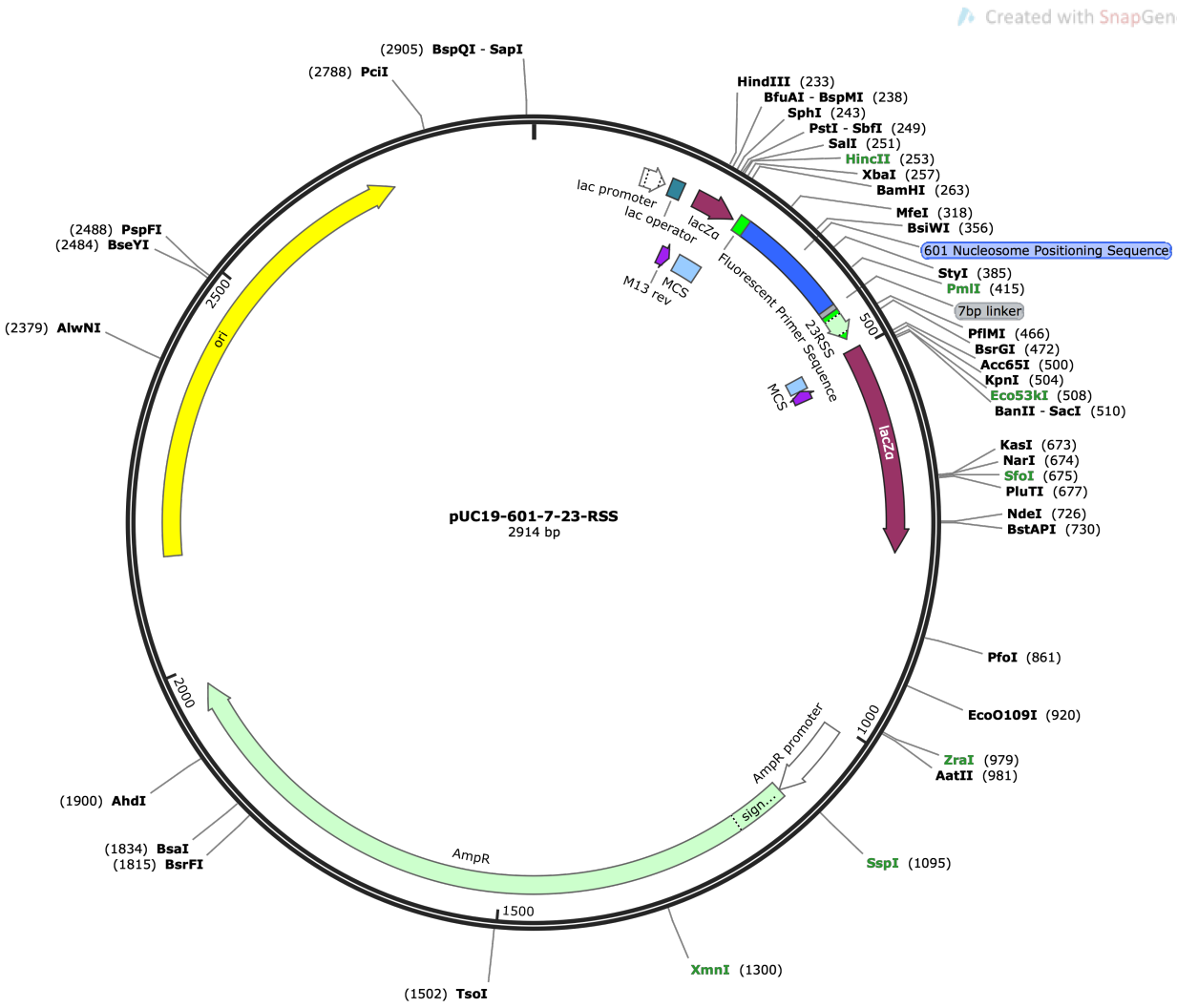


Figure 2.17 - Map of pUC19-601-7-23-RSS

A gene block containing a Widom 601 sequence adjacent to a 23-RSS was cloned into pUC19 digested with SmaI. Site-directed mutagenesis was used to to introduce a 7 bp linker sequence between the 601 sequence and RSS. All unique restriction enzyme sites are shown. AmpR = Ampicillin resistance gene (β-lactamase); ori = origin of replication; MCS = multiple cloning site.

Chapter 3 - Analysis of compound heterozygous RAG1 mutations identified in an adult patient with primary immunodeficiency

A) Introduction

It is well established that mutations within the RAG proteins that lead to a severe impairment in their function are a major cause of severe combined immunodeficiency (SCID) and Omenn syndrome, which both present very early in life. In more recent years, due to advances in high-throughput genome sequencing, hypomorphic RAG mutations are being identified in older immunodeficient patients with a less severe phenotype, such as combined immunodeficiency (CID). It is estimated that there are twice as many patients with hypomorphic RAG mutations (1:181,000) than those with inactivating mutations SCID/Omenn syndrome (1:336,000) (Kumánovics *et al.*, 2017). Therefore RAG mutations are emerging as a significant cause of many undiagnosed CID cases and there is growing interest in the study of such RAG mutations (Schuetz *et al.*, 2008; Lee *et* al., 2014; Notarangelo *et al.*, 2016; Kumánovics *et al.*, 2017; Lawless *et al.*, 2018; Tirosh *et al.*, 2018; Thwaites *et al.*, 2019; Dorna *et al.*, 2019).

In this chapter I analyse the effects of compound heterozygous RAG1 mutations from an adult patient with CID. The patient is a male who presented at the age of 34 with bronchiectasis and recurrent chest infections. Immunological assessment revealed that he had normal immunoglobulin levels but mild T cell lymphopenia. Comparison of the patient's TCRβ repertoire with that of a pool of healthy individuals revealed that usage of four TCR Vβ segments were significantly altered in the patient (Figure 3.1).

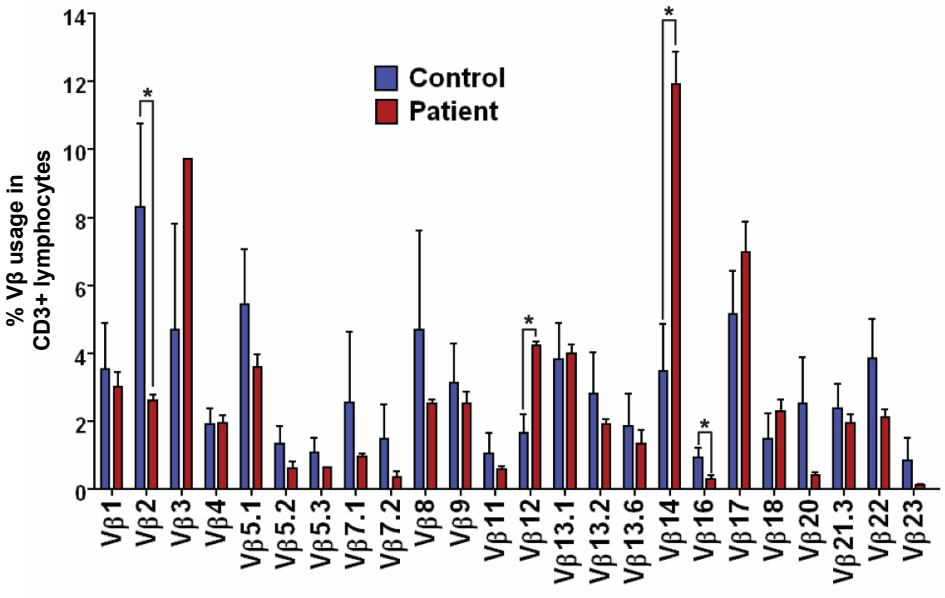


Figure 3.1 - Usage of TCRVβ gene segments in the patient

Frequency of TCRVβ gene segment use in 85 healthy control individuals (blue bars) and the immunodeficient patient (red bars). TCRVβ repertoire analysis was performed using the β Mark TCR Vβ Repertoire Kit (Beckman Coulter, IM3497), according to the manufacturer's instructions. Asterisks represent statistically significant differences between patient and control TCRVβ gene segment usage (Student's T-test with Bonferonni correction, $P = 0.01$). Error bars represent the standard deviation of three independent experiments for the patient sample. Data for this figure was provided by Dr Clive Carter.

Following genetic testing using a clinical panel of primary immunodeficiency genes, he was found to have biallelic missense mutations in RAG1: c.1210C>T p.Arg404Trp, and c.1520G>A p.Arg507Gln (Lawless *et al.*, 2018). I studied these mutations using mouse RAG1/2 as almost all previous studies of the RAG proteins use the mouse versions. There is 94% sequence similarity between human and mouse RAG1 but murine RAG1 has three fewer amino acids in the N-terminus and so the corresponding mutations in mouse RAG1 are R401W and R504Q (Figure 3.2).

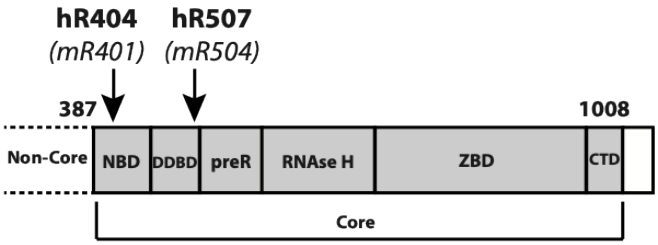


Figure 3.2 - RAG1 mutations identified in the patient

Schematic showing the main core domains of RAG1 and the position of the mutations identified in the immunodeficient patient. The equivalent amino acids in mouse are shown in brackets. NBD, nonamer binding domain; DDBD, dimerisation and DNA binding domain; preR, pre-RNase H domain; ZBD, zinc binding domain; CTD, carboxy-terminal domain.

The first part of this chapter describes the development of a qPCR based *in vivo* V(D)J recombination assay and subsequent analysis of the recombination activity of R401W and R504Q. One of these mutations, R401W, has negligible recombination activity, whereas R504Q retains more activity, accounting for the patient's relatively mild phenotype. This same pattern of decreased activity was observed in *in vitro* cleavage assays and further analysis revealed that both mutations negatively impact the binding step of V(D)J recombination. The latter part of this chapter mainly focuses on the biochemistry behind the R401W mutation and reveals a novel block in binding to the accessory proteins HMGB1. This observation suggested a vital role for HMGB1 during V(D)J recombination *in vivo*, something that has remained controversial for many years (Section 1.9). Analysis of V(D)J recombination in a HMGB1 deficient cell line revealed a severe drop in activity which was rescued by overexpression of either HMGB1 or HMGB2. Together, these data suggest that HMGB1/2 is indeed vital for V(D)J recombination *in vivo* and potentially ends a long standing controversy over its involvement in the reaction.

B) Results

3.1 Design and optimisation of an *in vivo* extra-chromosomal V(D)J recombination assay

To analyse the effect of the RAG mutations identified in the immunodeficient patient on recombination activity, I set out to develop a high-throughput quantitative V(D)J recombination assay. This assay uses an established recombination substrate, plasmid pJH299 (Hesse *et al.*, 1989), which contains consensus 12- and 23-RSSs located in the same orientation, meaning that recombination occurs by inversion (Section 1.7; Figure 3.3). Recombination of this plasmid was quantified by quantitative PCR (qPCR) and the assay is described in more detail in Section 2.26. To summarise, first round PCR was performed using a set of primers (1* and 2*; Figure 3.3) which only leads to successful PCR once recombination has occurred. The product of this reaction was then used as template in a qPCR assay using a second set of primers internal to the first (1 and 2; Figure 3.3). This nested PCR approach was required due to low levels of recombination products present in the samples. A hydrolysis probe complementary to the JH299 signal joint was used as the reporter, rather than SYBR Green, during the qPCR step to eliminate background from non-specific amplification. To account for differences during sample preparation a further qPCR assay was performed, using primers which remain unaltered by recombination (3 and 4; Figure 3.3); this allowed normalisation of the level of recombination to the total amount of pJH299 present in the sample.

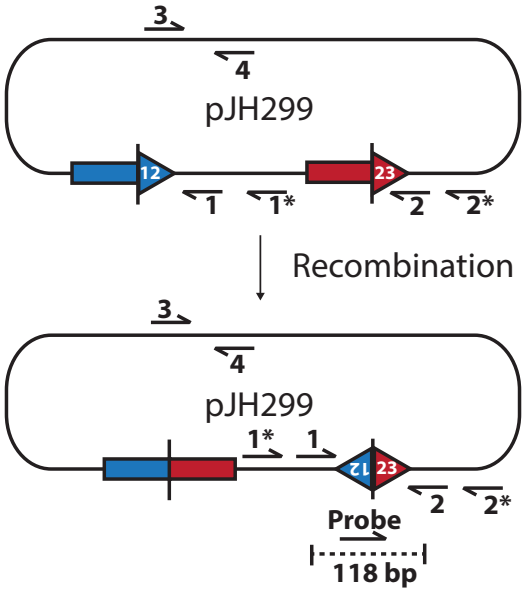


Figure 3.3 - qPCR based V(D)J recombination assay

Schematic of the *in vivo* V(D)J recombination assay. Plasmid pJH299 contains a 12- and 23-RSS (blue and red triangles) in the same orientation. Recombination can be quantified using first round primers, 1* and 2*, and second round primers, 1 and 2, plus a hydrolysis probe across the signal joint. The recombination results are normalised to the total amount of pJH299 present in the sample, which is quantified using primers 3 and 4.

Next, it was vital to determine if the assay is able to accurately quantify recombination resulting from different levels of RAG activity. To test this, NIH3T3 cells were transfected with increasing amounts of RAG1 expression vector, along with fixed amounts of pJH299 and RAG2 expression vector, pEFXC-RAG2, and the recombination activity determined. This showed that recombination activity indeed directly correlates with the amount of RAG1 expression vector, pCS2MT-RAG1, transfected and confirms that the assay can accurately quantify RAG activity (Figure 3.4).

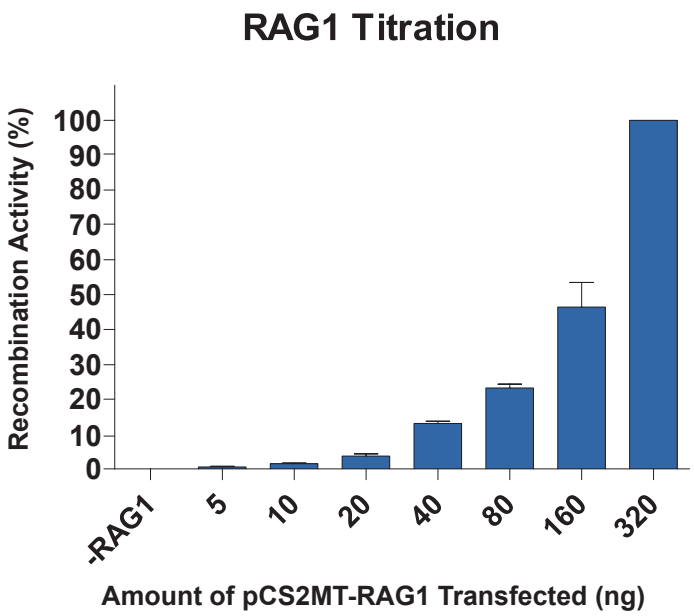


Figure 3.4 - Titration of RAG1 in the recombination assay

Increasing amounts of the RAG1 expression vector, pCS2MT-RAG1, were transfected into NIH3T3 cells and the recombination activity determined. There is a clear correlation between the amount of RAG1 present and recombination activity (*N* = 3, error bars represent SEM).

3.2 Analysis of V(D)J recombination activity of R401W and R504Q

To analyse the impact of the RAG1 mutations identified in the immunodeficient patient on recombination activity, the mutations were introduced into mouse RAG1 cDNA by Q5 mutagenesis (Section 2.21) and used in the recombination assay.

Results from the *in vivo* recombination assay revealed that the recombination activity of R401W is severely impaired with an almost complete loss of recombination (Figure 3.5). However, R504Q retains detectable levels of recombination although this is substantially reduced to around 5% of wild type. When both expression vectors were co-transfected, an intermediate level of recombination was observed (Figure 3.5).

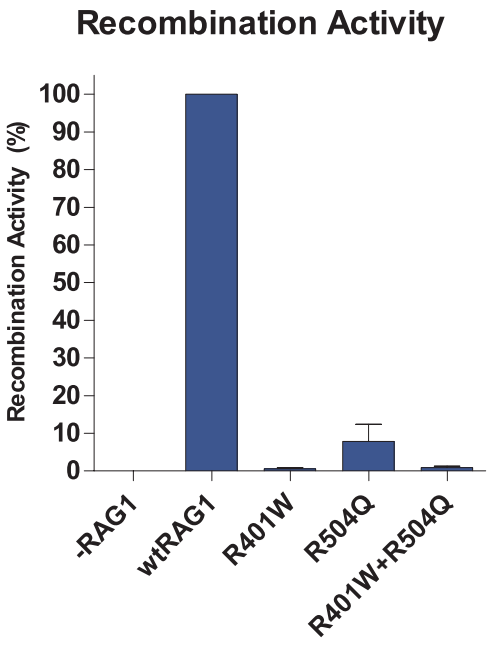


Figure 3.5 - Recombination activity of RAG1 mutants

Recombination activity of R401W and R504Q mutants relative to wild type RAG1 (*N* = 3, error bars represent SEM).

The impaired recombination activity of R401W and R504Q could be explained by decreased protein expression compared to wild type RAG1. Therefore, western blotting was performed to determine the relative expression levels of these proteins. The RAG1 expression vector, pCS2MT-RAG1, contains a 6x-Myc tag at the N-terminus of RAG1, which allowed an anti-Myc tag antibody to be used to probe for RAG1 expression. A plasmid expressing β-galactosidase (β-gal) was co-transfected to act as a transfection and loading control. Using a standard two-layer western blot protocol, β-gal expression was readily observed but no RAG1 expression could be detected in transfected NIH3T3 lysates. However, high levels of RAG1 expression could be detected in transfected COS7 cell lysates (Figure 3.6B). The transfection efficiency of COS7 cells is much higher than NIH3T3 cells (Figure 3.6C) and COS7 cells also express the SV40 large T antigen which enables replication of plasmids which contain an SV40 origin of replication, such as pCS2MT-RAG1, thus enabling higher levels of expression from such plasmids compared to NIH3T3 cells (Figure 3.6D). However, the recombination assay was not quantitative in COS7 cells. Due to these observations, a three layer

western blot approach was used to increase sensitivity and aid in detection of RAG1 expression in transfected NIH3T3 cells. Using this approach, along with long exposure times (~10 minutes), RAG1 expression was detected and this was similar for both mutants and wild type RAG1 (Figure 3.6A). This therefore implies that the reduced V(D)J recombination activity observed with R401W and R504Q is due to impaired RAG1 function and not altered protein expression.

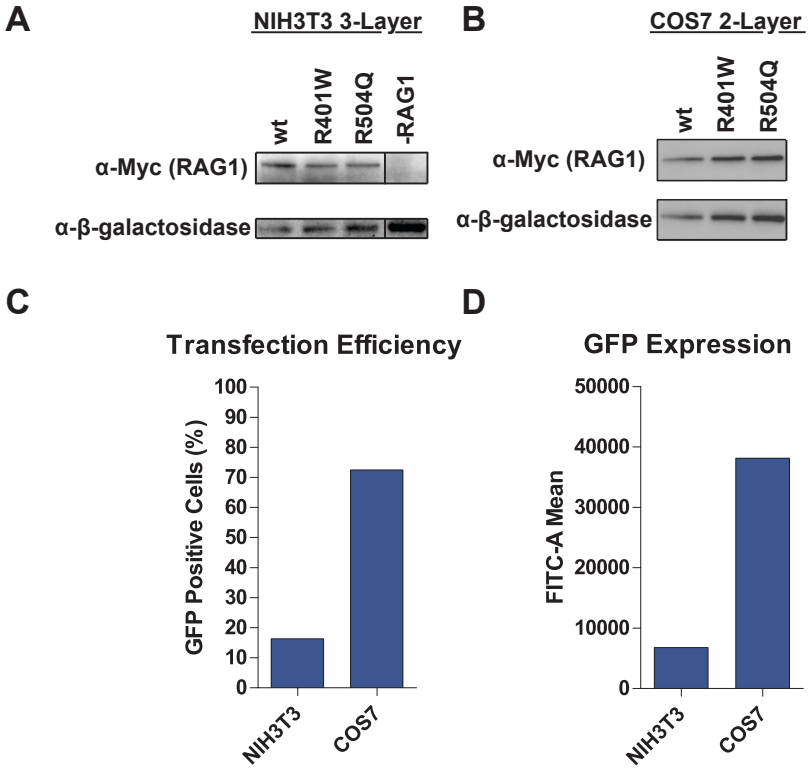


Figure 3.6 - Relative expression of RAG1 mutant proteins

(A) Western blot of NIH3T3 lysates transfected with wild type or mutant RAG1 and a β-galactosidase expression vector. RAG1 expression was only detected in NIH3T3 lysates when a 3-layer approach was used. Samples analysed on this western blot were from an independent transfection than those analysed for recombination in Figure 3.5. See Appendix 1 for the uncropped blot. (B) Western blot of COS7 lysates transfected with wild type or mutant RAG1 and a β-galactosidase expression vector. RAG1 expression was readily detected using a standard 2-layer western blot approach. (C) Transfection efficiency of NIH3T3 and COS7 cells was determined by transfection with pcDNA3-Clover and analysis of the number of GFP positive cells 48 hours post-transfection by FACs. (D) GFP expression of NIH3T3 and COS7 cells was determined by measurement of the mean fluorescence (FITC-A) of pcDNA3-Clover transfected cells by FACs.

3.3 Purification of wild type and mutant core RAG proteins for *in vitro* studies

V(D)J recombination can be divided into three distinct stages; (i) RAG binding, (ii) cleavage, and (iii) joining. Both binding and cleavage steps can be effectively analysed *in vitro* using established assays. To investigate the molecular basis of the reduced recombination levels with R401W and R504Q, *in vitro* RAG cutting and binding assays were performed. This required purified wild type and mutant RAG proteins. Due to solubility problems with full length RAG proteins, the catalytically active core RAG proteins were purified. Crucially, the mutations of interest are within the core region of RAG1.

The R401W and R504Q mutations were introduced into the RAG1 expression vector previously produced in the Boyes laboratory (pEF-McR1). This, along with pEF-McR2, contain the cDNA sequence of mouse core RAG1 (cRAG1, aa 384-1008) or mouse core RAG2 (cRAG1, aa 1-382) fused to an N-terminal maltose binding protein (MBP) amylose affinity tag (Figure 3.7A). These vectors were co-transfected into HEK293T cells which is a common mammalian cell line that is easily transfected and expresses the SV40 large T antigen, allowing high levels of expression of the SV40 origin containing pEF vectors. cRAG1 and cRAG2 were co-expressed and purified as this increases the yield and activity of the purified RAG proteins compared to when these proteins are individually expressed (Swanson, 2004).

HEK293T transfected cells were lysed according to an established core RAG purification protocol and MBP-RAG proteins were purified from the lysate via amylose affinity chromatography (Section 2.35). Fractions of eluted proteins were analysed by SDS-PAGE and those containing RAG proteins were pooled. A representative coomassie blue stained SDS-PAGE gel of fractions from a core RAG purification is shown in Figure 3.7B. After purification of all the required wild type and mutant RAG proteins, each preparation was separated on an SDS-PAGE gel and their relative concentrations were determined by 2D densitometry using ImageJ. Equal amounts of each RAG preparation were then used for *in vitro* studies (Figure 3.7C).

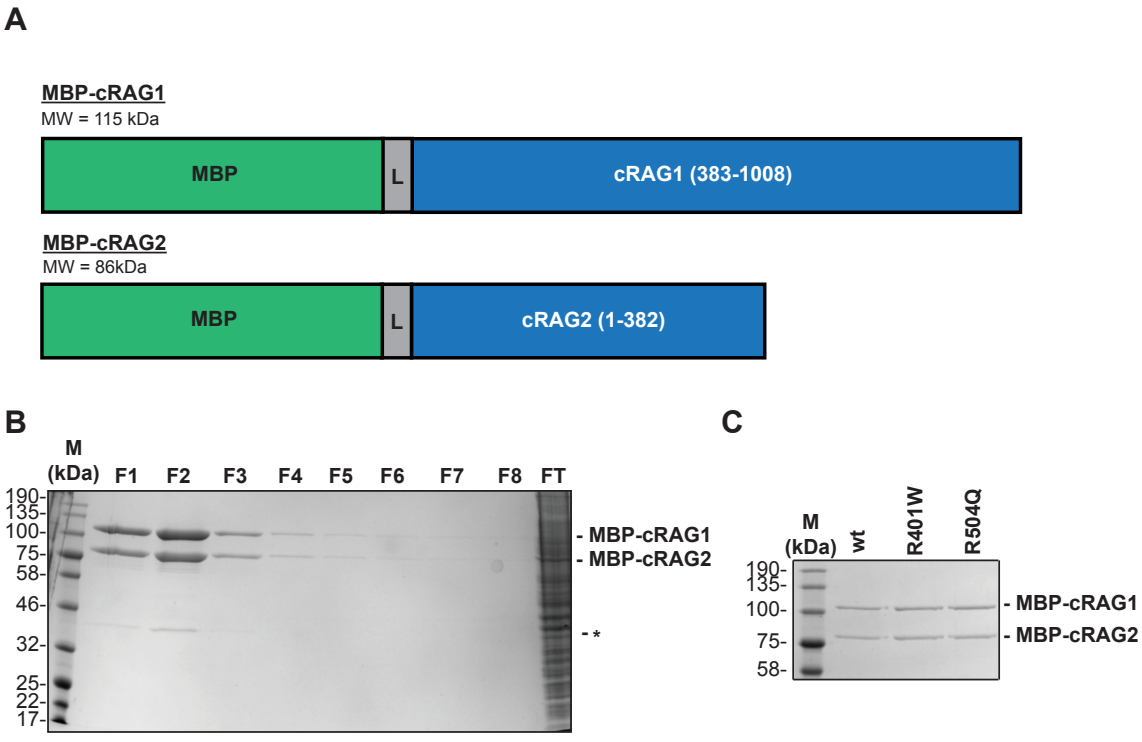


Figure 3.7 - Purification of core RAG proteins

(A) Diagram representing MBP-cRAG1 and MBP-cRAG2 fusion proteins. The predicted molecular weights (MW) of each protein are shown. Maltose binding protein (MBP) is shown in green, followed by a short glycine-serine linker (L) in grey and either core RAG1 (cRAG1) or core RAG2 (cRAG2) in blue. The amino acids of cRAG1 and cRAG2 present are shown in brackets. (B) Representative coomassie stained SDS-PAGE gel showing fractions (F1-F8) and flow through (FT) from a standard core RAG purification. The asterisk at ~40 kDa indicates an inert contaminant. (C) Coomassie stained SDS-PAGE gel showing equivalent levels of mutant and wild type RAG proteins from each purification.

3.4 Cleavage activity of RAG1 mutants is greatly impaired *in vitro*

Coupled RAG cleavage can be analysed by incubation of purified RAG proteins with an end-labelled 12- or 23-RSS containing oligonucleotide and a relevant unlabelled partner RSS, in the presence of magnesium (Section 2.37). The cleavage activity of both RAG1 mutants was analysed and compared with that of wild type. This revealed that cleavage of a 12- and 23-RSS was greatly reduced for both R401W and R504Q (Figure 3.8). R401W had almost undetectable levels of cleavage under these conditions whereas R504Q was reduced to around 15-fold lower than wild type RAG1.

Furthermore, the reduction in cleavage activities of both RAG1 mutants mirrors the decrease in recombination activity observed *in vivo* (Figure 3.5).

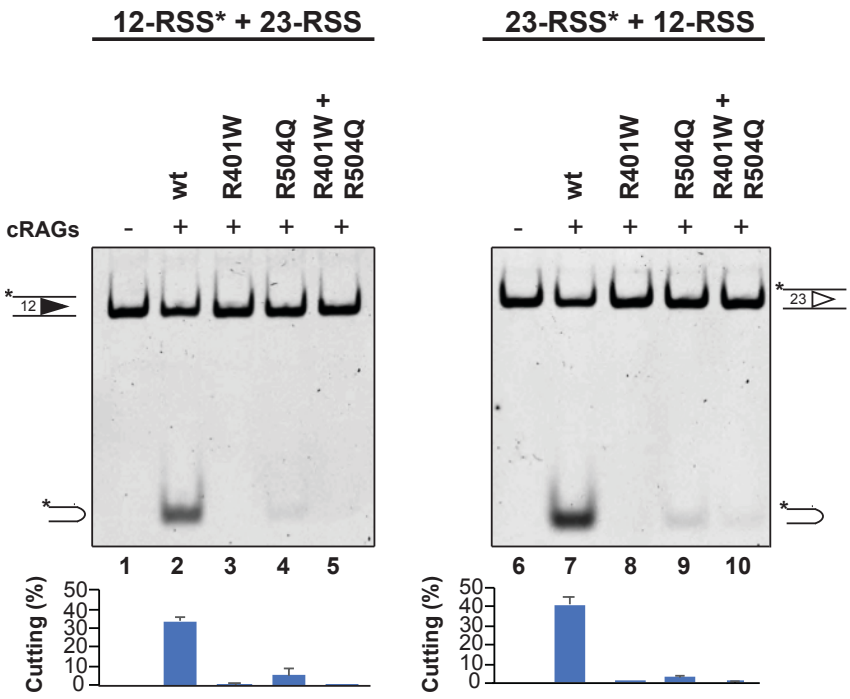


Figure 3.8 - Cleavage activity of wild type and mutant RAG1 proteins

Cleavage of a fluorescently labelled 12- (left) or 23-RSS (right) is shown. The labelled RSS oligonucleotide was incubated with equivalent amounts of wild type or mutant RAG proteins (Figure 3.7C) plus purified HMGB1, in a reaction buffer containing 1 mM magnesium chloride. Under these conditions a partner RSS is required for cleavage and so a 10-fold excess of unlabelled partner 12- or 23-RSS was included. Symbols next to gels represent a labelled (*) 12-RSS (black triangle), 23-RSS (white triangle) or a cleaved coding end (hairpin). The amount of cutting was determined using 2D densitometry in AIDA and is indicated on the graphs below each gel (*N* = 3, error bars represent SEM).

3.5 Binding of RAG1 mutants to RSSs is altered *in vitro*

As cleavage activity of the mutants mirrored *in vivo* recombination activity, RAG binding was analysed next to determine if this step is affected by the mutations. Binding of purified RAG proteins to end-labelled 12- and 23-RSS oligonucleotides is performed in a similar way to the RAG cleavage assay but the divalent cation used is calcium, which allows RAG binding but not cleavage (Section 2.38). Assembled RAG complexes are

separated by native PAGE and two major complexes are observed: SC1 and SC2 (Figure 3.9A, lanes 2 and 8). Both complexes contain two molecules of RAG1 but differ in the number of RAG2 molecules present, one or two respectively. Purified HMGB1 can be added to RAG reactions to increase activity (van Gent *et al.,* 1997) and this leads to the formation of complexes with slightly slower mobility, known as HSC1 and HSC2 (Figure 3.9A, lanes 3 and 9), where HSC2 is thought to form the paired complex with the partner RSS for cleavage (Swanson, 2004).

The binding assay was performed using wild type and mutant RAG proteins which were incubated with a labelled 12- or 23-RSS oligonucleotide and relevant unlabelled partner RSS, in the presence or absence of HMGB1. The expected 12- and 23-RSS complexes were formed with R504Q, although the amount of binding appears reduced when compared with wild type RAG1 (Figure 3.9A, compare lanes 3 and 9 with lanes 5 and 11). A surprising effect was observed however, with complexes formed with R401W. In the absence of HMGB1, R401W can form SC1 and SC2 complexes with a 12-RSS (Figure 3.9B). However, in the presence of HMGB1, HSC1 is barely formed and the majority of complexes appear similar to SC1, whilst HSC2 levels appear reduced (Figure 3.9A lane 4; Figure 3.10B lane 5). In contrast, R401W can form HSC1 and HSC2 complexes at a 23-RSS but appear to be at reduced levels when compared with wild type and have a slightly faster mobility (Figure 3.9A, compare lanes 9 and 10).

These observations suggest that both RAG1 mutations negatively impact the binding step of V(D)J recombination which could thus cause the subsequent decrease in cleavage activity and overall V(D)J recombination.

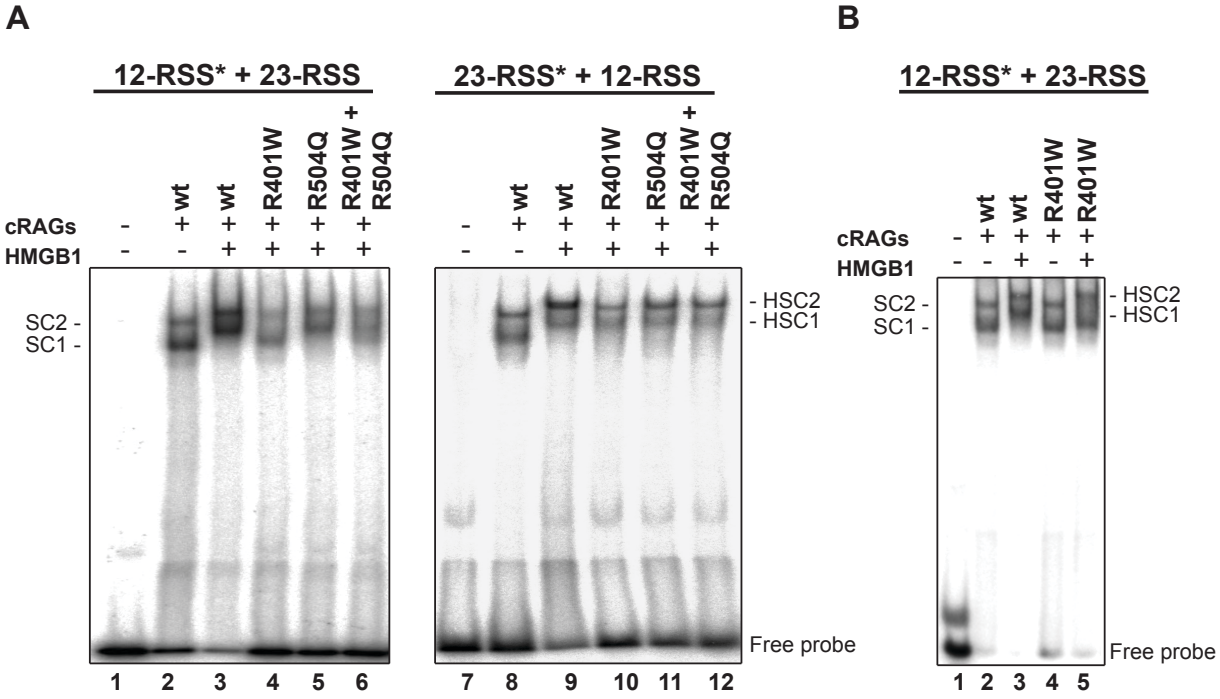


Figure 3.9 - Binding activity of wild type and mutant RAG proteins

A) Binding of wild type and mutant RAG proteins to a radioactively labelled 12- (left) or 23-RSS (right) is shown. Reactions were performed as for the cleavage assay but calcium replaced magnesium as the divalent cation. Under these conditions, RAG proteins can bind but not cleave the RSS, forming complexes SC1 and SC2 which are indicated next to the gels. Addition of HMGB1 leads to the formation of the slower migrating complexes HSC1 and HSC2. (B) Comparison of the complexes formed with wild type or R401W on a 12-RSS in the absence or presence of purified HMGB1. R401W forms complexes equivalent to wild type RAG1 in the absence, but not the presence, of HMGB1.

3.6 Location of R401 and R504 in the RAG complex

To better understand the molecular basis of the decreased binding of both RAG1 mutations, the positions of R401 and R504 were analysed using published high resolution structures of the RAG proteins in complex with a 12- and 23-RSS (Figure 3.10). R504 lies within the dimerisation and DNA binding domain (DDBD) of RAG1 and contacts the DNA backbone of the first two bases in the spacer immediately following the heptamer and, along with H501, is thought to anchor a three helix bundle of the DDBD to the major grove at the heptamer-spacer boundary (Kim *et al.*, 2018). Mutation

of R504 to glutamine causes loss of the positive charge, which would likely destabilise RAG interaction with this region of the RSS.

R401 is located within part of the NBD which contacts the compressed major groove found at the nonamer-spacer boundary. The spacer backbone is contacted by R401, along with other positively charged residues in this region such as K405 and K412 (Kim *et al.*, 2018). Similar to R504Q, mutation of R401 to tryptophan leads to the loss of a positive charge but also introduces a bulky group. Surprisingly, the position of R401 in the complex is directly opposite the binding site for HMGB1 Box A at both a 12- and 23-RSS. Two densities of HMGB1 binding, corresponding to Box A and Box B, can be observed at a 23-RSS whereas a single density, corresponding to Box A, is observed at a 12-RSS. The loss of positive charge or introduction of a bulky group in this region may affect binding of HMGB1 Box A and this may explain the observation that R401W inhibits HMGB1 binding at a 12-RSS but not at a 23-RSS, as Box B binding in the 23-RSS spacer (Figure 3.10, right) would be still able to occur.

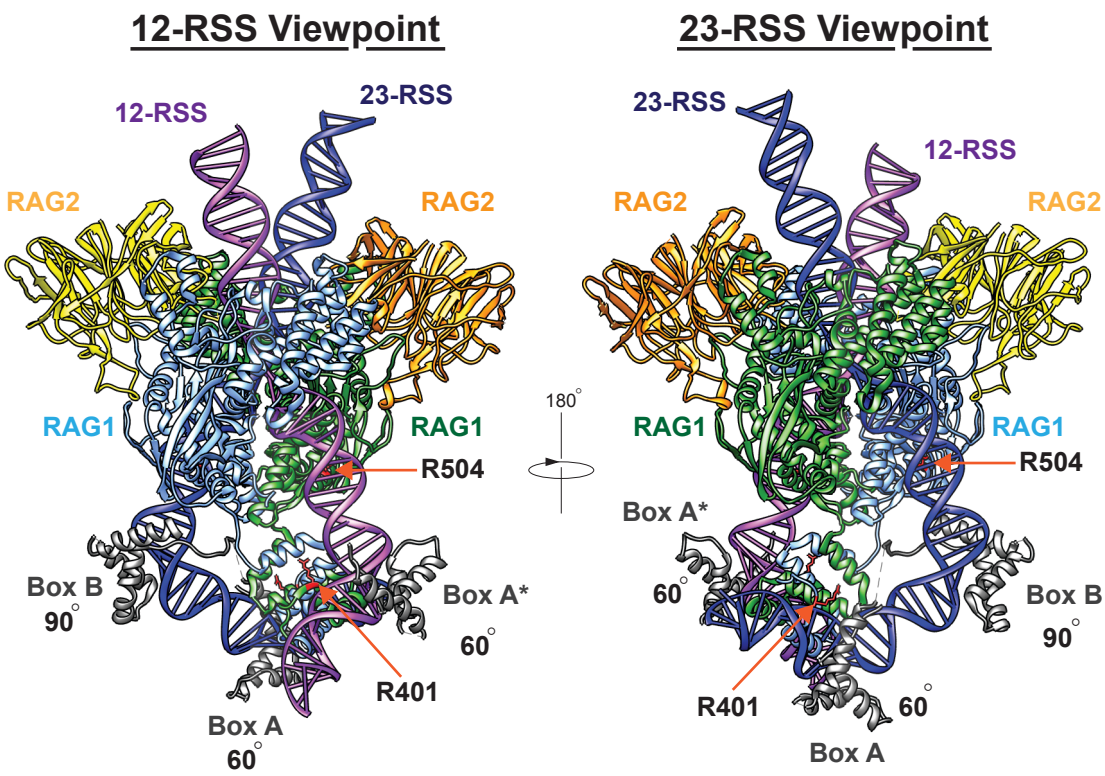


Figure 3.10 - Positions of R401 and R504 in the RAG:RSS complex

Structure of RAG:RSS complex from a 12- (left, pink DNA molecule) and 23-RSS (right, dark blue DNA molecule) viewpoint. The two monomers of RAG1 are represented as light blue or green cartoons whereas the two monomers of RAG2 are represented as yellow or orange cartoons. R401 and R504 residues in each RAG1 monomer are shown as red stick models and their positions are indicated with an orange arrow. R504 is located in the middle of the complex, near the first two bases of the spacer after the heptamer, for both the 12- and 23-RSS. R401 is found towards the base of the complex in the NBD for both the 12- and 23-RSS, directly opposite the binding site of HMGB1 Box A (grey cartoon). Box A at the 12-RSS was not fully resolved in the structures and so the structure of Box A was superimposed from the 23-RSS side. Data from PDB: 6CG0.

3.7 R401W does not bind HMGB1 correctly

To test if R401W indeed alters HMGB1 binding, an antibody super-shift assay was employed. HMGB1 that contains an N-terminal 6xHis tag was purified and the same amount of His-tagged HMGB1 was used as in previous RAG binding reactions (Figure 3.11A). RAG complexes formed at a 12-RSS with His-tagged HMGB1 are very similar to those formed with untagged HMGB1 (Figure 3.11B). Addition of an anti-His antibody

led to a super-shift of all HMGB1 containing complexes with wild type RAG1, whereas only the small amount of HSC2 formed with R401W was super-shifted. The dominant, faster migrating complex remained unchanged (Figure 3.11B). This supports the idea that HMGB1 binding at a 12-RSS is indeed inhibited by the R401W mutation.

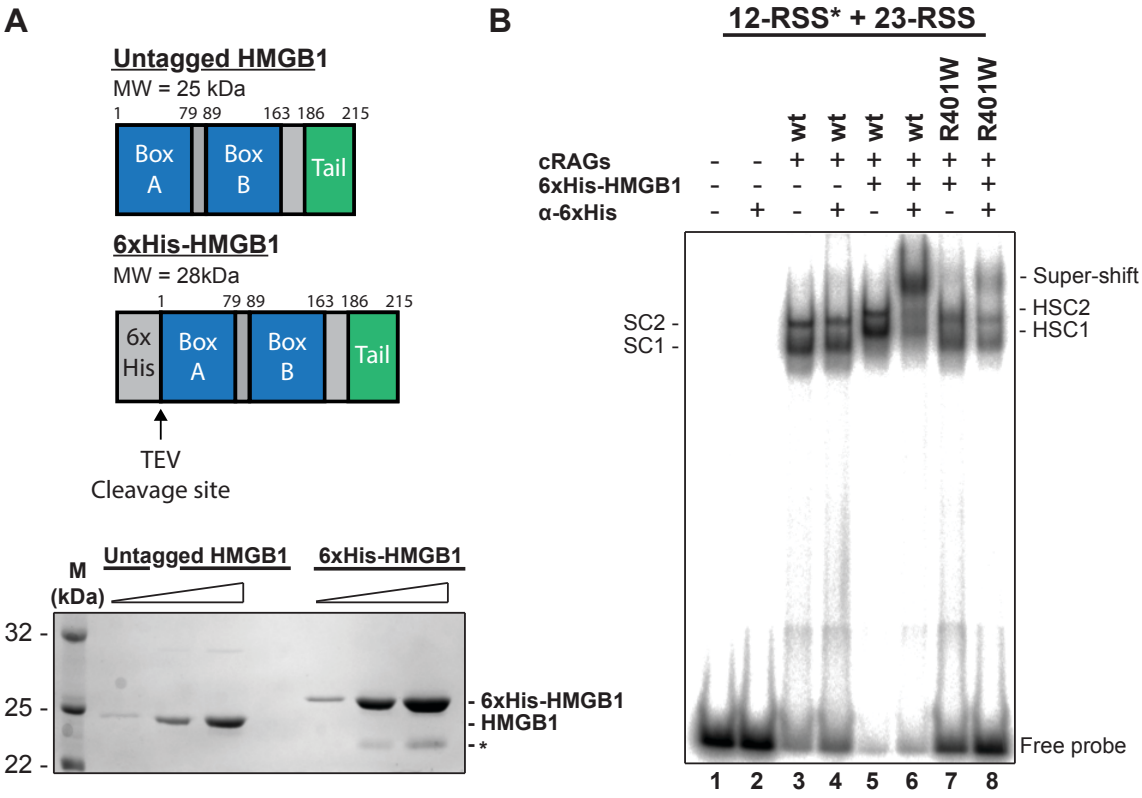


Figure 3.11 - Analysis of HMGB1 stoichiometry in wild type and R401W complexes

(A) Diagram of untagged and 6xHis tagged HMGB1 proteins and predicted molecular weights are shown (top). Coomassie blue stained SDS-PAGE gel showing a 3-fold serial dilution of purified untagged HMGB1 or 6xHis-HMGB1 (bottom). The asterisk at ~23 kDa represents an inert contaminant. (B) Super-shift experiment shows that HMGB1 binds less well to complexes formed with R401W. Wild type or R401W RAG1 proteins, together with wild type RAG2 were assembled on a labelled 12-RSS in the presence or absence of 6xHis HMGB1. Super-shifted complexes due to the addition of an anti-His tag antibody are indicated.

To test if weak HMGB1 binding by R401W can be overcome at higher concentrations of HMGB1 to restore cleavage activity, RAG cleavage assays were performed using increasing amounts of HMGB1. Whilst cleavage activity of wild type RAG1 increases as HMGB1 concentration increases, the activity of R401W remains unchanged (Figure 3.12A).

