Cut-and-Run: A New Mechanism by which V(D)J Recombination Could Trigger Leukaemia and Lymphoma

Christopher Michael Kirkham

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

September 2014

The candidate confirms that the work submitted is his/her 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 1 contains content adapted from a jointly authored review: "Kirkham CM, Scott JN, Boyes JM & Bevington S. (2014) The Molecular Basis of B Cell Development and the Role of Deregulated Transcription and Epigenetics in Leukaemia and Lymphoma. In: Bonifer C, Cockerill P, eds. *Transcriptional and Epigenetic Mechanisms Regulating Normal and Aberrant Blood Cell Development.* London: Springer, pp. 331-363." CM Kirkham wrote sections 13.7-13.15 and 13.17, made the figures and edited the manuscript; JN Scott wrote section 13.16 and made the figures; JM Boyes edited the manuscript, and S Bevington wrote sections 13.1-13.6.

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

The right of Christopher Michael Kirkham to be identified as Author of this work has been asserted by him in accordance with the Copyright, Designs and Patents Act 1988.

Acknowledgements

First and foremost I thank Dr Joan Boyes for her guidance and support, and for always making sure I saw the bigger picture. I must also thank Joan for the many motivational chats that kept spirits high, and for making sure my blood sugar levels were always too high. The world will never know a better-stocked snack drawer. I am especially grateful for the speed with which I received feedback on thesis drafts; given the characteristically "relaxed" manner in which this thesis was written, it would probably have gone unedited by almost any other supervisor.

I thank Dr Roman Tuma for his guidance and support, and for helpful discussions about RAG biochemistry and the finer points of brewing ale. I would like to thank all members of the Boyes, Tuma and Stockley labs for cake, helpful discussions, and for being wonderful people to work with. In particular, I thank Dr Mohamed Kamal Saleh for getting the qPCR assay off the ground in the early days, Dr Adam Kupinski for allowing me to include his qPCR data in Chapter 5, James Scott for his mouse DNA samples and for making sure the coffee supply never ran dry, Diana Monteiro for lending me her desk chair when I was writing up, and Dr Olya Platonova for her expert ÄKTA tutelage, although to this day it is a miracle if I finish a purification without it misbehaving. I also thank Dr Bernard Moss for the pTM1 vector, Prof Geoffrey Smith for the vaccinia virus and Hu TK⁻ cells, Dr John Barr for the vTF7-3 virus, Prof Marco Bianci for the pETM11-HMGB1 expression vector, Dr Chris Thomas for the pET-MBPL vector, Dr Amit Sharma for the monomeric streptavidin, and I thank the RAG proteins for being constantly capricious and never giving me an easy ride.

I thank my parents and my brothers for their constant love and support, and no, I still haven't cured it yet. I thank Griff Tilley, Darren Machin, Chris McKee, Matt Balmforth and Kat Horner for being great friends and keeping me sane. I am hugely indebted to Darren Machin and Grace Roberts for putting a roof over my head and feeding me whilst I wrote up, and to my father, John, for keeping me afloat for the last six months. Finally, I thank my long-suffering girlfriend, Charlotte, for her unending love and support, for (mostly) putting up with my unorthodox working hours, and for sticking with a long-distance relationship that hasn't always been easy. Who knows, maybe one day we will both be Dr C Kirkham.

Abstract

V(D)J recombination generates a hugely diverse repertoire of antigen receptors through the random joining of antigen receptor gene segments. Recombination is catalysed by the RAG proteins which bind to and cleave recombination signal sequences (RSSs) that flank each coding segment. Since this reaction involves the introduction of double-strand breaks in the genome, it poses a major risk to genome stability. Indeed, 35-40% of lymphoid cancers have chromosome translocations that bear the hallmarks of mistakes in the production of antigen receptor genes. Recently, V(D)J recombination was proposed to trigger genome instability by a new mechanism called reintegration, where the by-product of recombination, the excised signal circle (ESC), is reinserted back into the genome at an RSS. I examined the cleavage step of reintegration and found that an ESC stimulates cleavage of an RSS, but not vice versa. For reintegration to occur, simultaneous cleavage at an RSS and ESC is required, so this finding suggests that reintegration may not occur as frequently as others have predicted. Instead, the ESCs could trigger double-strand breaks at RSSs and cryptic RSSs throughout the genome, whilst remaining intact after each cleavage event, in a proposed new mechanism called "cut-and-run". I then investigated the molecular basis of asymmetric cleavage, and found that it is most likely caused by RAG proteins binding to both sides of an ESC simultaneously and obscuring the cleavage site. Using transfected extrachromosomal substrates, I found that RSS-ESC cleavage is also asymmetric *in vivo*. Finally, analysis of an endogenous ESC revealed that it is largely intact in a pro-B cell line, consistent with the cut-and-run hypothesis. Therefore, cut-and-run could be a potent source of genomic instability in developing lymphocytes that creates broken ends for chromosome translocations.