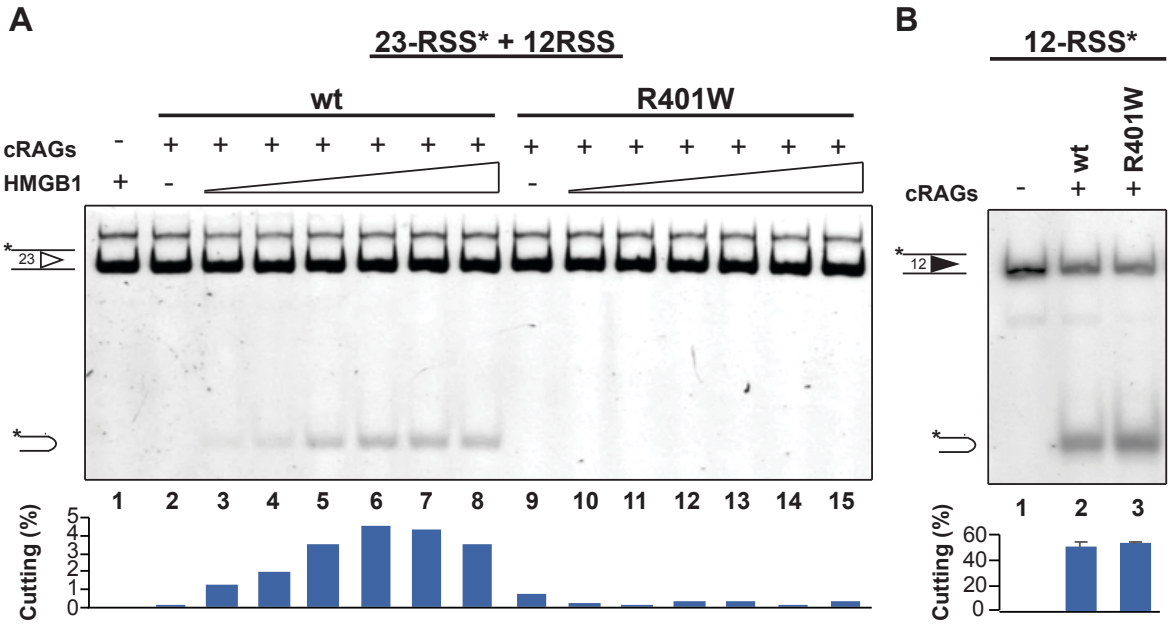


Figure 3.12 - Analysis of HMGB1 dependent and independent RAG cleavage by R401W

(A) Increasing amounts of HMGB1 does not restore cleavage activity of R401W. RAG cleavage assays were performed using wild type or R401W in the presence of increasing amounts of HMGB1 (0, 25, 50, 100, 200, 400, 800 nM). Reactions were performed as in Figure 3.9 but DMSO was omitted so the effect of HMGB1 on cleavage could be more reliably assayed (Bergeron *et al.*, 2006). Cleavage is negligible with wild type RAG in the absence of HMGB1 (lane 2) but increases as more HMGB1 is added. By contrast, R401W cleavage activity remains negligible regardless of HMGB1 concentration. (B) R401W has catalytic activity. Cleavage of a 12-RSS alone was performed using wild type RAG1 or R401W in the presence of manganese. Under these less stringent conditions, where a partner RSS or HMGB1 is not required, cleavage is readily observed with R401W at a level similar to wild type RAG1 (*N* = 3, error bars represent SEM).

Although it is clear that R401W is deficient in HMGB1 binding and HMGB1 is unable to stimulate cleavage by R401W, it remains possible that the lack of cleavage activity by R401W is independent of HMGB1 binding. To test this, RAG cleavage assays were performed in the presence of manganese. This allows cleavage of a single 12-RSS in the absence of a partner 23-RSS. Additionally, as HMGB1 is required to correctly position both RSSs in the complex for coupled cleavage, cleavage of a single 12-RSS will be independent of the requirement for HMGB1. Using these less stringent

conditions, cleavage activity by R401W was readily observed at similar levels to wild type RAG1 (Figure 3.12B). Therefore, R401W has catalytic activity but is unable to support coupled cleavage due to deficient HMGB1 binding.

3.8 Restored HMGB1 binding coincides with restored cleavage and V(D)J recombination activity

The R401W mutation could reduce HMGB1 binding in two potential ways. Firstly, introduction of a bulky tryptophan group close to the HMGB1 binding site could block its binding. Secondly, the loss of arginine's positive charge could disrupt interactions with the RSS spacer and reduce its bending by the RAG complex which is thought to be a prerequisite for HMGB1 binding. To investigate these possibilities, two further R401 mutations were generated. Firstly R401 was mutated to leucine, which has no charge but is a similar size to arginine, and secondly, R401 was mutated to lysine which retains the positive charge. The ability of R401L and R401K to bind HMGB1 was determined and the *in vitro* cleavage and *in vivo* recombination activity was analysed, to decipher which mutation, if any, would restore activity.

R401L partially restores HMGB1 binding to the 12-RSS, forming slightly higher levels of HSC2 but with a faster mobility than HSC2 formed with wild type RAG1 (Figure 3.13A). A smeary faster migrating complex is also formed with R401L which has intermediate mobility between SC1 and HSC1 (Figure 3.13A). This suggests that the bulky nature of R401W contributes partly to the block of HMGB1 binding. However, cleavage activity and *in vivo* recombination activity of R401L remains negligible (Figure 3.13B and C). Therefore, although HMGB1 binding is partially supported by R401L, it still leads to a non-functional complex, which is consistent with the smeary nature of the complexes formed and slightly faster mobility of HSC2 compared with wild type RAG1. This is indicative of less bent DNA, which would migrate faster through the gel.

In contrast, R401K forms complexes with HMGB1 that resemble those of wild type RAG1, although at a slightly lower levels (Figure 3.13A). This is accompanied by near

wild type levels of cleavage activity and *in vivo* recombination activity (Figure 3.13B and D). The restoration of HMGB1 binding and RAG activity with R401K suggests an important role for the positive charge of R401 in supporting HMGB1 binding. Additionally, the slightly lower binding and recombination activities of R401K could be explained by the weaker positive charge imposed by lysine which can form contacts in just one direction compared with three for arginine (Sokalingam *et al.*, 2012). These data therefore imply that V(D)J recombination depends on HMGB1 binding, and could shed light on the much debated role of HMGB1 during V(D)J recombination *in vivo*.

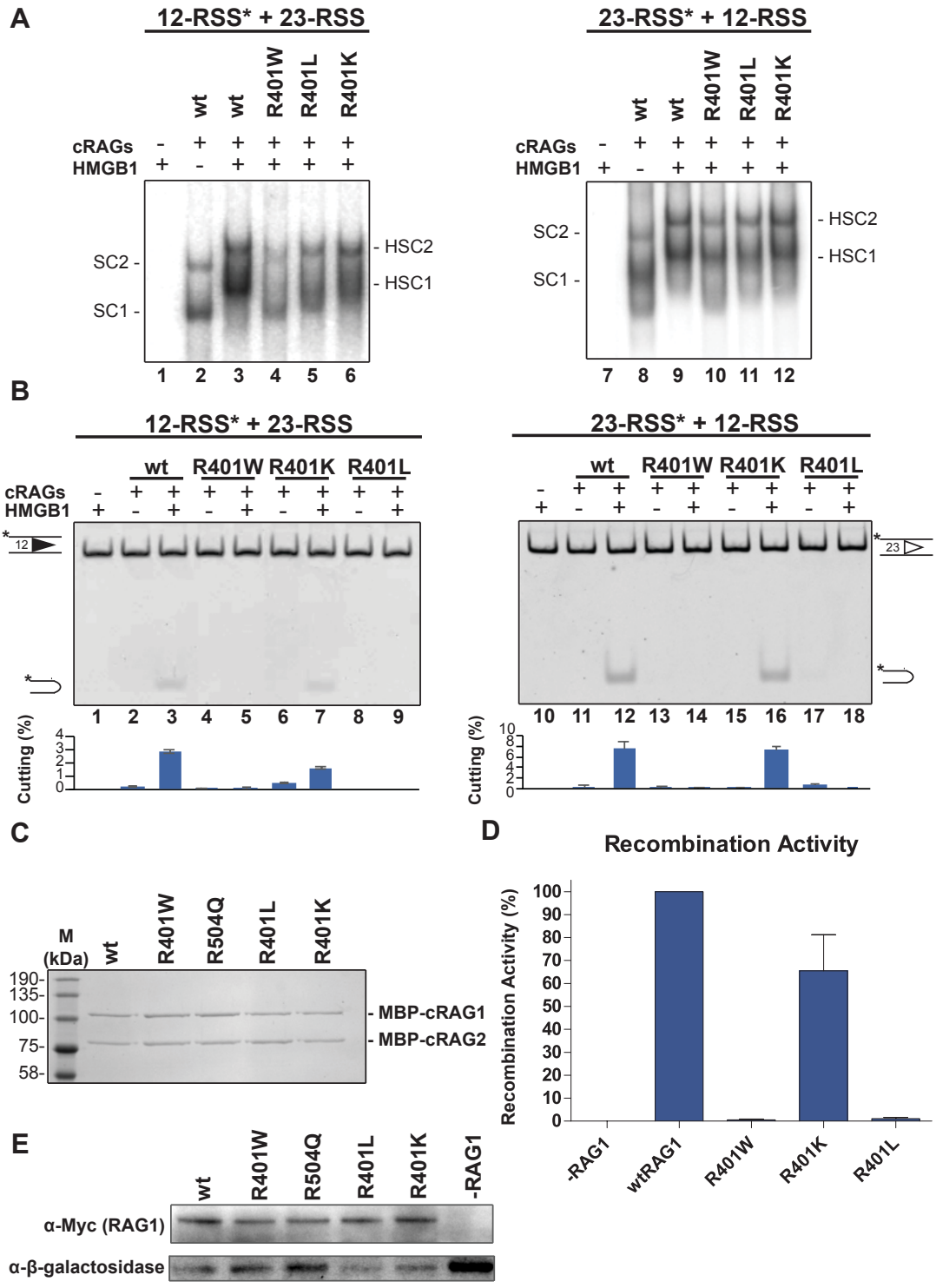


Figure 3.13 - Restored HMGB1 binding corresponds with restored RAG activity

(A) HMGB1 binding is progressively restored when R401 is mutated to leucine and then lysine. Binding reactions were performed as in Figure 3.9 but gels were run for a longer time to better resolve the complexes. (B) RAG cleavage is only restored with the R401K mutation. Cleavage reactions were performed as in Figure 3.8, without DMSO. (*N* = 3, error bars represent SEM). (C) Equivalent amounts of each separate protein preparation were used in cleavage and binding assays. Wild type or mutant RAG protein preparations were separated by SDS-PAGE

and stained using coomassie blue. (D) Recombination activity with R401K is also restored. The *in vivo* recombination assay was performed using R401W, R401L and R401K as in Figure 3.5 and recombination levels are shown relative to wild type RAG1 (*N* = 3, error bars represent SEM). (E) Western blot analysis of NIH3T3 lysate transfected with wild type or mutant RAG1 proteins shows similar levels of expression. Samples analysed on this western blot were from an independent transfection than those analysed for recombination in Figure 3.13D. See Appendix 1 for the uncropped blot.

3.9 Generation of a *Hmgb1* knockout cell line by CRISPR-Cas9

If HMGB1 is indeed vital for V(D)J recombination *in vivo*, then recombination in the absence of HMGB1 should be negligible. Furthermore, if HMGB1 expression is restored, then this should coincide with restoration of recombination in the presence of wild type RAG1 or R401K but not R401W. In order to test this idea, a HMGB1 deficient NIH3T3 cell line was required and to this end, CRISPR Cas9 DNA editing technology was used to delete the second exon of *Hmgb1.* This codes for the first 50 amino acids of the protein (Figure 3.14B) and deletion of this region can be achieved by simultaneous Cas9 cleavage at two sites in the genome, leading to deletion of the intervening sequence.

Therefore, two CRISPR single guide RNA (sgRNA) sequences were designed upstream and downstream of *Hmgb1* exon 2 and cloned into lenti-CRISPR-Puro V2 or lenti-sgRNA-Zeo respectively (Section 2.42) (Figure 3.14A and B). The resulting lentiviruses were used to transduce NIH3T3 cells and puromycin and zeocin treatment was used to select for transduced cells. Puromycin resistant cells express both the upstream sgRNA and the Cas9 protein whereas zeocin resistance cells express just the downstream sgRNA; therefore co-selection ensure all cells express both sgRNAs and the Cas9 protein. Lentiviral transduction was used rather than transfection due to the very low transfection efficiency of NIH3T3 cells (Figure 3.6C). Indeed, during antibiotic selection nearly all cells survived, suggesting a very high transduction efficiency of NIH3T3 cells by lentivirus.

To confirm successful deletion of *Hmgb1* exon 2, genomic DNA was extracted from the pool of transduced cells and PCR was performed using primers which span the deleted region (Figure 3.14B, primers 1 and 2; Section 2.42). The resulting PCR product was separated by agarose gel electrophoresis and bands corresponding to the wild type (867 bp) and deleted sequence (468 bp) were readily observed (Figure 3.14C). Subsequent gel extraction, cloning and Sanger sequencing of the ~468 bp band confirmed the presence of the correct deletion in the pool of transduced cells. HMGB1 protein expression in this population of cells was also reduced to around 30% of wild type NIH3T3 cells (Figure 3.14D). Monoclonal cell populations were isolated by plating a single cell into each well of a 96 well plate by flow cytometry. After three weeks of culturing, only seven wells contained cells, suggesting that NIH3T3 cells might not survive well as single cells. Genomic DNA was extracted and the copy number of *Hmgb1* exon two was determined by qPCR using primers located within the deleted region (Figure 3.14B, primers 3 and 4). Unfortunately, after normalising to the number of copies of *Hprt* present, all seven cell lines retained at least one copy of *Hmgb1* exon 2, implying that a homozygous *Hmgb1* knockout cell line had not been generated (Figure 3.14E).

I later discovered from the literature that NIH3T3 cells contain four *Hmgb1* alleles due to a genome duplication event (Leibiger *et al.*, 2013). This would therefore decrease the likelihood of generating a *Hmgb1* knockout and means a large number of monoclonal cell lines would need to be screened to identify a homozygous *Hmgb1* knockout. Due to difficulties encountered with single cell culture of NIH3T3 cells, this seemed impractical. Therefore a different approach was taken.

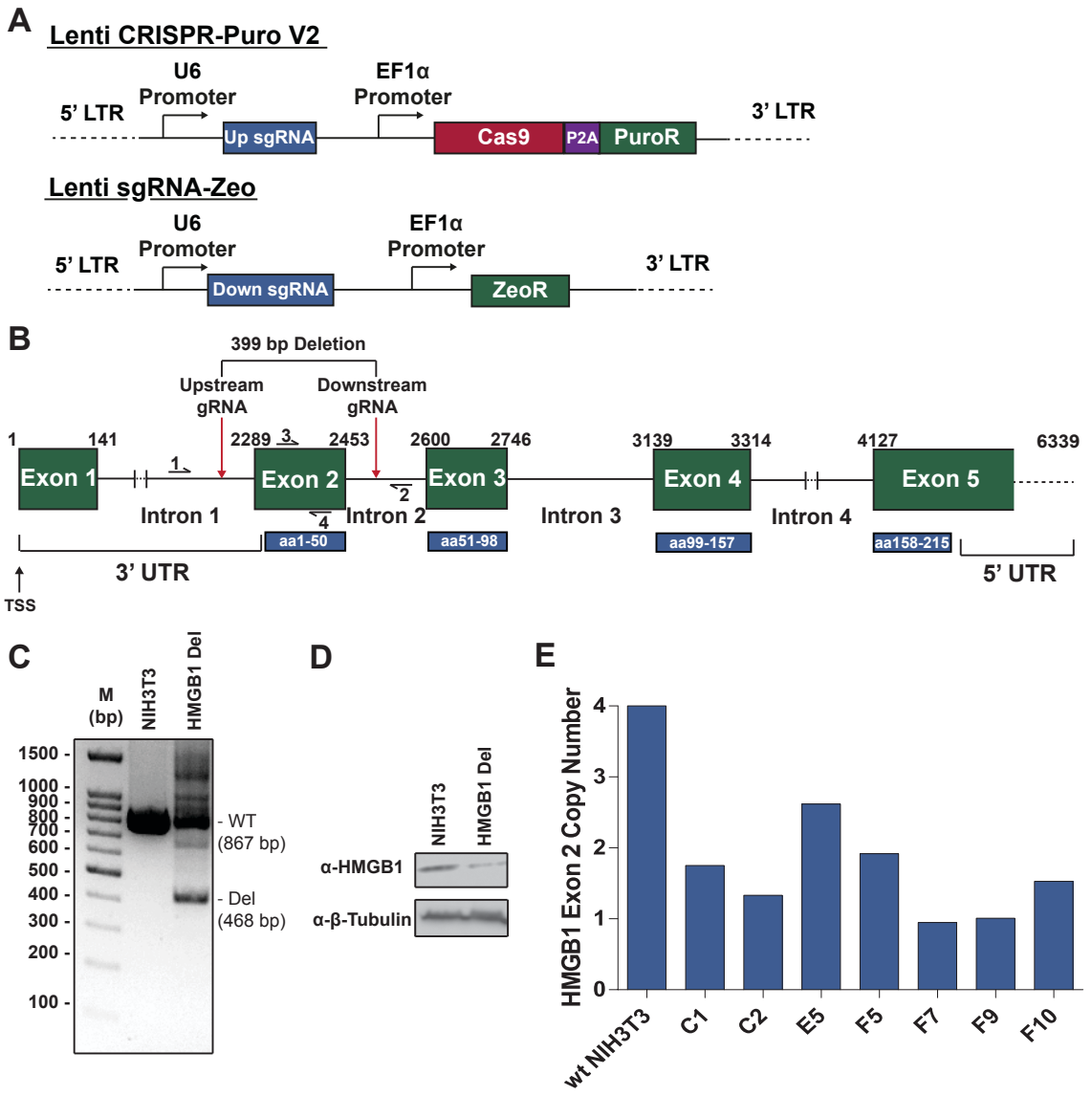


Figure 3.14 - Strategy to generate a *Hmgb1* knockout in NIH3T3 cells

(A) Diagram of the lentiviral cassettes used to transduce NIH3T3 cells. Lenti CRISPR-Puro V2 (top) contains both the *Hmgb1* exon 2 upstream sgRNA and the Cas9 cDNA sequence, along with the puromycin resistance gene. Lenti sgRNA-Zeo (bottom) contains the *Hmgb1* exon 2 downstream sgRNA and the zeocin resistance gene. (LTR, long terminal repeat). (B) Organisation of the murine HMGB1 locus. Exons are shown in green and the HMGB1 protein coding sequence is in blue. Red arrows denote the approximate location of the upstream and downstream sgRNA sequences. Numbered half arrows represent primer sequences. (UTR, untranslated region). (C) PCR analysis, using primers 1 and 2, of the polyclonal population of NIH3T3 cells expressing both guide RNA sequences and the Cas9 protein. A band corresponding to the unaltered sequence (867 bp) is observed in both wild type and transduced cells, whereas a band corresponding to a deletion of *Hmgb1* exon 2 (468 bp) is only seen in the

transduced cells. (D) HMGB1 protein expression is reduced in the population of cells lacking *Hmgb1* exon 2. Western blotting was performed using an anti-HMGB1 antibody. An anti-β-tubulin antibody was used as a loading control. (E) A complete *Hmgb1* knockout was not generated. The number of copies of *Hmgb1* exon 2 was determined by qPCR using primers 3 and 4 and normalised to *Hprt* copies. NIH3T3 cells contain four *Hmgb1* alleles and all screened cell lines contained at least one copy of *Hmgb1* exon 2.

3.10 *Hmgb1* knockdown using CRISPRi

CRISPR interference (CRISPRi) was instead used to knockdown HMGB1 protein expression in NIH3T3 cells. Using this system, efficient knockdown should occur in all successfully transduced cells, meaning that selection of individual cells is not required, thereby overcoming the main issue encountered previously. CRISPRi was used rather than RNAi as much higher knockdown levels are often achieved with the former, with up to 99% knockdown being reported (Gilbert *et al.*, 2014).

I therefore prepared a vector expressing a fusion protein comprised of a catalytically dead Cas9 (dCas9) and a Krüppel associated box (KRAB) transcriptional repressor domain which was then targeted to regions surrounding the *Hmgb1* transcription start site (Figure 3.15A). To generate the KRAB repressor vector, lenti-dCas9-KRAB-T2A-BLAST, the KRAB domain cDNA sequence from human ZNF10 was isolated from HEK293T cell total cDNA by PCR and ligated in frame with the cDNA sequence of dCas9 KRAB. This was then ligated into a lenti-T2A-BLAST vector, in frame with the T2A self cleaving peptide and blasticidin resistance gene (Section 2.43; Figure 3.15B). Three different sgRNA sequences which span the start of the *Hmgb1* gene were designed and individually cloned into lenti-sgRNA-Zeo (Section 2.44; Figure 3.15B). Two of the sgRNA sequences, TSS1 and TSS2, target dCas9-KRAB to within a 100 bp window of the transcription start site whereas the TLS sgRNA targets dCas9-KRAB to the translation start site of *Hmgb1* (Figure 3.15C). The resulting lentiviruses were used to transduce NIH3T3 cells which were then selected using a combination of blasticidin and zeocin. HMGB1 protein expression of cells transduced with the different sgRNA sequences was analysed by western blotting which revealed that the TSS1 sgRNA

gave the highest level of knockdown (Figure 3.15D). This sgRNA resulted in greater than 90% loss of HMGB1 protein expression and therefore was used for subsequent experiments. These cells will be referred to as HMGB1 KD. Expression of dCas9-KRAB in the absence of a sgRNA sequence led to no change in HMGB1 protein levels (Figure 3.15D), confirming that the knockdown is due to targeted transcriptional repression.

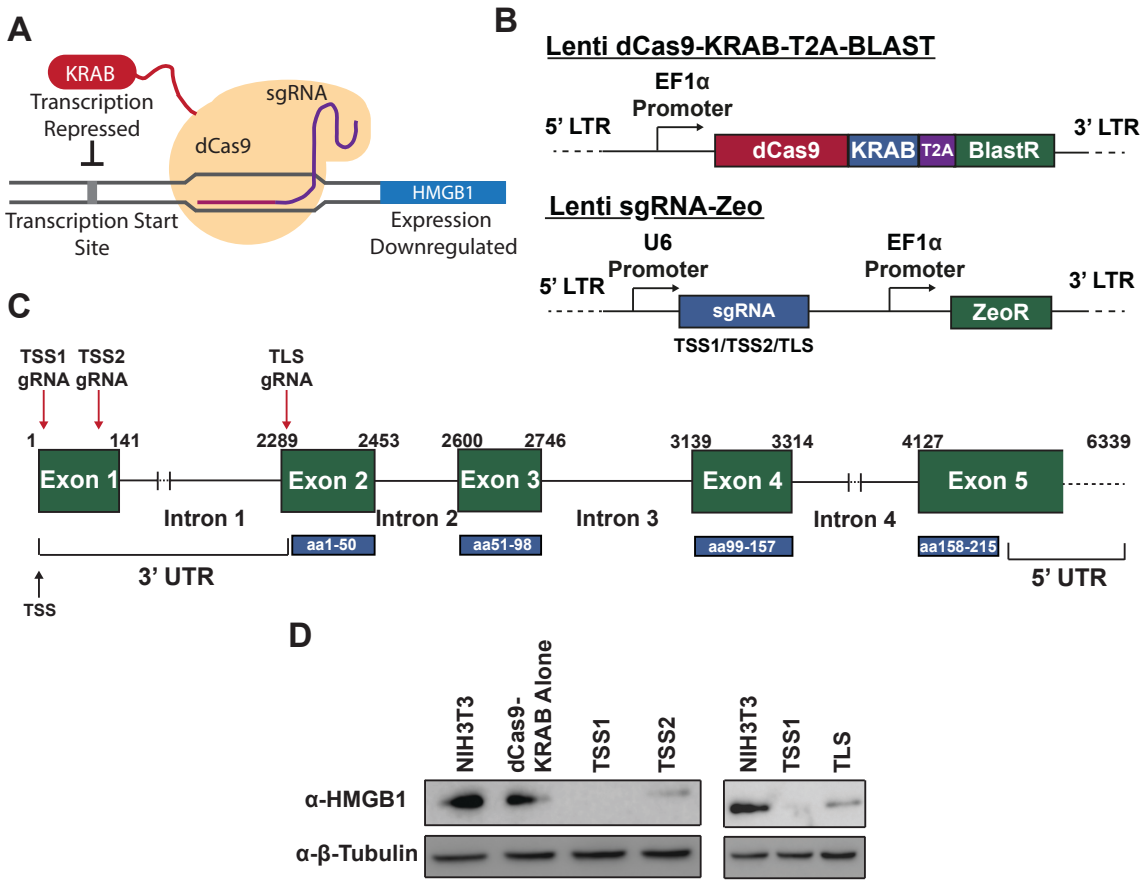


Figure 3.15 - Silencing of *Hmgb1* using CRISPRi in NIH3T3 cells

(A) Diagram showing dCas9-KRAB recruitment to the *Hmgb1* transcriptional start site (TSS). (B) Schematic of lentiviral cassettes introduced into NIH3T3 cells. Lenti dCas9-KRAB-T2A-BLAST (top) contains the cDNA sequence for a catalytically dead Cas9 (dCas; D10A, H841A) fused to the human ZNF10 KRAB domain, along with the blasticidin resistance gene. Lenti sgRNA-Zeo (bottom) contains the sgRNA sequences which target dCas9-KRAB to the *Hmgb1* locus, along with the zeocin resistance gene. (C) Genome organisation of the murine *Hmgb1* locus. Locations of gRNA sequences are represented by red arrows (TSS1, transcription start site 1; TSS2 transcription start site 2; TLS, translation start site). (D) Western blot showing HMGB1 protein expression levels in NIH3T3 cells transfected with dCas9-KRAB and the different gRNA sequences. See Appendix 2 for the uncropped blots.

3.11 Reduced HMGB1 protein expression coincides with reduced V(D)J recombination activity *in vivo*

Next, V(D)J recombination levels were measured in the HMGB1 KD cells in the presence of wild type RAG1, R401W or R401K. Following transfection of the recombination reporter plasmid, pJH299, and relevant RAG expression vectors, an average decrease in recombination of approximately 80% was observed in HMGB1 deficient cells compared to wild type NIH3T3 cells for both wild type RAG1 and R401K (Figure 3.16). When the same assays were performed in NIH3T3 cells expressing dCas9-KRAB alone, the results were not significantly different from wild type NIH3T3 cells (Figure 3.16). This confirmed that the drop in recombination activity was caused by HMGB1 deficiency and not due to expression of dCas9-KRAB. Together, these data support the hypothesis that HMGB1 is required for V(D)J recombination *in vivo*.

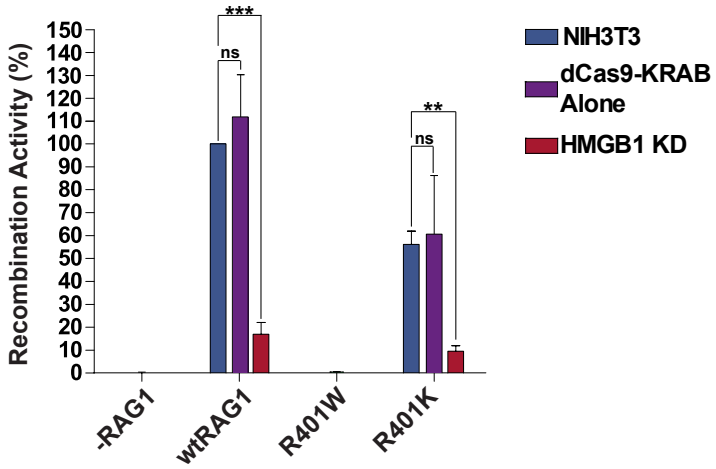


Figure 3.16 - Analysis of recombination activity in HMGB1 KD cells

Wild type NIH3T3 cells (blue bars), those expressing dCas9-KRAB alone (purple bars), or HMGB1 KD cells (red bars) were transfected with wild type RAG1, R401W or R401K and the recombination activity analysed. Only cells which retained HMGB1 protein expression were able to support efficient recombination. (N = 3, error bars represent SEM). The decrease in recombination activity in HMGB1 KD cells is statistically significant (Student's T-test; ns, not significant; ***, P = 0.001; **, P = 0.01).

3.12 Restoration of HMGB1 protein expression rescues V(D)J recombination activity in HMGB1 deficient cells

Although the recombination assay results from the HMGB1 KD cells strongly supports the essential role of HMGB1 during V(D)J recombination *in vivo*, it does not confirm if the negligible recombination activity of R401W *in vivo* is due to lack of HMGB1 binding. To test this idea, and further confirm if HMGB1 is required for V(D)J recombination *in vivo*, an expression vector for HMGB1 was introduced into the HMGB1 KD cells and the recombination activity of wild type RAG1, R401W and R401K was analysed.

To overexpress HMGB1 protein, HMGB1 KD cells were transduced with a lentivirus that expresses a HMGB1-P2A-Puro fusion protein under the control of the strong EF1α promoter (Section 2.45; Figure 3.17A). Additionally, a similar lentivirus was used which expresses a truncated HMGB1 protein lacking the C-terminal acidic tail (aa 1-172, HMGB1ΔC), as this has higher DNA and RAG binding activity (Bergeron *et al.*, 2005) (Figure 3.17A). Following selection with puromycin, HMGB1 protein expression was analysed by western blot and compared with HMGB1 KD cells. The retention of the P2A self cleaving peptide sequence on the C-terminus of HMGB1 allows ectoptic protein expression of full-length HMGB1 (HMGB1-P2A) to be distinguished from any endogenous HMGB1. Unfortunately, full-length HMGB1-P2A expression could not be detected, even after multiple lentiviral transductions. However, very high protein expression levels of HMGB1ΔC-P2A was readily detected (Figure 3.17B) and therefore these cells were used to perform the recombination assays.

As expected, exogenous expression of HMGB1 rescued V(D)J recombination activity with wild type RAG1 and R401K but remained negligible with R401W (Figure 3.17C). Indeed, recombination was greatly stimulated, up to ~400%, by overexpression of HMGB1 protein with wild type RAG1 and R401K, which mirrored the high expression levels of HMGB1ΔC in these cells. Despite this, R401W recombination was unchanged. Therefore, these data further confirm the vital role of HMGB1 during V(D)J

recombination *in vivo* and strongly imply that the inability of R401W to perform recombination is caused by inefficient HMGB1 binding.

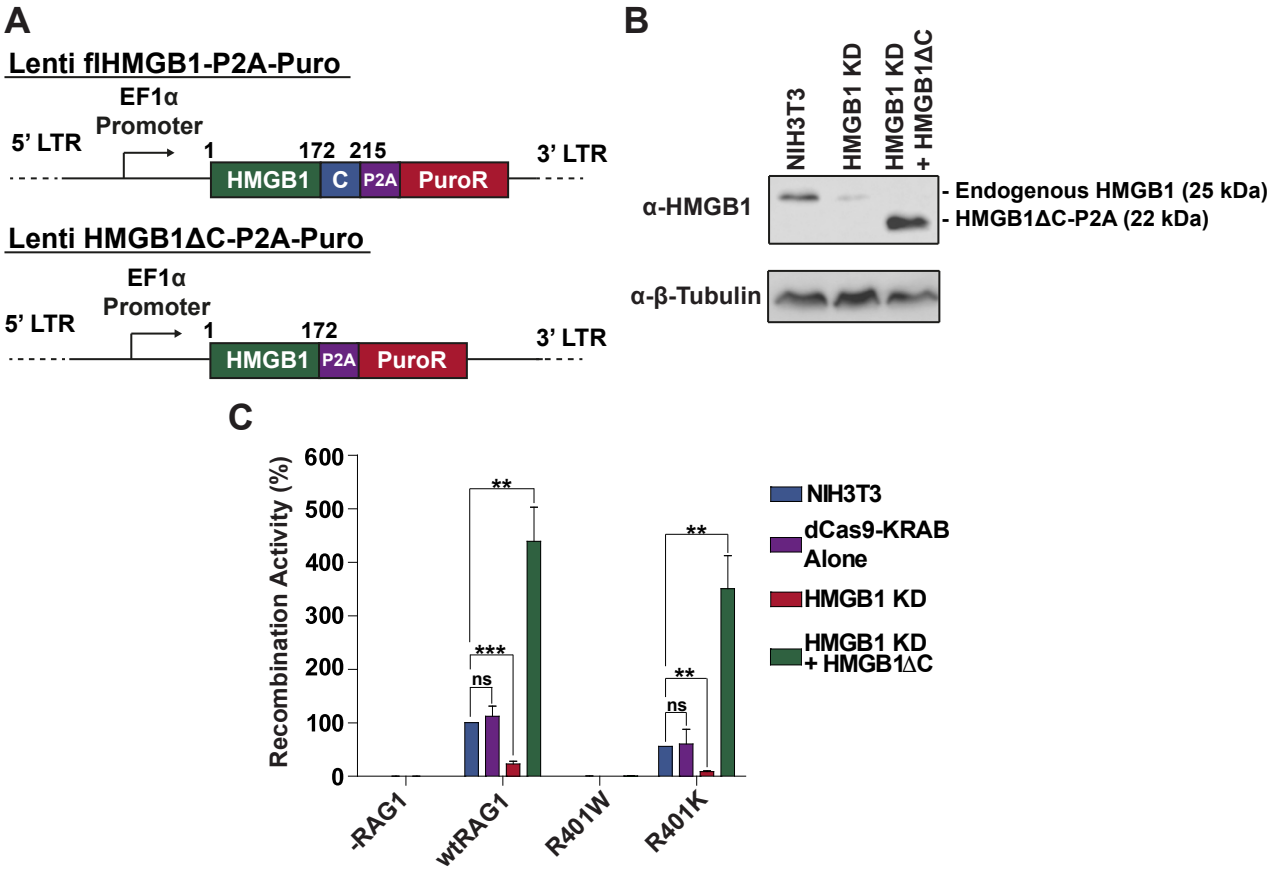


Figure 3.17 - Analysis of recombination in HMGB1 supplemented cells

(A) Schematic of lentiviral cassettes used to overexpress HMGB1 protein in the HMGB1 KD cells. Both cassettes contain the puromycin resistance gene along with either the full-length HMGB1 (Lenti flHMGB1-P2A-Puro, top), or truncated HMGB1 (Lenti HMGB1ΔC-P2A-Puro, bottom). (B) Western blot analysis shows high protein expression levels of HMGB1ΔC-P2A in HMGB1 KD cells transduced with Lenti HMGB1ΔC-P2A-Puro. See Appendix 3 for the uncropped blots. (C) Recombination activity in HMGB1 KD cells expressing HMGB1ΔC (green bars) transfected with wild type RAG1 and R401K is restored. However, activity with R401W remains negligible. Recombination activity is relative to NIH3T3 cells transfected with wild type RAG1 (*N* = 3, error bars represent SEM). The changes in recombination activity are statistically significant (Student's T-test; ns, not significant; ***, *P* = 0.001; **, *P* = 0.01).

3.13 Overexpression of HMGB2 also rescues V(D)J recombination activity in HMGB1 KD cells

Despite the above evidence for the involvement of HMGB1 during V(D)J recombination *in vivo,* it remains difficult to explain published data that show a normal immune system develops in *Hmgb1* knockout mice (Calogero *et al.*, 1999). Additionally, although the knockdown of *Hmgb1* in NIH3T3 cells resulted in very low levels of HMGB1 protein expression, these cells still retain ~20% recombination activity with wild type RAG1, which is substantially higher than R401W. Both of these outcomes could be explained if another bending factor can compensate for HMGB1 during V(D)J recombination. The most likely candidate is HMGB2 which has 79% sequence identity with HMGB1 and can substitute for HMGB1 in *in vitro* RAG cleavage assays (van Gent *et al.*, 1997; Figure 3.18A).

To test this idea, HMGB2 protein was overexpressed in the HMGB1 KD cells in a similar way as for HMGB1 (Figure 3.18B). However, unlike HMGB1, both full-length HMGB2-P2A and HMGB2ΔC-P2A expression could be readily detected in transduced cells (Figure 3.18C). The antibody used to detect HMGB2 recognises a unique epitope in the HMGB2 C-terminus which was removed in HMGB2ΔC and so this antibody did not detect HMGB2ΔC-P2A expression (Figure 3.18A). However, the anti-HMGB1 antibody used appeared to cross-react with HMGB2ΔC, allowing its detection (Figure 3.18C).

Recently it has been shown that many commercially available HMGB1 antibodies also detect HMGB2, which is likely due to their high homology (Davies *et al.* 2018; Figure 3.18A). Although a band corresponding to HMGB2ΔC-P2A was detected with the anti-HMGB1 antibody, endogenous HMGB2 and full-length HMGB2-P2A were not (Figure 3.18C). This maybe due to much higher expression of HMGB2ΔC than flHMGB2 in these cells, allowing detection of the former by the anti-HMGB1 antibody.

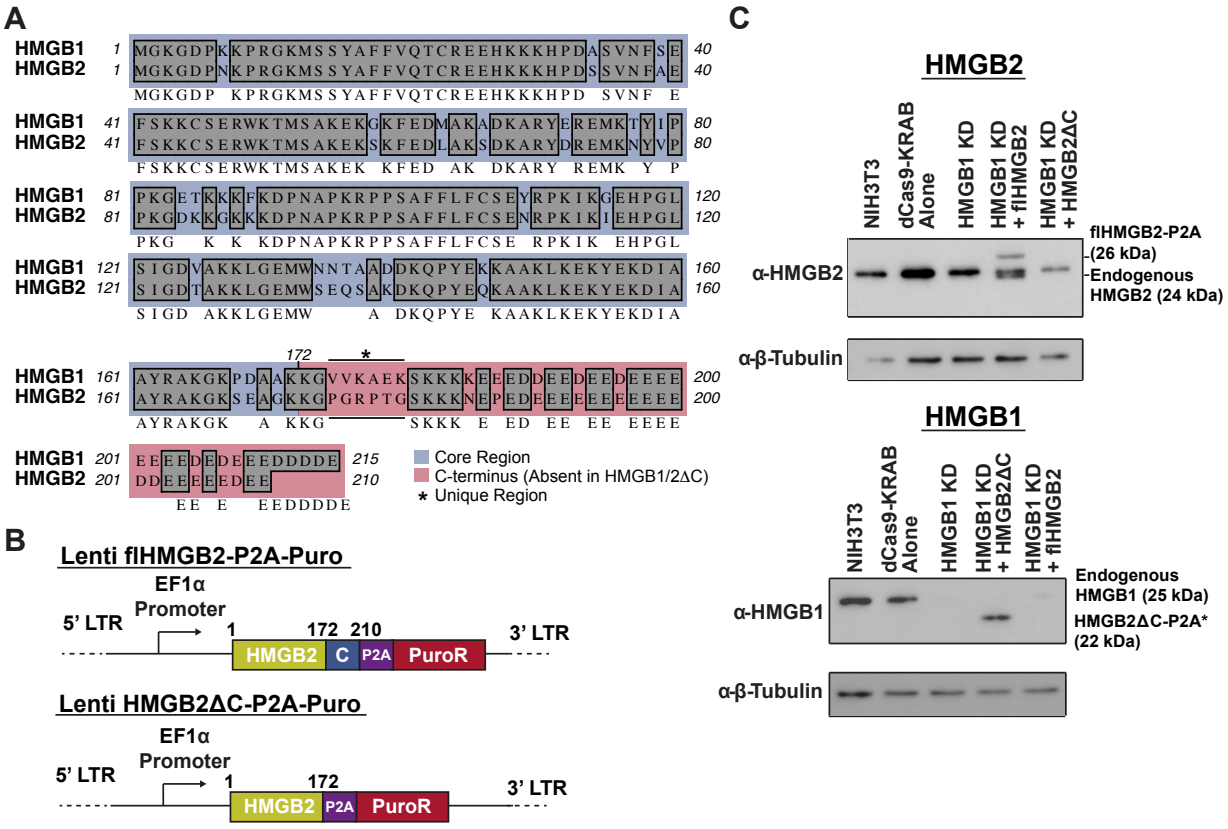


Figure 3.18 - Expression of HMGB2 protein in HMGB1 KD cells

(A) Alignment of murine HMGB1 and HMGB2 protein sequences. Grey shaded residues are those which are conserved between the two proteins, the consensus sequence is shown underneath. Blue shading denotes the core region of HMGB1/2 and red shading denotes the C-terminal tail which is removed in HMGB1/2ΔC. The asterisk represents a unique region between the two proteins which is likely the epitope for the HMGB2 antibody. (B) Schematic of the lentiviral cassettes used to overexpress HMGB2 protein. Both contain the puromycin resistance gene along with either full-length HMGB2 (Lenti flHMGB2-P2A-Puro, top), or truncated HMGB2 (Lenti HMGB2ΔC-P2A-Puro, bottom) cDNA sequence. (C) Western blot using an anti-HMGB2 (top) or anti-HMGB1 (bottom) antibody. HMGB2ΔC is not detected by the anti-HMGB2 antibody as the epitope is missing. However, a band corresponding to the same size as HMGB2ΔC-P2A is detected by the anti-HMGB1 antibody. Western blots were performed at a later date and with independent samples than those analysed for recombination in Figure 3.20. Since the exogenous expression of HMGB1/2 decreased gradually with time in culture, these blots may not accurately represent the levels of exogenous HMGB1/2 present in the samples in Figure 3.20 at the time of analysis. See Appendix 4 for the uncropped blots.

To confirm this cross-reactivity, vectors expressing full-length HMGB2 or HMGB2ΔC were transfected into HEK293T cells and western blots were performed using both anti-HMGB1 and anti-HMGB2 antibodies. This revealed that the anti-HMGB1 antibody does indeed cross react with HMGB2 as bands corresponding to both full-length HMGB2-P2A and HMGB2ΔC-P2A are visible on anti-HMGB1 western blots (Figure 3.19A). To eliminate any doubt that the HMGB2ΔC-P2A band detected by the anti-HMGB1 antibody in the HMGB1 KD cells was anything other than HMGB2ΔC, analysis of RNA expression was performed. PCR amplification, using HMGB1-P2A or HMGB2-P2A specific primers, of cDNA made from the knockdown HMGB1 NIH3T3 cells transduced with HMGB2ΔC revealed the presence of only HMGB2ΔC-P2A and not HMGB1ΔC-P2A (Figure 3.19B).

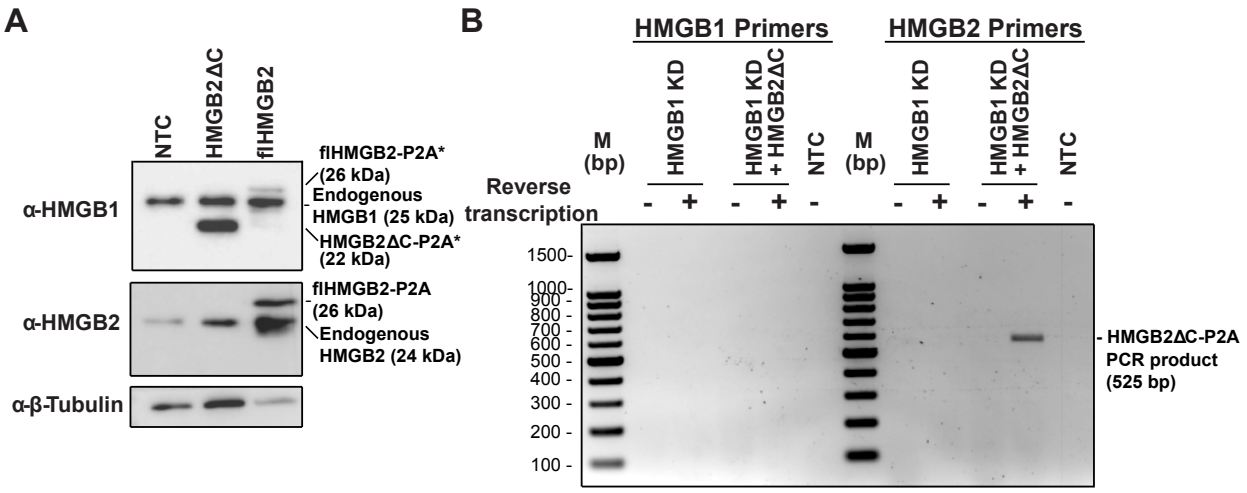


Figure 3.19 - The anti-HMGB1 antibody cross-reacts with HMGB2

(A) Western blot of HEK293T cells transfected with either Lenti flHMGB2-P2A-Puro or Lenti HMGB2ΔC-P2A-Puro using the anti-HMGB1 (top) or anti-HMGB2 (middle) antibody. A strong band corresponding to HMGB2ΔC-P2A was detected by the anti-HMGB1 antibody. See Appendix 5 for the uncropped blots. (B) Analysis of transcripts present in HMGB1 KD cells transduced with Lenti HMGB2ΔC-P2A-Puro detects HMGB2ΔC-P2A and not HMGB1ΔC-P2A.

Next, the recombination assay was performed in cell lines expressing either full-length HMGB2 or HMGB2ΔC with wild type RAG1, R401W and R401K. Similar to HMGB1ΔC, both forms of HMGB2 were able to rescue and greatly stimulate recombination activity with wild type RAG1 and R401K, but not R401W (Figure 3.20). HMGB2 protein

expression only occurs in embryonic and, notably, lymphoid cells and therefore the development of a normal immune system in *Hmgb1* knockout mice is most likely due to HMGB2 compensating for HMGB1 during V(D)J recombination. Additionally, as NIH3T3 cells are mouse embryonic fibroblasts, HMGB2 protein is detected in these cells (Figure 3.18C) and so the retained ~20% recombination activity observed in the HMGB1 KD cells could be due to stimulation via endogenous HMGB2 protein.

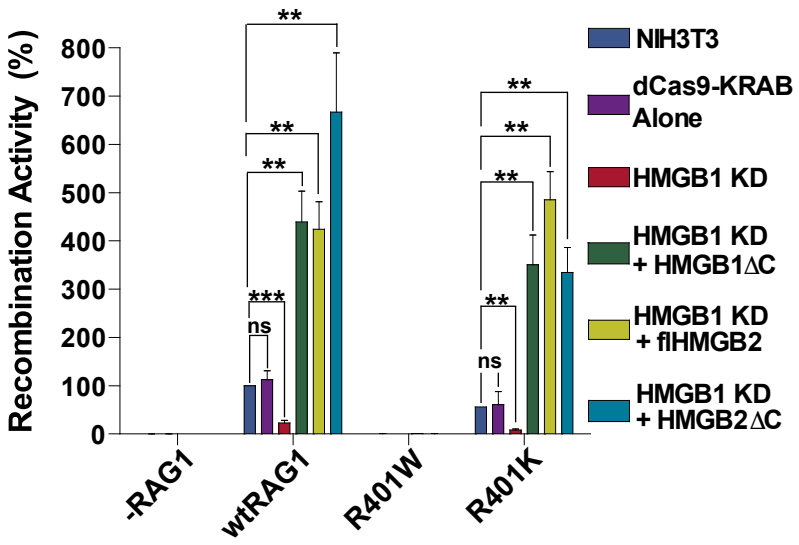


Figure 3.20 - Analysis of recombination activity in HMGB2 supplemented cells

Recombination activity was restored and stimulated in HMGB1 KD cells expressing either flHMGB2 (yellow bars) or HMGB2ΔC (cyan bars) transfected with wild type RAG1 or R401K. Similar to HMGB1 overexpression, recombination activity with R401W remained negligible. Recombination activity is relative to NIH3T3 cells transfected with wild type RAG1 (*N* = 3, error bars represent SEM). The changes in recombination activity are statistically significant (Student's T-test; ns, not significant; ***, *P* = 0.001; **, *P* = 0.01).

C) Discussion

In this chapter, I analysed two compound heterozygous mutations identified in an immunodeficient patient; R401W and R504Q. One of these mutations, R504Q, retained a low level of recombination activity, whereas the activity of the other, R401W, was negligible. Further biochemical analysis of the latter mutation surprisingly revealed a novel disruption in HMGB1 binding and this provided further insight into the involvement of this accessory protein during V(D)J recombination *in vivo*. Further, knockdown of HMGB1 protein expression coincided with greatly reduced recombination activity in NIH3T3 cells, which could be restored by the introduction of either HMGB1 or HMGB2. Together, these data help to resolve a long-standing controversy over the involvement of HMGB1/2 during V(D)J recombination *in vivo.*

3.14 Generation of an *in vivo* V(D)J recombination assay

To analyse the recombination activity of RAG mutations identified in immunodeficient patients, I initially developed a qPCR based *in vivo* V(D)J recombination assay.

A number of assays already existed to analyse V(D)J recombination *in vivo* but they have some disadvantages which make them unsuitable for accurate and efficient screening of large numbers of RAG mutations. The most established assays are those which were developed in the Gellert laboratory in the 1980s and involve the transfection of a recombination substrate plasmid, containing a 12- and 23-RSS, and RAG expression vectors into mouse fibroblasts (Lieber *et* al., 1987; Hesse *et al.*, 1989). Upon recombination, a transcriptional terminator upstream of a chloramphenicol resistance gene is either removed by deletion or inactivated by inversion. The plasmid DNA is then extracted and transformed into *E.coli* which are subsequently plated on chloramphenicol agar plates (Figure 3.21A). Recombination activity can then be scored by counting the number of chloramphenicol resistant colonies. Although informative, these types of assays are very time consuming and labour intensive and differences in transformation efficiency of the *E.coli* may lead to inaccuracies.

More recently, another recombination assay emerged which is similar to the one discussed above but involves an integrated cassette where recombination leads to inversion of an anti-sense GFP coding sequence which then allows GFP expression (Bredemeyer *et al.*, 2006). Recombination activity can then be measured by flow cytometry, with the number of GFP positive cells providing a readout of recombination activity (Figure 3.21B). This assay has benefits over the previous assay in that it allows much quicker analysis of recombination activity and is also useful in situations which require the isolation of living cells which have undergone recombination. However, as this is a binary assay between GFP positive and negative cells, some information could be missed. For example, a RAG mutation which retains relatively high levels of recombination (>50%) would be able to, although at a slower rate, express the same levels of GFP as wild type RAG if left for a long enough time. This increases the likelihood that some mutations may erroneously be scored as having the same recombination activity as wild type.

The qPCR recombination assay I developed (Figure 3.21C) overcomes many of the potential problems outlined above. The advantage of such an assay is that recombination activity can be very rapidly and accurately screened and in a high-throughput manner. Further, a qPCR based assay directly measures the product of recombination rather than a marker such as GFP or chloramphenicol resistance. The latter should allow for more accurate measurement of V(D)J recombination activity. However, as this assay utilises extra-chromosomal substrates, it may not faithfully detect recombination defects which involve RAG binding to chromatinised substrates.

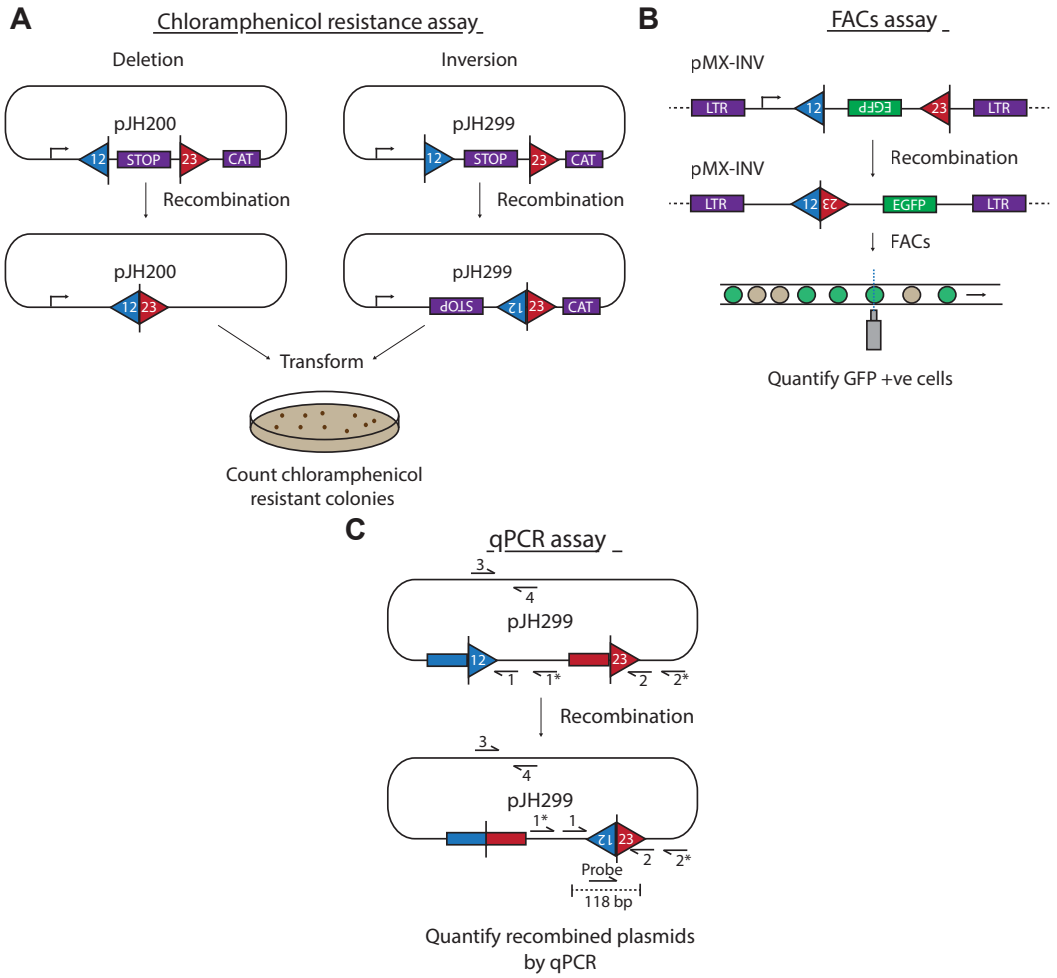


Figure 3.21 - Comparison of different V(D)J recombination assays

(A) Diagram of the chloramphenicol resistance recombination assay. Recombination substrate plasmids are transfected into mammalian cells, along with expression vectors for RAG1 and RAG2. These contain a 12- (blue triangles) and 23-RSS (red triangles) flanking a bacterial transcriptional terminator (purple 'STOP' rectangle). Recombination can occur by deletion (pJH200, left) or inversion (pJH299, right) which removes or inverts the transcriptional terminator. Once transformed into *E. coli,* this allows transcription of the chloramphenicol resistance gene (purple 'CAT' rectangle) and therefore confers chloramphenicol resistance. Recombination is scored by counting the number of resistant colonies. (B) Diagram of the FACs based recombination assay. A lentiviral cassette is integrated into the genome of a pro-B cell line which contains a 12- (blue triangles) and 23-RSS (red triangles), flanking an inverted EGFP cDNA sequence. Following inversional recombination, the EGFP cDNA is inverted and expressed and recombination activity is scored by FACs analysis of GFP positive cells. (C) Diagram of the qPCR based recombination assay used in this study. Successful PCR only occurs with primers 1 + 2 following recombination. This assay is described in detail in Figure 3.3.

The recombination assay developed and utilised in this chapter, and the accompanying publication (Thwaites *et al.*, 2019), has also been used for the analysis of RAG mutations identified in immunodeficient patients in two recent publications (Lawless *et al.*, 2018; Dorna *et al.*, 2019). Additionally, this assay was used in the work described in Chapter 4 to analyse recombination activity of RAG2 mutants with potential oncogenic effects. Therefore this assay is a valuable tool for the analysis of the RAG proteins.

3.15 R504Q is responsible for all recombination activity in the patient

By analysing the point mutations in each *RAG1* allele from the patient separately, the individual contribution of each mutation to V(D)J recombination can be dissected. This revealed that R504Q was the only mutation to retain any V(D)J recombination activity and therefore RAG complexes containing this mutation, and not R401W, are likely to perform all V(D)J recombination reactions in the patient (Figure 3.5). Additionally, when R401W and R504Q are mixed, intermediate recombination and RAG cutting was observed. This suggests that R401W:R504Q RAG complexes are inactive whereas R504Q:R504Q complexes retain partial activity. R504Q can form complexes with a 12- and 23-RSS which are similar to those formed with wild type RAG1, although at a much lower level (Figure 3.9). This explains the lower cleavage activity that is also observed with R504Q (Figure 3.8) and further translates to greatly decreased recombination activity.

Notably, comparison of the patient's TCRβ repertoire with that of a pool of healthy individuals revealed that the use of four TCR Vβ segments were significantly altered in the patient (Figure 3.1). As all recombination events occurring in the patient would be performed by R504Q RAG complexes, the skew in TCRβ usage maybe due to the altered binding activity caused by this mutation. Both of the TCR Vβ segments which are under-represented, TCRVβ2 and TCRVβ16, are associated with a 23-RSS which is less well conserved when compared to a consensus RSS, whilst TCRVβ14 more closely matches the consensus and is over-represented (Table 3.1). However,

TCRVβ12-1 is also over-represented in the patient but its RSS sequence does not seem to match the consensus as well as TCRVβ14 (Table 3.1). Like other hypomorphic core RAG1 mutations, such as R975Q which has preference for particular coding segment sequences (Wong *et al.*, 2008), R504Q may affect the overall quality of V(D)J recombination by restricting the repertoire of RSSs which can be used for recombination.

	Heptamer	Spacer	Nonamer	Change in patient
Consensus	CACAGTG	ATGNAANCNNANNGNGAACCTGT	ACAAAAACC	-
TCRVβ14	**CACAGTG**	C**T**TC**A**CAGTCGTGCCCTTG**CTGT**	G**CAAAA**C**C**A	Increase
TCRVβ12-1	**CACAG**CA	C**TGCA**GAATCTCCCCATCT**CTGT**	G**CAGAAA**C**C**	Increase
TCRVβ2	**CACAG**CC	T**TGCAA**AGAC**A**ACTCC**AGCCTGT**	G**CAAAAT**C**C**	Decrease
TCRVβ16	**CACAGTG**	T**T**AAA**A**TATTAGCTAATCTTAGGA	CAC**A**G**A**CT**C**	Decrease

Table 3.1 - Sequences of TCRVβ segment RSSs whose use is significantly altered in the patient

The sequences of the RSSs of the four TCRVβ segment whose use is significantly altered in the patient (Figure 3.1). The top row shows the consensus RSS sequence (Ramsden *et al.*, 1994). Nucleotides of the TCRVβ segment RSSs which match the consensus sequence are shown in dark blue.

3.16 R401W inhibits HMGB1 binding

Binding assays performed with R401W revealed a novel block in HMGB1 binding (Figure 3.9 and 3.11). Although R401W was able to cleave a single 12-RSS under less stringent cleavage conditions (Figure 3.12B), it was unable to perform coupled cleavage which is required for V(D)J recombination (Figure 3.8). This is consistent with structural studies which show that HMGB1 is required to correctly position both 12- and 23-RSS in the paired RAG complex to facilitate coupled cleavage (Kim *et al.*, 2018). R401 is located at a critical 60° bend in both the 12- and 23-RSS which is also the site for HMGB1 Box A binding (Figure 3.10). Mutation of R401 to leucine rather than tryptophan allows a slight increase in HMGB1 binding but this is not accompanied by restored cleavage or recombination activity (Figure 3.13). Instead, mutating R401 to

lysine leads to restored HMGB1 binding and near wild type levels of cleavage and recombination activity (Figure 3.13). This suggests that whilst the bulky nature of the tryptophan in R401W contributes to inhibiting HMGB1 binding, a positive charge at this position is required for complete HMGB1 binding and the formation of an active complex.

HMGB1 Box A has a preference for pre-bent DNA (Webb *et al.*, 2001) and is recruited, after RAG binding, to a 60° bend in the RSS found near R401 (Little *et al.*, 2013). Together, these data support a model whereby R401, via its positive charge, establishes the 60° bend in both the 12- and 23-RSS and this facilitates HMGB1 Box A binding. Box B can then establish an additional 90° bend in the 23-RSS and form a total 150° bend in the 23-RSS, which is absolutely vital for coupled cleavage (Kim *et al.*, 2018). The lack of positive charge and bulky nature of R401W could block proper formation of the 60° bend and therefore inhibit Box A binding. Consequently, even though HMGB1 can bind the 23-RSS via Box B, the required 150° bend of the 23-RSS cannot form and therefore recombination is blocked (Figure 3.22).

A previous study of human RAG1 mutations suggested that weaker RSS binding by mutation of R401 is due to the loss of a stacking interacting with R440 in the other RAG1 monomer which therefore disrupts the NBD dimer conformation (Lee *et al.*, 2016). Such interactions also support the effects described above as this stacking interaction may help to form the conformation which supports HMGB1 binding, and this would be disrupted in R401W.

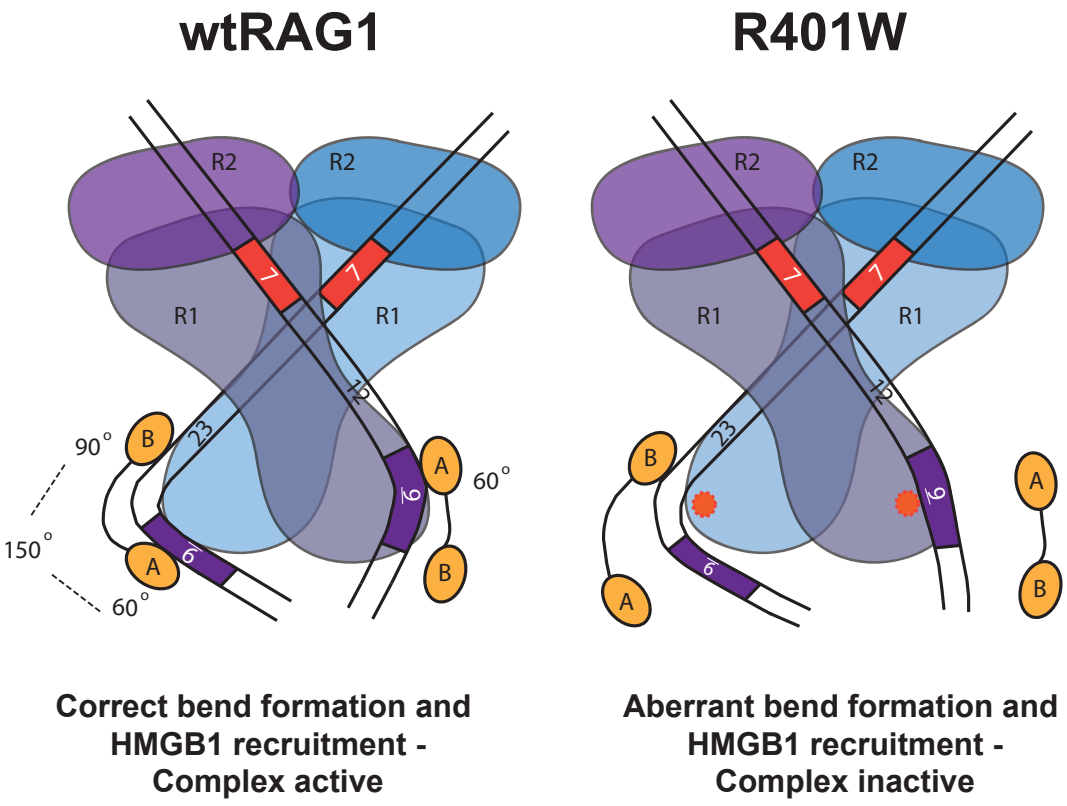


Figure 3.22 - Proposed model of R401W inactivity

Schematic representation of complex formation with wild type RAG1 and R401W. Each RAG1/2 dimer is shown in purple or blue and the box domains of HMGB1 are represented as yellow ovals. Wild type RAG1 enables correct HMGB1 and formation of the required bends at the 12- and 23-RSS to facilitate coupled-cleavage (left). The R401W mutation disrupts the conformation of the NBD and blocks HMGB1 Box A binding. This leads to aberrant bend formation and the inability to correctly position the different sized RSSs within the complex, blocking coupled-cleavage (right).

3.17 HMGB1/2 is required for V(D)J recombination *in vivo*

Many previous studies of RAG mutations revealed amino acids important in RAG structural integrity or catalytic activity. Here instead, a RAG1 mutation highlights the vital role of an accessory protein, HMGB1/2, during V(D)J recombination. HMGB1/2 has been known to stimulate RAG cleavage *in vitro* for some time (van Gent *et al.*, 1997) and a recent structural study revealed the molecular basis behind this (Kim *et al.*, 2018). However, the observation that HMGB1 knockout mice have a normal immune

system raised doubts in the role of HMGB1 during V(D)J recombination *in vivo*. Therefore, a long-standing question was whether HMGB1/2 is required for V(D)J recombination *in vivo*.

The discovery of a RAG1 mutation in an immunodeficient patient which disrupts HMGB1 binding and that V(D)J recombination in HMGB1 deficient cells is significantly reduced (Figure 3.16) provides strong evidence that HMGB1 is required for V(D)J recombination *in vivo*. Additionally, by showing that HMGB2 can compensate for HMGB1 *in vivo* (Figure 3.20), the data presented in this chapter solves a long-standing controversy by demonstrating that HMGB2 can substitute for HMGB1 to catalyse V(D)J recombination *in vivo*.

Chapter 4 - Identification of residues within the RAG2 C-terminus which inhibit reintegration

A) Introduction

Aberrant V(D)J recombination poses a threat to both efficient antigen receptor production and genomic integrity. The RAG proteins are capable of catalysing various reactions which involve the incorrect resolution of the DNA ends generated during V(D)J recombination (Figure 4.1A). Novel recombination products such as “open and shut” and hybrid joints were first detected in exogenous V(D)J recombination substrates introduced into cell lines expressing RAG proteins (Lewis *et al.*, 1988). Both of these outcomes involve the incorrect resolution of breaks generated by RAG cleavage, by the joining of hairpin sealed coding ends to blunt signal ends (Figure 4.1A). Hybrid joints can constitute 30% of standard recombination products in transient transfection experiments (Lewis *et* al., 1988; Roth, 2003), but they are much rarer at endogenous loci (Sollbach and Wu, 1995; VanDyk *et al.*, 1996). Excessive hybrid joining would impact the efficiency of V(D)J recombination and therefore must be suppressed. Indeed, the RAG2 C-terminus suppresses hybrid joint formation since there is an increase accumulation of these aberrant joins in cells expressing core RAG2 (cRAG2), which lacks the RAG2 C-terminus (Sekiguchi *et al.*, 2001).

Another aberrant V(D)J recombination reaction is transposition, which was initially thought to pose a significant risk to genomic stability by insertion of the by-product of V(D)J recombination, the excised signal circle (ESC), into the genome (Figure 4.1A). The ESC can become rebound by the RAG proteins and cleaved by a nick-nick mechanism, generating two blunt signal ends (Neiditch *et al.*, 2001). During transposition, RAGs catalyse the nucleophilic attack of the DNA phosphodiester backbone by the 3’OH of the signal end (Agrawal *et al.*, 1998; Hiom *et al.*, 1998; Elkin *et al.*, 2003). This means the signal ends can become inserted into potentially any

target sequence in the genome. Similar to hybrid joint formation, *in vitro* transposition is inhibited by the presence of the RAG2 C-terminus (Elkin *et al.*, 2003), meaning transposition *in vivo* is rare due to the presence of full length RAGs. Notably, the only transposition event identified *in vivo* occurred in at a site in the genome which contains inverted repeats and is thought to form a cruciform structure that terminates in a hairpin (Roth, 2003). Therefore, like hybrid joints, transposition may involve the aberrant joining of the blunt RSS signal end to a hairpin DNA end and this maybe what is suppressed by the RAG2 C-terminus. Since transposition is rare *in vivo*, it is no longer thought to pose a significant threat to genome integrity.

Signal ends can also become reinserted into the genome via a process known as reintegration (Figure 4.1A). This reaction differs from transposition in that the site of insertion is an RSS partner. During reintegration, RAGs catalyse the insertion of the ESC into a target RSS, producing a new signal joint and a pseudo hybrid joint (Vanura *et al.*, 2007; Figure 4.1A). This is a potentially dangerous outcome as strong promoter elements could become juxtaposed to proto-oncogenes, leading to their activation. Similar to the aberrant reactions described above, the RAG2 C-terminus suppresses this reaction (Curry *et al.*, 2007).

The RAG2 C-terminus, aside from the PHD finger, is a relatively poorly understood region, primarily due to difficulties in purifying full length RAG proteins (Schatz and Swanson, 2011). Many studies of RAG proteins are performed using truncated versions of the proteins, that are easier to purify but still retain catalytic activity. However, as discussed above, the RAG2 C-terminus plays a vital role in inhibition of aberrant V(D)J recombination reactions, highlighting the importance of studying this region. Three major domains make up the RAG2 C-terminus (Figure 4.1B): (i) the acidic hinge, which contains an autoinhibitory domain that suppresses the binding and cleavage activities of the RAG complex, (ii) the PHD finger, which binds to H3K4me3 and also relieves the autoinhibition imposed by the acidic hinge, and (iii) the C-terminal extension, which contains a degradation signal that couples RAG2 expression to the

cell cycle. The specific regions of the RAG2 C-terminus required to inhibit aberrant V(D)J recombination reactions and the mechanisms of these suppressive activities remain largely unexplored.

In this chapter I set out to map the RAG2 C-terminus to identify the region that is vital to inhibit one of these aberrant reactions, reintegration. After identification of a region vital to suppress reintegration, I proceeded to analyse the properties of this suppression and explore if this region has a wider role in inhibiting other aberrant joining reactions following V(D)J recombination.

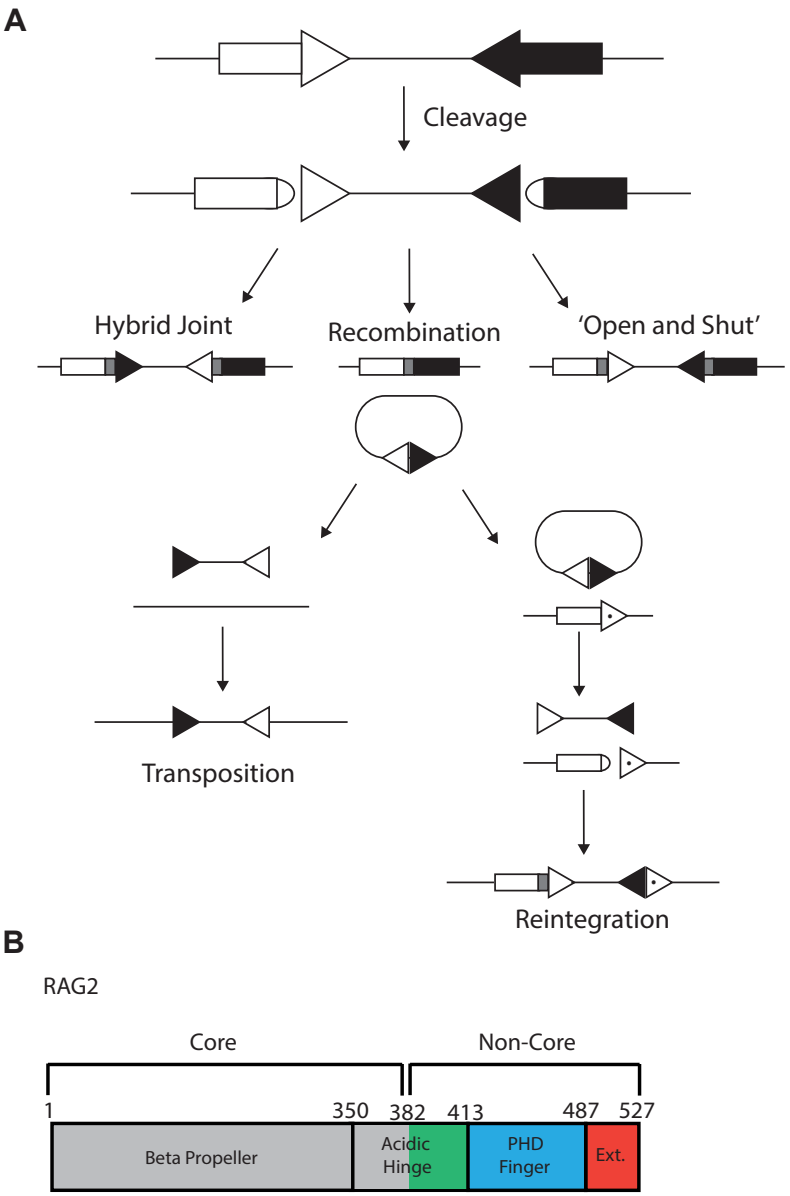


Figure 4.1 - Aberrant V(D)J recombination reactions and the RAG2 C-terminus

(A) Diagram of possible standard and non-standard products formed during V(D)J recombination. Triangles represent RSSs whereas rectangles represent gene segments. (B) Schematic of the RAG2 protein. The main domains of the RAG2 C-terminus and their amino acid positions are shown. The core of RAG2 (grey) consists of a six bladed beta propeller (aa 1-350) and part of the acidic hinge (aa 352-382). The non-core C-terminus consists of the second part of the acidic hinge (aa 383-413; green), PHD finger (aa 414-487; blue) and a C-terminal extension (aa 488-527; red).

B) Results

4.1 Development of an *in vivo* extra-chromosomal reintegration assay

Firstly, an assay to analyse the reintegration activity of RAG2 mutants *in vivo* was required. This assay must be able to quantitatively detect differences in reintegration activity of various RAG2 mutants to enable the domain that suppresses reintegration to be mapped. To this end, a qPCR extra-chromosomal reintegration assay was designed and performed in a similar way to the V(D)J recombination assay described in Chapter 3 (Section 3.1).

In brief, two substrate plasmids containing either a consensus 12-RSS (pcDNA3.1-12-RSS), that acts as the target RSS, or a signal joint (SJ; pcDNA3.1-ESC-Reint), that mimics the ESC, are transfected into COS7 cells, along with expression vectors for RAG1 and RAG2 (Section 2.27). After reintegration of the ESC plasmid into the 12-RSS plasmid, the level of reintegration can be determined via qPCR using primers that amplify the resulting pseudo-hybrid joint (ΨHJ) (Figure 4.2A, primers 1 and 2). A hydrolysis probe complementary to the sequence upstream of the ΨHJ is used to increase specificity as some non-specific amplification occurs in samples without RAG2 or with low reintegration activity (data not shown). The results from the reintegration qPCR are normalised to the total amount of pcDNA3.1-12-RSS present, which was determined using a set of primers complementary to a region of pcDNA3.1-12-RSS that is unaltered by reintegration (Figure 4.2A, primers 3 and 4).

To test if the assay can quantitatively detect differences in reintegration activity, an increasing amount of core RAG2 (cRAG2) expression vector was transfected and the reintegration activity was determined. Reintegration is efficiently detected in the presence of cRAG2 and this activity increases with the levels of cRAG2 present (Figure 4.2B), confirming that the assay should quantitively detect reintegration activity of different RAG2 mutants.

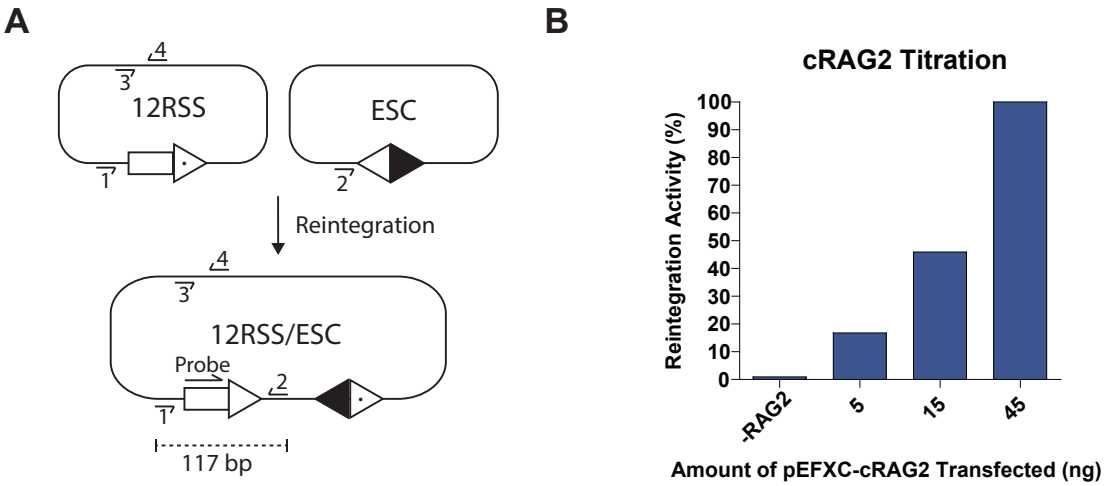


Figure 4.2 - Generation of an *in vivo* extra-chromosomal reintegration assay

(A) Diagram of the reintegration assay. White triangles represent 12-RSSs whereas black triangles represent 23-RSSs. Primer and probe binding sites and shown along with the expected size of a ΨHJ. (B) Increasing amounts of pEFXC-cRAG2 plasmid were transfected into COS7 cells and the reintegration activity determined. There is a clear correlation between the amount of RAG2 present and reintegration activity.

4.2 Analysis of reintegration activity of RAG2 C-terminal truncation mutants

To identify the specific region of the RAG2 C-terminus which suppresses reintegration, a number of RAG2 truncation and deletion mutants were analysed.

Initially, three RAG2 truncation mutants were generated: (i) RAG2$^{487\Delta}$, which lacks the C-terminal extension, (ii) RAG2$^{413\Delta}$, which lacks the PHD finger and the C-terminal extension, and (iii) cRAG2 (382Δ), which lacks the entire C-terminus (Figure 4.3A). The reintegration assay was performed using wild type RAG1 and wild type or mutant RAG2 to determine which of the main domains of the RAG2 C-terminus might be involved in the suppression of reintegration. To account for differences in protein expression, the levels of the different RAG2 mutants were determined by western blotting (Figure 4.3C). Both RAG2$^{487\Delta}$ and RAG2$^{413\Delta}$ show similar levels of reintegration to wild type RAG2, whereas cRAG2 has around seven-fold greater activity (Figure 4.3B). This is consistent with previous observations in mice expressing cRAG2 (Curry *et* al., 2007). Further, since RAG2$^{413\Delta}$ only differs from cRAG2 by the presence of 31

amino acids from the acidic hinge but has wild type levels of reintegration, this suggests that the suppressive region lies entirely within residues spanning 383-413. This is further supported by a similar increase in reintegration activity with RAG2$^{383\text{-}413\Delta}$, where amino acids 383-413 have been deleted, but the PHD finger and C-terminal extension are retained (Figure 4.3A and B).

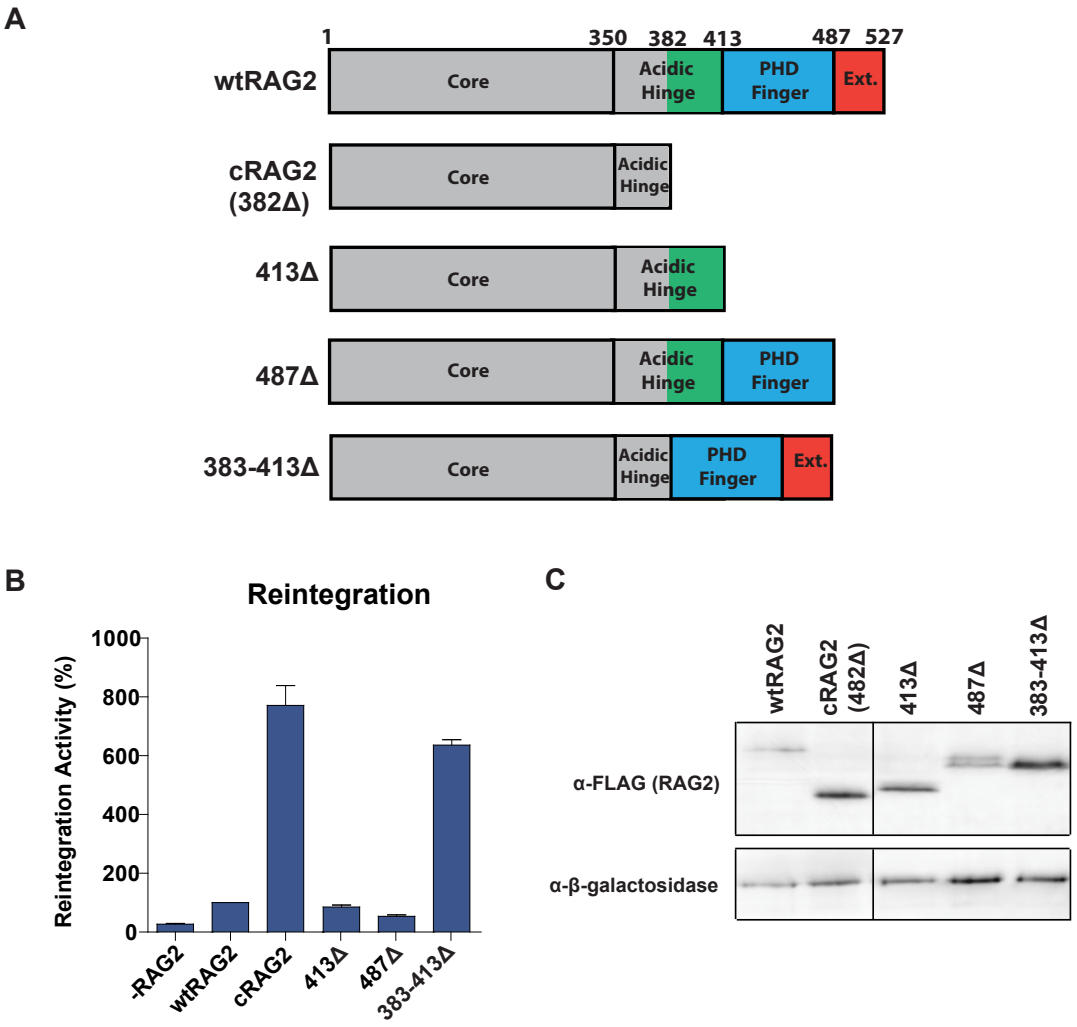


Figure 4.3 - Mapping the region of the RAG2 C-terminus that suppresses reintegration

(A) Schematic of RAG2 mutants used in the reintegration assay. (B) Reintegration activity of RAG2 mutants relative to wtRAG2 (*N* = 3, error bars represent SEM). (C) Western blot of COS7 lysates transfected with wild type or mutant RAG2 and a β-galactosidase expression vector. All RAG2 proteins contained an N-terminal FLAG tag which allows their detection using an anti-FLAG antibody. Samples analysed on this western blot were from an independent transfection than those analysed for reintegration. See Appendix 6 for the uncropped blots.

4.3 Substitution of the acidic hinge with an alternate acidic sequence does not suppress reintegration

The acidic hinge is a largely unstructured stretch of acidic amino acids which bridges the globular core and PHD finger of RAG2 (Callebaut and Mornon, 1998). A region of the core acidic hinge (370-383) has previously been shown to be important in correct repair pathway choice: mutations which neutralise the negative charge, and therefore the flexibility of this region, lead to increased repair of ends via the error-prone alternative NHEJ pathway (Coussens *et al.*, 2013). The region of the acidic hinge potentially involved in suppressing reintegration, 383-413, also contains a high density of negatively charged residues (Figure 4.4A). To determine if the acidic, flexible nature of 383-413 is important in the suppression of reintegration, this sequence was replaced with that of a region of the human alpha synuclein (α-syn) C-terminus (aa 107-137). Like the RAG2 acidic hinge, the α-syn C-terminus is intrinsically disordered and highly acidic (Emamzadeh, 2016; Figure 4.4A). The reintegration activity of $RAG2^{\alpha\text{-syn}}$ is similar to both cRAG2 and $RAG2^{383\text{-}413\Delta}$ (Figure 4.4B), suggesting that something specific to the intrinsic sequence of residues 383-413 of RAG2, and not the negative charge, is responsible for the suppression of reintegration.

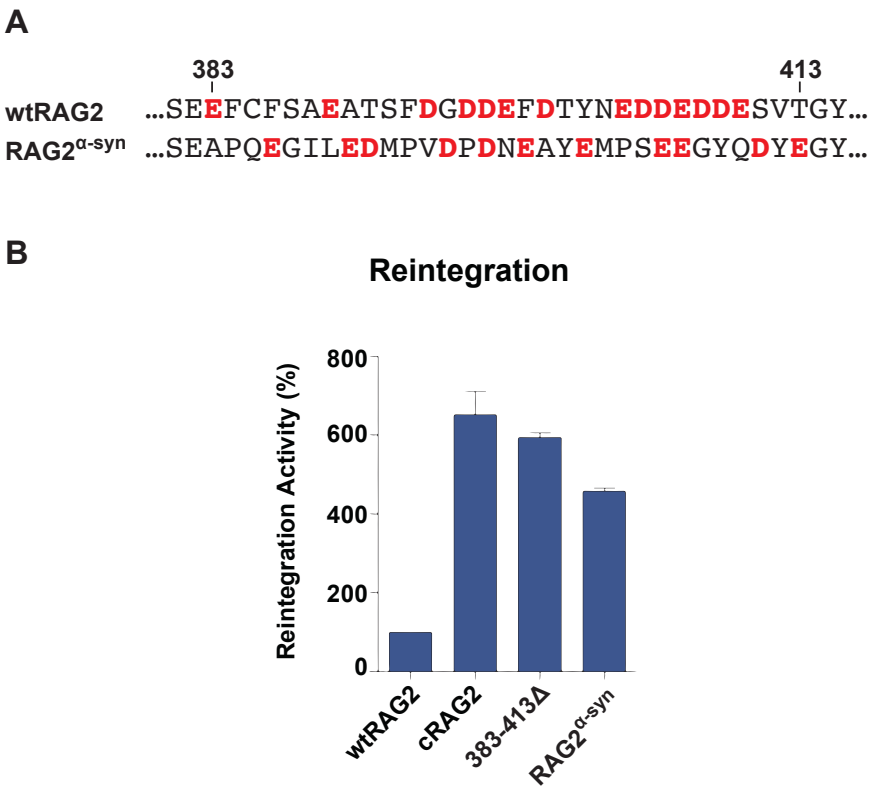


Figure 4.4 - An α-synuclein acidic region cannot substitute for the RAG2 acidic hinge

(A) Sequence alignment of wtRAG2 and RAG2$^{\alpha\text{-syn}}$. The region 383-413 of RAG2 was replaced with amino acids 107-137 of human alpha synuclein. (B) Reintegration activity of RAG2$^{\alpha\text{-syn}}$ and RAG2$^{383\text{-}413\Delta}$ relative to wtRAG2 (*N* = 3, error bars represent SEM).

Next, structural disorder prediction of the RAG2 C-terminus amino acid sequence using the MFDp2 algorithm was performed (Mizianty *et al.*, 2013). This approach predicts the disorder of a given amino acid sequence by scoring the number of order or disorder promoting residues, accompanied by alignments against known disordered protein chains (Mizianty *et al.*, 2013). This revealed that most of the acidic hinge is predicted to be very disordered (Figure 4.5). Surprisingly, however a more ordered region is predicted between residues 380-395 (Figure 4.5, O). Notably, the first part of residues 383-413 lies within this region. Since such order infers a structured region, it is possible that this part of the hinge forms a structure which is vital for suppressing reintegration. Unfortunately, this region is missing from all published high resolution structures of the RAG complex, which is likely due to the disordered nature of the rest of the acidic hinge.

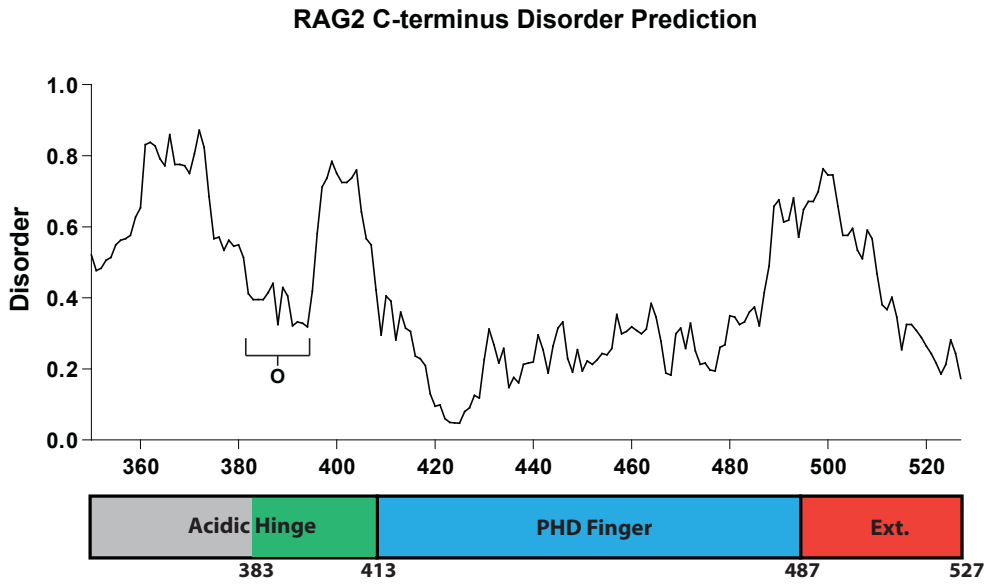


Figure 4.5 - Structural disorder prediction of the RAG2 C-terminus

The sequence of the RAG2 C-terminus was analysed using the MFDp2 algorithm (Mizianty *et al.*, 2013). The locations of the main regions of the RAG2 C-terminus are displayed under the graph. O denotes region of the hinge, spanning residues 380-395 that is predicted to be the most ordered.

4.4 Identification of the amino acids in the RAG2 acidic hinge that are vital to inhibit reintegration

Next, to identify which part of the acidic hinge is required to inhibit reintegration, two additional RAG2 truncation mutants were generated, RAG2$^{403\Delta}$ and RAG2$^{393\Delta}$, which lack the final 10 or 20 amino acids of the acidic hinge, respectively (Figure 4.6A). The reintegration activity of these mutants was compared with that of wild type RAG2 and other acidic hinge mutants (Figure 4.6A). A modest increase in reintegration activity is observed with RAG2$^{403\Delta}$, which is similar to when RAG2 is truncated further to generate RAG2$^{393\Delta}$ (Figure 4.6B). However, the reintegration activity of these truncation mutants is about four-fold lower than the activity observed when the entire acidic hinge is removed with cRAG2 or RAG2$^{383\text{-}413\Delta}$. This suggests that residues spanning 383-393 are crucial for the suppression of reintegration.