Contents

List of Figures

List of Tables

Abbreviations

AID	Activation-induced cytidine deaminase
βME	β-mercaptoethanol
BHK-21	Baby hamster kidney 21 cells
bp	Base pair
BrdU	Bromodeoxyuridine
BSA	Bovine serum albumin
CE	Coding end
CDR	Complementarity determining region
CSR	Class switch recombination
CTCF	CCCTC-binding factor
CV	Column volume
dH$_2$O	Deionised water
DIP	Defective interfering particle
DMEM	Dulbecco's modified Eagle's medium
DMSO	Dimethyl sulphoxide
DNA	Deoxyribonucleic acid
DNA-PKcs	DNA-dependent protein kinase catalytic subunit
DSB	Double strand break
DTT	Dithiothreitol
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
EMSA	Electrophoretic mobility shift assay
ESC	Excised signal circle
FACS	Fluorescence-activated cell sorting
FCS	Foetal calf serum
H3K4me3	Trimethylated lysine 4 of histone H3
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HMGB1	High mobility group protein B1
HSC	Haematopoietic stem cell
HSC1/2	HMGB1 single-RSS containing complex 1 or 2
HuTK⁻ 143B	Human thymidine kinase-deficient 143B cells
Ig	Immunoglobulin
IRF4	Interferon regulatory factor 4
kDa	Kilodalton
KOAc	Potassium acetate

LM-PCR	Ligation-mediated PCR
MBP	Maltose binding protein
MCS	Multiple cloning site
MEME	Minimum essential medium Eagle's
MOPS	3-(N-morpholino)propanesulfonic acid
NaOAc	Sodium acetate
NBD	Nonamer-binding domain
NHEJ	Non-homologous end joining
Ni-NTA	Nickel-nitrilotriacetic acid
NS	Nonspecific
NTG	Nontransgenic
PAGE	Polyacrylamide gel electrophoresis
PC	Paired complex
PCC	Post-cleavage complex
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PEI	Polyethyleneimine
P_i	Inorganic phosphate
PMSF	Phenylmethylsulphonyl fluoride
PNK	Polynucleotide kinase
qPCR	Quantitative polymerase chain reaction
RAG	Recombination activating gene
RSS	Recombination signal sequence
RT	Room temperature
SC	Single RSS-containing complex
SDS	Sodium dodecyl sulphate
SE	Signal end
SEC	Signal end complex
SFM	Serum-free medium
SHM	Somatic hypermutation
SJ	Signal joint
smTIRFm	Single molecule total internal reflection fluorescence microscopy
ssDNA	Single-stranded deoxyribonucleic acid
TAE	Tris-acetate-EDTA
TCR	T cell receptor
TdT	Terminal deoxynucleotidyl transferase
TE	Tris-EDTA

TK	Thymidine kinase
V(D)J	Variable (diversity) joining
XRCC4	X-ray repair cross complementing protein

Chapter 1

Introduction

A)   The adaptive immune system

1.1   Innate and adaptive immunity

All multicellular organisms are under constant attack from a plethora of different pathogens, including viruses, bacteria, fungi and parasites. To defend themselves from such attacks, vertebrates have evolved a complex immune system capable of recognising and fighting off these pathogens. The first line of defence against infection is the innate immune system. Although its complexity can differ greatly between species, a form of innate immune system can be found in all classes of plant and animal life, including vertebrates and invertebrates. The innate immune system works by recognising common features of pathogens and destroys them via non-specific mechanisms. For example, specialised cells such as macrophages recognise pathogens and infected cells and ingest them by phagocytosis, breaking them down into harmless products.

Whilst the innate immune system can be very effective, it offers no protection against pathogens that are capable of evolving and adapting to evade the generic, unchanging defence mechanisms of the innate immune system. Thus, an adaptive immune system has evolved in vertebrates to counteract this, allowing vertebrates to survive and thrive in the face of rapidly changing pathogens that the innate immune system cannot recognise. The importance of adaptive immunity is evident in children that are born without a fully functioning adaptive immune system: their chances of survival are vanishingly small unless extraordinary measures are taken to shield them from all invading pathogens (Alberts et al., 2002).