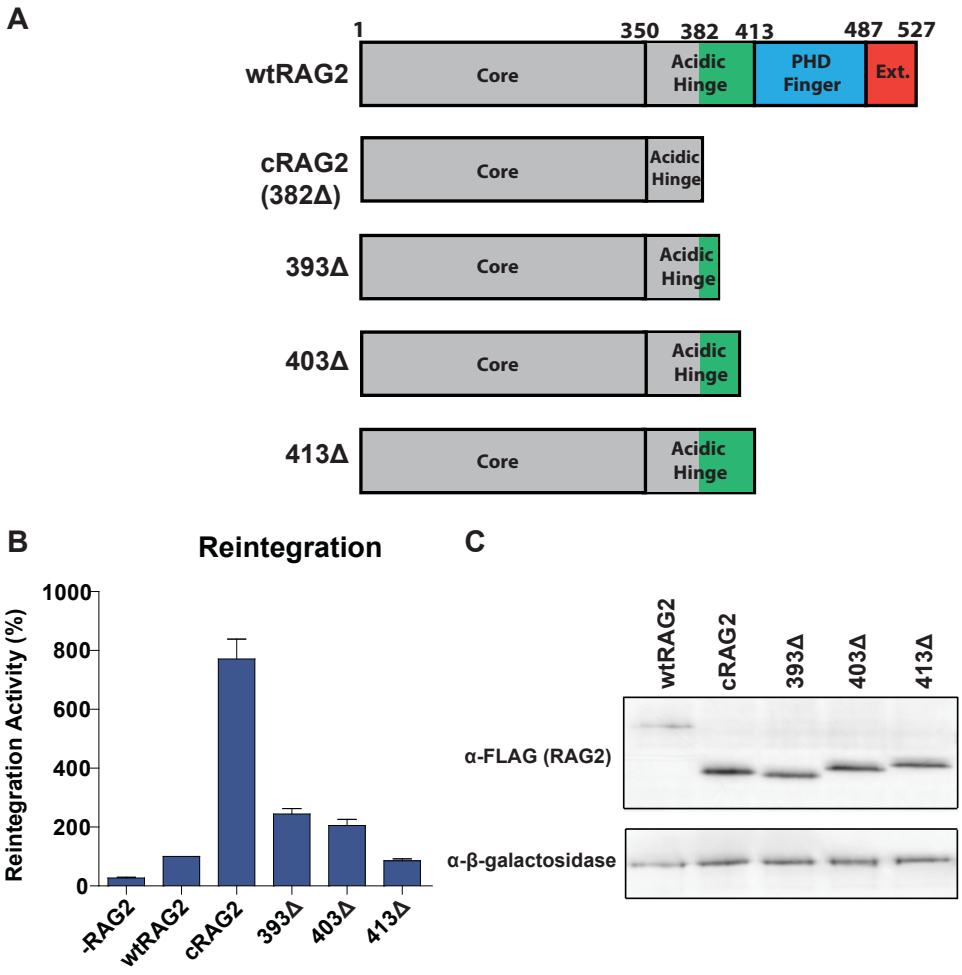


Figure 4.6 - Mapping the RAG2 acidic hinge for the suppression of reintegration

(A) Schematic of RAG2 mutants used in the reintegration assay. (B) Reintegration activity of RAG2 mutants relative to wild type RAG2 (*N* = 3, error bars represent SEM). (C) Western blot of COS7 lysates transfected with wild type or mutant RAG2 and a β-galactosidase expression vector. All RAG2 proteins contained an N-terminal FLAG tag which allows their detection using an anti-FLAG antibody. Samples analysed on this western blot were from an independent transfection than those analysed for reintegration. See Appendix 6 for uncropped blots.

To pinpoint the specific residues within this region that are important for suppressing reintegration, three alanine patch mutants were produced where 3-4 adjacent residues are mutated to alanine (Figure 4.7A; $RAG2^{383\text{-}386A}$, $RAG2^{387\text{-}390A}$, $RAG2^{391\text{-}393A}$). Analysis of the reintegration activity of these alanine patch mutants reveals that only $RAG2^{383\text{-}386A}$ has increased reintegration activity, which is similar to that of cRAG2 and

RAG2$^{383\text{-}413\Delta}$ (Figure 4.7B). This suggests a distinct role for the amino acids 383-386 in suppressing reintegration. However, mutation of the individual residues spanning 383-386 to alanine leads to only a small increase in reintegration (data not shown), suggesting that these individual residues contribute additively to the suppression of reintegration and that no single residue is critical for the suppressive activity.

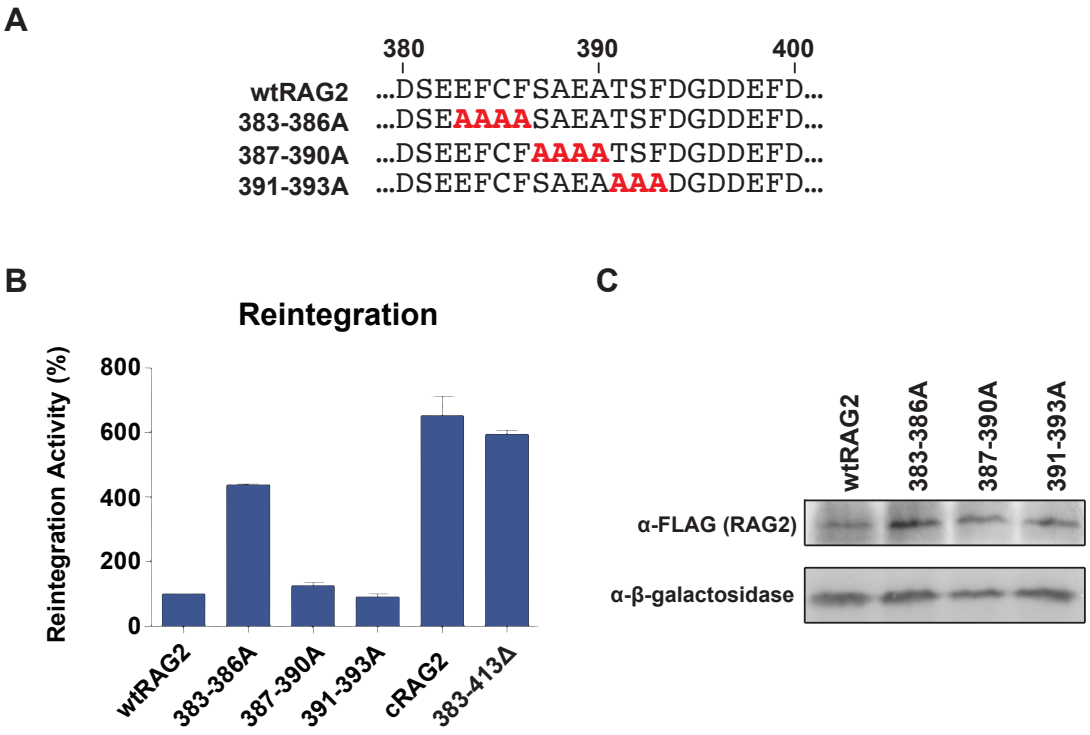


Figure 4.7 - Analysis of reintegration activity of acidic hinge alanine patch mutants

(A) Sequence alignment of wild type RAG2 with the alanine patch mutants. Residues mutated to alanine are shown in red. (B) Reintegration activity of RAG2 mutants relative to wild type RAG2 (*N* = 3, error bars represent SEM). (C) Western blot of COS7 lysates transfected with wild type RAG2 or acidic hinge patch mutants and a β-galactosidase expression vector. All RAG2 proteins contained an N-terminal FLAG tag which allows their detection using an anti-FLAG antibody. Samples analysed on this western blot were from an independent transfection than those analysed for reintegration. See Appendix 7 for uncropped blot.

4.5 Analysis of recombination activity of RAG2 acidic hinge mutants

An autoinhibitory domain has been previously mapped to the residues 388-405 of the RAG2 acidic hinge (Lu *et al.*, 2015; Figure 4.8A), which allosterically inhibits the binding and cleavage activities of RAG1. Once the RAG2 PHD finger engages with

H3K4me3, a number of conformational changes occur in both RAG1 and RAG2 which relieves this autoinhibition (Bettridge *et al.*, 2017). Mutation of all (388-405) or part (388-396 or 397-405) of the autoinhibitory domain to alanine leads to around a three-fold increase in recombination activity compared with wild type RAG2 (Lu *et al.*, 2015). Additionally, since binding of the RAG2 PHD finger to H3K4me3 is required to overcome autoinhibition, mutation of the autoinhibitory domain is able to rescue the recombination activity of PHD finger mutants (Lu *et al.*, 2015; Bettridge *et al.*, 2017).

The increase in reintegration observed with RAG2$^{383\text{-}386A}$ could be due to an increase in overall RAG activity, possibly by disruption of the previously mapped autoinhibitory domain, which is located just C-terminal to residues 383-386 (Figure 4.8A). However, both RAG2$^{387\text{-}390A}$ and RAG2$^{391\text{-}393A}$ have wild type levels of reintegration even though they mutate residues within the autoinhibitory domain. This suggests that increased reintegration activity may not rely on disrupting the autoinhibitory domain. Instead, it is possible that residues 383-386 are involved in a separate inhibitory mechanism. To explore this idea, the recombination activity of the acidic hinge mutants was determined using the *in vivo* V(D)J recombination assay described previously (Section 3.1).

As expected, RAG2$^{383\text{-}413\Delta}$, where the autoinhibitory domain is removed, has increased recombination activity compared with wild type RAG2 (Figure 4.8B). This is consistent with the three-fold increase in recombination observed when the autoinhibitory domain is disrupted (Lu *et al.*, 2015). Both RAG2$^{387\text{-}390A}$ and RAG2$^{391\text{-}393A}$ have recombination activity that is similar to wild type RAG2, even though part of the autoinhibitory domain is mutated (Figure 4.8B). This suggests that although these mutants are able to bind H3K4me3 and overcome autoinhibition like wild type RAG2, these mutations are insufficient to cause activity above wild type levels (as observed with RAG2$^{383\text{-}413\Delta}$). This might be because mutating just three or four amino acids is insufficient to disrupt the autoinhibitory domain and lead to the increased activity observed for larger mutations of the autoinhibitory domain (388-396A, 397-405A or 388-405A; Lu *et al.*, 2015).

In contrast, RAG2$^{383\text{-}386A}$ has increased recombination activity compared to wild type RAG2 (Figure 4.8B). This suggests that the increased reintegration activity observed when residues 383-386 are mutated is indeed due to increased RAG activity, likely caused by disruption of an inhibitory mechanism. However, from this data it is unclear whether the effect of residues 383-386 is an extension of the autoinhibitory domain or an entirely different inhibitory mechanism within the RAG2 acidic hinge.

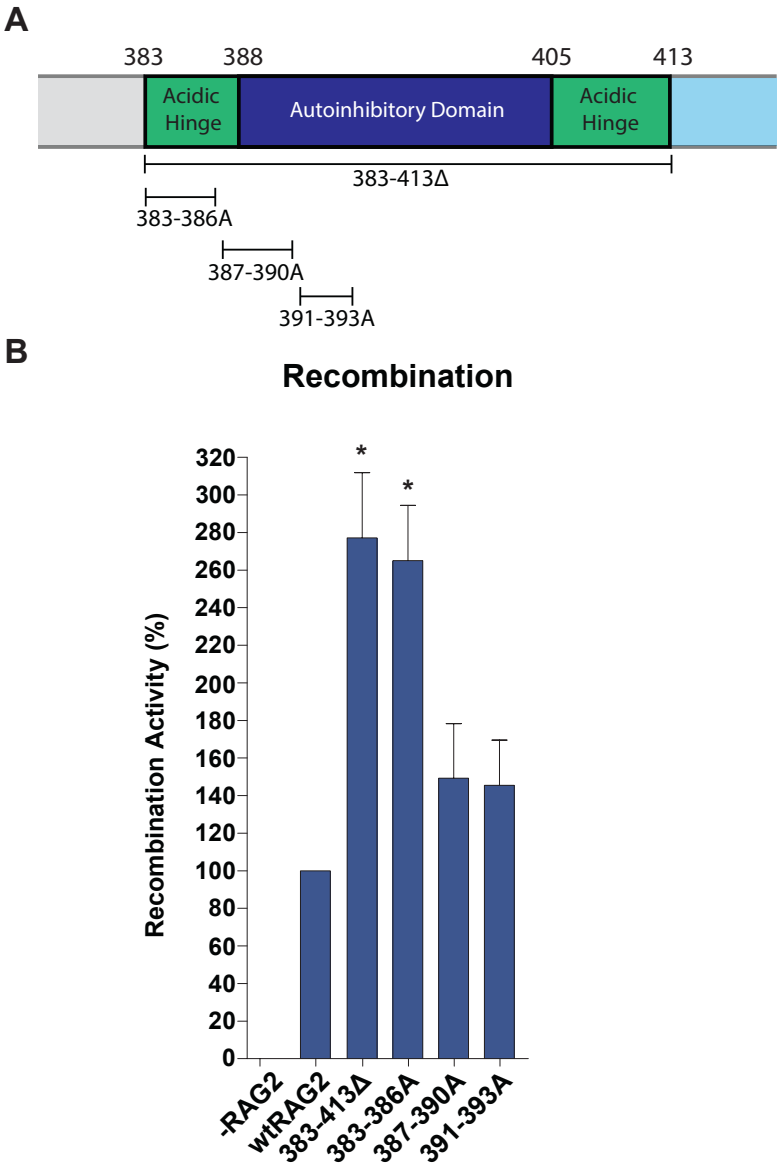


Figure 4.8 - Analysis of recombination activity of RAG2 acidic hinge mutants

(A) Diagram of the RAG2 acidic hinge. The previously mapped autoinhibitory domain (388-405) is shown in blue and the locations of mutated or deleted residues are represented beneath. (B) Recombination activity of RAG2 mutants relative to wild type RAG2 (*N* = 3; error bars represent

SEM). The recombination assay is performed in NIH3T3 cells whereas the reintegration assay is performed in COS7 cells. However, the reintegration activity of RAG2 mutants in NIH3T3 is comparable to the activity observed in COS7 cells (data not shown). Unfortunately, low expression levels in NIH3T3 cells meant that it is not possible to detect RAG2 expression by western blotting. Since the results of the reintegration assay are consistent between COS7 and NIH3T3 cells, it is highly likely that the relative protein expression levels of the RAG2 mutants are also very similar. Asterisks above error bars represent a statistically significant difference in recombination activity compared to wild type RAG2 (Student's T-test; *, $P = 0.05$; **, $P = 0.01$).

4.6 Analysis of *in vivo* recombination activity of acidic hinge mutants with a second mutation in the PHD finger

Next, to decipher if residues 383-386 are part of the autoinhibitory domain, I examined the recombination activity of acidic hinge mutations when the RAG2 PHD finger is mutated in tandem. Mutation of the autoinhibitory domain (388-405A) leads to a constitutively active RAG complex in the absence of the PHD finger binding to H3K4me3, since no allosteric inhibition is imposed upon RAG1 (Bettridge *et al.*, 2017). Therefore, mutation of the autoinhibitory domain is capable of rescuing the recombination activity of RAG2 proteins that also have a mutated PHD finger (W453A; Lu *et al.*, 2015). The W453A mutation prevents engagement of the RAG2 PHD finger with H3K4me3 by disrupting the binding pocket and this stops the accompanying conformational changes that overcome autoinhibition, leading to a severe decrease in recombination (Matthews *et al.*, 2007; Lu *et al.*, 2015; Bettridge *et al.*, 2017). Therefore, if residues 383-386 are part of the previously mapped autoinhibitory domain we would expect that mutation of these residues would rescue the recombination activity of W453A, in the same way as mutating the autoinhibitory domain (Lu *et al.*, 2015). To test this idea, the W453A mutation was introduced into constructs already carrying the acidic hinge mutations ($RAG2^{383\text{-}413\Delta}$, $RAG2^{383\text{-}386A}$, $RAG2^{387\text{-}390A}$, $RAG2^{391\text{-}393A}$), and the recombination activity of the respective proteins was determined.

As expected, relatively high levels of recombination are observed when the entire acidic hinge is removed, including the autoinhibitory domain, in the presence of the

W453A mutation (Figure 4.9, RAG2$^{383\text{-}413\Delta}$). In contrast, recombination activity of RAG2$^{383\text{-}386A}$ is severely decreased when W453A is present, in a similar manner to wild type RAG2 (Figure 4.9). This suggests that the autoinhibitory domain is not disrupted by mutating residues 383-386. Nonetheless, increased recombination and reintegration activity is observed. This suggests that residues 383-386 may be part of a distinct inhibitory mechanism.

A caveat to this conclusion, however, is that both RAG2$^{387\text{-}390A}$ and RAG2$^{391\text{-}393A}$ are also unable to rescue recombination, even though residues in the autoinhibitory domain are mutated (Figure 4.9). Together with the observation that these mutations do not cause increased recombination (Figure 4.8), this suggests that mutating 3-4 amino acids may insufficient to disrupt the autoinhibitory domain and that a much larger region needs to be mutated. Because of this, it is not possible to fully conclude that residues 383-386 are part of a distinct inhibitory mechanism.

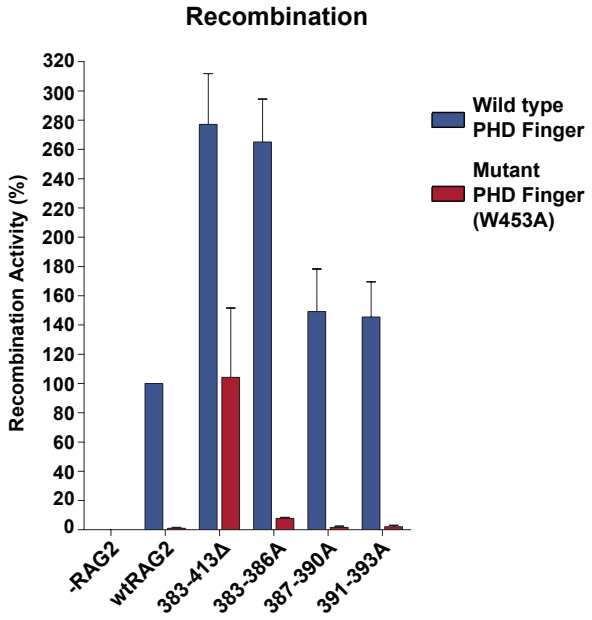


Figure 4.9 - Analysis of recombination of acidic hinge mutants with an additional PHD finger mutation

RAG2$^{383\text{-}386A}$ does not rescue *in vivo* recombination of a RAG2 PHD finger mutant. The *in vivo* V(D)J recombination reaction was performed using the different RAG2 mutants indicated, in the presence (red bars) or absence (blue bars) of the W453A mutation. (*N* = 3, error bars represent SEM).

To address this further, two additional mutations were generated where blocks of eight amino acids of the RAG2 acidic hinge were mutated: RAG2$^{383\text{-}390A}$ and RAG2$^{387\text{-}394A}$. Previously, Lu *et* al. (2015) showed that mutation of nine amino acidic of the autoinhibitory domain (388-396A or 397-405A) enables recombination activity of PHD finger mutants at wild type levels. Therefore, we may expect RAG2$^{387\text{-}394A}$, which mutates eight residues located entirely within the autoinhibitory domain, to have increased recombination as well as the ability to rescue recombination when W453A is present. In contrast, if residues 383-386 are indeed part of a distinct inhibitory mechanism, and mutating four residues is not enough to disrupt the autoinhibitory domain, we may expect RAG2$^{383\text{-}390A}$ to have increased recombination activity but will be unable to rescue recombination when W453A is also present.

However, RAG2$^{387\text{-}394A}$ has recombination activity which is similar to wild type RAG2 and is unable to rescue recombination activity when the PHD finger is also mutated (Figure 4.10). This is similar to the activity observed with RAG2$^{387\text{-}390A}$ and RAG2$^{391\text{-}393A}$ (Figure 4.9), but slightly more recombination activity is retained with RAG2$^{387\text{-}394A}$ when the PHD finger is mutated. Since Lu *et al.* (2015) demonstrated that mutating residues 388-396 to alanine is capable of rescuing recombination when the PHD finger is mutated, this may suggest that residues 395 and 396, which are not mutated in RAG2$^{387\text{-}394A}$, maybe crucial for the activity of the autoinhibitory domain. RAG2$^{383\text{-}390A}$ has increased recombination activity compared with wild type RAG2 and is able to support slightly higher levels of recombination when the PHD finger is mutated (Figure 4.10). However, this rescue of recombination activity is much lower than that observed when the entire acidic hinge is removed (Figure 4.9, RAG2$^{383\text{-}413\Delta}$) or when nine or 18 residues of the autoinhibitory domain are mutated to alanine (388-396A, 397-405A, 388-405A; Lu *et al.*, 2015). Unfortunately, therefore it is not possible to conclude from this data if residues 383-386 are part of or distinct from the activities of the RAG2 autoinhibitory domain.

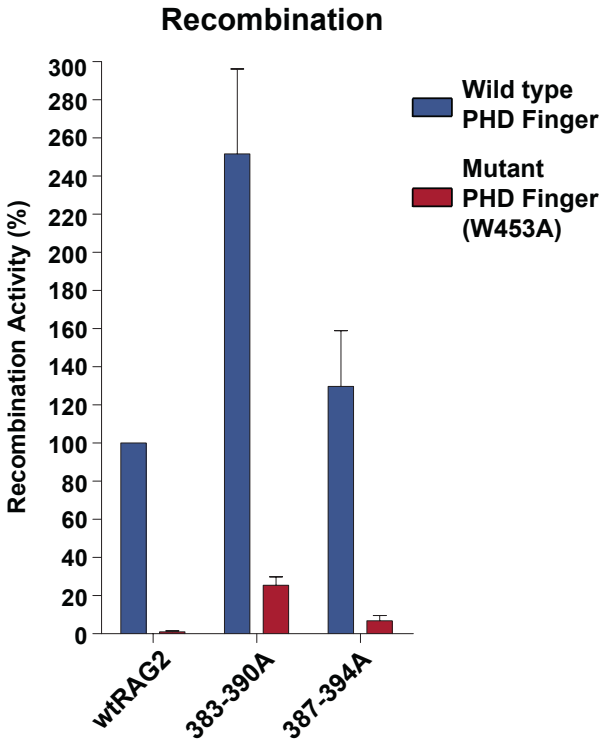


Figure 4.10 - Analysis of recombination of acidic hinge mutants with a large number of residues mutated to alanine.

The *in vivo* V(D)J recombination reaction was performed using the different RAG2 mutants indicated, in the presence (red bars) or absence (blue bars) of the W453A mutation. (*N* = 3, error bars represent SEM).

4.7 Purification of full-length RAG2 proteins for *in vitro* analysis of 383-386A

To investigate the molecular basis of the inhibitory mechanism of residues 383-386, I next performed *in vitro* assays involving purified RAG proteins. Given that residues 383-386 reside in the non-core region of RAG2, it is necessary to purify full-length RAG2 (flRAG2). Since purification of full-length RAG1 is notoriously problematic, all flRAG2 proteins were co-purified with core RAG1.

An alternative protocol was used with milder lysis buffers than those used for core RAG purification (Section 3.3; Section 2.35) to generate flRAG2 protein that is more active (Bergeron et al., 2006). The 383-386A mutation was introduced into the flRAG2 expression vector (pEF-MfR2) which contains the cDNA sequence of mouse flRAG2 (aa 1-527), fused to an N-terminal maltose binding protein (MBP) amylose affinity tag (Figure 4.11A). This was co-transfected into HEK293T cells along with the expression vector for core RAG1 (pEF-McR1). MBP-RAG proteins were purified from transfected HEK293T lysates using amylose affinity chromatography (Section 2.35) and fractions containing RAG proteins were pooled (Figure 4.11B). Following purification of all the required proteins, each preparation was analysed by SDS-PAGE and coomassie staining. The concentration of each preparation was determined by 2D densitometry using ImageJ and equal amounts were used for *in vitro* studies (Figure 4.11C).

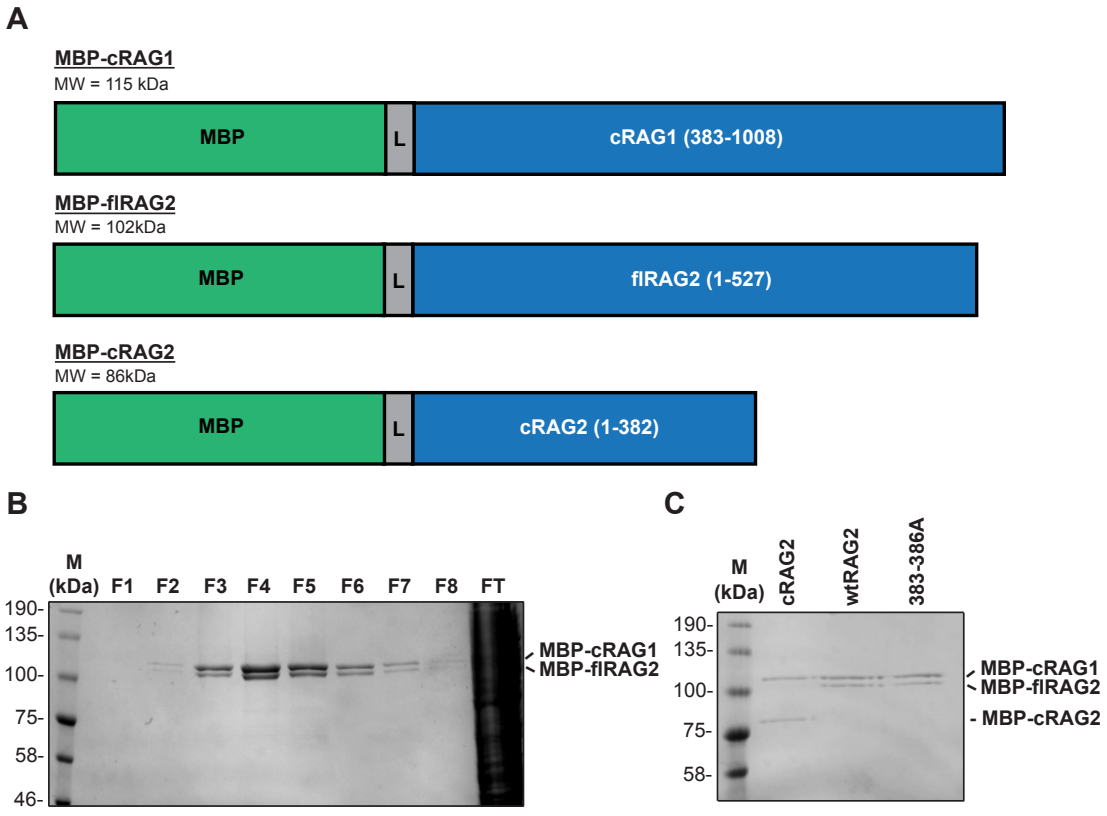


Figure 4.11 - Purification of full-length RAG2 proteins

(A) Diagram representing MBP-cRAG1, MBP-flRAG2 and MBP-cRAG2 fusion proteins. The predicted molecular weights (MW) of each protein are shown. Maltose binding protein (MBP) is shown in green, followed by a short glycine-serine linker (L) in grey and either RAG1 or RAG2 in blue. The amino acids of cRAG1, flRAG2 and cRAG2 present are shown in brackets. (B) Representative coomassie stained SDS-PAGE gel of 8 fractions from a full-length RAG2 purification. The positions of MBP-cRAG1 and MBP-flRAG2 and indicated next to the gel. (C) Coomassie stained SDS-PAGE gel showing equivalent levels of cRAG2, wtRAG2 and RAG2$^{383\text{-}386A}$ proteins from each purification.

4.8 Cleavage and binding activities of RAG2$^{383\text{-}386A}$ are similar to wild type RAG2

To test if the suppressive effect of residues 383-386 occurs at the binding or cleavage stages of V(D)J recombination, I next performed *in vitro* cleavage and binding assays using purified RAG2$^{383\text{-}386A}$. Cleavage and binding reactions were performed as described in Section 3.4 and 3.5 but with minor alterations. Firstly, a short H3K4me3 peptide was included since reactions with full-length RAG2 contain the RAG2 PHD

finger, which must bind H3K4me3 for efficient cleavage and binding. Secondly, the DMSO concentration was decreased to 10% v/v since reactions performed with the usual DMSO concentration (20% v/v) proceeded too quickly and caused excessive cleavage and binding with full-length RAG2 in the absence of the H3K4me3 peptide.

As expected, reactions containing core RAG2 are able to efficiently bind and cleave a 12- or 23-RSS, regardless of the inclusion of the H3K4me3 peptide (Figure 4.12A and B, cRAG2). In contrast, reactions with wild type full-length RAG2 depend on the presence of the H3K4me3 peptide for both efficient binding and cleavage (Figure 4.12A and B, wtRAG2). This is consistent with similar *in vitro* studies performed by Lu *et al.*, (2015). Interestingly, $RAG2^{383\text{-}386A}$ has cleavage and binding activities that are similar to wild type full-length RAG2 (Figure 4.12A and B, $RAG2^{383\text{-}386A}$). The requirement of the H3K4me3 peptide for efficient binding and cleavage in reactions containing $RAG2^{383\text{-}386A}$ is consistent with inability of this mutant to rescue recombination activity *in vivo* when the PHD finger is mutated (Figure 4.9, $RAG2^{383\text{-}386A}$). Notably, since the cleavage activity and complexes formed when residues 383-386 are mutated to alanine are very similar to wild type full-length RAG2, this suggests that the suppressive role of these residues does not effect the binding or cleavage steps of V(D)J recombination.

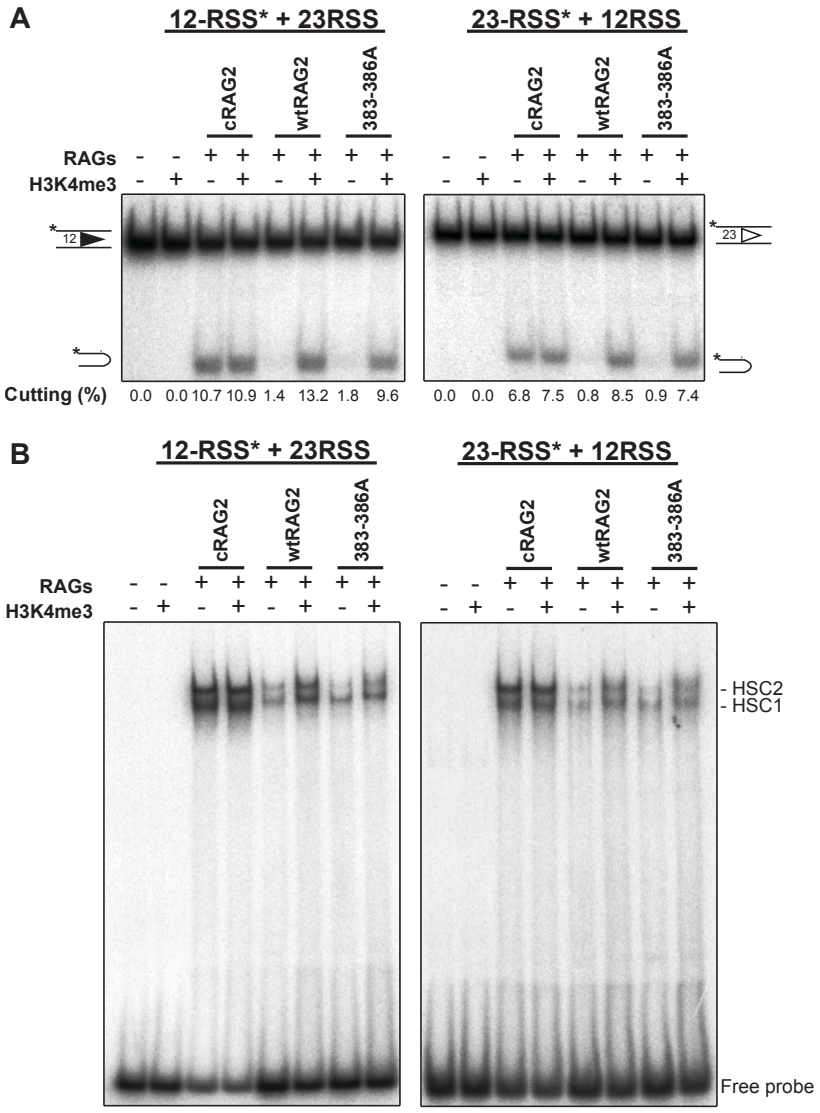


Figure 4.12 - Analysis of RAG binding and cleavage activities of RAG2$^{383\text{-}386A}$

(A) Cleavage activity of RAG2$^{383\text{-}386A}$ is similar to wild type full-length RAG2. RAG cleavage reactions were performed as in Figure 3.9 but DMSO concentration was decreased to 10%. A H3K4me3 peptide, which contains the first 21 amino acids of histone H3 and trimethylated lysine 4, was included where indicated at a concentration of 2 μM. (B) Complexes formed with a 12- or 23-RSS by RAG2$^{383\text{-}386A}$ are similar to wild type full-length RAG2. Binding reactions were identical to cleavage reactions but calcium replaced magnesium as the divalent cation.

4.9 Analysis of hybrid joint formation *in vivo*

Given that mutating residues 383-386 does not appear to effect the binding and cleavage steps of V(D)J recombination, it is possible that these residues play a role in the joining step. The RAG2 C-terminus suppresses several aberrant V(D)J recombination reactions, such as; hybrid joining, reintegration and transposition (Sekiguchi *et al.*, 2001; Elkin *et al.*, 2003; Curry *et al.*, 2007). These aberrant reactions share one common feature; incorrect blunt-to-hairpin end joining. Since my data suggests that residues 383-386 are involved in suppressing one of these reactions, reintegration, I next set out to test if these residues are also involved in suppressing other similar aberrant reactions, namely hybrid joint formation. It is possible that residues 383-386 may suppress joining in a way that ensures the correct resolution of the DNA ends, by avoiding aberrant blunt-to-hairpin end joining. To test this idea, the ability of the RAG2 acidic hinge mutants to catalyse hybrid joints was analysed. This was performed in a similar way to the *in vivo* recombination assay, using pJH299 as a substrate but with primers designed to analyse hybrid joint formation (Section 2.28; Figure 4.13A, primers 5 and 6).

Unfortunately, hybrid joint formation could not be quantified by qPCR as the primers used also amplify a larger product from pJH299 (Figure 4.13A, primers 5 and 6). Instead, a semi-quantitative PCR method was used which is similar to other published methods for the analysis of hybrid joints (Coussens *et al.*, 2013). The PCR conditions were optimised to ensure that they can detect increases in hybrid joints (Figure 4.13B), and under the conditions used, hybrid joints increase according to increased amounts of pEFXC-RAG2$^{383\text{-}413\Delta}$ transfected (Figure 4.13C).

The level of hybrid joints formation with the different RAG2 mutants was determined using 2D densitometry in ImageJ (Figure 4.13D) and a representative gel is shown in Figure 4.13E. Removal of the entire RAG2 acidic hinge (Figure 4.13D, RAG2$^{383\text{-}413\Delta}$) or mutation of residues 383-386 (Figure 4.13D, RAG2$^{383\text{-}386A}$) leads to increased hybrid joint formation compared to wild type RAG2. The levels are highest with RAG2$^{383\text{-}413\Delta}$

which may be due to the removal of both the autoinhibitory domain, which increases the cutting and binding activities of the RAG complex, and residues 383-386. By contrast, mutation of the adjacent residues of the acidic hinge appear to have no effect on hybrid joint formation since the levels are very similar to wild type RAG2 (Figure 4.13D, $RAG2^{387\text{-}390A}$ and $RAG2^{391\text{-}393A}$). Notably, the ability of these acidic hinge mutants to catalyse hybrid joint formation is remarkably similar to their ability to catalyse reintegration (Figure 4.7B). This is consistent with the idea that a mechanism to inhibit aberrant hairpin-to-blunt joining may exist and the residues 383-386 are involved with this.

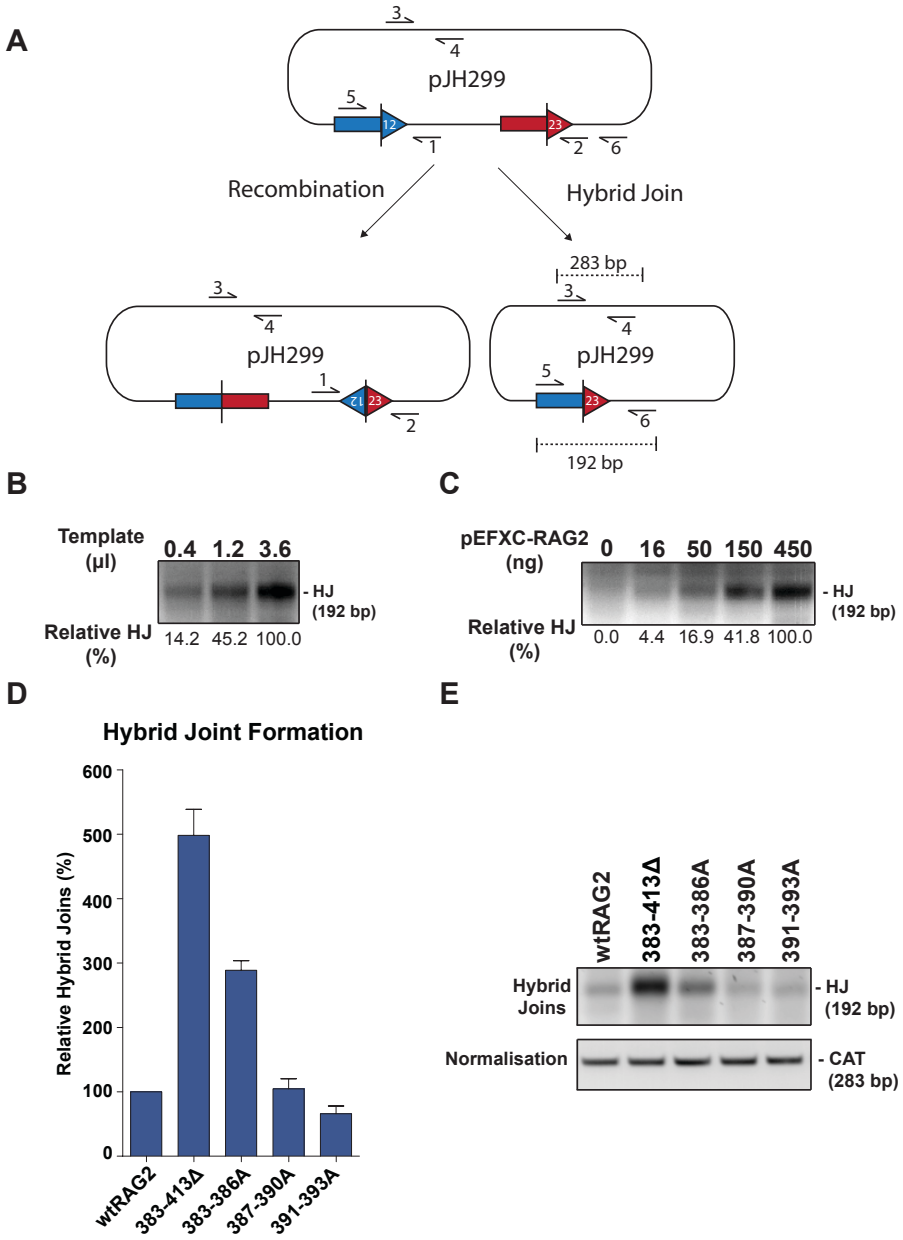


Figure 4.13 - Analysis of hybrid joint formation by RAG2 acidic hinge mutants

(A) Schematic of hybrid joint formation using pJH299. (B) The PCR reaction is semi-quantitative. Hybrid joint PCR was performed using a three fold serial dilution of template. The level of hybrid joint formation is shown under the gel. (C) Hybrid joint formation correlates with the amount of RAG2 present. Increasing amounts of pEFXC-RAG2$^{383\text{-}413\Delta}$ were transfected and the level of hybrid joint formation determined. (D) The levels of hybrid joint formation with the different RAG2 mutants was quantified by 2D densitometry in ImageJ and is shown relative to wtRAG2 (*N* = 3, error bars represent SEM). (E) Representative gel from the analysis of hybrid joints formed using the different RAG2 mutants (top). A separate PCR using primers 3 and 4 which, amplify a region of the chloramphenicol resistance (CAT) gene, was performed (bottom) and shows a similar level of template is present in each sample.

C) Discussion

In this chapter, I describe the generation of an assay to quantitatively detect reintegration *in vivo* (Figure 4.2) and its use to identify the region of the RAG2 C-terminus which is responsible for inhibiting reintegration (Figure 4.3). This led to the identification of a patch of four residues in the RAG2 acidic hinge (aa 383-386) that are vital to suppress reintegration (Figure 4.7) and appear to also suppress hybrid joint formation (Figure 4.13). Mutation of residues 383-386 does not affect the binding and cleavage steps of V(D)J recombination (Figure 4.12) but does cause an increase in recombination activity *in vivo* (Figure 4.8). Together, these data suggest a novel inhibitory region may exist in the RAG2 acidic hinge that suppresses aberrant end joining reactions during V(D)J recombination.

4.10 Generation of an *in vivo* extra-chromosomal reintegration assay

Existing assays to analyse reintegration were not suitable to accurately quantify the levels of reintegration of a large number of different RAG2 mutants. One previous study took advantage of a reintegration hotspot (b196 hotspot) in developing T cells to determine the frequency of reintegration (Curry *et al.*, 2007). The authors performed nested PCR on DNA extracted from 85 mice thymi and determined reintegration frequency by counting reactions which were positive for reintegration as determined by the presence of bands on an agarose gel. Although informative, this assay is very time and labour intensive and would be impractical to screen a large number of RAG2 mutants. In a separate study, reintegration was readily detected in fibroblasts which had been transfected with separate plasmids which act as a target RSS and an ESC, together with RAG expression vectors, but this assay was only semi-quantitative (Vanura *et al.*, 2007). Therefore, I developed a qPCR based assay to quantify the amount of reintegration from transfected fibroblasts (Figure 4.2A).

The reintegration assay I generated is able to measure the differences in reintegration activity from cells transfected with increasing amounts of pEFXC-cRAG2 (Figure 4.2B).

Additionally, the nature of qPCR means that a large number of samples can be analysed in a fairly high-throughput manner. However, a drawback of this assay is that is uses extra-chromosomal substrates and consequently it may not be suitable to analyse the effects of chromatin on reintegration. This could be addressed by performing this assay using a target RSS which is located in the genome. Alternatively, reintegration at the b196 hotspot in developing T cells could be quantified by qPCR.

4.11 A patch of amino acids within the RAG2 acidic hinge is vital to inhibit reintegration and hybrid joint formation

Since reintegration is seven-fold higher in mice expressing cRAG2 compared to wild type RAG2 (Curry *et al.*, 2007), I set out to determine which part of the RAG2 C-terminus is responsible for suppressing reintegration. By analysing the reintegration activity of RAG2 C-terminus truncation mutants, the regions involved in suppressing reintegration could be mapped.

Removal of the entire RAG2 C-terminus leads to increased reintegration, consistent with the study by Curry *et al.* (2007) (Figure 4.3B, cRAG2). In contrast, reintegration activity with $RAG2^{413\Delta}$ is similar to wild type RAG2 (Figure 4.3B). This mutant extends cRAG2 by 30 amino acids and these residues span the acidic hinge (Figure 4.3A), suggesting this region is completely responsible for inhibiting reintegration. Indeed, this is confirmed by deleting the acidic hinge alone which leads to a similar increase in reintegration activity as cRAG2 (Figure 4.3B, $RAG2^{383\text{-}413\Delta}$).

Reintegration activity is high when the acidic hinge is replaced with a sequence which has similar properties (Figure 4.4, $RAG2^{\alpha\text{-}syn}$), suggesting that the negative charge or intrinsic disorder of this region may not be involved in suppressing reintegration. Instead, part of the hinge may form a structure which acts via an unknown mechanism to suppress reintegration. Notably, a distinct ordered region of the acidic hinge is predicted to span the amino acids involved in inhibiting reintegration (Figure 4.5, aa ~380-395). Truncating the acidic hinge further reveals that the suppressive residues

indeed fall within this more ordered region (Figure 4.6B). Alanine scanning mutations in the residues spanning 383-393 pinpointed the residues vital for suppressing reintegration to residues 383-386 (Figure 4.7A and B). Additionally, analysis of hybrid joint formation by the acidic hinge mutants revealed that residues 383-386 are also involved in the suppression this aberrant joining reaction (Figure 4.13).

4.12 Are residues spanning 383-386 part of in the RAG2 autoinhibitory domain?

Similar to reintegration activity, the recombination activity of RAG2$^{383\text{-}386A}$ is also increased compared to wild type RAG2 (Figure 4.8). This suggests that mutation of residues 383-386 disrupts an inhibitory mechanism, leading to a more active RAG complex. Given that these residues are adjacent to a previously mapped autoinhibitory domain (Figure 4.8A; 388-405), this opened the possibility that mutation of residues 383-386 disrupts this domain, leading to increased recombination activity and to possibly also cause the increased reintegration activity and hybrid joint formation observed with RAG2$^{383\text{-}386A}$.

To test this idea, the ability of RAG2 acidic hinge mutants to rescue the recombination activity of a PHD finger mutant was determined. Unlike removal of the autoinhibitory domain (Figure 4.9, RAG2$^{383\text{-}413\Delta}$), mutation of residues 383-386 is unable to overcome autoinhibition when the PHD finger is mutated in tandem (Figure 4.9, RAG2$^{383\text{-}386A}$). Therefore, mutating residues 383-386 does not appear to disrupt the autoinhibitory domain and instead suggests that the effect observed when mutating these residues is due to disruption of a separate inhibitory mechanism. This is supported by analysis of cleavage and binding *in vitro* which appear to be unaffected by mutating residues 383-386 (Figure 4.12A and B), suggesting they are involved in the joining step of V(D)J recombination. Additionally, the suppressive mechanism of residues 383-386 appears to act after autoinhibition has been relieved since increased recombination is only observed with RAG2$^{383\text{-}386A}$ when the PHD finger is wild type .

However, a caveat to this conclusion is that mutating a similar number of residues within the autoinhibitory domain is unable to disrupt autoinhibition (Figure 4.9, $RAG2^{387\text{-}390A}$ and $RAG2^{391\text{-}393A}$). A further two RAG2 acidic hinge mutants where a larger number of amino acids were mutated ($RAG2^{383\text{-}390A}$ and $RAG2^{387\text{-}394A}$) also appeared to not disrupt the autoinhibitory domain. Therefore, it can not be unequivocally concluded that the effects of residues 383-386 are separate from the autoinhibitory domain and further mutational analysis, such as mutating residues 395-403A, is required to determine the contribution of residues 383-386 to the autoinhibitory domain.

4.13 How might residues 383-386 inhibit aberrant V(D)J recombination?

The suppression of reintegration by residues 383-386 could occur in a number of ways. It may play a role in the suppression of *trans* V(D)J recombination, which has been shown to be inhibited at the joining step (Han *et al.,* 1999), or it could prevent the ESC from becoming re-opened. However, it has been demonstrated that ESCs can become readily reopened by RAGs via a nick-nick mechanism and also the signal ends could reintegrate before the ESC is ligated, as the latter occurs only when RAGs have been down-regulated (Neiditch *et al.*, 2002). A third and more likely possibility, as discussed below, is that the residues 383-386 acts to promote correct blunt-to-blunt and hairpin-to-hairpin DNA end joining.

The outcome of reintegration is similar to other aberrant V(D)J recombination reactions; the incorrect joining of a hairpin coding end to a blunt signal end (Figure 4.1A). Most notably, hybrid joint formation involves this same erroneous resolution of the ends generated during recombination. Another aberrant reaction is transposition which is mechanistically similar to hybrid joint formation but instead results in the joining of the blunt signal end to any region of the genome. Despite this, the only detected transposition event *in vivo* occurred at a sequence that is thought to form a cruciform structure, terminating in a hairpin (Roth, 2003). This raises the possibility that, at least *in vivo*, transposition also involves this same aberrant joining of blunt and hairpin ends.

The RAG2 C-terminus has been demonstrated to inhibit all of these reactions (Sekiguchi *et al.*, 2001; Elkin *et al.*, 2003; Curry *et al.,* 2007). Therefore, it seems likely that RAGs include an inherent mechanism to suppress this broader aberrant reaction, rather than a novel mechanism which suppresses just the reintegration reaction.

Indeed, mutating residues 383-386 leads to an increase in both reintegration activity (Figure 4.7B) and hybrid joint formation (Figure 4.13). Since the inhibitory effects of residues 383-386 potentially during the joining step (Figure 4.12) and both reintegration and hybrid joint formation involve incorrect hairpin-to-blunt end joining, it is possible that residues 383-386 act to suppress this aberrant joining of ends during V(D)J recombination. Additionally, it has been proposed that the RAG2 C-terminus includes a binding pocket which has a higher affinity for hairpin DNA ends over other DNA targets, due to the ability of flRAG2 to inhibit transposition but only in the presence of coding ends (Elkin *et al.*, 2003). Notably, residues 383-386 lie with a region of the hinge which is predicted to be the most ordered area (Figure 4.5), therefore potentially forming a structure which might be the proposed hairpin binding pocket. This may act to isolate the coding ends from the signal ends to help guide their joining by the NHEJ machinery, enforcing blunt-to-blunt and hairpin-to-hairpin joining. Notably, a region of the acidic hinge upstream of residues 383-386 has shown to be involved in regulating repair pathway choice, promoting repair of ends by the classical NHEJ machinery (Coussens *et al.*, 2013; Figure 4.14).

Interestingly, the conformational changes that occur in RAG1 and RAG2 once the PHD finger binds to H3K4me3, leads to residues F384 and F386 of the RAG2 acidic hinge becoming more susceptible to thermolysin cleavage (Bettridge *et al.*, 2017), suggesting that this region becomes more accessible. Notably, this occurs at the same time that autoinhibition is relieved which causes enhanced DNA binding and cleavage activity (Bettridge *et al.*, 2017). Therefore, the 383-386 hairpin binding pocket could move the coding ends into closer proximity and into a region that is more accessible to the NHEJ machinery, to coordinate their proper joining. Mutation of residues 383-386 would

disrupt this mechanism, leaving coding and signal ends equally accessible to the NHEJ machinery. The future directions for this project, including testing this proposed model of inhibiting blunt-to-hairpin joining, are discussed in Chapter 6.

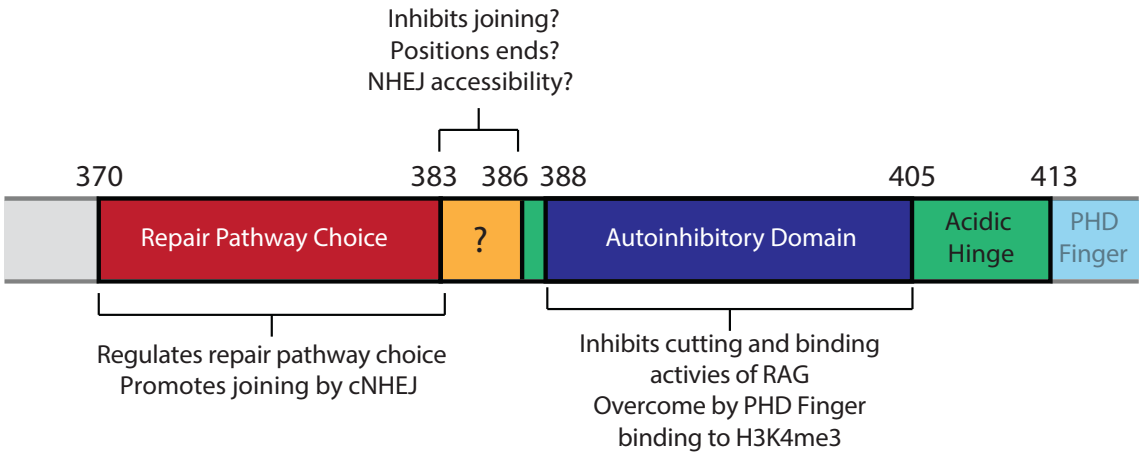


Figure 4.14 - Regulatory functions of the RAG2 acidic hinge

Diagram of the RAG2 acidic hinge. The roles of specific regions which have been mapped are shown. The autoinhibitory domain (388-405) is shown in blue (Lu *et al.*, 2015), a region important in regulating repair pathway choice (370-383) is shown in red (Coussens *et al.*, 2013) and the region spanning residues 383-386 is shown in yellow. Parts of the hinge with no defined function are shown in green.

Chapter 5 - Does an interaction between the RAG2 C-terminus and the nucleosome stabilise the PCC?

A) Introduction

The creation of a break in our genome is a hazardous event which can lead to chromosome translocations and genome instability. However, V(D)J recombination absolutely requires the generation of double-stranded DNA breaks (DSBs) and therefore improper repair of these breaks could compromise the immune system or even lead to cancer. After cleavage the RAG proteins remain associated with the broken ends in a post-cleavage complex (PCC), and the transfer of broken ends from the RAG PCC to the NHEJ machinery is a crucial, yet unresolved step in the reaction that must be correctly regulated to prevent genome instability.

Within the PCC, RAGs remain stably associated with the signal ends both *in vitro* and *in vivo*, due to the binding of RAG1 to the RSS (Agrawal and Schatz, 1997; Cortes *et al.*, 1996). In contrast, the coding ends are rapidly lost from the PCC *in vitro*, however, this is not thought to occur *in vivo* and their joining is thought to be coordinated by RAG proteins (Schatz and Swanson, 2011). Indeed, there is substantial evidence to suggest a role for the RAG2 C-terminus in guiding repair (Coussens *et al.*, 2013; Gigi *et al.*, 2014), as removal of this region leads to genomic instability and lymphomagenesis in mice (Deriano *et al.*, 2011). Currently published structures of the RAG complex lack two important features: (i) the PCC structure, due to its instability and (ii) the RAG2 C-terminus, due to its inherent flexibility. Consequently, vital information is missing regarding how the RAG complex ensures correct repair following cleavage and how this recruits the NHEJ machinery.

V(D)J recombination *in vivo* occurs in the context of chromatin packaging, yet the majority of biochemical studies of the RAG proteins use naked DNA substrates.

Therefore, these studies may be missing a crucial binding parter, the nucleosome, which could aid in the retention of coding ends within the PCC. In pro-B cells, at the *Igh* locus, the RSS heptamer was shown to be located at the entry point of the nucleosome, placing the rest of the RSS within the accessible linker DNA; this would place the coding segment within the nucleosome (Pulivarthy *et al.*, 2016). It is possible that within the PCC *in vivo,* RAG1 holds the signal ends via its interaction with the RSS whereas RAG2, which is positioned near the coding end (Ru *et al.,* 2015; Kim *et al.*, 2018), holds the coding end via an unresolved interaction with the nucleosome.

Indeed, the RAG2 C-terminus has been demonstrated to interact with nucleosomes in two different ways: (i) the binding of the RAG2 PHD finger to H3K4me3 (Matthews et al., 2007; Ramon-Maiques et al., 2007; Liu et al., 2007a; Morshead et al., 2003; Schatz and Ji, 2011), and (ii) the direct interaction of an intrinsically disordered domain (IDD) of the acidic hinge with the core histones (West et al., 2005). I propose a model whereby the RAG2 C-terminus holds the coding ends in the PCC, prior to joining, through these interactions with the nucleosome (Figure 5.1A). Without these interactions, the coding end may be readily released *in vivo*, as they are *in vitro* with non-chromatin substrates, and cause genomic instability (Figure 5.1B). This is consistent with the genomic instability observed in mice lacking the RAG2 C-terminus (Deriano et al., 2011). Crucially, if this model is correct, a system utilising chromatin substrates could stabilise the PCC *in vitro* and allow its structure to be solved for the first time. Additionally, IDDs are thought to fold upon interaction with their target (Warren and Shechter, 2017; Mollica *et al.*, 2016), and consequently, binding of the IDD within the RAG2 acidic hinge to core histones may cause this to adopt a stable conformation and also allow the structure of the RAG2 C-terminus to be resolved. Together, these structures would provide a wealth of knowledge into the elusive joining step of V(D)J recombination. In this chapter, I set out to develop chromatin substrates for use in biochemical studies of the RAG proteins, that could be used to test the model outlined above (Figure 5.1).

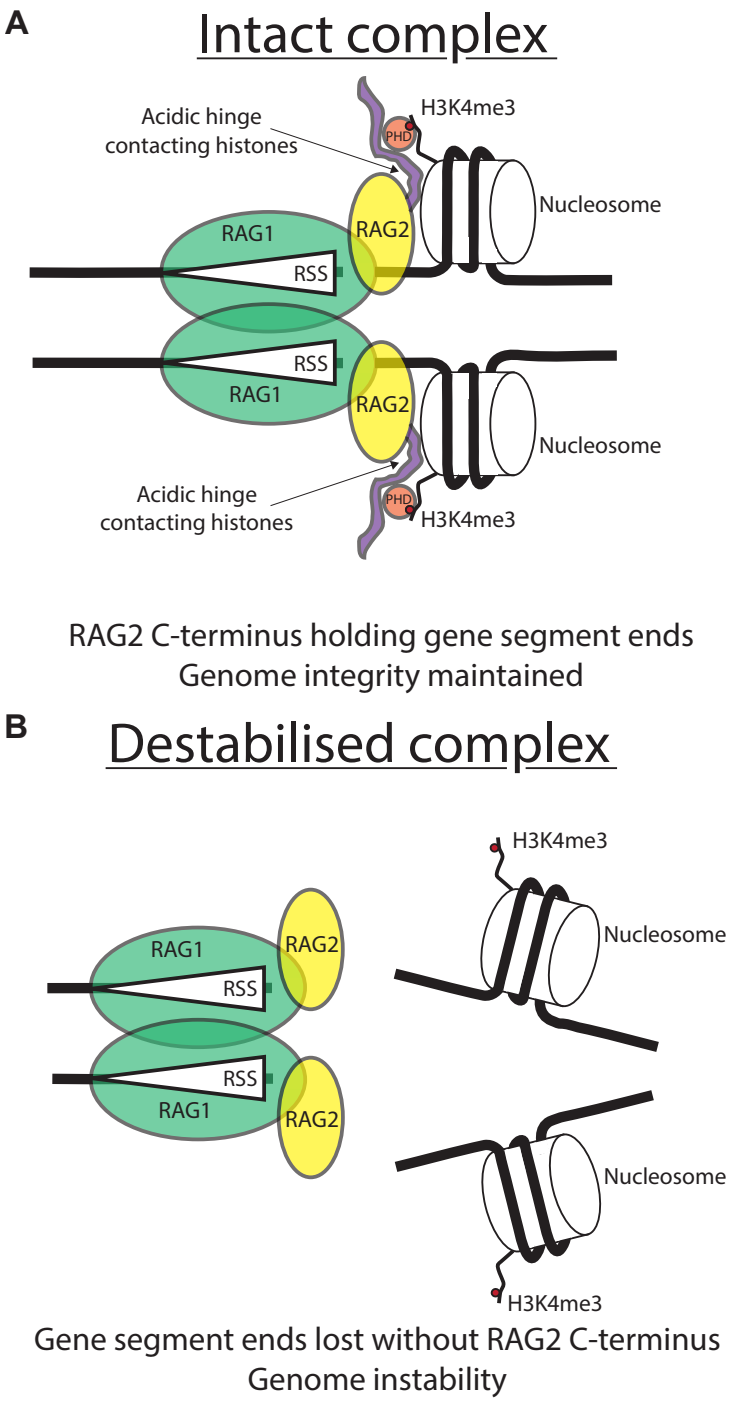


Figure 5.1 - Proposed model of stabilisation of the PCC by an interaction of the RAG2 C-terminus with the nucleosome

(A) Schematic showing an intact PCC facilitated by RAG1 (green ovals) binding to the RSS (white triangles) on the signal ends and RAG2 (yellow ovals) binding to the nucleosome on the coding ends, via the acidic hinge (purple) and PHD finger (orange circle). (B) Schematic showing a destabilised PCC in the absence of the RAG2 C-terminus. The core RAG complex remains bound to the RSS on the signal ends but the coding ends are released due to the lack of the RAG2 C-terminus binding to the nucleosome.

B) Results

5.1 Generation of 601-RSS templates for nucleosome reconstitution

To study RAG activity on chromatin substrates *in vitro*, a template which allows the uniform positioning of a nucleosome next to an RSS is required. To achieve this, a strong nucleosome positioning element, the Wisdom 601 sequence (Lowary and Widom, 1998), was cloned 5' of the heptamer sequence both for a 12- and a 23-RSS (Figure 5.2A). The 601 sequence contains 147 bp of DNA and was identified from a pool of fragments by SELEX (Systematic Evolution of ligands by Exponential Enrichment), selecting those which have the highest propensity to position a nucleosome (Lowary and Widom, 1998). Therefore the 601 sequence should precisely position a nucleosome in close proximity to the RSS and, after RAG cleavage, the nucleosome remains on the coding end, which therefore mimics the likely position at RSSs undergoing recombination *in vivo* (Pulivarthy *et al.*, 2016).

Initially the 601 sequence was positioned directly next to the RSS heptamer, but after nucleosome reconstitution, no RAG cleavage or binding activity was observed (data not shown). This is likely due to the nucleosome inhibiting RAG access to the RSS. Instead, a 7 bp linker was placed between the end of the 601 sequence and the start of the RSS, giving 601-7-12-RSS and 601-7-23-RSS (Figure 5.2A). These fragments were individually cloned into the pUC19 cloning vector for their large scale production (Section 2.47), which was performed by two different methods.

For RAG binding and cleavage assays labelled versions of these fragments were required. This was achieved by PCR using a forward primer containing a 5' Alexa Fluor 488 dye molecule and an unlabelled reverse primer (Figure 5.2A; Section 2.49). The labelled 601-7-12-RSS and 601-7-23-RSS fragments were isolated from the PCR products by native PAGE purification. 10 µg of purified labelled fragment, the minimum amount required for a single nucleosome reconstitution, could be generated from

approximately 48 PCR reactions. A representative gel showing purified fluorescently labelled 601-7-12-RSS and 601-7-23-RSS fragments is shown in Figure 5.2B.

Unlabelled fragments were required to act as partner RSSs. For RAG binding and cleavage assays, the partner RSS is used at a ten fold excess over the labelled RSS. Therefore, it is not practical to generate the large amount of unlabelled fragment required by PCR. Instead, large scale plasmid purification of pUC19-601-7-12-RSS and pUC19-601-7-23-RSS was performed, followed by digestion with EcoRV to release the fragment from the vector backbone (Figure 5.2A; Section 2.48). PEG precipitation was then used to separate the 601-7-12-RSS and 601-7-23-RSS fragments from the vector backbone, since their small size causes them to remain in the supernatant (Figure 5.2C). Using this approach, approximately 200 µg of unlabelled fragment was generated from 1L of bacterial culture.

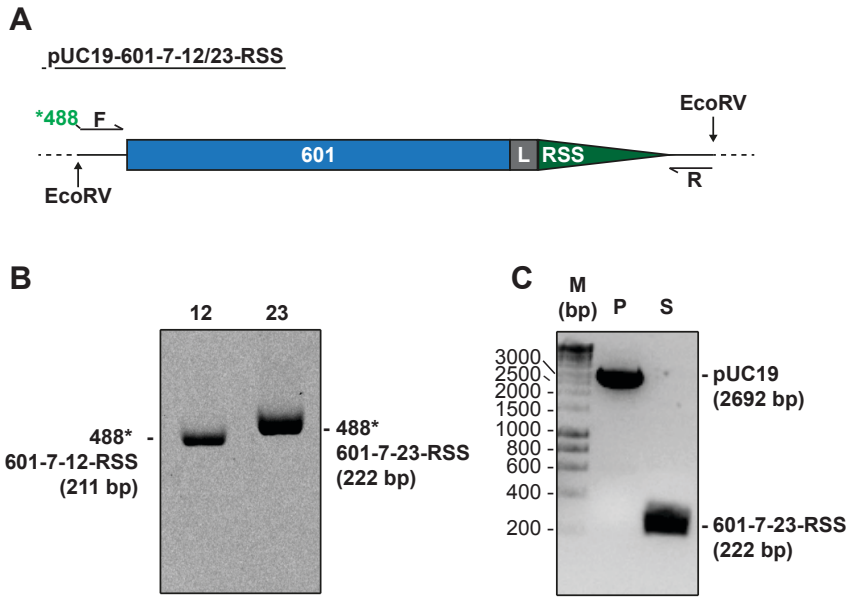


Figure 5.2 - Purification of 601-RSS templates

(A) Schematic showing the 601-7-RSS cassette of pUC19-601-7-12-RSS and pUC19-601-7-23-RSS. The positions of the 601 sequence (blue box) and 12- or 23-RSS (green box) are shown. The grey box (L) represents the 7 bp linker sequence. Primer locations and EcoRV restriction sites are shown. (B) PAGE gel showing purified Alexa Fluor 488 labelled 601-7-12-RSS and 601-7-23-RSS fragments. Imaging was performed using a Fujifilm FLA-5100 gel imager with the 473 nm laser at 1000V. (C) Representative agarose gel of PEG precipitation used to isolate 601-7-23-RSS from the vector backbone after digestion with EcoRV (P, pellet; S, supernatant).

5.2 Refolding the histone octamer from recombinant histone proteins

Histone octamers are required for nucleosome reconstitutions and these were generated by the refolding of purified recombinant *Xenopus laevis* histone proteins (H2A, H2B, H3 and H4) into octamers. Large stocks of all four histone proteins had been previously purified by Dr Joan Boyes (Bevington and Boyes, 2013; Section 2.50). Firstly, each histone is denatured in an unfolding buffer and, after being mixed at equimolar ratios, is refolded by extensive dialysis against a refolding buffer (Section 2.52). This allows the histones to refold and facilitates the formation of histone octamers. However, further purification is required as the preparation often contains a mix of unwanted histone complexes, such as; H2A/H2B dimers, H3/H4 tetramers and high molecular weight aggregates.

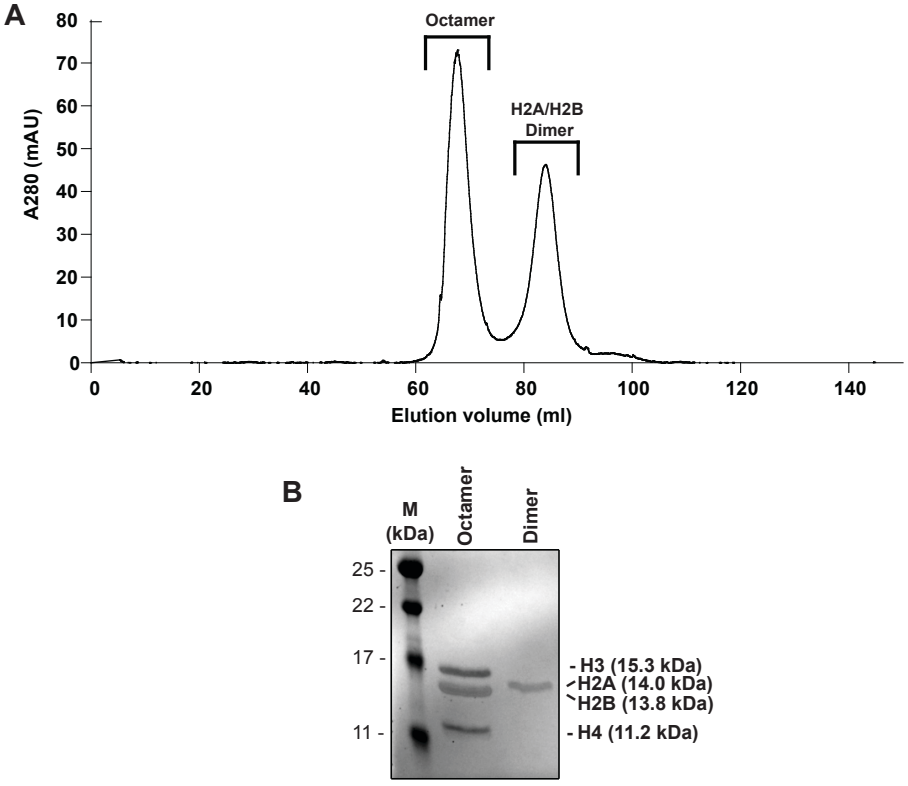


Figure 5.3 - Purification of histone octamers by size exclusion chromatography

(A) Elution profile from size exclusion chromatography of refolding reactions using recombinant *Xenopus laevis* histone H2A, H2B, H3 and H4. Peaks corresponding to histone octamers or H2A/H2B dimers are labelled. (B) Fractions from each elution peak were analysed by SDS-PAGE and stained with coomassie blue. Bands corresponding to each core histone and their approximate molecular weight are labelled on the side of the gel. H2A and H2B bands are not fully resolved due to their very similar molecular weights.

To isolate histone octamers, the refolding reaction was subjected to size exclusion chromatography (SEC) and fractions were collected. Two major elution peaks were observed (Figure 5.3A) and fractions from these peaks were analysed by SDS-PAGE and stained with coomassie blue. This revealed that the first peak contained purified octamers whereas the second contained H2A/H2B dimers (Figure 5.3B). Octamer fractions containing equal amounts of each histone were pooled and used for nucleosome reconstitutions.

5.3 Nucleosome reconstitution

Reconstitution of nucleosomes occurs by the sequential binding of a H3/H4 tetramer and two H2A/H2B dimers to DNA and this can be achieved *in vitro* by salt gradient deposition. This is possible due to the ability of H3/H4 tetramers to bind DNA at higher ionic strengths than H2A/H2B dimers (Thomas and Butler, 1977).

Typically, reconstitution is performed by salt gradient dialysis where a mix of octamers and template DNA is dialysed against buffers of decreasing salt concentration (Dyer *et al.*, 2004; Section 2.54). To achieve the highest efficiency of reconstitution, the octamer:DNA ratio must be optimised. This can be performed using a microscale reconstitution protocol, which involves step dilution of the salt concentration rather than dialysis (Dyer *et al.*, 2004; Section 2.53). Microscale reconstitution is much quicker and uses less material than salt gradient dialysis, allowing efficient optimisation of the octamer:DNA ratio before a larger scale reconstitution is performed.

Microscale reconstitution was performed using 500 ng of purified 601-7-12-RSS or 601-7-23-RSS template DNA and increasing amounts of histone octamer. A representative gel from the microscale reconstitution of Alexa Fluor 488 labelled 601-7-12-RSS is shown in Figure 5.4A. Reconstitutions which contain lower amounts of octamer have high amounts of free DNA, whereas those with higher amounts contain aggregates. The ratio which resulted in the highest amount of reconstituted

nucleosomes, but little aggregation, was used for larger scale salt gradient dialysis reconstitution (Figure 5.4A, lane 9).

Although the reconstitution efficiency was high, further purification was required to remove unwanted free DNA and hexasomes, that contain only one H2A/H2B dimer. Nucleosomes readily dissociate at dilute concentrations and therefore labelled and unlabelled nucleosomes had to be purified differently due to differences in the amount of starting material. Unlabelled nucleosome reconstitutions were performed using high amounts of template DNA (100 µg), allowing nucleosomes to be purified by sucrose density gradient without causing excessive dilution (Section 2.55). Fractions from the gradient were collected and analysed by PAGE and ethidium bromide staining. A representative gel from the purification of unlabelled 601-7-23-RSS nucleosomes is shown in Figure 5.4B. The fractions containing just nucleosomes were pooled (Figure 4.5B, fractions 10 and 11).

Fluorescently labelled nucleosome reconstitutions were performed using a lower amount of template DNA (10 µg) and since purification by sucrose density gradient led to excessive dissociation, a gel extraction method was utilised (Section 2.55). Fluorescent nucleosomes can be identified directly on the gel without staining and then eluted from the gel in a small volume to keep the nucleosome concentration high. A representative gel showing nucleosomes, hexasomes and unincorporated DNA purified by gel extraction of an Alexa Fluor 488 labelled 601-7-12-RSS nucleosome reconstitution is shown in Figure 5.4C. A small amount of contaminating hexasomes and unincorporated DNA is present in the purified nucleosome preparation but these contaminating bands are less visible during RAG cutting and binding assays due to the lower amounts of nucleosome used during these reactions.

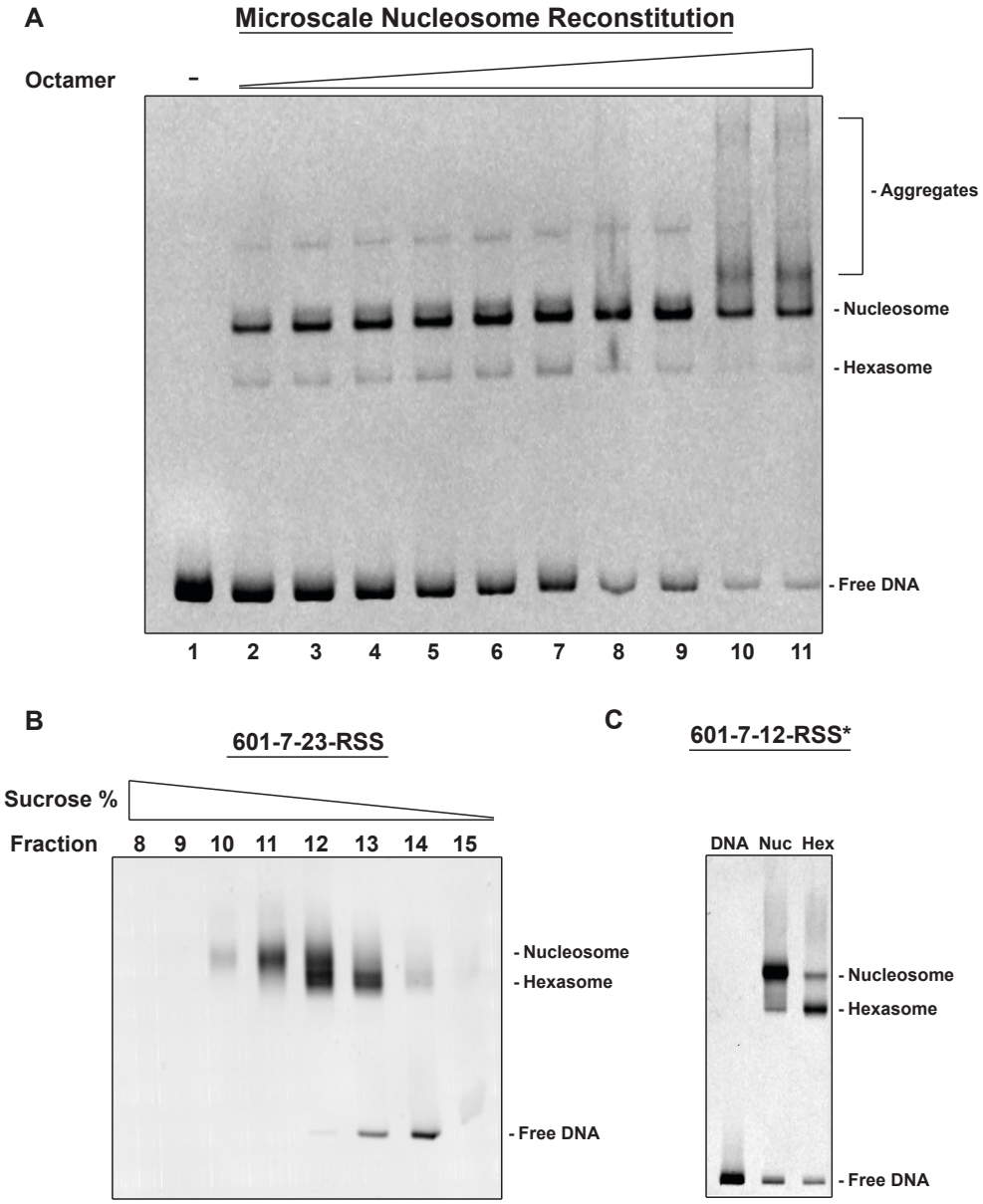


Figure 5.4 - Nucleosome reconstitution and purification

(A) Representative gel of microscale reconstitution of Alexa Fluor 488 labelled 601-7-12-RSS nucleosomes using different DNA:Octamer ratios (1:0, 1:0.7, 1:0.8, 1:0.9, 1:1, 1:1.1, 1:1.2, 1:1.3, 1:1.14, 1:1.15 w/w). (B) Sucrose density gradient purification of unlabelled 601-7-23-RSS nucleosomes. A large scale nucleosome reconstitution containing 100 μg of DNA was separated on a 4 ml 10-40% sucrose gradient and 200 μl fractions were collected and analysed by PAGE. Fractions 10 and 11 were combined to give the nucleosome samples whereas fraction 13 was used for hexasomes. (C) A small scale nucleosome reconstitution containing 10 μg of labelled DNA was separated by native PAGE; nucleosomes, hexasomes and unincorporated DNA were eluted from the corresponding gel slices. The PAGE gel shows complexes purified from a gel extraction of an Alexa Fluor 488 labelled 601-7-12-RSS nucleosome reconstitution.

5.4 Nucleosomes are uniformly positioned at the 601 sequence

Next, it was vital to test if the nucleosome had been uniformly positioned on the 601 sequence and was not covering the RSS. The 147 bp of DNA that is bound by the histone octamer to form the nucleosome is protected from digestion by micrococcal nuclease (MNase). Following digestion, the protected fragment can be isolated and the nucleosome position mapped using restriction enzymes (Figure 5.5A). MNase digestion of a purified 601-7-12-RSS nucleosome generated a band with the expected size of 147 bp (Figure 5.5B). Digestion of this protected fragment with AccI led to the formation of two bands with approximate sizes of 38 bp and 109 bp (Figure 5.5B). The generated bands matched the sizes expected by digestion with AccI (Figure 5.5A), and confirms that the nucleosome is uniformly positioned on the 601 sequence. The same digestion pattern was also observed with 601-7-23-RSS nucleosomes (data not shown).

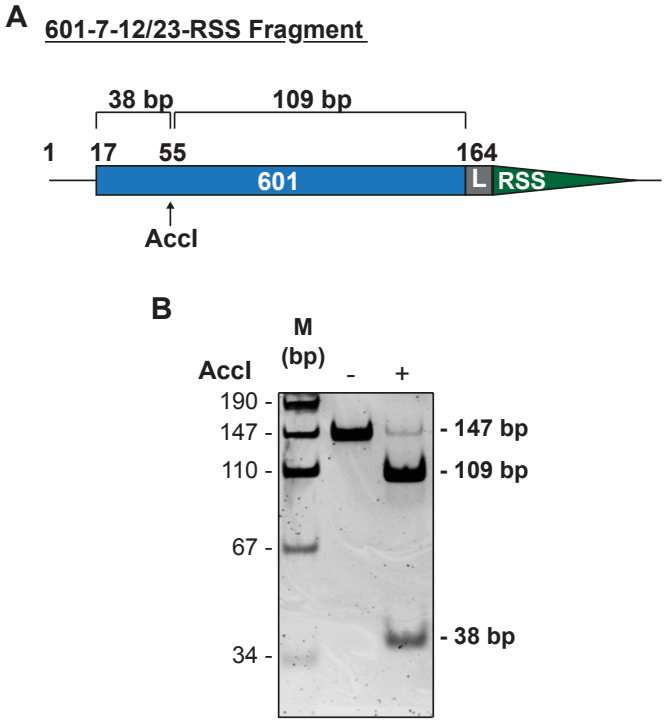


Figure 5.5 - Mapping of nucleosome position

(A) Schematic of the 601-7-12/23-RSS fragment. The AccI cleavage site is denoted by an arrow and the sizes of the resulting fragments are shown. (B) Nucleosomes are positioned at the 601 sequence. Agarose gel of a 147 bp generated from MNase digestion of 601-7-12-RSS nucleosomes, with or without digestion with AccI. Sizes of bands are denoted next to the gel.

5.5 The RAG2 C-terminus inhibits RAG activity on nucleosome substrates in the absence of a H3K4me3 peptide

Next, the binding and cleavage activity of RAGs with nucleosome substrates was analysed. Cleavage reactions were performed with core (cRAG2) or full-length RAG2 (flRAG2), together with core RAG1, using labelled 601-7-12-RSS nucleosomes and a ten-fold excess of unlabelled 601-7-23-RSS nucleosomes to provide a partner RSS. As these nucleosomes are produced from recombinant histones, they do not contain any post-translational modifications; therefore, a H3K4me3 peptide was included, where indicated, to activate flRAG2. Cleavage of 601-7-12-RSS nucleosomes by RAGs is readily observed, but at a lower efficiency than with free DNA substrates (Figure 5.6). This confirms that the nucleosome is positioned at a sufficient distance away from the RSS to allow cleavage. However, only low levels of RAG complexes are visible with the nucleosome substrates (Figure 5.6, lanes 2-5), whereas RAG complexes with free DNA are readily observed (Figure 5.6, lanes 7-10). This suggests that although RAGs can cleave the nucleosome substrates, the proximity of the nucleosome to the RSS may destabilise binding.

To determine if the nucleosome coding ends are held in the post-cleavage complex (PCC) by RAGs, the cleavage reactions were analysed without deproteinisation on native gels. Since the H3K4me3 peptide was supplied in *trans*, coding ends retained in complexes formed with these substrates will not be due to the PHD finger binding to H3K4mes on the nucleosome. Instead, this would be due to a direct interaction between RAG and the nucleosome, such as the residues of the RAG2 acidic hinge which have been shown to interact with core histones (West *et al.* 2005; Figure 5.1A). However, due to the RAG:nucleosome complexes being present at very low levels, coding ends retained in these complexes could not be analysed. Higher levels of RAG:nucleosome complexes were achieved in later experiments by adjusting the electrophoresis conditions and the coding ends retained within these were successfully analysed (Section 5.7 and 5.8).

These experiments showed a surprising effect in cleavage reactions containing flRAG2. It is well established that binding of the RAG2 PHD to H3K4me3 relieves the autoinhibition of cleavage and binding activity imposed on RAG1 (Section 1.8 iv). However, cleavage of a DNA template can still occur, albeit at a lower efficiency, in the absence of H3K4me3 (Lu *et al.*, 2015; Bettridge *et al.*, 2017). Under the conditions used here, reactions containing full-length RAG2 are able to cleave DNA substrates in the absence or presence of H3K4me3 (Figure 5.6, lanes 9 and 10), whereas cleavage of nucleosome substrates is markedly reduced in the absence of H3K4me3. This suggests a novel role for the RAG2 C-terminus in inhibiting activity on nucleosome substrates in the absence of at least one RAG2 PHD finger binding to H3K4me3.

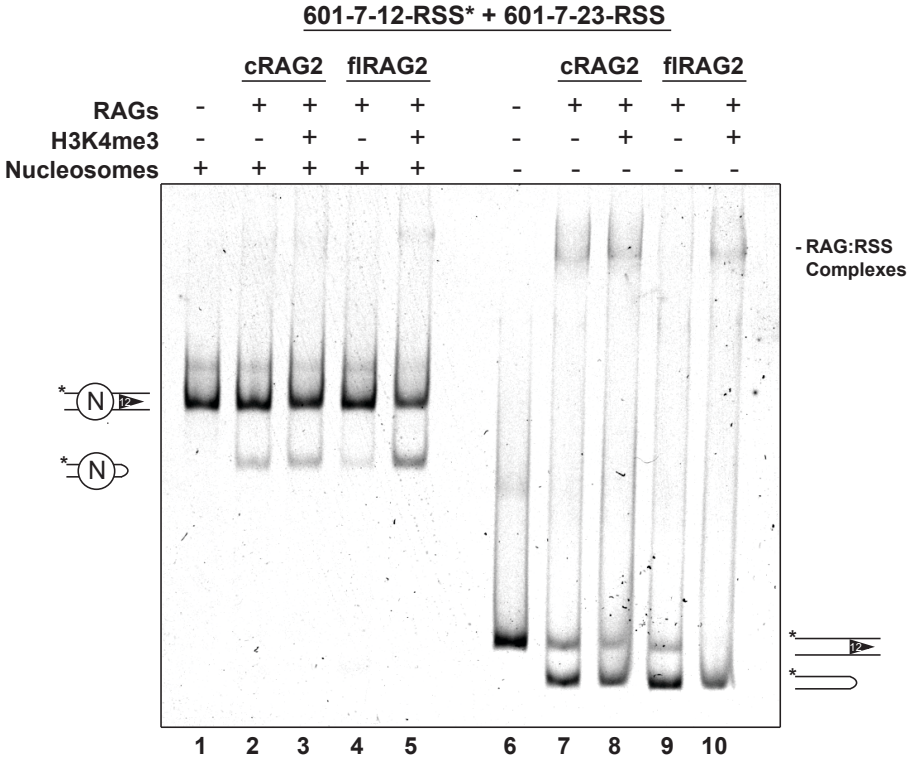


Figure 5.6 - Analysis of RAG binding and cleavage of nucleosome or free DNA substrates

Purified cRAG or flRAG2, together with cRAG1, was incubated with fluorescently labelled 601-7-12-RSS nucleosome (lanes 1-5) or free DNA substrates (lanes 6-10) in the presence of magnesium. Unlabelled partner 601-7-23-RSS nucleosome (lanes 1-5) or free DNA substrates (lanes 6-10) were present in a ten-fold excess over 601-7-12-RSS. H3K4me3 was supplied, where indicated, as a peptide which contains the first 21 amino acids of histone H3 and trimethylated lysine 4. Complexes were separated on a 6% polyacrylamide gel (37.5:1 acrylamide:bis, 0.2 X TBE) at 170V for 2 hours.

5.6 Generation of chromatin substrates containing a H3K4me3 analogue

The nucleosomes used in the above experiments were unmodified. To analyse RAG activity on chromatin substrates that better match those found *in vivo*, nucleosomes which bear the H3K4me3 modification are required. Such substrates would enable the hypothesis to be tested that binding of the RAG2 PHD finger to H3K4me3 allows RAGs to retain the coding ends after cleavage.

Recombinant histones purified from *E. coli* do not contain post translational modifications. Therefore, a methyl-lysine analogue (MLA) was introduced into histone H3 at position 4. This is achieved by *in vitro* chemical alkylation of cysteine residues, which, depending on the alkylating agent used, can introduce mono-, di-, or tri-MLAs into proteins (Simon *et al.*, 2007; Figure 5.6A). Fortuitously, histone H3 only contains a single cysteine residue at position 110 whose mutation to alanine is well tolerated, allowing the site specific introduction of a MLA at any position in histone H3. A lysine to cysteine mutation was introduced into histone H3 cDNA at position 4, along with the C110A mutation (H3K4$_C$). This modified protein was expressed in and purified from *E. coli* via size exclusion chromatography followed by ion exchange chromatography (Figure 5.7B, lane 1; Section 2.50).

Next, to introduce the MLA, the lyophilised mutated histone H3K4$_C$ was resuspended in a buffer which contained the alkylation agent (Section 2.51). Following alkylation the protein was analysed by SDS-PAGE using an 18% gel. This revealed that alkylated histone H3K4$_C$ (H3K4$_C$me3) has a slightly slower mobility than proteins that have not undergone alkylation (Figure 5.7B). Additionally, the presence of a single band suggests highly efficient alkylation has occurred. Indeed, this method has been optimised to allow greater than 90% of cystine residues to be converted to the desired MLA (Simon *et al.,* 2007).

MLAs are very similar to their natural counterparts, except for the lysine γ-methylene which is replaced with a sulphide (Figure 5.7A). The similarity enables their recognition by antibodies raised against the corresponding natural modification (Simon et al., 2007). Western blotting using an anti-H3K4me3 antibody reveals a band of the correct size that is present only in preparations which have undergone alkylation, confirming the successful generation of H3K4$_C$me3 (Figure 5.7C). Next, H3K4$_C$me3 was used to refold the histone octamer as described previously (Section 5.2), in place of standard histone H3. The introduction of the MLA does not appear to affect refolding as the octamers produced are consistent with those generated by the standard preparation (compare Figure 5.7D to Figure 5.3B).

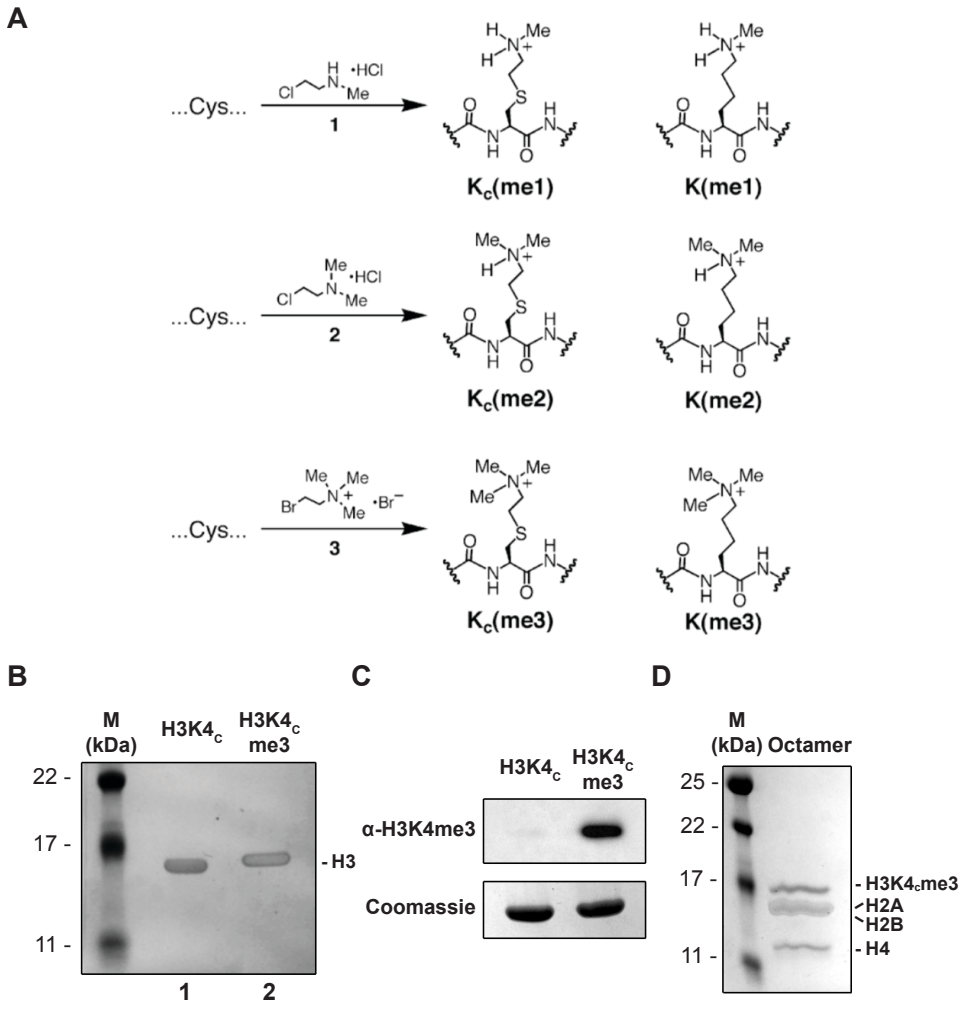


Figure 5.7 - Generation of a H3K4me3 analogue

(A) Diagram showing the conversion of cysteine residues into analogues of mono-, di-, or tri-methylated lysine via treatment with different alkylating agents (1, 2-chloroethyl-methylammonium chloride; 2, 2-chloroethyl-dimethylammonium chloride; 3, 2-bromoethyl trimethylammonium bromide). Taken from Simon *et al.*, 2007. (B) Alkylation of purified H3K4$_c$. 18% SDS-PAGE gel stained with coomassie blue showing purified H3K4$_c$ treated with (lane 2) or without (lane 1) alkylating agent. (C) H3K4$_c$me3 is recognised by an anti-H3K4me3 antibody. Western blot (top) was performed using 100 ng of purified protein. A coomassie blue stained gel (bottom) was performed in parallel as a loading control. (D) Refolding of octamers containing H3K4$_c$me3. 18% SDS-PAGE gel stained with coomassie blue, showing that all histones are present in the same relative proportions as seen in Figure 5.3.

5.7 RAG activity is stimulated by chromatin templates containing H3K4$_C$me3

The octamers containing H3K4$_C$me3 were used in nucleosome reconstitutions with Alexa Fluor 488 labelled 601-7-12-RSS and unlabelled 601-7-23-RSS DNA as

described previously (Section 5.3), and RAG activity assays were performed using these substrates.