1.2   B and T lymphocytes

Central to the adaptive immune system are the B and T lymphocytes (more commonly called B and T cells), which express receptors on their surface that recognise a specific part of a pathogen, called an antigen. In B cells these receptors are antibodies, also referred to as immunoglobulins (Igs), and in T cells they are known as T cell receptors (TCRs).

When a B cell successfully recognises a cognate antigen via its cell-surface Ig, it proliferates to generate many identical B cells that secrete huge numbers of the same

antibody molecule in a soluble form, which circulate the body and bind to the same antigen that originally stimulated its production. These extracellular antibodies can neutralise viruses and toxins by completely coating them and preventing them from infecting other cells. In addition, antibody binding targets the pathogen for destruction by other cells of the immune system such as macrophages and natural killer cells. The interaction between an antibody and an antigen is highly specific. Remarkably, antibodies are capable of distinguishing between proteins that differ by as little as one amino acid.

After binding to a cognate antigen, the response of a T cell depends on whether the cell expresses the CD4 or CD8 cell surface marker. CD8+ T cells, also known as cytotoxic T cells, are capable of inducing the death of infected cells by releasing perforin proteins that form lethal pores in the membrane of the infected cell. The majority of CD4+ T cells, also called T helper cells, do not have the capacity to directly induce apoptosis. Instead, T helper cells expand the immune response by releasing cytokines that direct the functions of other cells of the immune system. However, the recent identification of a subset of CD4+ T cells that are directly capable of inducing apoptosis shows that cytotoxic effects are not strictly limited to CD8+ T cells (Cheroutre and Husain, 2013; Mucida et al., 2013).

T helper cells are classically divided into T helper 1 (T_H1) cells and T helper 2 (T_H2) cells. T_H1 cells respond to infections by intracellular bacteria and protozoa by releasing interferon-γ to recruit macrophages that engulf the pathogen, whereas T_H2 cells are mainly involved in the response to extracellular parasites such as helminths (Murphy et al., 2007). More recently, the subsets of T helper cells have been expanded to include T_H17, T follicular helper (T_{FH}), T regulatory (T_{reg}) and T_H9 cells (Yamane and Paul, 2013).

T_H17 cells are defined by the expression of IL-17, and help to fight bacteria and fungi through recruitment of neutrophils, although aberrant T_H17 regulation plays a role in autoimmune and inflammatory disorders (Korn et al., 2009). T_{FH} cells interact with B cells and help direct the production of antibodies (Ma and Deenick, 2014) and Th9 cells are thought to be important in the intestinal response to parasitic worm infection (Kaplan, 2013). Unlike all other T helper cell subtypes, which activate the functions of other immune cells, T_{reg} cells function by attenuating the activity of other cells of the immune system. This functionality includes preventing the recognition of self antigens, which is crucial in preventing autoimmune diseases (Vignali et al., 2008). The fact that

each T helper subset is equipped with a specific function allows the immune system to respond effectively to a variety of different threats.

After a pathogen has been cleared by the immune system, most of the B and T cells that were made during the immune response will die, but a small number remain as memory B and T cells. In the event of re-infection by the same pathogen, these memory B and T cells will recognise the pathogen and rapidly differentiate and proliferate into effector cells, dramatically reducing the response time of the immune system. It is this immunological memory that forms the basis of vaccinations (Murphy et al., 2007).