Although H3K4$_C$me3 was recognised by the anti-H3K4me3 antibody, it was not known if the RAG2 PHD would also interact with the mark in the same way as a genuine H3K4me3. To this end, flRAG2 was incubated with fluorescently labelled 601-7-12-RSS nucleosomes, in the presence of magnesium, to analyse binding and cleavage activity. This revealed that more efficient RAG complex formation and cleavage occurs with those nucleosomes containing H3K4$_C$me3 (Figure 5.8, compare lanes 3 and 6). By contrast, reactions containing cRAG2 are able to bind and cleave nucleosome substrates regardless of whether H3K4$_C$me3 is present (Figure 5.8, compare lanes 2 and 5). Not only does this suggest that the RAG2 PHD finger can indeed bind to H3K4$_C$me3, it is also consistent with the previous observation that efficient cleavage of nucleosome templates by flRAG2 occurs only in the presence of H3K4me3 (Section 5.5). In reactions with methylated nucleosomes, both the labelled 601-7-12-RSS nucleosomes and the unlabelled partner 601-7-23-RSS nucleosomes contained H3K4$_C$me3 (Figure 5.8, lanes 4-6). Therefore, an open question is whether one or both PHD fingers of the RAG heterotetramer must bind to H3K4me3 to overcome the inhibition.

Analysis of RAG activity with fluorescently labelled 601-7-12 hexasomes, with or without H3K4$_C$me3, was also performed. In contrast to nucleosome substrates, flRAG2 is able to support efficient binding and cleavage of hexasome substrates in the absence of H3K4$_C$me3 (Figure 5.8, compare lanes 9 and 12). Together, these data further suggest a novel role for the RAG2 C-terminus in inhibiting binding to, and cleavage of, nucleosome substrates lacking H3K4me3.

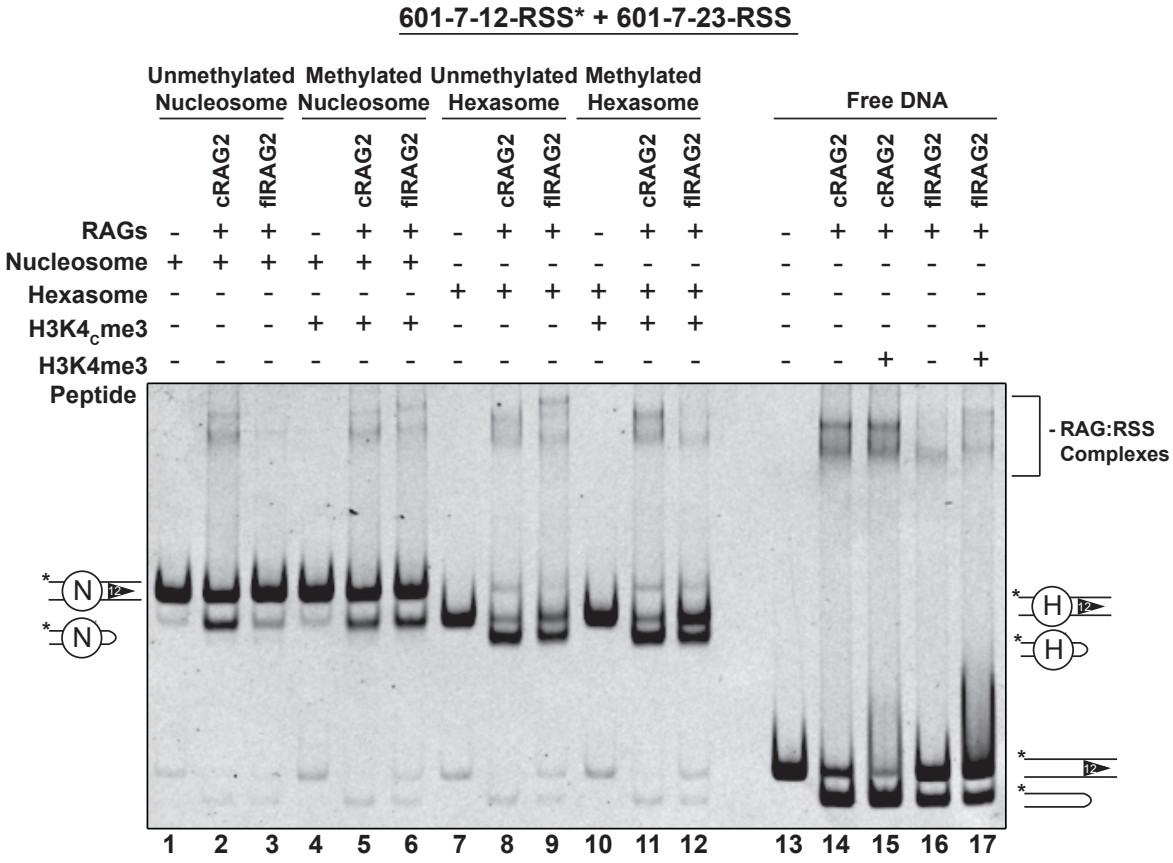


Figure 5.8 - Analysis of RAG binding and cleavage activity with unmethylated or methylated chromatin substrates.

Purified cRAG or flRAG2, together with cRAG1, was incubated with fluorescently labelled 601-7-12-RSS nucleosome (lanes 1-6), hexasome (lanes 7-12) or free DNA substrates (lanes 13-17), in the presence of magnesium. Unlabelled partner 601-7-23-RSS nucleosome (lanes 1-6), hexasome (lanes 7-12) or free DNA substrates (lanes 13-17) were present in a ten-fold excess. H3K4$_C$me3 is present on the nucleosome or hexasome, where indicated, whereas H3K4me3 is supplied as a small peptide for reactions containing free DNA. Milder electrophoresis condition were use than those in Figure 5.6. Complexes were separated by 6% PAGE (59:1 acrylamide:bis, 20 mM HEPES pH 8.0) at 100V for 4 hours.

5.8 Analysis of coding end retention in PCCs formed using nucleosome substrates

The RAG:nucleosome complex appears to be very unstable as evidenced by the low levels of complexes present during experiments performed (Figure 5.6). However,

using milder electrophoresis conditions, higher levels of RAG complex formation were observed (Figure 5.8), and therefore the presence of coding ends within these complexes could be analysed.

This was achieved by the gel extraction of DNA from bands corresponding to RAG complexes formed with either methylated nucleosome substrates or free DNA substrates, from reactions containing either cRAG2 or flRAG2 (Section 2.58). As expected, bands corresponding to 12- and 23-RSS signal ends are present in all samples (Figure 5.9A and B). However, 23-RSS signal ends appear to be much more abundant in complexes formed with cRAG2. These complexes were formed in reactions containing a ten-fold excess of 601-7-23-RSS nucleosome or free DNA substrates, but in the presence of magnesium which should facilitate only coupled cleavage. Therefore the levels of 12- and 23-RSS signal ends should be similar. The increased abundance of 23-RSS signal ends could be due to cRAG2 reactions using a cryptic 12-RSS found elsewhere on 601-7-23-RSS as partner. The unknown bands (Figure 5.9B, bands marked with an asterisk) could be cryptic 12-RSS signal ends.

Surprisingly, DNA extracted from each of the reactions contained a band of approximately 171 bp, which is the expected size for coding ends (Figure 5.9A and B). In complexes formed with free DNA substrates, the number of coding ends present is low, at approximately 15-20% the level of signal ends (Figure 5.9B and C). This is consistent with the idea that coding ends are rapidly released by RAGs after cleavage *in vitro* and explains why they are not present in published high resolution structures. A modest increase in retained coding ends is observed when complexes are formed with nucleosome templates, where around 30-35% of the level of signal ends are retained (Figure 5.9B and C). This may suggest some mild stabilisation of the PCC by the presence of a nucleosome. However, this is independent of H3K4me3 as a similar level of coding end retention is observed with complexes formed using cRAG2 and flRAG2.

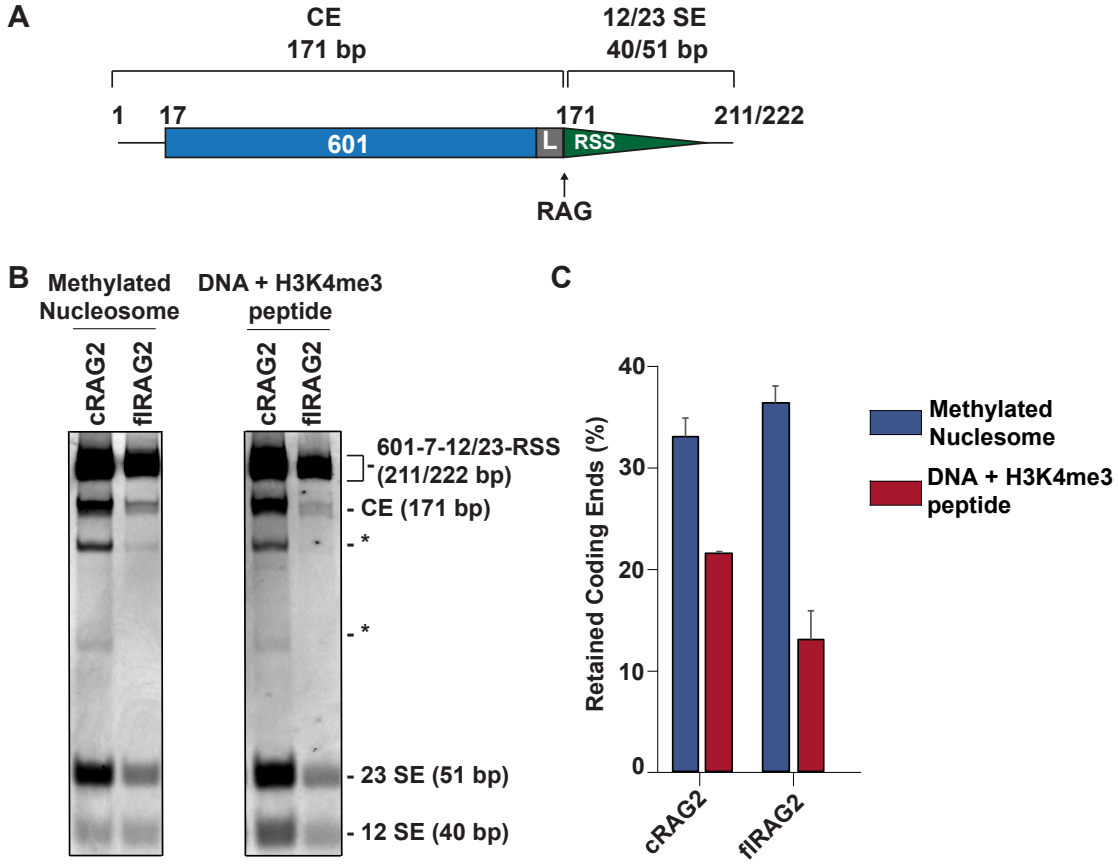


Figure 5.9 - Analysis of coding end retention in RAG complexes formed with nucleosome or DNA substrates.

(A) Diagram of 601-7-RSS substrate. The expected sizes of the coding end (CE) and signal end (12/23 SE) generated by RAG cleavage of 601-7-12-RSS and 601-7-23-RSS are shown. (B) RAG cleavage reactions were performed using nucleosome or free DNA templates as for Figure 5.8. DNA from RAG complexes formed in these reactions was gel extracted and subjected to proteinase K treatment to remove RAG and histone proteins. DNA was separated by PAGE and gels were stained with dsGreen gel staining solution. Imaging was performed by a Fujifilm FLA-5100 gel imager using the 473 nm laser at 1000V. (C) Levels of retained coding ends in RAG PCCs. The levels of coding ends and 12 and 23 signal ends were determined using 2D densitometry via ImageJ. Levels of retained coding ends are shown as a percentage of total signal ends present in each sample (*N* = 2, error bars represent SEM).

C) Discussion

In this chapter, I describe the successful development and generation of RAG substrates which contain a nucleosome harbouring a H3K4me3 analogue positioned in close proximity to an RSS. Initially, I planned to use these substrates to test the hypothesis that the coding ends generated by RAG cleavage are retained in the PCC via an interaction between the RAG proteins and the nucleosome positioned at the start of the coding segment (Figure 5.1). Generation of a stabilised RAG PCC *in vitro* would be a vital tool for the biochemical analysis of the joining step of V(D)J recombination. Unfortunately, the experiments performed using these nucleosome substrates did not demonstrate that the RAG PCC is substantially stabilised by the nucleosome. However, I made the unexpected discovery that the RAG2 C-terminus inhibits RAG binding and cleavage on nucleosome substrates in the absence of H3K4me3. This reflects the potential of these nucleosome substrates as a valuable tool for the biochemical analysis of the RAG proteins with substrates which better resemble those *in vivo*.

5.9 Generation of RAG substrates containing a nucleosome with a H3K4me3 analogue

To analyse if an interaction between a nucleosome positioned on the coding segment and the RAG proteins could aid the retention of coding ends in the PCC, RAG substrates containing nucleosomes were generated. The nonamer sequence of the RSS has been shown to inherently position the nucleosome over the RSS (Baumann *et al.*, 2003) and this would completely inhibit RAG activity (Golding *et al.*, 1999). To overcome this, a Widom 601 sequence was used as a strong nucleosome positioning element to uniformly position recombinant histone octamers 5’ of the heptamer of a 12- or 23-RSS (Figure 5.5). The nucleosome had to be positioned relative to the RSS in a manner which would allow RAG accessibility but also retain any potential interactions with the nucleosome. Initially, the nucleosome was positioned to directly abut the RSS heptamer. However, no RAG binding and cleavage was observed with these substrates

(data not shown). This was likely due to the close proximity of the nucleosome to the RSS inhibiting RAG accessibility. Indeed, introduction of a 7 bp linker DNA sequence between the 601 sequence and the RSS heptamer allowed RAG binding and cleavage of the resulting nucleosome substrates (Figure 5.6; Figure 5.8). However, the complex formed between RAGs and the nucleosome substrates appeared to be very labile as mild electrophoresis conditions were required to detect appreciable amounts of the RAG:nucleosome complex (Figure 5.6; Figure 5.8).

Since recombinant histones purified from *E. coli* do not contain any post-translational modifications, initially H3K4me3 was supplied as a small peptide in *trans* to activate flRAG2 (Figure 5.6). However, this system does not enable the idea that RAG2 PHD finger facilitates coding end retention in the PCC via binding to nucleosomal H3K4me3. To address this, histone H3 containing an analogue of H3K4me3 was generated by the alkylation of an introduced cysteine residue at position four of the protein (Figure 5.7). Histone H3 containing the MLA was able to form histone octamers (Figure 5.7D) and H3K4$_C$me3 was recognised by an antibody which recognises genuine H3K4me3. Further, H3K4$_C$me3 containing nucleosomes were able to stimulate RAG binding and cleavage in reactions containing flRAG2 (Figure 5.8). This suggests that the RAG2 PHD finger can bind to H3K4$_C$me3 in a similar manner to its natural counterpart.

Development of this system set the groundwork for testing the hypothesis of coding end retention in the RAG PCC with nucleosomal substrates (Figure 5.1). Crucially, a stabilised PCC could be used in structural studies to gain insight into the poorly understood joining step of V(D)J recombination. Additionally, this system enables other important questions regarding the RAG proteins and their interaction with chromatin to be answered. For example, it is not currently known if both RAG2 PHD fingers present in the RAG heterotetramer must engage with H3K4me3 to completely overcome autoinhibition. This system could address this as it allows the assembly of RAG complexes with nucleosomal substrates where the presence of H3K4$_C$me3 at the 12- or 23-RSS can be controlled. Unfortunately, time constraints prevented these questions

from being fully explored. The progress I have made and the potential future directions with this system are discussed below and in Chapter 6.

5.10 PCCs formed with nucleosome substrates may retain more coding ends than those formed with DNA substrates

To determine if an interaction between the RAG2 C-terminus and the nucleosome holds the coding ends in the PCC, RAG complexes formed in cleavage reactions, containing cRAG2 or flRAG2 and nucleosome or free DNA substrates, were analysed. This was achieved by gel extraction of bands corresponding to RAG complexes, followed by deproteinisation and analysis of the DNA by PAGE (Figure 5.9).

Unexpectedly, all complexes analysed contained some coding ends (Figure 5.9B and C). This was particularly surprising for complexes formed using free DNA substrates as coding ends are thought to be rapidly lost from the PCC after cleavage *in vitro* and they are missing from published structures (Ru *et al.*, 2015; Kim *et al.*, 2018). Despite this, coding ends retained in PCCs formed with free DNA substrates are still very low, present at approximately 15-20% the level of signal ends (Figure 5.9C). Therefore, such RAG PCCs would not be useful for structural studies due to extensive coding end loss and heterogeneity. However, the coding ends generated by the RAG cleavage of 601-7-12-RSS and 601-7-23-RSS are much larger (171 bp; Figure 5.9A) than coding ends generated by the most commonly used RSSs (16 bp), including those used in published high resolution structures (Ru *et al.*, 2015; Kim *et al.*, 2018). It is possible that the longer coding ends are less mobile and persist longer in the PCC. This phenomenon would be even more pronounced *in vivo* where coding ends would encompass an entire chromosome. This opens the possibility that a stabilised RAG PCC could be reconstituted *in vitro* simply by the use of longer DNA substrates.

A moderate increase in the levels of retained coding ends was observed in PCCs formed using nucleosome substrates (Figure 5.9C), but this was still relatively low and the levels are similar for complexes formed in reactions containing either cRAG2 or

flRAG2 (Figure 5.9C). This is not consistent with the proposed model that the RAG2 C-terminus holds the coding ends in the PCC due to an interaction with the nucleosome. However, the method used to analyse coding end retention in the PCC maybe unsuitable due to complexes dissociating during electrophoresis. Indeed, RAG complex formation was greatly influenced by the electrophoresis conditions used (Section 5.8), with more complex formed when less current is used. An alternative approach using mild fixation to preserve the labile RAG complexes during analysis may be more suitable. Alternatively, a single molecule approach, such as fluorescence anisotropy (Gradinaru *et al.*, 2010), could be used to analyse RAG complexes in solution and would circumvent the need for damaging electrophoresis.

The eventual aim of this project is to generate a stabilised RAG PCC, containing the coding ends, to resolve its structure for this first time. Unfortunately, the current conditions do not provide this stabilised RAG PCC, as proposed in Figure 5.1. Nonetheless, a different approach could be used to generate a stable RAG PCC. One possibility is GraFix, which has been used to successfully stabilise labile macromolecule complexes for use in single particle cryo-electron microscopy (Stark, 2010). During GraFix, complexes are simultaneously fixed and separated by size and conformation via density gradient centrifugation in the presence of a cross-linking agent. This procedure could be used in conjunction with nucleosome or longer DNA substrates if these are found to be important factors in stabilising the RAG PCC. Therefore, GraFix could generate a stable RAG PCC containing coding ends and allow its structure to be solved for the first time.

If a stable RAG PCC can be generated and its structure solved, this could be used as a platform for future work involving NHEJ proteins which would reveal a wealth of knowledge about the vital joining step of V(D)J recombination.

5.11 The RAG2 C-terminus inhibits binding and cleavage of nucleosome substrates in the absence of H3K4me3

During the analysis of RAG activity on nucleosome substrates, a novel role of the RAG2 C-terminus was observed. Although reactions containing flRAG2 can bind and cleave free DNA substrates in the absence of a H3K4me3 peptide, the same was not observed with nucleosome substrates (Figure 5.6). Instead, efficient cleavage and binding of nucleosome substrates is dependent on the presence of the H3K4me3 peptide (Figure 5.6). In contrast, reactions containing cRAG2 are able to bind and cleave nucleosome substrates, independent of H3K4me3 (Figure 5.6). This same pattern of cleavage and binding is observed during experiments using nucleosomes which harbour the H3K4me3 analogue (Figure 5.8).

However, binding and cleavage of substrates containing a hexasome is similar to that of free DNA substrates (Figure 5.8). A hexasome can form by the transient eviction of a H2A/H2B dimer from the nucleosome during transcription. This releases 35-40 bp of nucleosomal DNA which has been shown to facilitate RAG accessibility to an RSS which would otherwise be blocked by the nucleosome (Bevington and Boyes, 2013). Therefore, in the hexasome substrates, it is possible that up to 35-40 bp of additional accessible DNA exists between the RSS and the hexasome. This distance may stop any inhibitory interaction between the RAG2 C-terminus and the hexasome, allowing binding and cleavage in the absence of H3K4me3, similar to with free DNA substrates. However, to confirm this, the position of the hexasome needs to be mapped.

Together, these data suggest a new mechanism whereby an interaction between the RAG2 C-terminus and the histone octamer inhibits RAG binding and cleavage which can be overcome by the PHD finger binding to H3K4me3. Although the purpose of this mechanism is not understood, it may exist as an additional regulatory mechanism to inhibit RAG activity at cryptic RSSs in the genome where the correct epigenetic modification may not be present. Alternatively, this mechanism may be involved in the temporal control of V(D)J recombination by inhibiting activity on RSSs which lack

H3K4me3. Indeed, the RAG2 C-terminus has been demonstrated to be involved in the ordered regulation of V(D)J recombination and accessibility to the *Igh* locus (Curry and Schlissel, 2008; Ward *et al.*, 2018).

Since this mechanism would not have been discovered without the use of nucleosomal RSS substrates, this finding emphasises the need for future study of the RAG proteins in a chromatin environment.

Chapter 6 - General Discussion and Future Plans

In this thesis I investigated the RAG proteins in a number of different ways; using naturally occurring or artificial mutations, and novel *in vitro* substrates. These approaches led to three new discoveries, enhancing our understanding of RAG protein biochemistry.

6.1 A novel naturally occurring RAG1 mutation reveals the critical role of HMGB1/2 during V(D)J recombination *in vivo*

In Chapter 3, I analysed two naturally occurring mutations that were identified in an immunodeficient patient. This led to the discovery of a novel *RAG1* mutation (R401W) that causes a deficiency in binding to the accessory protein HMGB1 (Figure 3.9), which is accompanied by a severe decrease in cleavage activity *in vitro* (Figure 3.8) and recombination activity *in vivo* (Figure 3.5). Mutation of R401 to leucine or lysine revealed that efficient HMGB1 binding only occurs with the latter mutation, and correlates with increased cleavage *in vitro* and recombination *in vivo* (Figure 3.13). Therefore, the positive charge of R401 appears to be important to allow efficient binding by HMGB1. However, the bulky nature of the tryptophan introduced with R401W could also cause steric clashes and hinder HMGB1 binding. Since HMGB1 Box A has a preference to bind pre-bent DNA and is found within the RAG complex at a 60° bend in the RSS, close to where R401 is located (Figure 3.10), this suggests that the positive charge of R401 maybe involved in the formation of this bend and allow HMGB1 Box A recruitment.

Given that this mutation occurs in an immunodeficient patient, this suggested a critical role for HMGB1 during V(D)J recombination *in vivo*. Although HMGB1 has been known to stimulate RAG binding and cleavage *in vitro* for some time (van Gent *et al.*, 1997), there were conflicting ideas about its role *in vivo* (Callogero *et al.*, 1999; Aidinis *et al.*, 1999). However, the observation presented in Chapter 3 that recombination is greatly reduced in HMGB1 deficient cells suggests that HMGB1 is indeed vital for V(D)J

recombination *in vivo* (Figure 3.16). Furthermore, exogenous HMGB1 or HMGB2 is capable of rescuing recombination in HMGB1 deficient cells (Figure 3.20). Together, these data provide strong evidence that either of the DNA bending proteins, HMGB1 or HMGB2, are required for V(D)J recombination *in vivo*.

Previous studies of RAG mutations provided insights into the function of the proteins themselves. Instead, the research presented in Chapter 3 demonstrates a critical role for an accessory protein during V(D)J recombination. The RAG complex, like the ancestral transposons it is related to, usurps cellular factors during V(D)J recombination (Meek *et al.*, 2019). Here, RAGs co-opt HMGB1 for efficient cleavage and binding. Given that the RAG complex also uses the NHEJ machinery to facilitate the final joining step, it is possible that analysis of other RAG mutations might shed light on how RAGs interact with these other accessory proteins.

6.2 Using artificially generated RAG2 mutations to identify a region important to suppress aberrant V(D)J recombination

In the pursuit of generating an effective adaptive immune system, V(D)J recombination must break and rejoin the genome several times to generate a functional antigen receptor variable exon. Since up to a million antibody genes are generated each day in humans, the chances for error are very high. Although mistakes do occur, and this can lead to cancer (Papaemmanuil *et al.*, 2014; Roth, 2014), these are rare since the RAG proteins contain built in regulatory mechanisms to suppress aberrant V(D)J recombination reactions. It is therefore important to understand how these mechanisms work.

Most notably, the RAG2 C-terminus plays a vital role in protecting genomic integrity (Deriano *et al.*, 2011; Coussens *et al.*, 2013), and suppresses three aberrant reactions; transposition, reintegration and hybrid joint formation (Sekiguchi *et al.*, 2001; Elkin *et al.*, 2003; Curry *et al.*, 2007). Despite this, the mechanisms by which the RAG2 C-terminus suppresses these reactions is poorly understand, including which specific

regions within the 144 amino acids of the C-terminus are responsible. This is largely due to the fact that core versions of the RAG proteins, which in the case of RAG2 lacks the C-terminus, have been used in most biochemical studies of the RAG proteins, due to being easier to purify but still able to catalyse V(D)J recombination (Schatz and Swanson, 2010).

Naturally occurring disease-causing mutations in the RAG2 C-terminus are rare, especially within the acidic hinge (Notarangelo *et al.*, 2016). Therefore, in Chapter 4, I describe the generation of artificial mutations in the RAG2 C-terminus to identify which region is required to inhibit reintegration. Initially, this led to the discovery that the RAG2 acidic hinge is completely responsible for inhibiting reintegration (Figure 4.3B) and analysis of further mutations pinpointed this to the four residues spanning 383-386 (Figure 4.7B). These residues lie within the region of the acidic hinge which is predicted to have the most order (Figure 4.5), suggesting that they may be more structured than the rest of the hinge which is intrinsically disordered. However, since the RAG2 C-terminus is missing from all published high resolution structures (Kim *et al.*, 2015; Ru *et al.*, 2015; Kim *et al.*, 2018; Ru *et al.*, 2018), it is not possible to confirm this.

Mutation of residues 383-386 also leads to increased recombination (Figure 4.8) which suggests this is an inhibitory mechanism of the acidic hinge. Furthermore, the inability of these residues to rescue recombination of a PHD finger mutant suggests that this mechanism may be distinct from the previously mapped autoinhibitory domain (Figure 4.9; Lu *et al.*, 2015). Given that RAG2$^{383\text{-}386A}$ has *in vitro* binding and cleavage activity that is similar to wild type RAG2 (Figure 4.12), this suggests that the inhibitory effects of residues 383-386 likely act within the joining step of V(D)J recombination.

Hybrid joint formation is also increased when residues 383-386 are mutated (Figure 4.13), and since reintegration also involves the formation of a hybrid joint it is possible that residues 383-386 suppress joining in a manner which avoids aberrant blunt-to-

hairpin joining. Thus, these studies have potentially identified a region that appears crucial to regulate correct end joining.

6.3 Towards a structure of the RAG2 C-terminus and post-cleavage complex

As highlighted in Chapter 4, the RAG2 C-terminus is important in the joining step of V(D)J recombination, yet little is known about this stage of the reaction. Understanding the structure of the RAG2 C-terminus and the post-cleavage complex could unlock a wealth of knowledge regarding this phase of the reactions. However, all published high resolution RAG structures lack both the structure of the RAG2 C-terminus and the PCC. Therefore, in Chapter 5 I set out to develop a system capable of generating a stable RAG PCC *in vitro* for use in structural studies.

Although the structure of the RAG2 PHD finger has been solved in isolation (Matthews *et al.*, 2007), due to excessive conformational heterogeneity caused by the dynamic nature of the rest of the C-terminus, the structure of the RAG2 PHD finger is not resolved in the context of the full-length protein. The RAG2 acidic hinge is an intrinsically disordered domain (IDD) and therefore it is entirely possible that it forms no structure. However, other IDDs are known to fold upon interaction with their target ligand (Mollica *et al.*, 2016), such as histone chaperones (Warren and Shechter, 2017), and since residues within the RAG2 acidic hinge have been demonstrated to interact with the core histones (West *et al.*, 2005), is it possible that the acidic hinge may form a stable structure when they interact with a nucleosome. Additionally, if the RAG2 PHD finger is bound to H3K4me3 found on this same nucleosome, this could limit the movement of the PHD finger in 3D space. Therefore, by assembling the RAG complex onto an RSS with a nucleosome, which better resembles *in vivo* substrates, it is possible that a structure of the RAG2 C-terminus could be solved for the first time.

The structure of the RAG post-cleavage complex also remains elusive, since the coding ends are readily lost from the complex after cleavage *in vitro* (Ru *et al.*, 2015). Solving this structure is equally as important as the structure of the RAG2 C-terminus

since it would improve understanding of the fate of the ends after cleavage and how they become correctly joined by the NHEJ machinery. Since the nucleosome is thought to be positioned on the coding end in pro-B cells (Pulivarthy *et al.*, 2016) and that the coding ends are unlikely to be readily released from the PCC *in vivo*, I hypothesised that the interactions I outlined above between the RAG2 C-terminus and the nucleosome may also function to retain the coding ends in the PCC before joining (Figure 5.1).

The established *in vitro* RAG binding and cleavage assays use naked DNA substrates for RSSs, therefore I had to generate substrates which contained a nucleosome. To this end, a Widom 601 sequence was positioned 5' of the RSSs heptamer (Figure 5.2) and this enabled uniform positioning of a purified histone octamer at this location during nucleosome reconstitution (Figure 5.5). A seven base pair spacer had to be placed between the end of the RSS and start of the 601 to allow RAG cleavage, most likely because the nucleosome blocks RAG access to the RSS (Golding *et al.*, 1999).

Recombinant histones purified from *E.coli* were used to produce histone octamers for nucleosome reconstitution, meaning they lack any post-translational modifications. This meant that any nucleosome would not contain the H3K4me3 epigenetic modification, which is required for the RAG2 PHD finger to bind (Mathews *et al.*, 2007). To address this, I used a chemical alkylation approach to modify a cysteine residues at position four of histone H3, which leads to the formation of a methyl lysine analogue, H3K4$_C$me3, that mimics H3K4me3 (Figure 5.7). Nucleosomes containing this modification are able to stimulate RAG binding and cleavage (Figure 5.8), however the RAG PCCs formed were not stabilised since the majority of coding ends were released (Figure 5.9).

6.4 Identification of a role of the RAG2 C-terminus in inhibiting binding to and cleavage of RSSs positioned close to a nucleosome

Although the aim of Chapter 5 was to produce a stabilised RAG PCC containing the coding ends for use in structural studies, a role for the RAG2 C-terminus in inhibiting RAG activity on nucleosome substrates lacking H3K4me3 was identified. Reactions containing cRAG2 are able to bind and cleave nucleosome substrates regardless of the presence of H3K4me3, whereas those with flRAG2 rely on the presence of H3K4me3 for efficient activity (Figure 5.6 and Figure 5.8). In contrast, under these conditions reactions, the complex formed with flRAG2 is able to efficiently bind and cleave naked DNA substrates in the absence or presence of H3K4me3, with only a small increase in activity when the latter is added (Figure 5.6 and Figure 5.8). This is due to overcoming the autoinhibitory domain once the PHD finger binds H3K4me3 (Lu *et al.*, 2015; Bettridge *et al.*, 2017).

In the studies by Lu *et al.* (2015) and Bettridge *et al.* (2017) using naked DNA templates, some RAG binding and cleavage is also observed in the absence of H3K4me3, but they observe a greater stimulation in RAG activity when H3K4me3 is added compared with reactions shown in Chapter 5 (Figure 5.6 and Figure 5.8). However, reactions in these previous studies were allowed to proceed for a much shorter time than those in Chapter 5 (Lu *et al.*, 2015; Bettridge *et al.*, 2017; Section 2.57), suggesting that although the reactions kinetics are slower in the absence of H3K4me3, with extended times a similar level of binding and cleavage of naked DNA substrates can occur regardless of whether the autoinhibitory domain is relieved. Since RAG activity with nucleosome substrates is greatly suppressed in the absence of H3K4me3 (Figure 5.6 and Figure 5.8), despite the long incubation times, this suggests a role for the RAG2 C-terminus in inhibiting RAG activity, possibly by enhancing the effects of the autoinhibitory domain.

Interestingly, it has been demonstrated in a recent study that disruption of the autoinhibitory domain allows association of RAG with the *IgH* locus in the absence of

H3K4me3 binding, which is normally a requirement (Ward *et al.*, 2018). The authors suggest that the autoinhibitory domain acts as a gate to prevent RAG accessibility to the *IgH* locus unless the RAG2 PHD finger binds to H3K4me3. The effect of the RAG2 C-terminus inhibiting activity on nucleosome substrates which lack H3K4me3 observed in Chapter 4 could indeed be due to this function of the autoinhibitory observed by Ward *et al*. (2018).

This unexpected finding highlights the importance of studying the RAG proteins using nucleosome substrates. *In vivo*, the RAG proteins function in a chromatin environment and many aspects of the RAG proteins are linked to chromatin (Matthews *et al.*, 2007; West *et al.*, 2005; Lu *et al.*, 2015; Leavy *et al.*, 2010). The system I have developed here could be used to increase our understanding of the RAG proteins and their interactions with chromatin. For example, the non-core regions of RAG1 are a largely unexplored part of the protein but is known to have a role in the ubiquitination of histone H3 (Deng *et al*, 2015). This role could be studied at greater depth using the nucleosome substrates developed in Chapter 5, including the site specific introduction of ubiquitinated lysine, rather than methylated lysine, analogs in histone H3 using a similar chemical modification (Qi *et al.*, 2018).

6.5 Future work

i) Chapter 4

The data in Chapter 4 clearly demonstrate a role for residues 383-386 of the RAG2 C-terminus in inhibiting reintegration and hybrid joint formation. These inhibitory effects appear to function during the joining phase of V(D)J recombination, rather than the preceding binding and cleavage steps. However, it is unclear how residues 383-386 suppress the joining step and therefore any future work should focus on this area.

Residues 383-386 could be involved in any of a number of functions during the joining step; i) maintaining stability of the RAG PCC, ii) controlling accessibility of DNA repair

enzymes to the signal or coding ends, or iii) they are directly involved in the recruitment of the NHEJ proteins.

The involvement of residues 383-386 in maintaining association with the signal ends after RAG cleavage could be determined using an established *in vitro* signal end complex (SEC) stability assay (Coussens *et al.*, 2013). In this assay, purified RAG proteins are incubated with 12- and 23-RSS oligonucleotides followed by thermal challenge at increasing temperatures (37-70°C). The level of signal ends released from the SEC at each temperature is determined by native PAGE. Should residues 383-386 play an important role in maintaining the signal ends within the PCC, we would expect signal ends to be released more readily at lower temperatures with RAG2$^{383\text{-}386A}$ than wild type RAG2. Association with the coding ends is more difficult to analyse *in vitro* since these are rapidly released from the PCC after cleavage (Ru *et al.*, 2015). This could be explored *in vivo* using a Southern blot approach to analyse the levels of coding ends present in cells (Lescale *et al.*, 2016).

If residues 383-386 are involved in governing signal or coding ends accessibility to direct joining by the NHEJ proteins, this could be explored using hydrogen-deuterium exchange mass spectrometry (HDX-MS). HDX-MS measures the exchange of hydrogen with deuterium in proteins, which occurs more readily in solvent exposed areas (Kochert *et al.*, 2018). This method has been successfully used in the analysis of protein-DNA complexes (Graham *et al.*, 2016) and therefore would be ideal for studying the RAG proteins. The solvent exposure of the broken DNA ends in complexes formed with wild type RAG2 or RAG$^{383\text{-}386A}$ could be analysed by HDX-MS to determine if residues 383-386 are involved in controlling accessibility to the ends.

A co-immunoprecipitation approach, such as SPINE (strep-protein interaction experiment; Herzberg *et al.*, 2007), could be used to determine if residues 383-386 are involved in the recruitment of NHEJ. A strep-tag would be fused to the N or C-terminus of RAG2 and expressed, along with RAG1, in cells undergoing recombination. Strep-

Tactin beads would be used to pull down step-tagged RAG2 plus any interacting proteins and mass spectrometry would reveal the profile of proteins present. This could be compared between wild type RAG2 and RAG2$^{383\text{-}386A}$ to determine if residues 383-386 are involved in the recruitment of NHEJ or other DNA repair enzymes.

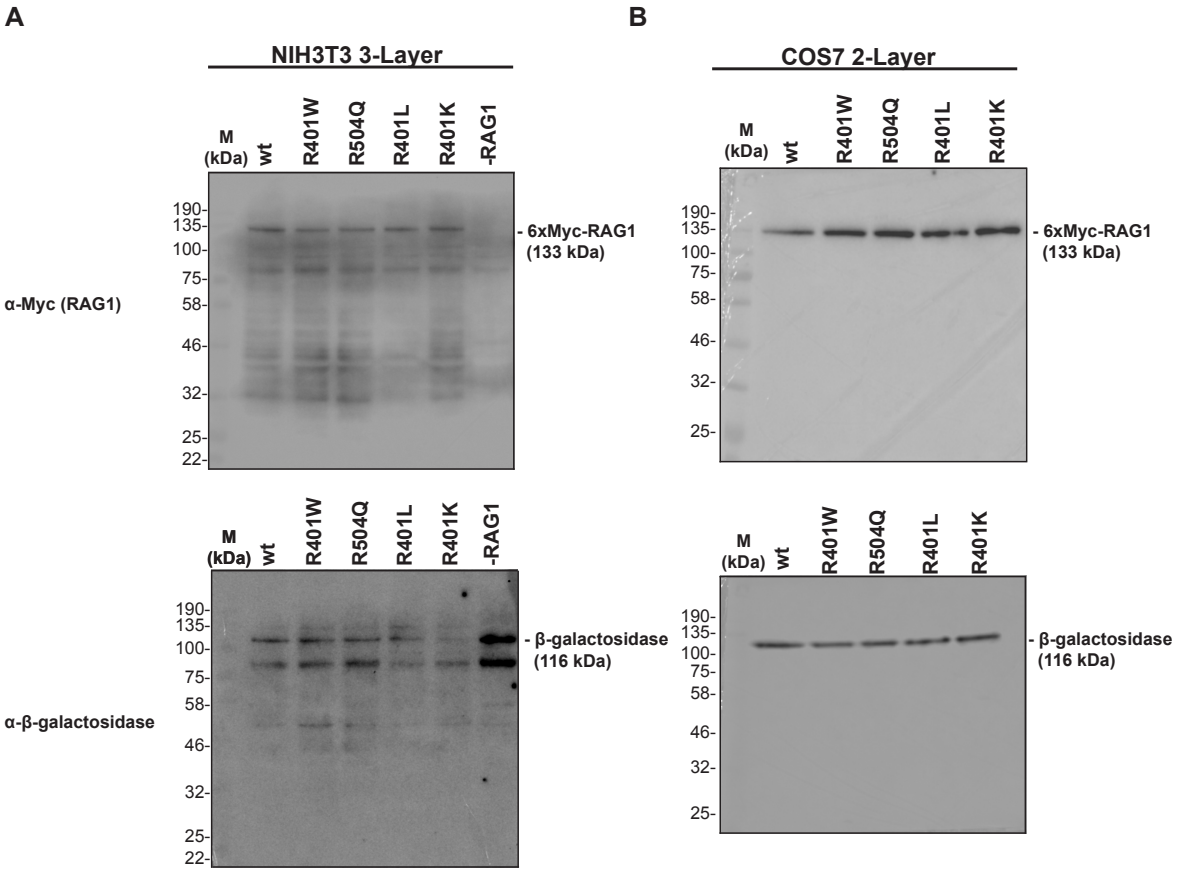
i) Chapter 5

Although a stabilised RAG PCC was not generated in Chapter 5, this is still worth pursuing. Since the RAG PCC is very labile *in vitro*, the use of a crosslinking agent may help to stabilise the complex. GraFix is an increasingly popular technique that is being used to stabilise labile macromolecular complexes for single particle cryo-electron microscopy (Stark, 2010). During GraFix complexes undergo weak crosslinking whilst simultaneously moving through a density gradient by ultracentrifugation. This not only keeps the fragile complex intact but also purifies the complex based on size and conformation. Since the RAG complex is known to undergo many conformational changes during catalysis (Ru *et al.*, 2018b), GraFix would allow different conformations to be captured. If the structure of the RAG PCC could be solved this work could reveal a wealth of knowledge about the elusive repair step of V(D)J recombination and would lay to ground work for future structural studies involving the RAG PCC and its interaction with NHEJ proteins.

A novel role for the RAG2 C-terminus in inhibiting binding and cleavage activity of nucleosome substrates was observed in Chapter 5. This could be explored further by using the RAG2 C-terminus mutants generated in Chapter 4 to identify which residues are involved in this inhibition. In particular, residues Y402, N403, D406 and E407 have been shown previously to directly interact with the core histones (West *et al.*, 2005). However, the role of this interaction is not understood. It is possible that these residues are responsible for contacting the nucleosome and inhibiting binding and cleavage in the absence of the PHD finger binding to H3K4me3. This could be explored using the nucleosome substrates I developed in Chapter 5 in cleavage and binding assays with purified versions of RAG2 containing mutations in the residues identified to contact

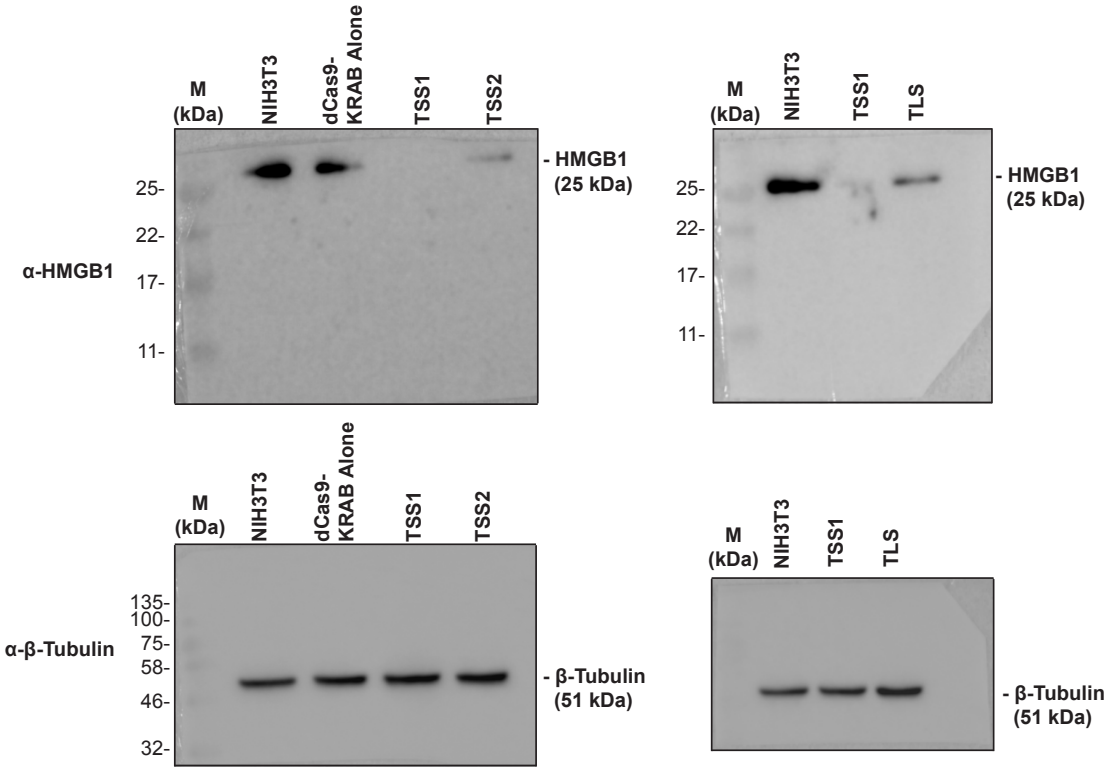
histones. As described above for the analysis of residues 383-386, HDX-MS could also be employed here to analyse interactions between the RAG2 C-terminus and the nucleosome.

Appendix



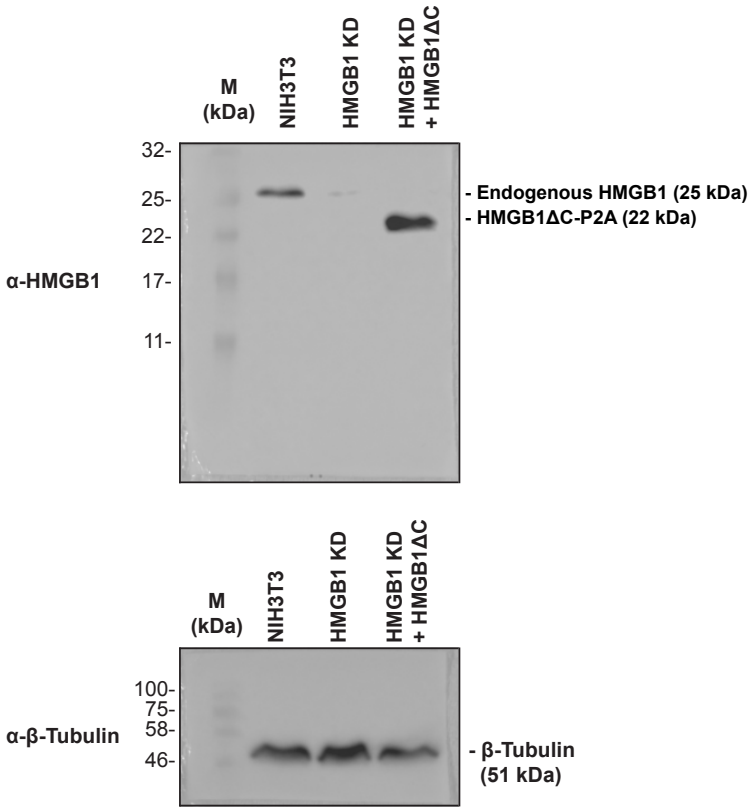
Appendix 1 - Expression of RAG1 in transfected NIH3T3 and COS7 cell lysates

(A) Uncropped anti-Myc (RAG1) and anti-beta-galactosidase western blot images of transfected NIH3T3 cell lysates from Figures 3.6A and 3.13E. (B) Uncropped anti-Myc (RAG1) and anti-beta-galactosidase western blot images of transfected COS7 cell lysates from Figure 3.6B.