1.3   Immunoglobulin and TCR structure

Immunoglobulins are composed of four polypeptide chains linked by disulphide bonds: two identical heavy (IgH) chains and two identical light (IgL) chains (Fig. 1.1A). Igs can be divided into two distinct domains: a variable region that binds an antigen and a constant region that provides a scaffold for the variable region and also signals an immune response upon antigen binding. The variable region contains three hypervariable or complementarity-determining regions (CDRs), the sequence of which varies greatly between B cells, thus enabling recognition of a wide range of pathogens. Between the three CDRs are framework regions, which do not vary in sequence but provide a necessary scaffold for the CDRs.
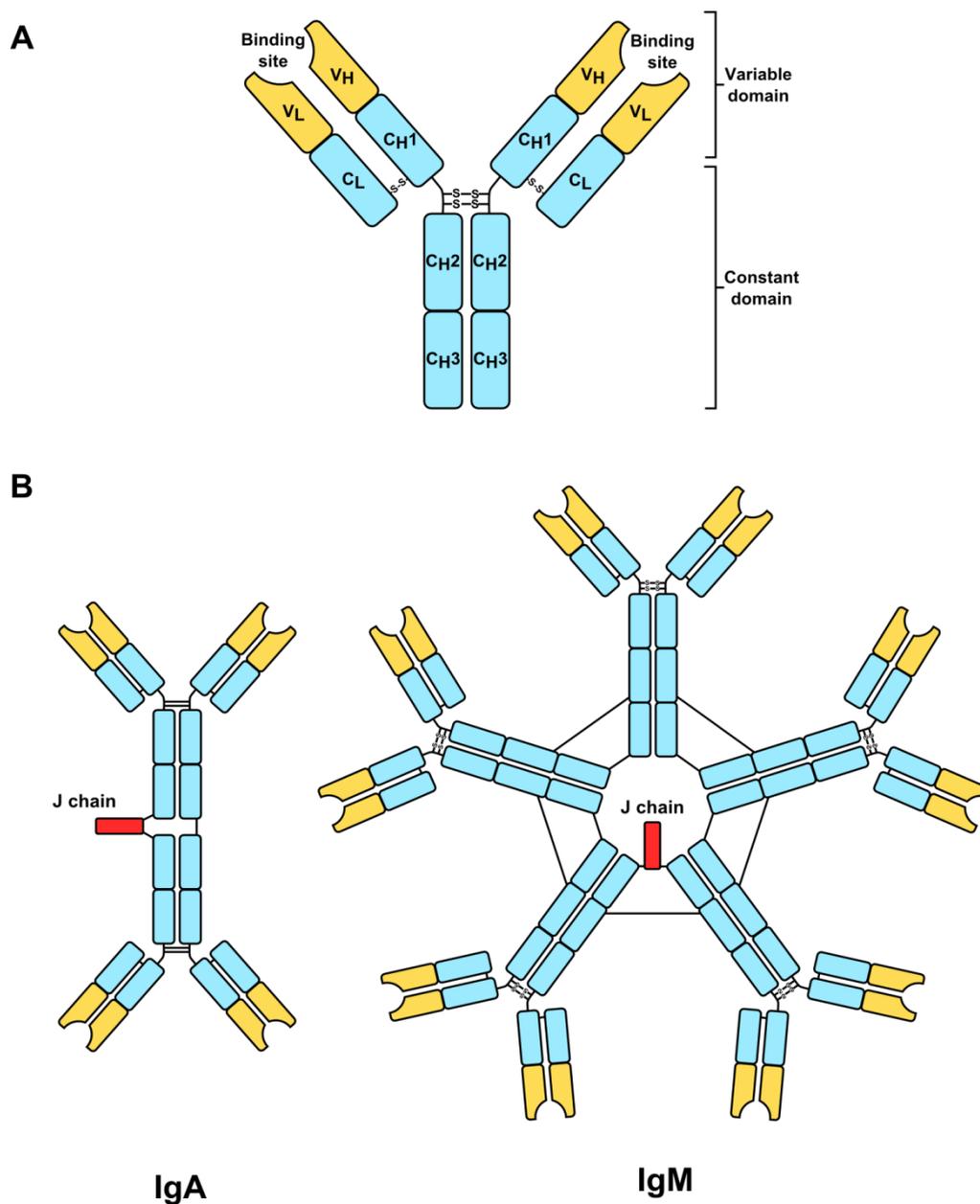
Figure 1.1 – Immunoglobulin structure. **(A)** Antibodies consist of four polypeptide chains, two heavy and two light, connected by disulphide bonds. Each chain contains variable and constant regions. The variable regions of the light and heavy chains form the antigen binding site, and the constant region of the heavy chain dictates the effector function of the antibody. The schematic depicts the domain architecture of a soluble IgA or IgG antibody, which contain three tandem constant domains and a flexible hinge region. IgD also contains three constant domans, but is always membrane-bound, and IgM and IgE both contain four tandem constant domains. **(B)** IgA and IgM antibodies form dimers and pentamers, respectively, of the basic antibody structure shown in part (A) via cross-linking with an additional J chain.