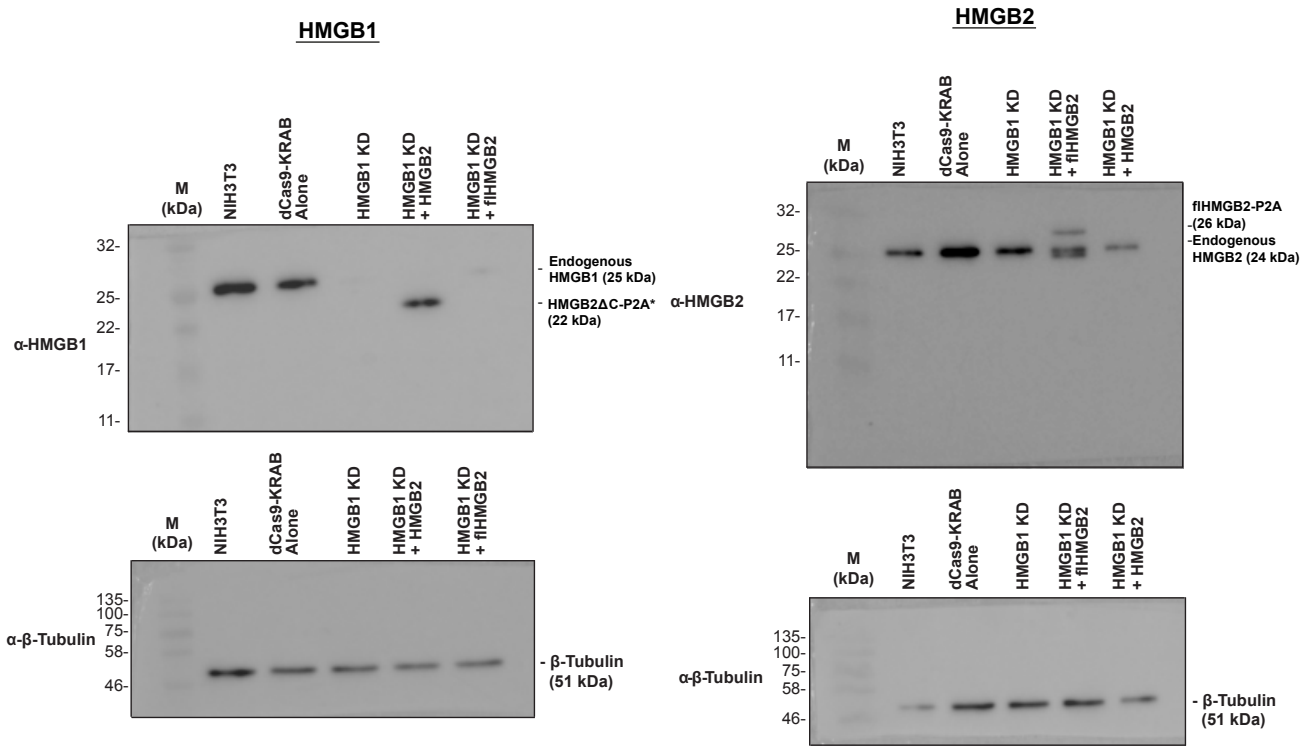
Appendix 2 - HMGB1 knockdown western blots

Uncropped anti-HMGB1 (top panels) and anti-beta-tubulin (bottom panels) western blot western blot images from Figure 3.15D.



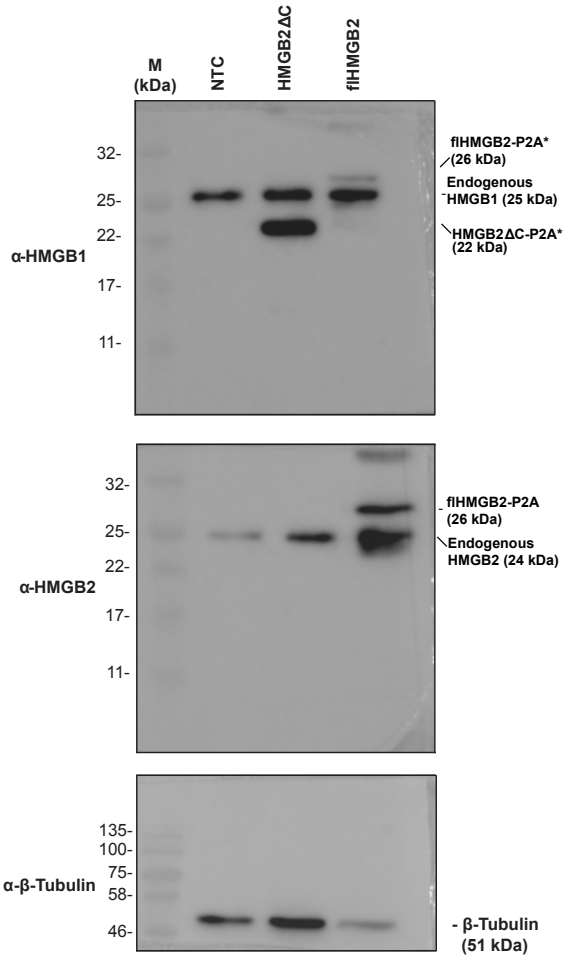
Appendix 3 - Expression of exogenous HMGB1 in HMGB1 KD cells

Uncropped anti-HMGB1 (top panels) and anti-beta-tubulin (bottom panels) western blot western blot images from Figure 3.17B.



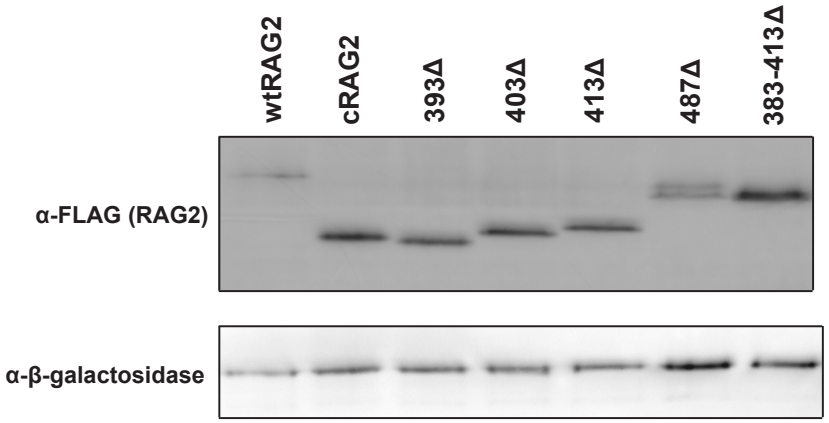
Appendix 4 - Expression of exogenous HMGB1/2 in HMGB1 KD cells

Uncropped anti-HMGB1 (top left) and anti-HMGB2 (top right) western blots images with anti-beta-tubulin (bottom panels) western blot images from Figure 3.18C.



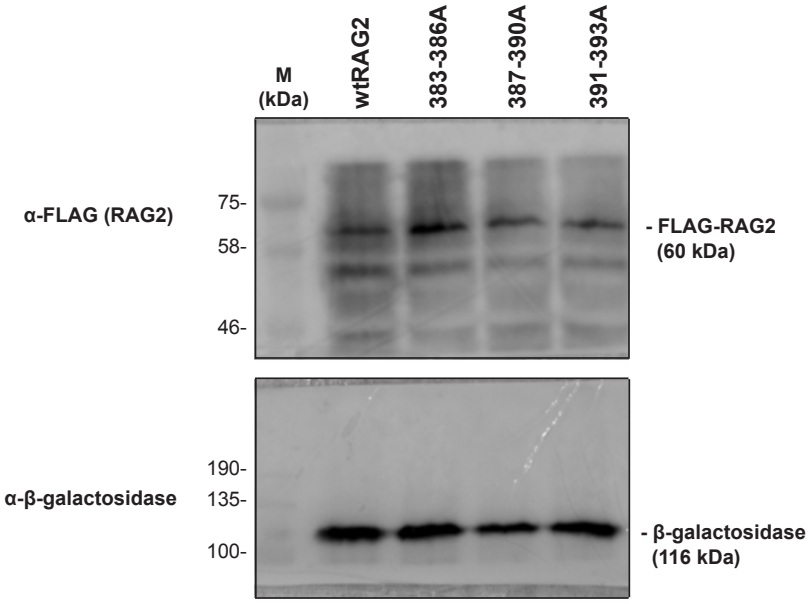
Appendix 5 - Expression of HMGB2 in transfected HEK293T cell lysates

Uncropped anti-HMGB1 (top), anti-HMGB2 (middle), and anti-beta-tubulin (bottom) western blot images from Figure 3.19A.



Appendix 6 - Expression of RAG2 truncation mutants in transfected COS7 cell lysates

Uncropped anti-FLAG (top) and anti-beta-galactosidase (bottom) western blot images from Figures 4.3C and 4.6C.



Appendix 7 - Expression of RAG2 acidic hinge mutants in transfected COS7 cell lysates

Uncropped anti-FLAG (top) and anti-beta-galactosidase (bottom) western blot images from Figure 4.7.

References

Abarrategui, I. and Krangel, M.S. 2009. Germline transcription: a key regulator of accessibility and recombination. *Advances in experimental medicine and biology*. **650**,pp.93–102.

Abarrategui, I. and Krangel, M.S. 2007. Noncoding transcription controls downstream promoters to regulate T−cell receptor α recombination. *The EMBO Journal*. **26**,pp. 4380–4390.

Abarrategui, I. and Krangel, M.S. 2006. Regulation of T cell receptor-α gene recombination by transcription. *Nature Immunology*. **7**,pp.1109–1115.

Adams, J.M., Harris, A.W., Pinkert, C.A., Corcoran, L.M., Alexander, W.S., Cory, S., Palmiter, R.D. and Brinster, R.L. 1985. The c-myc oncogene driven by immunoglobulin enhancers induces lymphoid malignancy in transgenic mice. *Nature*. **318**,pp.533-538.

Agrawal, A. and Schatz, D.G. 1997. RAG1 and RAG2 Form a Stable Postcleavage Synaptic Complex with DNA Containing Signal Ends in V(D)J Recombination. *Cell*. **89**,pp.43-53.

Agrawal, A., Eastman, Q.M. and Schatz, D.G. 1998. Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. *Nature*. **394**,pp.744-751.

Aidinis, V., Bonaldi, T., Beltrame, M., Santagata, S., Bianchi, M.E. and Spanopoulou, E. 1999. The RAG1 homeodomain recruits HMG1 and HMG2 to facilitate recombination signal sequence binding and to enhance the intrinsic DNA-bending activity of RAG1-RAG2. *Molecular and Cellular Biology.* **19**,pp.6532-6542.

Aidinis, V., Dias, D.C., Gomez, C.A., Bhattacharyya, D., Spanopoulou, E. and Santagata, S. 2000. Definition of Minimal Domains of Interaction Within the

Recombination-Activating Genes 1 and 2 Recombinase Complex. *Journal of Immunology* **164**,pp.5826-5832.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter, P. 2002. Molecular Biology of the Cell, 4th edition.

Allan, J., Hartman, P.G., Crane-Robinson, C. and Aviles, F.X. 1980. The structure of histone H1 and its location in chromatin. *Nature*. **288**,pp.675-679.

Aravind, L. and Landsman, D. 1998. AT-hook motifs identified in a wide variety of DNA-binding proteins. *Nucleic Acids Research*. **26**,p.4413-4421.

Arbuckle, J.L., Rahman, N.S., Zhao, S., Rodgers, W. and Rodgers, K.K. 2011. Elucidating the domain architecture and functions of non-core RAG1: the capacity of a non-core zinc-binding domain to function in nuclear import and nucleic acid binding. *BMC Biochemistry*. **12**,pp.23-41.

Arnal, S.M., Holub, A.J., Salus, S.S. and Roth, D.B. 2010. Non-consensus heptamer sequences destabilize the RAG post-cleavage complex, making ends available to alternative DNA repair pathways. *Nucleic Acids Research*. **38**,pp.2944–2954.

Bakhshi, A., Wright, J.J., Graninger, W., Seto, M., Owens, J., Cossman, J., Jensen, J.P., Goldman, P. and Korsmeyer, S.J. 1987. Mechanism of the t(14;18) chromosomal translocation: structural analysis of both derivative 14 and 18 reciprocal partners. *PNAS*. **84**,pp.2396–2400.

Barreto, V. and Cumano, A. 2000. Frequency and characterization of phenotypic Ig heavy chain allelically included IgM-expressing B cells in mice. *Journal of Immunology.* **164**,pp.893–899.

Bassing, C.H., Swat, W. and Alt, F.W. 2002. The Mechanism and Regulation of Chromosomal V(D)J Recombination. *Cell*. **109**,pp.45–55.

Baumann, M., Mamais, A., McBlane, F., Xiao, H. and Boyes, J. 2003. Regulation of V(D)J recombination by nucleosome positioning at recombination signal sequences. *The EMBO Journal*. **22**,pp.5197–5207.

Bergeron, S., Madathiparambil, T. and Swanson, P.C. 2005. Both high mobility group (HMG)-boxes and the acidic tail of HMGB1 regulate recombination activating gene (RAG)-mediated recombination signal synapsis and cleavage in vitro. *Journal of Biological Chemistry.* **280**,pp.31314-31324.

Bergeron, S., Anderson, D.K. and Swanson, P.C. 2006. RAG and HMGB1 proteins: purification and biochemical analysis of recombination signal complexes. *Methods in Enzymology.* **408**,pp.511-528.

Bettridge, J., Na, C.H., Pandey, A. and Desiderio, S. 2017. H3K4me3 induces allosteric conformational changes in the DNA-binding and catalytic regions of the V(D)J recombinase. *PNAS*. **114**,pp.1904-1909.

Bevington, S. and Boyes, J. 2013. Transcription−coupled eviction of histones H2A/H2B governs V(D)J recombination. *The EMBO Journal*. **32**,pp.1381–1392.

Bischerour, J., Lu, C., Roth, D.B. and Chalmers, R. 2009. Base Flipping in V(D)J Recombination: Insights into the Mechanism of Hairpin Formation, the 12/23 Rule, and the Coordination of Double-Strand Breaks. *Molecular and Cellular Biology.* **29**,pp.5889–5899.

Brandt, V.L. and Roth, D.B. 2009. Recent insights into the formation of RAG-induced chromosomal translocations. *Advances in experimental medicine and biology*. **650**,pp.32–45.

Bransteitter, R., Pham, P., Scharff, M.D. and Goodman, M.F. 2003. Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. *PNAS*. **100**,pp.4102–4107.

Brändle, D., Müller, C., Rülicke, T., Hengartner, H. and Pircher, H. 1992. Engagement of the T-cell receptor during positive selection in the thymus down-regulates RAG-1 expression. *PNAS*. **89**,pp.9529–9533.

Bredemeyer, A.L., Sharma, G.G., Huang, C.Y., Helmink, B.A., Walker, L.M., Khor, K.C., Nuskey, B., Sullivan, K.E., Pandita, T.K., Bassing, C.H. and Sleckman, B.P. 2006. ATM stabilizes DNA double-strand-break complexes during V(D)J recombination. *Nature.* **27**,pp.466-470.

Brownlie, R.J. and Zamoyska, R. 2013. T cell receptor signalling networks: branched, diversified and bounded. *Nature Reviews Immunology*. **13**,p.257-269.

Burton, D.R. and Hangartner, L. 2016. Broadly Neutralizing Antibodies to HIV and Their Role in Vaccine Design. *Annual Review of Immunology.* **34**,pp.635–659.

Bustin, M. 2001. Revised nomenclature for high mobility group (HMG) chromosomal proteins. *Trends in biochemical sciences*. **26**,pp.152-153.

Callebaut, I. and Mornon, J.P. 1998. The V(D)J recombination activating protein RAG2 consists of a six-bladed propeller and a PHD fingerlike domain, as revealed by sequence analysis. *Cellular and Molecular Life Sciences CMLS*. **54**,pp.880-891.

Calogero, S., Grassi, F., Aguzzi, A., Voigtländer, T., Ferrier, P., Ferrari, S. and Bianchi, M.E. 1999. The lack of chromosomal protein Hmg1 does not disrupt cell growth but causes lethal hypoglycaemia in newborn mice. *Nature Genetics*. **22**,pp.276-280.

Carmona, L.M. and Schatz, D.G. 2017. New insights into the evolutionary origins of the recombination-activating gene proteins and V(D)J recombination. *The FEBS journal*. **284**,pp.1590-1605.

Casellas, R., Zhang, Q., Zheng, N.-Y., Mathias, M.D., Smith, K. and Wilson, P.C. 2007. Igkappa allelic inclusion is a consequence of receptor editing. *The Journal of Experimental Medicine*. **204**,pp.153–160.

Cattoretti, G., Büttner, M., Shaknovich, R., Kremmer, E., Alobeid, B. and Niedobitek, G. 2006. Nuclear and cytoplasmic AID in extrafollicular and germinal center B cells. *Blood*. **107**,pp.3967–3975.

Chang, H.H.Y. and Lieber, M.R. 2016. Structure-Specific nuclease activities of Artemis and the Artemis: DNA-PKcs complex. *Nucleic Acids Research*. **44**,pp.4991-4997.

Chen, C.K.M., Chan, N.-L. and Wang, A.H.J. 2011. The many blades of the β-propeller proteins: conserved but versatile. *Trends in biochemical sciences*. **36**,pp.553-561.

Chowdhury, D. and Sen, R. 2001. Stepwise activation of the immunoglobulin μ heavy chain gene locus. *The EMBO Journal*. **20**,pp.6394–6403.

Clayton, A.L., Hazzalin, C.A. and Mahadevan, L.C. 2006. Enhanced Histone Acetylation and Transcription: A Dynamic Perspective. *Molecular Cell*. **23**,pp.289–296.

Coleclough, C., Perry, R.P., Karjalainen, K. and Weigert, M. 1981. Aberrant rearrangements contribute significantly to the allelic exclusion of immunoglobulin gene expression. *Nature*. **290**,pp.372-378.

Conticello, S.G., Thomas, C.J.F., Petersen-Mahrt, S.K. and Neuberger, M.S. 2005. Evolution of the AID/APOBEC Family of Polynucleotide (Deoxy)cytidine Deaminases. *Molecular Biology and Evolution*. **22**,pp.367–377.

Cortes, P., Ye, Z.S. and Baltimore, D. 1994. RAG-1 interacts with the repeated amino acid motif of the human homologue of the yeast protein SRP1. *PNAS*. **91**,pp. 7633-7637.

Cortes, P., Weis-Garcia, F., Misulovin, Z., Nussenzweig, A., Lai, J.S., Li, G. Nussenzweig, M.C. and Baltimore, D. 1996. *In vitro* V(D)J recombination: signal joint formation. *PNAS.* **93**,pp.14008-14013.

Couëdel, C., Roman, C., Jones, A., Vezzoni, P., Villa, A. and Cortes, P. 2010. Analysis of mutations from SCID and Omenn syndrome patients reveals the central role of

the Rag2 PHD domain in regulating V(D)J recombination. *Journal of Clinical Investigation*. **120**,pp.1337–1344.

Coussens, M.A., Wendland, R.L., Deriano, L., Lindsay, C.R., Arnal, S.M. and Roth, D.B. 2013. RAG2's acidic hinge restricts repair-pathway choice and promotes genomic stability. *Cell reports*. **4**,pp.870-878.

Crotty, S. 2014. T follicular helper cell differentiation, function, and roles in disease. *Immunity*. **41**,pp.529-542.

Cuomo, C.A. and Oettinger, M.A. 1994. Analysis of regions of RAG-2 important for V(D)J recombination. *Nucleic Acids Research*. **22**,pp.1810-1814.

Curry, J.D., Schulz, D., Guidos, C.J., Danska, J.S., Nutter, L., Nussenzweig, A. and Schlissel, M.S. 2007. Chromosomal reinsertion of broken RSS ends during T cell development. *The Journal of Experimental Medicine*. **204**,pp.2293–2303.

Davies, J.E., Apta, B.H.R. and Harper, M.T. 2018. Cross-reactivity of anti-HMGB1 antibodies for HMGB2. *Journal of Immunological Methods.* **456**,pp.72-76.

De, P., Peak, M.M. and Rodgers, K.K. 2004. DNA Cleavage Activity of the V(D)J Recombination Protein RAG1 Is Autoregulated. *Molecular and Cellular Biology* **24**,pp.6850-6860.

Deng, Z., Liu, H. and Liu, X. 2015. RAG1-mediated ubiquitylation of histone H3 is required for chromosomal V(D)J recombination. *Cell research*. **25**,pp.181-192.

Deriano, L. and Roth, D.B. 2013. Modernizing the Nonhomologous End-Joining Repertoire: Alternative and Classical NHEJ Share the Stage. *Annual Review of Genetics.* **47**,pp.433–455.

Deriano, L., Chaumeil, J., Coussens, M., Multani, A., Chou, Y., Alekseyenko, A.V., Chang, S., Skok, J.A. and Roth, D.B. 2011. The RAG2 C terminus suppresses genomic instability and lymphomagenesis. *Nature*. **471**,pp.119-123.

Di Noia, J.M. and Neuberger, M.S. 2007. Molecular Mechanisms of Antibody Somatic Hypermutation. *Annual Review of Biochemistry.* **76**,pp.1–22.

Dickerson, S.K., Market, E., Besmer, E. and Papavasiliou, F.N. 2003. AID Mediates Hypermutation by Deaminating Single Stranded DNA. *The Journal of Experimental Medicine*. **197**,pp.1291–1296.

Dorna, M.D., Alves, P.F., Rangel-Santos, A., Csomos, K., Ujhazi, B., Dasso, J., Thwaites, D., Boyes, J., Savic, S. and Walter, J.E. 2019. Combined immunodeficiency with late-onset progressive hypogammaglobulinemia and normal B cell count in a patient with RAG2 deficiency. *Frontiers in Pediatrics*. **7**,pp. 122-128.

Dudley, D.D., Sekiguchi, J., Zhu, C., Sadofsky, M.J., Whitlow, S., DeVido, J., Monroe, R.J., Bassing, C.H. and Alt, F.W. 2003. Impaired V(D)J Recombination and Lymphocyte Development in Core RAG1-expressing Mice. *The Journal of Experimental Medicine*. **198**,pp.1439-1450.

Düber, S., Engel, H., Rolink, A., Kretschmer, K. and Weiss, S. 2003. Germline transcripts of immunoglobulin light chain variable regions are structurally diverse and differentially expressed. *Molecular Immunology*. **40**,pp.509–516.

Dyer, P.N., Edayathumangalam, R.S., White, C.L., Bao, Y., Chakravarthy, S., Muthurajan, U.M. and Luger, K. 2004. Reconstitution of nucleosome core particles from recombinant histones and DNA. *Methods in Enzymology*. **375**,pp.23-44.

Eastman, Q.M. and Schatz, D.G. 1997. Nicking is asynchronous and stimulated by synapsis in 12/23 rule-regulated V(D)J cleavage. *Nucleic Acids Research*. **25**,pp. 4370–4378.

Ebert, A., Medvedovic, J., Tagoh, H., Schwickert, T.A. and Busslinger, M. 2013. Control of Antigen Receptor Diversity through Spatial Regulation of V(D)J Recombination. *Cold Spring Harbor Symposia on Quantitative Biology*. **78**,pp.11–21.

Elkin, S.K., Ivanov, D., Ewalt, M., Ferguson, C.G., Hyberts, S.G., Sun, Z.-Y.J., Prestwich, G.D., Yuan, J., Wagner, G., Oettinger, M.A. and Gozani, O.P. 2005. A PHD Finger Motif in the C Terminus of RAG2 Modulates Recombination Activity. *Journal of Biological Chemistry*. **280**,pp.28701-28710.

Elkin, S.K., Matthews, A.G. and Oettinger, M.A. 2003. The C−terminal portion of RAG2 protects against transposition in vitro. *The EMBO Journal*. **22**,pp.1931–1938.

Elnour, I.B., Ahmed, S., Halim, K. and Nirmala, V. 2007. Omenn's Syndrome: A rare primary immunodeficiency disorder. *Sultan Qaboos University medical journal*. **7**,pp.133–138.

Emamzadeh, F.N. 2016. Alpha-synuclein structure, functions, and interactions. *Journal of Research in Medical Sciences.* **21**.pp,29-36.

Espinoza, C.R. and Feeney, A.J. 2005. The Extent of Histone Acetylation Correlates with the Differential Rearrangement Frequency of Individual VH Genes in Pro-B Cells. *Journal of Immunology* **175**,pp.6668–6675.

Ferrando, A.A., Neuberg, D.S., Staunton, J., Loh, M.L., Huard, C., Raimondi, S.C., Behm, F.G., Pui, C.-H., Downing, J.R., Gilliland, D.G., Lander, E.S., Golub, T.R. and Look, A.T. 2002. Gene expression signatures define novel oncogenic pathways in T cell acute lymphoblastic leukemia. *Cancer Cell*. **1**,pp.75–87.

Fischer, A., Notarangelo, L.D., Neven, B., Cavazzana, M. and Puck, J.M. 2015. Severe combined immunodeficiencies and related disorders. *Nature Reviews Disease Primers*. **1**,p.15061.

Fitzsimmons, S.P., Bernstein, R.M., Max, E.E., Skok, J.A. and Shapiro, M.A. 2007. Dynamic Changes in Accessibility, Nuclear Positioning, Recombination, and Transcription at the Igκ Locus. *Journal of Immunology* **179**,pp.5264–5273.

Fugmann, S.D. 2010. The origins of the Rag genes--from transposition to V(D)J recombination. *Seminars in immunology*. **22**,pp.10-16.

Fugmann, S.D., Messier, C., Novack, L.A., Cameron, R.A. and Rast, J.P. 2006. An ancient evolutionary origin of the *Rag1\2* gene locus. *PNAS*. **103**,pp.3728-3733.

Fukushima, T., Takata, M., Morrison, C., Araki, R., Fujimori, A., Abe, M., Tatsumi, K., Jasin, M., Dhar, P.K., Sonoda, E., Chiba, T. and Takeda, S. 2001. Genetic Analysis of the DNA-dependent Protein Kinase Reveals an Inhibitory Role of Ku in Late S–G2 Phase DNA Double-strand Break Repair. *Journal of Biological Chemistry*. **276**,pp.44413–44418.

Fuller, K. and Storb, U. 1997. Identification and characterization of the murine Rag1 promoter. *Molecular Immunology*. **34**,pp.939–954.

Fuxa, M., Skok, J., Souabni, A., Salvagiotto, G., Roldan, E. and Busslinger, M. 2004. Pax5 induces V-to-DJ rearrangements and locus contraction of the immunoglobulin heavy-chain gene. *Genes & Development*. **18**,pp.411–422.

Gauss, G.H. and Lieber, M.R. 1992. The basis for the mechanistic bias for deletional over inversional V(D)J recombination. *Genes \& Development*. **6**,pp.1553–1561.

Geier, J.K. and Schlissel, M.S. 2006. Pre-BCR signals and the control of Ig gene rearrangements. *Seminars in Immunology*. **18**,pp.31–39.

Gellert, M. 2002. V(D)J Recombination: RAG Proteins, Repair Factors, and Regulation. *Annual Review of Biochemistry.* **71**,pp.101-132.

Gigi, V., Lewis, S., Shestova, O., Mijušković, M., Deriano, L., Meng, W., Prak, E.T.L. and Roth, D.B. 2014. RAG2 mutants alter DSB repair pathway choice in vivo and illuminate the nature of ‘alternative NHEJ’. *Nucleic Acids Research*. **42**,pp. 6352-6364.

Gilbert, L.A., Horlbeck, M.A., Adamson, B., Villalta, J.E., Chen, Y., Whitehead, E.H., Guimaraes, C., Panning, B., Ploegh, H.L., Bassik, M.C., Qi, L.S., Kampmann, M. and Weissman J.S. 2014 Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell*. **159**,pp.647-661.

Glimcher, L.H. and Singh, H. 1999. Transcription Factors in Lymphocyte Development — T and B Cells Get Together. *Cell*. **96**,pp.13–23.

Golding, A., Chandler, S., Ballestar, E., Wolffe, A.P. and Schlissel, M.S. 1999. Nucleosome structure completely inhibits in vitro cleavage by the V(D)J recombinase. *The EMBO Journal*. **18**,pp.3712–3723.

Goldmit, M., Ji, Y., Skok, J., Roldan, E., Jung, S., Cedar, H. and Bergman, Y. 2004. Epigenetic ontogeny of the Igk locus during B cell development. *Nature Immunology*. **6**,pp.198-203.

Goodwin, G.H., Sanders, C. and Johns, E.W. 1973. A New Group of Chromatin-Associated Proteins with a High Content of Acidic and Basic Amino Acids. *European Journal of Biochemistry*. **38**,pp.14-19.

Gorman, J.R. and Alt, F.W. 1998. Regulation of immunoglobulin light chain isotype expression. *Advances in immunology*. **69**,pp.113–181.

Graham, B.W, Tao, Y., Dodge, K.L., Thaxton, C.T., Olaso, D., Young, N.L., Marshall, A.G. and Tarkselis, M.A. 2016. DNA Interactions Probed by Hydrogen-Deuterium Exchange (HDX) Fourier Transform Ion Cyclotron Resonance Mass Spectrometry Confirm External Binding Sites on the Minichromosomal Maintenance (MCM) Helicase. *The Journal of Biological Chemistry.* **291**,pp.12467-12480.

Grawunder, U., Leu, T.M.J., Schatz, D.G., Werner, A., Rolink, A.G., Melchers, F. and Winkler, T.H. 1995. Down-regulation of RAG1 and RAG2 gene expression in PreB cells after functional immunoglobulin heavy chain rearrangement. *Immunity*. **3**,pp. 601–608.

Grazini, U., Zanardi, F., Citterio, E., Casola, S., Goding, C.R. and McBlane, F. 2010. The RING Domain of RAG1 Ubiquitylates Histone H3: A Novel Activity in Chromatin-Mediated Regulation of V(D)J Joining. *Molecular Cell*. **37**,pp.282-293.

Guo, C., Yoon, H.S., Franklin, A., Jain, S., Ebert, A., Cheng, H.-L., Hansen, E., Despo, O., Bossen, C., Vettermann, C., Bates, J.G., Richards, N., Myers, D., Patel, H., Gallagher, M., Schlissel, M.S., Murre, C., Busslinger, M., Giallourakis, C.C. and Alt, F.W. 2011. CTCF-binding elements mediate control of V(D)J recombination. *Nature*. **477**,pp.424-430.

Gwyn, L.M., Peak, M.M., De, P., Rahman, N.S. and Rodgers, K.K. 2009. A zinc site in the C-terminal domain of RAG1 is essential for DNA cleavage activity. *Journal of molecular biology*. **390**,pp.863-378.

Han, J.O., Steen, S.B. and Roth, D.B. 1999. Intermolecular V(D)J recombination is prohibited specifically at the joining step. *Molecular Cell*. **3**,pp,331-338.

Hardy, R.R., Carmack, C.E., Shinton, S.A., Kemp, J.D. and Hayakawa, K. 1991. Resolution and characterization of pro-B and pre-pro-B cell stages in normal mouse bone marrow. *The Journal of Experimental Medicine*. **173**,pp.1213–1225.

Helmink, B.A. and Sleckman, B.P. 2012. The Response to and Repair of RAG-Mediated DNA Double-Strand Breaks. *Annual Review of Immunology* **30**,pp. 175-202.

Hencken, C.G., Li, X. and Craig, N.L. 2012. Functional characterization of an active Rag-like transposase. *Nature Structural & Molecular Biology*. **19**,pp.834-836.

Herzberg, C., Weidinger, L.A., Dorrbecker, B., Hubner, S., Stulke, J. and Commichau, F.M. 2007. SPINE: a method for the rapid detection and analysis of protein-protein interactions *in vivo*. *Proteomics.* **7**,pp.4032-4035.

Hesse, J.E., Lieber, M.K., Mizuuchi, K. and Gellert, M. 1989. V(D)J recombination: a functional definition of the joining signals. *Genes and Development* **3**,pp. 1053-1061.

Hewitt, S.L., Wong, J.B., Lee, J.-H., Nishana, M., Chen, H., Coussens, M., Arnal, S.M., Blumenberg, L.M., Roth, D.B., Paull, T.T. and Skok, J.A. 2017. The Conserved ATM

Kinase RAG2-S365 Phosphorylation Site Limits Cleavage Events in Individual Cells Independent of Any Repair Defect. *Cell reports*. **21**,pp.979–993.

Hewitt, S.L., Yin, B., Ji, Y., Chaumeil, J., Marszalek, K., Tenthorey, J., Salvagiotto, G., Steinel, N., Ramsey, L.B., Ghysdael, J., Farrar, M.A., Sleckman, B.P., Schatz, D.G., Busslinger, M., Bassing, C.H. and Skok, J.A. 2009. RAG-1 and ATM coordinate monoallelic recombination and nuclear positioning of immunoglobulin loci. *Nature Immunology*. **10**,pp.655–664.

Hiom, K. and Gellert, M. 1998. Assembly of a 12/23 Paired Signal Complex: A Critical Control Point in V(D)J Recombination. *Molecular Cell*. **1**,pp.1011-1019.

Hiom, K., Melek, M. and Gellert, M. 1998. DNA Transposition by the RAG1 and RAG2 Proteins A Possible Source of Oncogenic Translocations. *Cell*. **94**,pp.463–470.

Hirt, B. 1967. Selective extraction of polyoma DNA from infected mouse cell cultures. *Journal of Molecular Biology*. **26**,pp.365-369.

Hizume, K., Araki, S., Hata, K., Prieto, E., Kundu, T.K., Yoshikawa, K. and Takeyasu, K. 2011. Nano-scale analyses of the chromatin decompaction induced by histone acetylation. *Archives of Histology and Cytology*. **73**,pp.149–163.

Hsu, L.-Y., Lauring, J., Liang, H.-E., Greenbaum, S., Cado, D., Zhuang, Y. and Schlissel, M.S. 2003. A Conserved Transcriptional Enhancer Regulates RAG Gene Expression in Developing B Cells. *Immunity*. **19**,pp.105–117.

Huang, J. and Muegge, K. 2001. Control of chromatin accessibility for V(D)J recombination by interleukin−7. *Journal of Leukocyte Biology*. **69**,pp.907–911.

Huang, S., Tao, X., Yuan, S., Zhang, Y., Li, P., Beilinson, H.A., Zhang, Y., Yu, W., Pontarotti, P., Escriva, H., Le Petillon, Y., Liu, X., Chen, S., Schatz, D.G. and Xu, A. 2016. Discovery of an Active RAG Transposon Illuminates the Origins of V(D)J Recombination. *Cell*. **166**,pp.102-114.

Huye, L.E., Purugganan, M.M., Jiang, M.-M. and Roth, D.B. 2002. Mutational Analysis of All Conserved Basic Amino Acids in RAG-1 Reveals Catalytic, Step Arrest, and Joining-Deficient Mutants in the V(D)J Recombinase. *Molecular and Cellular Biology.* **22**,pp.3460–3473.

Jamieson, E.R., Jacobson, M.P., Barnes, C.M., Chow, C.S. and Lippard, S.J. 1999. Structural and Kinetic Studies of a Cisplatin-modified DNA Icosamer Binding to HMG1 Domain B. *Journal of Biological Chemistry*. **274**,pp.12346–12354.

Janeway, C.A., Travers, P., Walport, M. and Shlomchik, M.J. 2001. *Immunobiology: The Immune System in Health and Disease. 5th edition.*

Jenuwein, T. and Allis, C.D. 2001. Translating the Histone Code. *Science*. **293**,pp. 1074–1080.

Ji, Y., Resch, W., Corbett, E., Yamane, A., Casellas, R. and Schatz, D.G. 2010. The in vivo pattern of binding of RAG1 and RAG2 to antigen receptor loci. *Cell*. **141**,pp. 419–431.

Jiang, H., Chang, F.-C., Ross, A.E., Lee, J., Nakayama, K., Nakayama, K. and Desiderio, S. 2005. Ubiquitylation of RAG-2 by Skp2-SCF links destruction of the V(D)J recombinase to the cell cycle. *Molecular Cell*. **18**,pp.699–709.

Jones, J.M. and Gellert, M. 2003. Autoubiquitylation of the V(D)J recombinase protein RAG1. *PNAS*. **100**,pp.15446-15451.

Jones, J.M. and Gellert, M. 2001. Intermediates in V(D)J recombination: A stable RAG1/2 complex sequesters cleaved RSS ends. *PNAS*. **98**,pp.12926-12931.

Jones, J.M. and Gellert, M. 2002. Ordered assembly of the V(D)J synaptic complex ensures accurate recombination. *The EMBO Journal*. **21**,pp.4162–4171.

Jones, J.M. and Gellert, M. 2004. The taming of a transposon: V(D)J recombination and the immune system. *Immunological Reviews*. **200**,pp.233-248.

Kapitonov, V.V. and Jurka, J. 2005. RAG1 Core and V(D)J Recombination Signal Sequences Were Derived from Transib Transposons. *PLoS Biology*. **3**,p.e181.

Kim, D.R., Dai, Y., Mundy, C.L., Yang, W. and Oettinger, M.A. 1999. Mutations of acidic residues in RAG1 define the active site of the V(D)J recombinase. *Genes & Development*. **13**,pp.3070-3080.

Kim, M.-S., Chuenchor, W., Chen, X., Cui, Y., Zhang, X., Zhou, Z.H., Gellert, M. and Yang, W. 2018. Cracking the DNA Code for V(D)J Recombination. *Molecular Cell*. **70**,pp.358-370.

Kim, M.-S., Lapkouski, M., Yang, W. and Gellert, M. 2015. Crystal structure of the V(D)J recombinase RAG1–RAG2. *Nature*. **518**,pp.507-511.

King, C.Y. and Weiss, M.A. 1993. The SRY high-mobility-group box recognizes DNA by partial intercalation in the minor groove: a topological mechanism of sequence specificity. *PNAS*. **90**,pp.11990–11994.

Kireeva, M.L., Walter, W., Tchernajenko, V., Bondarenko, V., Kashlev, M. and Studitsky, V.M. 2002. Nucleosome Remodeling Induced by RNA Polymerase II Loss of the H2A/H2B Dimer during Transcription. *Molecular Cell*. **9**,pp.541–552.

Kirkham, C.M., Scott, J.N.F., Wang, X., Smith, A.L., Kupinski, A.P., Ford, A.M., Westhead, D.R., Stockley, P.G., Tuma, R. and Boyes, J. 2019. Cut-and-Run: A Distinct Mechanism by which V(D)J Recombination Causes Genome Instability. *Molecular Cell*. **74**,pp.584-597.

Klein, F., Diskin, R., Scheid, J.F., Gaebler, C., Mouquet, H., Georgiev, I.S., Pancera, M., Zhou, T., Incesu, R.-B., Fu, B.Z., Gnanapragasam, P.N.P., Oliveira, T.Y., Seaman, M.S., Kwong, P.D., Bjorkman, P.J. and Nussenzweig, M.C. 2013. Somatic Mutations of the Immunoglobulin Framework Are Generally Required for Broad and Potent HIV-1 Neutralization. *Cell*. **153**,pp.126–138.

Ko, J.E., Kim, C.W. and Kim, D.R. 2004. Amino Acid Residues in RAG1 Responsible for the Interaction with RAG2 during the V(D)J Recombination Process. *Journal of Biological Chemistry*. **279**,pp.7715–7720.

Kochert, B.A., Iacob, R.E., Wales, T.E., Makriyannis, A. and Engen, J.R. 2018. Hydrogen-Deuterium Exchange Mass Spectrometry to Study Protein Complexes. *Methods in Molecular Biology*. **1764**,pp.153-171.

Kosak, S.T., Skok, J.A., Medina, K.L., Riblet, R., Beau, M.M.L., Fisher, A.G. and Singh, H. 2002. Subnuclear Compartmentalization of Immunoglobulin Loci During Lymphocyte Development. *Science*. **296**,pp.158–162.

Krajewski, W.A. and Becker, P.B. 1998. Reconstitution of hyperacetylated, DNase I-sensitive chromatin characterized by high conformational flexibility of nucleosomal DNA. *PNAS*. **95**,pp.1540–1545.

Krangel, M.S. 2015. Beyond Hypothesis: Direct Evidence That V(D)J Recombination Is Regulated by the Accessibility of Chromatin Substrates. *Journal of Immunology* **195**,pp.5103–5105.

Kridel, R., Sehn, L.H. and Gascoyne, R.D. 2012. Pathogenesis of follicular lymphoma. *Journal of Clinical Investigation*. **122**,pp.3424–3431.

Krokan, H.E. and Bjørås, M. 2013. Base Excision Repair. *Cold Spring Harbor Perspectives in Biology*. **5**,p.a012583.

Kumánovics, A., Lee, Y.N., Close, D.W., Coonrod, E.M., Ujhazi, B., Chen, K., MacArthur, D.G., Krivan, G., Notarangelo, L.D. and Walter, J.E. 2017. Estimated disease incidence of RAG1/2 mutations: A case report and querying the Exome Aggregation Consortium. *Journal of Allergy and Clinical Immunology*. **139**,pp.690–692.

Küppers, R. and Dalla-Favera, R. 2001. Mechanisms of chromosomal translocations in B cell lymphomas. *Oncogene*. **20**,pp.5580-5594.

Kwan, A., Abraham, R.S., Currier, R., Brower, A., Andruszewski, K., Abbott, J.K., Baker, M., Ballow, M., Bartoshesky, L.E., Bonilla, F.A., Brokopp, C., Brooks, E., Caggana, M., Celestin, J., Church, J.A., Comeau, A.M., Connelly, J.A., Cowan, M.J., Cunningham-Rundles, C., Dasu, T., Dave, N., La Morena, De, M.T., Duffner, U., Fong, C.-T., Forbes, L., Freedenberg, D., Gelfand, E.W., Hale, J.E., Hanson, I.C., Hay, B.N., Hu, D., Infante, A., Johnson, D., Kapoor, N., Kay, D.M., Kohn, D.B., Lee, R., Lehman, H., Lin, Z., Lorey, F., Abdel-Mageed, A., Manning, A., McGhee, S., Moore, T.B., Naides, S.J., Notarangelo, L.D., Orange, J.S., Pai, S.-Y., Porteus, M., Rodriguez, R., Romberg, N., Routes, J., Ruehle, M., Rubenstein, A., Saavedra-Matiz, C.A., Scott, G., Scott, P.M., Secord, E., Seroogy, C., Shearer, W.T., Siegel, S., Silvers, S.K., Stiehm, E.R., Sugerman, R.W., Sullivan, J.L., Tanksley, S., Tierce, M.L.4., Verbsky, J., Vogel, B., Walker, R., Walkovich, K., Walter, J.E., Wasserman, R.L., Watson, M.S., Weinberg, G.A., Weiner, L.B., Wood, H., Yates, A.B., Puck, J.M. and Bonagura, V.R. 2014. Newborn screening for severe combined immunodeficiency in 11 screening programs in the United States. *JAMA*. **312**,pp. 729–738.

Kwon, J., Imbalzano, A.N., Matthews, A. and Oettinger, M.A. 1998. Accessibility of Nucleosomal DNA to V(D)J Cleavage Is Modulated by RSS Positioning and HMG1. *Molecular Cell*. **2**,pp.829–839.

Laemmli, U.K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*. **227**,pp.680-685.

Landree, M.A., Wibbenmeyer, J.A. and Roth, D.B. 1999. Mutational analysis of RAG1 and RAG2 identifies three catalytic amino acids in RAG1 critical for both cleavage steps of V(D)J recombination. *Genes & Development*. **13**,pp.3059-3069.

Larijani, M., Petrov, A.P., Kolenchenko, O., Berru, M., Krylov, S.N. and Martin, A. 2007. AID Associates with Single-Stranded DNA with High Affinity and a Long Complex Half-Life in a Sequence-Independent Manner. *Molecular and Cellular Biology.* **27**,pp.20–30.

Lauring, J. and Schlissel, M.S. 1999. Distinct Factors Regulate the Murine RAG-2 Promoter in B- and T-Cell Lines. *Molecular and Cellular Biology.* **19**,pp.2601–2612.

Lawless, D., Geier, C.B., Farmer, J.R., Lango, H.A., Thwaites, D., Atschekzei, F., Brown, M., Buchbinder, D., Burns, S.O., Butte, M.J., Csomos, K., Deevi, S.V.V., Egner, W., Ehl, S., Eibl, M.M., Fadugba, O., Foldvari, Z., Green, D.M., Henrickson, S.E., Holland, S.M., John, T., Klemann, C., Kuijpers, T.W., Moreira, F., Piller, A., Rayner-Matthews, P., Romberg, N.D., Sargur, R., Schmidt, R.E., Schröder, C., Schuetz, C., Sharapova, S.O., Smith, K.G.C., Sogkas, G., Speckmann, C., Stirrups, K., Thrasher, A.J., Wolf, H.M., Notarangelo, L.D., Anwar, R., Boyes, J., Ujhazi, B., Consortium, N.B.R.D., Thaventhiran, J., Walter, J.E. and Savic, S. 2018. Prevalence and clinical challenges among adult primary immunodeficiency patients with RAG deficiency. *Journal of Allergy and Clinical Immunology*. **141**,pp. 2303-2306.

Leavy, O. 2010. V(D)J recombination: RAG recombination centres. *Nature Reviews in Immunology.* **10**,p.383.

Lee, G.S., Neiditch, M.B., Salus, S.S. and Roth, D.B. 2004. RAG Proteins Shepherd Double-Strand Breaks to a Specific Pathway, Suppressing Error-Prone Repair, but RAG Nicking Initiates Homologous Recombination. *Cell*. **117**,pp.171–184.

Lee, Y.N., Frugoni, F., Dobbs, K., Walter, J.E., Giliani, S., Gennery, A.R., Al-Herz, W., Haddad, E., LeDeist, F., Bleesing, J.H., Henderson, L.A., Pai, S.-Y., Nelson, R.P., El-Ghoneimy, D.H., El-Feky, R.A., Reda, S.M., Hossny, E., Soler-Palacin, P., Fuleihan, R.L., Patel, N.C., Massaad, M.J., Geha, R.S., Puck, J.M., Palma, P., Cancrini, C., Chen, K., Vihinen, M., Alt, F.W. and Notarangelo, L.D. 2014. A systematic analysis of recombination activity and genotype-phenotype correlation in human recombination-activating gene 1 deficiency. *Journal of Allergy and Clinical Immunology*. **133**,pp.1099–1108.

Lee, N.Y., Frugoni, F., Dobbs, K., Tirosh, I., Du, L., Ververs, F.A., Ru, H., Ott de Bruin, L., Adeli, M., Bleesing, J.H., Buchbinder, D., Butte, M.J., Cancrini, C., Chen, K.,

Choo, S., Elfeky, R.A., Finocchi, A., Fuleihan, R.L., Gennery, A.R., El-Ghoneimy, D.H., Henderson, L.A., Al-Herz, W., Hossny, E., Nelson, R.P., Pai, S., Patel, N.C., Reda, S.M., Soler-Palacin, P., Somech, R., Palma, P., Wu, H., Giliani, S., Walter, J.E. and Notarangelo L.D. 2016. Characterization of T and B cell repertoire diversity in patients with RAG deficiency. *Science Immunology*. **1**(6).

Leibiger, C., Kosyakova, N., Mkrtchyan, H., Glei, M., Trifonov, V. and Liehr, T. 2013. First Molecular Cytogenetic High Resolution Characterization of the NIH3T3 Cell Line by Murine Multicolor Banding. *Journal of Histochemistry and Cytochemistry*. **61**,pp. 306-312.

Lescale, C., Abramowski, V., Bedora-Faure, M., Murigneux, V., Vera, G., Roth, D.B., Revy, P., Villartay, J.-P. de and Deriano, L. 2016. RAG2 and XLF/Cernunnos interplay reveals a novel role for the RAG complex in DNA repair. *Nature Communications*. **7**,p.10529.

Leu, T.M.J., Eastman, Q.M. and Schatz, D.G. 1997. Coding Joint Formation in a Cell-Free V(D)J Recombination System. *Immunity*. **7**,pp.303–314.

Lewis, S.M., Hesse, J.E., Mizuuchi, K. And Gellert M. 1988. Novel strand exchanges in V(D)J recombination. *Cell*. **55**,pp.1099-1107.

Lewis, S.M. 1994. The mechanism of V(D)J joining: lessons from molecular, immunological, and comparative analyses. *Advances in immunology*. **56**,pp.27–150.

Lewis, S.M., Agard, E., Suh, S. and Czyzyk, L. 1997. Cryptic signals and the fidelity of V(D)J joining. *Molecular and Cellular Biology.* **17**,pp.3125–3136.

Li, Z., Dordai, D.I., Lee, J. and Desiderio, S. 1996. A Conserved Degradation Signal Regulates RAG-2 Accumulation during Cell Division and Links V(D)J Recombination to the Cell Cycle. *Immunity*. **5**,pp.575–589.

Liang, H.-E., Hsu, L.-Y., Cado, D. and Schlissel, M.S. 2004. Variegated Transcriptional Activation of the Immunoglobulin κ Locus in Pre-B Cells Contributes to the Allelic Exclusion of Light-Chain Expression. *Cell*. **118**,pp.19–29.

Lieber, M.R., Hesse, J.E., Mizuuchi, K. and Gellert, M. 1987. Developmental stage specificity of the lymphoid V(D)J recombination activity. *Genes & Development.* **1**,pp.751-761.

Lieber, M.R. 2010. The Mechanism of Double-Strand DNA Break Repair by the Nonhomologous DNA End-Joining Pathway. *Annual Review of Biochemistry.* **79**,pp. 181–211.

Lieber, M.R. 1993. The role of site-directed recombinases in physiologic and pathologic chromosomal rearrangements. *In*: *The Causes and Consequences of Chromosomal Aberrations*.

Lin, W.-C. and Desiderio, S. 1993. Regulation of V(D)J Recombination Activator Protein RAG-2 by Phosphorylation. *American Association for the Advancement of Science*. **260**,p.953.

Little, A.J., Corbett, E., Ortega, F. and Schatz, D.G. 2013. Cooperative recruitment of HMGB1 during V(D)J recombination through interactions with RAG1 and DNA. *Nucleic Acids Research*. **41**,pp.3289–3301.

Liu, Y., Subrahmanyam, R., Chakraborty, T., Sen, R. and Desiderio, S. 2007a. A plant homeodomain in RAG-2 that binds Hypermethylated lysine 4 of histone H3 is necessary for efficient antigen-receptor-gene rearrangement. *Immunity*. **27**,pp. 561-571.

Liu, H., Schmidt-Supprian, M., Shi, Y., Hobeika, E., Barteneva, N., Jumaa, H., Pelanda, R., Reth, M., Skok, J., Rajewsky, K. and Shi, Y. 2007b. Yin Yang 1 is a critical regulator of B-cell development. *Genes & Development*. **21**,pp.1179–1189.

Lovely, G.A., Brewster, R.C., Schatz, D.G., Baltimore, D. and Phillips, R. 2015. Single-molecule analysis of RAG-mediated V(D)J DNA cleavage. *PNAS*. **112**,pp. 1715-1723.

Lowary, P.T. and Widom, J. 1998. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *Journal of Molecular Biology.* **276**,pp.19-42.

Lu, C., Ward, A., Bettridge, J., Liu, Y. and Desiderio, S. 2015. An autoregulatory mechanism imposes allosteric control on the V(D)J recombinase by histone H3 methylation. *Cell reports*. **10**,pp.29-38.

Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F. and Richmond, T.J. 1997. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*. **389**,pp.251-260.

Malarkey, C.S. and Churchill, M.E.A. 2012. The high mobility group box: the ultimate utility player of a cell. *Trends in biochemical sciences*. **37**,pp.553-562.

Marculescu, R., Le, T., Simon, P., Jaeger, U. and Nadel, B. 2002. V(D)J-mediated Translocations in Lymphoid Neoplasms. *The Journal of Experimental Medicine*. **195**,pp.85–98.

Matthews, A.G.W., Kuo, A.J., Ramón-Maiques, S., Han, S., Champagne, K.S., Ivanov, D., Gallardo, M., Carney, D., Cheung, P., Ciccone, D.N., Walter, K.L., Utz, P.J., Shi, Y., Kutateladze, T.G., Yang, W., Gozani, O. and Oettinger, M.A. 2007. RAG2 PHD finger couples histone H3 lysine 4 trimethylation with V(D)J recombination. *Nature*. **450**,pp.1106-1110.

McBlane, F. and Boyes, J. 2000. Stimulation of V(D)J recombination by histone acetylation. *Current Biology*. **10**,pp.483–486.

McBlane, J.F., van Gent, D.C., Ramsden, D.A., Romeo, C., Cuomo, C.A., Gellert, M. and Oettinger, M.A. 1995. Cleavage at a V(D)J recombination signal requires only RAG1 and RAG2 proteins and occurs in two steps. *Cell*. **83**,pp.387-395.

McMahan, C.J., Difilippantonio, M.J., Rao, N., Spanopoulou, E. and Schatz, D.G. 1997. A basic motif in the N-terminal region of RAG1 enhances V(D)J recombination activity. *Molecular and Cellular Biology.* **17**,pp.4544-4552.

McMurry, M.T. and Krangel, M.S. 2000. A Role for Histone Acetylation in the Developmental Regulation of V(D)J Recombination. *Science*. **287**,pp.495–498.

Medvedovic, J., Ebert, A., Tagoh, H., Tamir, I.M., Schwickert, T.A., Novatchkova, M., Sun, Q., Huis in t Veld, P.J., Guo, C., Yoon, H.S., Denizot, Y., Holwerda, S.J.B., de Laat, W., Cogné, M., Shi, Y., Alt, F.W. and Busslinger, M. 2013. Flexible Long-Range Loops in the VH Gene Region of the Igh Locus Facilitate the Generation of a Diverse Antibody Repertoire. *Immunity*. **39**,pp.229–244.

Meek, K. 2019. RAGs usurp cellular factors for both breaking and repairing. *Blood*. **133**,pp.773-774.

Messier, T.L., O'Neill, J.P., Hou, S.M., Nicklas, J.A. and Finette, B.A. 2003. In vivo transposition mediated by V(D)J recombinase in human T lymphocytes. *The EMBO Journal*. **22**,pp.1381–1388.

Methot, S.P. and Noia, J.M.D. 2017. Advances in Immunology. *Advances in immunology*. **133**,pp.37–87.

Mizianty, M.J., Peng, Z. and Kurgan, L. 2013. MFDp2: Accurate predictor of disorder in proteins by fusion of disorder probabilities, content and profiles. *Intrinsically Disordered Proteins.* **1**(11).

Mizushima, S. and Nagata, S. 1990. pEF-BOS, a powerful mammalian expression vector. *Nucleic Acids Research.* **18**.p,5322.

Mollica, L., Bessa, L.M., Hanoulle, X., Jensen, M.R., Blackedge, M. and Schneider, R. 2016. Binding Mechanisms of Intrinsically Disordered Proteins: Theory, Simulation, and Experiment. *Frontiers in Molecular Biosciences.* **3**:52.

Morshead, K.B., Ciccone, D.N., Taverna, S.D., Allis, C.D. and Oettinger, M.A. 2003. Antigen receptor loci poised for V(D)J rearrangement are broadly associated with BRG1 and flanked by peaks of histone H3 dimethylated at lysine 4. *PNAS*. **100**,pp. 11577-11582.

Mostoslavsky, R., Singh, N., Tenzen, T., Goldmit, M., Gabay, C., Elizur, S., Qi, P., Reubinoff, B.E., Chess, A., Cedar, H. and Bergman, Y. 2001. Asynchronous replication and allelic exclusion in the immune system. *Nature*. **414**,pp.221-225.

Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y. and Honjo, T. 2000. Class Switch Recombination and Hypermutation Require Activation-Induced Cytidine Deaminase (AID), a Potential RNA Editing Enzyme. *Cell*. **102**,pp.553–563.

Murphy, K. and Weaver, C. 2017. *Janeway's Immunobiology 9th Edition*.

Nambiar, M. and Raghavan, S.C. 2012. Mechanism of Fragility at BCL2 Gene Minor Breakpoint Cluster Region during t(14;18) Chromosomal Translocation. *Journal of Biological Chemistry*. **287**,pp.8688–8701.

Nambu, Y., Sugai, M., Gonda, H., Lee, C.-G., Katakai, T., Agata, Y., Yokota, Y. and Shimizu, A. 2003. Transcription-Coupled Events Associating with Immunoglobulin Switch Region Chromatin. *Science*. **302**,pp.2137–2140.

Neiditch, M.B., Lee, G.S., Huye, L.E., Brandt, V.L. and Roth, D.B. 2002. The V(D)J Recombinase Efficiently Cleaves and Transposes Signal Joints. *Molecular Cell*. **9**,pp.871–878.

Nemazee, D. 2006. Receptor editing in lymphocyte development and central tolerance. *Nature Reviews Immunology*. **6**,pp.728-740.

Ng, H.H., Robert, F., Young, R.A. and Struhl, K. 2003. Targeted Recruitment of Set1 Histone Methylase by Elongating Pol II Provides a Localized Mark and Memory of Recent Transcriptional Activity. *Molecular Cell*. **11**,pp.709–719.

Nicolas, N., Moshous, D., Cavazzana-Calvo, M., Papadopoulo, D., Chasseval, R. de, Deist, F.L., Fischer, A. and Villartay, J.-P. de 1998. A Human Severe Combined Immunodeficiency (SCID) Condition with Increased Sensitivity to Ionizing Radiations and Impaired V(D)J Rearrangements Defines a New DNA Recombination/Repair Deficiency. *The Journal of Experimental Medicine*. **188**,pp. 627–634.

Nightingale, K.P., Baumann, M., Eberharter, A., Mamais, A., Becker, P.B. and Boyes, J. 2007. Acetylation increases access of remodelling complexes to their nucleosome targets to enhance initiation of V(D)J recombination. *Nucleic Acids Research*. **35**,pp.6311–6321.

Notarangelo, L.D., Kim, M.-S., Walter, J.E. and Lee, Y.N. 2016. Human RAG mutations: biochemistry and clinical implications. *Nature Reviews Immunology*. **16**,pp. 234-246.

Nussenzweig, A. and Nussenzweig, M.C. 2010. Origin of Chromosomal Translocations in Lymphoid Cancer. *Cell*. **141**,pp.27–38.

Nutt, S.L., Heavey, B., Rolink, A.G. and Busslinger, M. 1999. Commitment to the B-lymphoid lineage depends on the transcription factor Pax5. *Nature*. **402**,pp. 556-562.

Oettinger, M.A., Schatz, D.G., Gorka, C. and Baltimore, D. 1990. RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. *Science*. **248**,pp. 1517-1523.

Ott, J.A., Castro, C.D., Deiss, T.C., Ohta, Y., Flajnik, M.F. and Criscitiello, M.F. 2018. Somatic hypermutation of T cell receptor α chain contributes to selection in nurse shark thymus. *eLife*. **7**:e28477.

Oudelaar, A.M., Hanssen, L.L.P., Hardison, R.C., Kassouf, M.T., Hughes, J.R. and Higgs, D.R. 2017. Between form and function: the complexity of genome folding. *Human Molecular Genetics*. **26**,pp.208–215.

Pai, S.-Y., Logan, B.R., Griffith, L.M., Buckley, R.H., Parrott, R.E., Dvorak, C.C., Kapoor, N., Hanson, I.C., Filipovich, A.H., Jyonouchi, S., Sullivan, K.E., Small, T.N., Burroughs, L., Skoda-Smith, S., Haight, A.E., Grizzle, A., Pulsipher, M.A., Chan, K.W., Fuleihan, R.L., Haddad, E., Loechelt, B., Aquino, V.M., Gillio, A., Davis, J., Knutsen, A., Smith, A.R., Moore, T.B., Schroeder, M.L., Goldman, F.D., Connelly, J.A., Porteus, M.H., Xiang, Q., Shearer, W.T., Fleisher, T.A., Kohn, D.B., Puck, J.M., Notarangelo, L.D., Cowan, M.J. and O'Reilly, R.J. 2014. Transplantation Outcomes for Severe Combined Immunodeficiency, 2000–2009. *The New England Journal of Medicine*. **371**,pp.434–446.

Pan, X., Papasani, M., Hao, Y., Calamito, M., Wei, F., Quinn, W.J., Basu, A., Wang, J., Hodawadekar, S., Zaprazna, K., Liu, H., Shi, Y., Allman, D., Cancro, M. and Atchison, M.L. 2013. YY1 controls Igκ repertoire and B−cell development, and localizes with condensin on the Igκ locus. *The EMBO Journal*. **32**,pp.1168–1182.

Papaemmanuil, E., Rapado, I., Li, Y., Potter, N.E., Wedge, D.C., Tubio, J., Alexandrov, L.B., Loo, P.V., Cooke, S.L., Marshall, J., Martincorena, I., Hinton, J., Gundem, G., Delft, F.W.V., Nik-Zainal, S., Jones, D.R., Ramakrishna, M., Titley, I., Stebbings, L., Leroy, C., Menzies, A., Gamble, J., Robinson, B., Mudie, L., Raine, K., O'Meara, S., Teague, J.W., Butler, A.P., Cazzaniga, G., Biondi, A., Zuna, J., Kempski, H., Muschen, M., Ford, A.M., Stratton, M.R., Greaves, M. and Campbell, P.J. 2014. RAG-mediated recombination is the predominant driver of oncogenic rearrangement in ETV6-RUNX1 acute lymphoblastic leukemia. *Nature Genetics*. **46**,pp.116–125.

Parelho, V., Hadjur, S., Spivakov, M., Leleu, M., Sauer, S., Gregson, H.C., Jarmuz, A., Canzonetta, C., Webster, Z., Nesterova, T., Cobb, B.S., Yokomori, K., Dillon, N.,

Aragon, L., Fisher, A.G. and Merkenschlager, M. 2008. Cohesins Functionally Associate with CTCF on Mammalian Chromosome Arms. *Cell*. **132**,pp.422–433.

Peled, J.U., Kuang, F.L., Iglesias-Ussel, M.D., Roa, S., Kalis, S.L., Goodman, M.F. and Scharff, M.D. 2008. The Biochemistry of Somatic Hypermutation. *Annual Review of Immunology.* **26**,pp.481–511.

Perkins, E.J., Kee, B.L. and Ramsden, D.A. 2004. Histone 3 lysine 4 methylation during the pre−B to immature B−cell transition. *Nucleic Acids Research*. **32**,pp. 1942–1947.

Perry, R.P., Kelley, D.E., Coleclough, C., Seidman, J.G., Leder, P., Tonegawa, S., Matthyssens, G. and Weigert, M. 1980. Transcription of mouse kappa chain genes: implications for allelic exclusion. *PNAS*. **77**,pp.1937–1941.

Pertea, M., Shumate, A., Pertea, G., Varabyou, A., Chang, Y.-C., Madugundu, A.K., Pandey, A. and Salzberg, S. 2018. Thousands of large-scale RNA sequencing experiments yield a comprehensive new human gene list and reveal extensive transcriptional noise. *bioRxiv*.p.332825.

Peters, A. and Storb, U. 1996. Somatic Hypermutation of Immunoglobulin Genes Is Linked to Transcription Initiation. *Immunity*. **4**,pp.57–65.

Pham, P., Bransteitter, R., Petruska, J. and Goodman, M.F. 2003. Processive AID-catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation. *Nature*. **424**,pp.103-107.

Pui, J.C., Allman, D., Xu, L., DeRocco, S., Karnell, F.G., Bakkour, S., Lee, J.Y., Kadesch, T., Hardy, R.R., Aster, J.C. and Pear, W.S. 1999. Notch1 Expression in Early Lymphopoiesis Influences B versus T Lineage Determination. *Immunity*. **11**,pp.299–308.

Pulivarthy, S.R., Lion, M., Kuzu, G., Matthews, A.G.W., Borowsky, M.L., Morris, J., Kingston, R.E., Dennis, J.H., Tolstorukov, M.Y. and Oettinger, M.A. 2016.

Regulated large-scale nucleosome density patterns and precise nucleosome positioning correlate with V(D)J recombination. *PNAS*. **113**,pp.6427–6436.

Qi, Y.K., Ai, H.S., Li, Y.M. and Yan.B. 2018. Total Chemical Synthesis of Modified Histones. *Frontiers in Chemistry.* **6**,pp.19-26.

Raghavan, S.C., Houston, S., Hegde, B.G., Langen, R., Haworth, I.S. and Lieber, M.R. 2004. Stability and Strand Asymmetry in the Non-B DNA Structure at the bcl-2 Major Breakpoint Region. *Journal of Biological Chemistry*. **279**,pp.46213–46225.

Raghavan, S.C., Kirsch, I.R. and Lieber, M.R. 2001. Analysis of the V(D)J Recombination Efficiency at Lymphoid Chromosomal Translocation Breakpoints. *Journal of Biological Chemistry*. **276**,pp.29126–29133.

Raghavan, S.C., Swanson, P.C., Ma, Y. and Lieber, M.R. 2005. Double-Strand Break Formation by the RAG Complex at the Bcl-2 Major Breakpoint Region and at Other Non-B DNA Structures In Vitro. *Molecular and Cellular Biology.* **25**,pp.5904–5919.

Raghavan, S.C., Swanson, P.C., Wu, X., Hsieh, C.-L. and Lieber, M.R. 2004. A non-B-DNA structure at the Bcl-2 major breakpoint region is cleaved by the RAG complex. *Nature*. **428**,pp.88-93.

Rajewsky, K., Forster, I. and Cumano, A. 1987. Evolutionary and somatic selection of the antibody repertoire in the mouse. *Science*. **238**,pp.1088–1094.

Ramiro, A.R., Stavropoulos, P., Jankovic, M. and Nussenzweig, M.C. 2003. Transcription enhances AID-mediated cytidine deamination by exposing single-stranded DNA on the nontemplate strand. *Nature Immunology*. **4**,pp.452–456.

Ramón-Maiques, S., Kuo, A.J., Carney, D., Matthews, A.G.W., Oettinger, M.A., Gozani, O. and Yang, W. 2007. The plant homeodomain finger of RAG2 recognizes histone H3 methylated at both lysine-4 and arginine-2. *PNAS*. **104**,pp.18993-18998.

Ramsden, D.A. and Gellert, M. 1995. Formation and resolution of double-strand break intermediates in V(D)J rearrangement. *Genes & Development*. **9**,pp.2409–2420.

Ramsden, D.A., Baetz, K. and Wu, G.E. 1994. Conservation of sequence in recombination signal sequence spacers. *Nucleic Acids Research*. **22**,pp. 1785-1796.

Ramsden, D.A., Paull, T.T. and Gellert, M. 1997. Cell-free V(D)J recombination. *Nature*. **388**,pp.488-491.

Raval, P., Kriatchko, A.N., Kumar, S. and Swanson, P.C. 2008. Evidence for Ku70/Ku80 association with full-length RAG1. *Nucleic Acids Research*. **36**,pp.2060-2072.

Repasky, J.A.E., Corbett, E., Boboila, C. and Schatz, D.G. 2004. Mutational Analysis of Terminal Deoxynucleotidyltransferase- Mediated N-Nucleotide Addition in V(D)J Recombination. *Journal of Immunology.* **172**,pp.5478-5488.

Ribeiro de Almeida, C., Hendriks, R.W. and Stadhouders, R. 2015. Advances in Immunology. *Advances in immunology*. **128**,pp.183–271.

Rodgers, K.K. 2017. Riches in RAGs: Revealing the V(D)J Recombinase through High-Resolution Structures. *Trends in biochemical sciences*. **42**,pp.72-84.

Rodgers, K.K., Bu, Z., Fleming, K.G., Schatz, D.G., Engelman, D.M. and Coleman, J.E. 1996. A Zinc-binding Domain Involved in the Dimerization of RAG1. *Journal of Molecular Biology*. **260**,pp.70-84.

Roldan, E., Fuxa, M., Chong, W., Martinez, D., Novatchkova, M., Busslinger, M. and Skok, J.A. 2004. Locus ‘decontraction’ and centromeric recruitment contribute to allelic exclusion of the immunoglobulin heavy-chain gene. *Nature Immunology*. **6**,pp.31-41.

Ronfani, L., Ferraguti, M., Croci, L., Ovitt, C.E., Scholer, H.R., Consalez, G.G. and Bianchi, M.E. 2001. Reduced fertility and spermatogenesis defects in mice lacking chromosomal protein Hmgb2. *Development*. **128**,pp.1265-1273.

Roth, D.B. 2003. Restraining the V(D)J recombinase. *Nature Reviews Immunology*. **3**,pp.656–666.

Roth, D.B. and Roth, S.Y. 2000. Unequal Access Regulating V(D)J Recombination through Chromatin Remodeling. *Cell*. **103**,pp.699–702.

Ru, H., Chambers, M.G., Fu, T.-M., Tong, A.B., Liao, M. and Wu, H. 2015. Molecular Mechanism of V(D)J Recombination from Synaptic RAG1-RAG2 Complex Structures. *Cell*. **163**,pp.1138-1152.

Ru, H., Mi, W., Zhang, P., Alt, F.W., Schatz, D.G., Liao, M. and Wu, H. 2018a. DNA melting initiates the RAG catalytic pathway. *Nature structural & molecular biology*. **25**,pp.732-742.

Ru, H., Zhang, P. and Wu, H. 2018b. Structural gymnastics of RAG-mediated DNA cleavage in V(D)J recombination. *Current opinion in structural biology*. **53**,pp.178–186.

Rupp, R.A., Snider, L. And Weintraub, H. 1994. Xenopus embryos regulate the nuclease localisation of XMyoD. *Genes & Development*. **8**,pp.1311-1323.

Ruthenburg, A.J., Allis, C.D. and Wysocka, J. 2007. Methylation of Lysine 4 on Histone H3: Intricacy of Writing and Reading a Single Epigenetic Mark. *Molecular Cell*. **25**,pp.15-30.

Sadofsky, M.J., Hesse, J.E. and Gellert, M. 1994. Definition of a core region of RAG-2 that is functional in V(D)J recombination. *Nucleic Acids Research*. **22**,pp. 1805-1809.

Saint-Basile, G. de, Deist, F.L., Villartay, J.P. de, Cerf-Bensussan, N., Journet, O., Brousse, N., Griscelli, C. and Fischer, A. 1991. Restricted heterogeneity of T lymphocytes in combined immunodeficiency with hypereosinophilia (Omenn's syndrome). *Journal of Clinical Investigation*. **87**,pp.1352–1359.

Sakano, H., Hüppi, K., Heinrich, G. and Tonegawa, S. 1979. Sequences at the somatic recombination sites of immunoglobulin light-chain genes. *Nature*. **280**,pp.288-294.

Sambrook, J., Fritsch, E.F. and Maniatis, T. 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbour Laboratory, Cold Spring Harbour, N.Y.

Santagata, S., Besmer, E., Villa, A., Bozzi, F., Allingham, J.S., Sobacchi, C., Haniford, D.B., Vezzoni, P., Nussenzweig, M.C., Pan, Z.-Q. and Cortes, P. 1999. The RAG1/ RAG2 Complex Constitutes a 3′ Flap Endonuclease Implications for Junctional Diversity in V(D)J and Transpositional Recombination. *Molecular Cell*. **4**,pp.935–947.

Sayegh, C.E., Sayegh, C., Jhunjhunwala, S., Riblet, R. and Murre, C. 2005. Visualization of looping involving the immunoglobulin heavy-chain locus in developing B cells. *Genes & Development*. **19**,pp.322–327.

Schatz, D.G. and Ji, Y. 2011. Recombination centres and the orchestration of V(D)J recombination. *Nature Reviews Immunology*. **11**,pp.251-263.

Schatz, D.G. and Swanson, P.C. 2011. V(D)J Recombination: Mechanisms of Initiation. *Annual Review of Genetics.* **45**,pp.167-202.

Schatz, D.G., Oettinger, M.A. and Baltimore, D. 1989. The V(D)J recombination activating gene, RAG-1. *Cell*. **59**,pp.1035–1048.

Schatz, D.G., Oettinger, M.A. and Schlissel, M.S. 1992. V(D)J Recombination: Molecular Biology and Regulation. *Annual Review of Immunology.* **10**,pp.359-383.

Schlimgen, R.J., Reddy, K.L., Singh, H. and Krangel, M.S. 2008. Initiation of allelic exclusion by stochastic interaction of Tcrb alleles with repressive nuclear compartments. *Nature Immunology*. **9**,pp.802-809.

Schlissel, M.S. and Baltimore, D. 1989. Activation of immunoglobulin kappa gene rearrangement correlates with induction of germline kappa gene transcription. *Cell*. **58**,pp.1001–1007.

Schuetz, C., Huck, K., Gudowius, S., Megahed, M., Feyen, O., Hubner, B., Schneider, D.T., Manfras, B., Pannicke, U., Willemze, R., Knüchel, R., Göbel, U., Schulz, A.,

Borkhardt, A., Friedrich, W., Schwarz, K. and Niehues, T. 2008. An Immunodeficiency Disease with RAG Mutations and Granulomas. *The New England Journal of Medicine*. **358**,pp.2030–2038.

Schultz, H.Y., Landree, M.A., Qiu, J.-X., Kale, S.B. and Roth, D.B. 2001. Joining-Deficient RAG1 Mutants Block V(D)J Recombination In Vivo and Hairpin Opening In Vitro. *Molecular Cell*. **7**,pp.65–75.

Schwarz, K., Gauss, G.H., Ludwig, L., Pannicke, U., Li, Z., Lindner, D., Friedrich, W., Seger, R.A., Hansen-Hagge, T.E., Desiderio, S., Lieber, M.R. and Bartram, C.R. 1996. RAG Mutations in Human B Cell-Negative SCID. *Science*. **274**,pp.97–99.

Seitan, V.C., Hao, B., Tachibana-Konwalski, K., Lavagnolli, T., Mira-Bontenbal, H., Brown, K.E., Teng, G., Carroll, T., Terry, A., Horan, K., Marks, H., Adams, D.J., Schatz, D.G., Aragon, L., Fisher, A.G., Krangel, M.S., Nasmyth, K. and Merkenschlager, M. 2011. A role for cohesin in T-cell-receptor rearrangement and thymocyte differentiation. *Nature*. **476**,pp.467-471.

Sekiguchi, J., Whitlow, S. and Alt, F.W. 2001. Increased Accumulation of Hybrid V(D)J Joins in Cells Expressing Truncated versus Full-Length RAGs. *Molecular Cell*. **8**,pp.1383–1390.

Shogren-Knaak, M., Ishii, H., Sun, J.-M., Pazin, M.J., Davie, J.R. and Peterson, C.L. 2006. Histone H4-K16 Acetylation Controls Chromatin Structure and Protein Interactions. *Science*. **311**,pp.844–847.

Shurtleff, S.A., Buijs, A., Behm, F.G., Rubnitz, J.E., Raimondi, S.C., Hancock, M.L., Chan, G.C., Pui, C.H., Grosveld, G. and Downing, J.R. 1995. TEL/AML1 fusion resulting from a cryptic t(12;21) is the most common genetic lesion in pediatric ALL and defines a subgroup of patients with an excellent prognosis. *Leukemia*. **9**,pp. 1985–1989.

Simkus, C., Anand, P., Bhattacharyya, A. and Jones, J.M. 2007. Biochemical and Folding Defects in a RAG1 Variant Associated with Omenn Syndrome. *Journal of Immunology.* **179**,pp.8332-8340.

Simkus, C., Makiya, M. and Jones, J.M. 2009a. Karyopherin alpha 1 is a putative substrate of the RAG1 ubiquitin ligase. *Molecular Immunology*. **46**,pp.1319-1325.

Simkus, C., Bhattacharyya, A., Zhou, M., Veenstra, T.D. and Jones, J.M. 2009b. Correlation between recombinase activating gene 1 ubiquitin ligase activity and V(D)J recombination. *Immunology*. **128**,pp.206-217.

Simon, M.D., Chu, F., Racki, L.R., de la Cruz, C.C., Burlingame, A.L., Panning, B., Narlikar, G.J. and Shokat, K.M. 2007. The Site-Specific Installation of Methyl-Lysine Analogs into Recombinant Histones. *Cell*. **128**,pp,1003-1012.

Skok, J.A., Gisler, R., Novatchkova, M., Farmer, D., Laat, W. de and Busslinger, M. 2007. Reversible contraction by looping of the Tcra and Tcrb loci in rearranging thymocytes. *Nature Immunology*. **8**,pp.378-387.

Sokalingam, S., Raghunathan, G., Soundrarajan, N., Lee, S.G. 2012. A study on the effect of surface lysine to arginine mutagenesis on protein stability and structure using green fluorescent protein. *PLoS One.* **7**:e40410.

Sollbach, A.E. and Wu, G.E. 1995. Inversions produced during V(D)J rearrangement at IgH, the immunoglobulin heavy-chain locus. *Molecular Cell Biology.* **15**,pp.671-681.

Sonoda, E., Pewzner-Jung, Y., Schwers, S., Taki, S., Jung, S., Eilat, D. and Rajewsky, K. 1997. B Cell Development under the Condition of Allelic Inclusion. *Immunity*. **6**,pp.225–233.

Spanopoulou, E., Cortes, P., Shih, C., Huang, C.-M., Silver, D.P., Svec, P. and Baltimore, D. 1995. Localization, interaction, and RNA binding properties of the V(D)J recombination-activating proteins RAG1 and RAG2. *Immunity*. **3**,pp.715-726.

Stanhope-Baker, P., Hudson, K.M., Shaffer, A.L., Constantinescu, A. and Schlissel, M.S. 1996. Cell Type–Specific Chromatin Structure Determines the Targeting of V(D)J Recombinase Activity In Vitro. *Cell*. **85**,pp.887–897.

Stark, H. 2010. GraFix: stabilization of fragile macromolecular complexes for single particle cryo-EM. *Methods in Enzymology.* **481**,pp.109-126.

Stavnezer, J., Guikema, J.E.J. and Schrader, C.E. 2008. Mechanism and Regulation of Class Switch Recombination. *Annual Review of Immunology.* **26**,pp.261–292.

Steen, S.B., Gomelsky, L., Speidel, S.L. and Roth, D.B. 1997. Initiation of V(D)J recombination in vivo: role of recombination signal sequences in formation of single and paired double−strand breaks. *The EMBO Journal*. **16**,pp.2656–2664.

Steen, S.B., Han, J.O., Mundy, C., Oettinger, M.A. and Roth, D.B. 1999. Roles of the 'dispensable' portions of RAG-1 and RAG-2 in V(D)J recombination. *Molecular and Cellular Biology.* **19**,pp.3010-3017.

Stott, K., Watson, M., Bostock, M.J., Mortensen, S.A., Travers, A., Grasser, K.D. and Thomas, J.O. 2014. Structural insights into the mechanism of negative regulation of single-box high mobility group proteins by the acidic tail domain. *Journal of Biological Chemistry*. **289**,pp.29817–29826.

Swanson, P.C. 2002. A RAG-1/RAG-2 Tetramer Supports 12/23-Regulated Synapsis, Cleavage, and Transposition of V(D)J Recombination Signals. *Molecular and Cellular Biology.* **22**,pp.7790–7801.

Swanson, P.C. 2004. The bounty of RAGs: recombination signal complexes and reaction outcomes. *Immunological Reviews*. **200**,pp.90-114.

Swanson, P.C. and Desiderio, S. 1999. RAG-2 Promotes Heptamer Occupancy by RAG-1 in the Assembly of a V(D)J Initiation Complex. *Molecular and Cellular Biology.* **19**,p.3674-3683.

Štros, M. 2010. HMGB proteins: Interactions with DNA and chromatin. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*. **1799**,pp.101-113.

Takata, M., Sasaki, M.S., Sonoda, E., Morrison, C., Hashimoto, M., Utsumi, H., Yamaguchi Iwai, Y., Shinohara, A. and Takeda, S. 1998. Homologous recombination and non−homologous end−joining pathways of DNA double−strand break repair have overlapping roles in the maintenance of chromosomal integrity in vertebrate cells. *The EMBO Journal*. **17**,pp.5497-5508.

Tchernaenko, V., Halvorson, H.R., Kashlev, M. and Lutter, L.C. 2008. DNA Bubble Formation in Transcription Initiation. *Biochemistry*. **47**,pp.1871–1884.

Teitell, M.A. and Pandolfi, P.P. 2009. Molecular Genetics of Acute Lymphoblastic Leukemia. *Pathology: Mechanisms of Disease*. **4**,pp.175–198.

Thomas, J.O. and Butler, P.J. 1977. Characterization of the octamer of histones free in solution. *Journal of Molecular Biology.* **116**,pp.769-781.

Thomas, J.O. and Travers, A.A. 2001. HMG1 and 2, and related 'architectural' DNA-binding proteins. *Trends in biochemical sciences*. **26**,pp.167-174.

Thompson, C.B. 1995. New insights into V(D)J recombination and its role in the evolution of the immune system. *Immunity*. **3**,pp.531-539.

Thwaites, D.T., Carter, C., Lawless, D., Savic, S. and Boyes, J.M. 2019. A novel *RAG1* mutation reveals a critical in vivo role for HMGB1/2 during V(D)J recombination. *Blood*. **133**,pp.820-829.

Tiller, T., Tsuiji, M., Yurasov, S., Velinzon, K., Nussenzweig, M.C. and Wardemann, H. 2007. Autoreactivity in Human IgG+ Memory B Cells. *Immunity*. **26**,pp.205–213.

Tirosh, I., Yamazaki, Y., Frugoni, F., Ververs, F.A., Allenspach, E.J., Zhang, Y., Burns, S., Al-Herz, W., Noroski, L., Walter, J.E., Gennery, A.R., Burg, M.V.D., Notarangelo, L.D. and Lee, Y.N. 2018. Recombination activity of human recombination-activating

gene 2 (RAG2) mutations and correlation with clinical phenotype. *Journal of Allergy and Clinical Immunology*. **143**,pp.726–735.

Tolhuis, B., Palstra, R.-J., Splinter, E., Grosveld, F. and Laat, W. de 2002. Looping and Interaction between Hypersensitive Sites in the Active β-globin Locus. *Molecular Cell*. **10**,pp.1453–1465.

Tonegawa, S. 1983. Somatic generation of antibody diversity. *Nature*. **302**,pp.575-581.

Tsai, A.G. and Lieber, M.R. 2010. Mechanisms of chromosomal rearrangement in the human genome. *BMC Genomics*. **11**,p.S1.

Tycko, B. and Sklar, J. 1990. Chromosomal translocations in lymphoid neoplasia: a reappraisal of the recombinase model. *Cancer cells (Cold Spring Harbor, N.Y. : 1989)*. **2**,pp.1–8.

VanDyk, L.F., Wise, T.W., Moore, B.B. and Meek, K. 1996. Immunoglobulin D(H) recombination signal sequence targeting: effect of D(H) coding and flanking regions and recombination partner. *Journal of Immunology*. **157**,pp.4005-4015.

van Gent, D.C., Hiom, K., Paull, T.T. and Gellert, M. 1997. Stimulation of V(D)J cleavage by high mobility group proteins. *The EMBO Journal*. **16**,pp.2665-2670.

van Gent, D.C., Mizuuchi, K. and Gellert, M. 1996. Similarities Between Initiation of V(D)J Recombination and Retroviral Integration. *Science*. **271**,pp.1592-1594.

Vanasse, G.J., Concannon, P. and Willerford, D.M. 1999. Regulated genomic instability and neoplasia in the lymphoid lineage. *Blood*. **94**,pp.3997–4010.

Vanura, K., Montpellier, B., Le, T., Spicuglia, S., Navarro, J.-M., Cabaud, O., Roulland, S., Vachez, E., Prinz, I., Ferrier, P., Marculescu, R., Jäger, U. and Nadel, B. 2007. In Vivo Reinsertion of Excised Episomes by the V(D)J Recombinase: A Potential Threat to Genomic Stability. *PLoS Biology*. **5**:e43.

Vaughan, E.E., DeGiulio, J.V. and Dean, D.A. 2006. Intracellular trafficking of plasmids for gene therapy: mechanisms of cytoplasmic movement and nuclear import. *Current gene therapy*. **6**,pp.671-681.

Vettermann, C. and Schlissel, M.S. 2010. Allelic exclusion of immunoglobulin genes: models and mechanisms. *Immunological Reviews*. **237**,pp.22–42.

Vignali, D.A.A., Collison, L.W. and Workman, C.J. 2008. How regulatory T cells work. *Nature Reviews Immunology*. **8**,pp.523-532.

Villa, A., Santagata, S., Bozzi, F., Giliani, S., Frattini, A., Imberti, L., Gatta, L.B., Ochs, H.D., Schwarz, K., Notarangelo, L.D., Vezzoni, P. and Spanopoulou, E. 1998. Partial V(D)J Recombination Activity Leads to Omenn Syndrome. *Cell*. **93**,pp.885–896.

Wabl, M. and Steinberg, C. 1992. Allelic exclusion model questioned. *Nature*. **359**,pp. 370–371.

Ward, A., Kumari, G., Sen, R. and Desiderio, S. 2018. The RAG-2 Inhibitory Domain Gates Accessibility Of The V(D)J Recombinase To Chromatin. *Molecular and Cellular Biology*. **38**,pp.159–18.

Warren, C. and Shechter, D. 2017. Fly Fishing for Histones Catch and Release by Histone Chaperone Intrinsically Disordered Regions and Acidic Stretches. *Journal of Molecular Biology.* **429**,pp.2401-2426.

Wang, Y., Prosen, D.E., Mei, L., Sullivan, J.C., Finney, M. and Vander Horn, P.B. 2004. A novel strategy to engineer DNA polymerases for enhanced processivity and improved performance *in vitro*. *Nucleic Acids Research.* **32**,pp,1197-1207.

Webb, M., Payet, D., Lee, K.B., Travers, A.A. and Thomas, J.O. 2001. Structural requirements for cooperative binding of HMG1 to DNA minicircles. *Journal of Molecular Biology*. **309**.pp,79-88.

West, K.L., Singha, N.C., Ioannes, P.D., Lacomis, L., Erdjument-Bromage, H., Tempst, P. and Cortes, P. 2005. A Direct Interaction between the RAG2 C Terminus and the Core Histones Is Required for Efficient V(D)J Recombination. *Immunity*. **23**,pp. 203-212.

Willmann, K.L., Milosevic, S., Pauklin, S., Schmitz, K.-M., Rangam, G., Simon, M.T., Maslen, S., Skehel, M., Robert, I., Heyer, V., Schiavo, E., Reina-San-Martin, B. and Petersen-Mahrt, S.K. 2012. A role for the RNA pol II–associated PAF complex in AID-induced immune diversification. *The Journal of Experimental Medicine*. **209**,pp.2099–2111.

Wilson, A., Held, W. and MacDonald, H.R. 1994. Two waves of recombinase gene expression in developing thymocytes. *The Journal of Experimental Medicine*. **179**,pp.1355–1360.

Wilson, D.R., Norton, D.D. and Fugmann, S.D. 2008. The PHD domain of the sea urchin RAG2 homolog, SpRAG2L, recognizes dimethylated lysine 4 in histone H3 tails. *Developmental & Comparative Immunology*. **32**,pp.1221-1230.

Wittschieben, B.Ø., Otero, G., Bizemont, T. de, Fellows, J., Erdjument-Bromage, H., Ohba, R., Li, Y., Allis, C.D., Tempst, P. and Svejstrup, J.Q. 1999. A Novel Histone Acetyltransferase Is an Integral Subunit of Elongating RNA Polymerase II Holoenzyme. *Molecular Cell*. **4**,pp.123–128.

Wong, S.-Y., Lu, C.P. and Roth, D.B. 2008. A RAG1 Mutation Found in Omenn Syndrome Causes Coding Flank Hypersensitivity: A Novel Mechanism for Antigen Receptor Repertoire Restriction. *Journal of Immunology.* **181**,pp.4124–4130.

Woof, J.M. and Russell, M.W. 2011. Structure and function relationships in IgA. *Mucosal Immunology*. **4**,pp.590-597.

Xu, C.-R. and Feeney, A.J. 2009. The Epigenetic Profile of Ig Genes Is Dynamically Regulated during B Cell Differentiation and Is Modulated by Pre-B Cell Receptor Signaling. *Journal of Immunology.* **182**,pp.1362–1369.

Xu, Z., Fulop, Z., Wu, G., Pone, E.J., Zhang, J., Mai, T., Thomas, L.M., Al-Qahtani, A., White, C.A., Park, S.-R., Steinacker, P., Li, Z., III, J.Y., Herron, B., Otto, M., Zan, H., Fu, H. and Casali, P. 2010. 14-3-3 adaptor proteins recruit AID to 5′-AGCT-3′–rich switch regions for class switch recombination. *Nature structural & molecular biology*. **17**,pp.1124-1135.

Xu, Z., Zan, H., Pone, E.J., Mai, T. and Casali, P. 2012. Immunoglobulin class-switch DNA recombination: induction, targeting and beyond. *Nature Reviews Immunology*. **12**,pp.517-531.

Yamane, H. and Paul, W.E. 2013. Early signaling events that underlie fate decisions of naive CD4+ T cells toward distinct T-helper cell subsets. *Immunological Reviews*. **252**,pp.12-23.

Yancopoulos, G.D. and Alt, F.W. 1985. Developmentally controlled and tissue-specific expression of unrearranged VH gene segments. *Cell*. **40**,pp.271–281.

Yeap, L.-S., Hwang, J.K., Du, Z., Meyers, R.M., Meng, F.-L., Jakubauskaitė, A., Liu, M., Mani, V., Neuberg, D., Kepler, T.B., Wang, J.H. and Alt, F.W. 2015. Sequence-Intrinsic Mechanisms that Target AID Mutational Outcomes on Antibody Genes. *Cell*. **163**,pp.1124–1137.

Yin, F.F., Bailey, S., Innis, C.A., Ciubotaru, M., Kamtekar, S., Steitz, T.A. and Schatz, D.G. 2009. Structure of the RAG1 nonamer binding domain with DNA reveals a dimer that mediates DNA synapsis. *Nature structural & molecular biology*. **16**,pp. 499-508.

Yurchenko, V., Xue, Z. and Sadofsky, M. 2003. The RAG1 N-terminal domain is an E3 ubiquitin ligase. *Genes & Development*. **17**,pp.581-585.

Zagelbaum, J., Shimazaki, N., Esguerra, Z.A., Watanabe, G., Lieber, M.R. and Rothenberg, E. 2016. Real-time analysis of RAG complex activity in V(D)J recombination. *PNAS*. **113**,pp.11853–11858.