Humans and mice contain two different types of light chain: kappa (Igκ) and lambda (Igλ). The two light chains of each antibody are always identical, being either both κ or λ, and never one of each. Each light chain is made of one variable region and one constant region. There are also five different types of heavy chain constant region that define the five classes of immunoglobulin: IgA, IgD, IgE, IgG, and IgM, with each constant region imparting a different functional effect on the antibody. Before binding its cognate antigen all antibodies are of the membrane-bound IgD or IgM subtype. Antigen binding, followed by T_H2 cell interaction, triggers switching to one of the other classes. The constant region of an antibody also affects its oligomerisation state: IgD, IgE and IgG are monomers, IgA is a dimer and IgM forms a pentamer (Fig. 1.1B). Pentameric IgM and IgA contain cysteine residues in the C-terminus of their constant regions that form disulphide bridges with an additional polypeptide chain called the joining (J) chain (not to be confused with the J gene segment). Crosslinking with the J chain promotes multimerisation into pentamers (IgM) and dimers (IgA), which bestows important properties upon the Ig molecules (Mestecky et al., 1971; Murphy et al., 2007). In the case of IgA, dimerisation is essential for secretion into the mucosal linings, acting as a first line of defence against pathogens that enter through the mucous membrane. Multimerisation of IgM creates many more antigen binding sites per molecule, thus increasing the avidity (binding strength) of the antibody, and enables neutralisation of bacteria and viruses through agglutination (Johansen et al., 2000).

TCRs consist of two polypeptides, either an α and a β chain in αβ T cell , or a δ and a γ chain in δγ T cells (Fig. 1.2). 95% of T cells in the human body express the αβ form, whilst the other 5% express the δγ form (Holtmeier and Kabelitz, 2005). Unlike antibodies, TCRs are always membrane-bound, and unlike antibodies, which are capable of recognising any form of organic antigen (protein, carbohydrate, lipid, etc.), TCRs can only recognise small peptide antigens that are presented on the surface of other cells via the major histocompatibility complex (MHC) proteins. Furthermore, the affinity and specificity of TCR-antigen interactions is usually much lower than antibody-antigen interactions. T cells that express the δγ TCR are found mostly in the intestinal lining and play an important role in linking the adaptive and innate immune systems (Holtmeier and Kabelitz, 2005; Born et al., 2006). Like antibodies, both chains of the TCR are joined by a disulphide bond and contain a variable domain and a constant domain, followed by a transmembrane domain and a short cytoplasmic tail (Fig. 1.2).
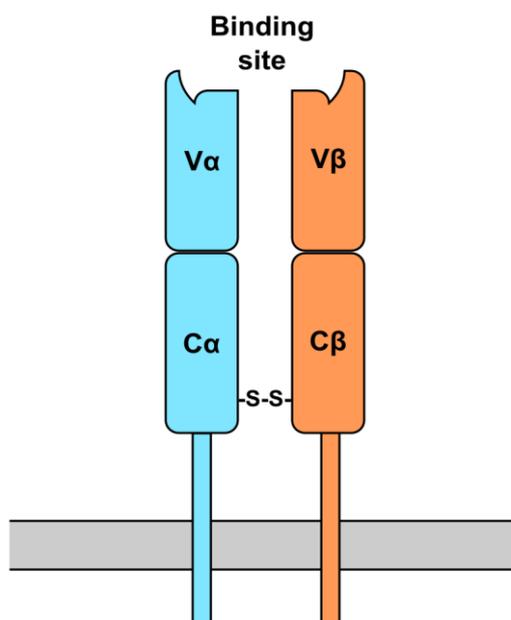
Figure 1.2 – T cell receptor structure. TCRs contain two disulphide-linked polypeptide chains, α and β (equivalent to light and heavy chains of an antibody, respectively), or δ and γ, and as such possess only one antigen binding site. TCRs are always membrane-bound. Each chain contains one variable domain and one constant domain, followed by a transmembrane domain and a short intracellular tail.

B)   V(D)J recombination

1.4   Generation of antigen receptor diversity

The number of unique antibodies and TCRs that can be found within the human body at any given time is upwards of one billion (Murphy et al., 2007). Given that there are only around 20-25,000 protein coding genes in the entire human genome, how is it possible to make such a vast repertoire of antibodies? The answer lies in the mixing and matching of different gene segments in the genomes of B and T cells, by the process of V(D)J recombination. The variable domains of antibodies and TCRs are encoded by multiple different variable (V), diversity (D) and joining (J) segments (Fig. 1.3A). During V(D)J recombination, one of each gene segment is chosen at random and recombined to form a complete variable domain coding exon of an antibody or TCR. It is the stochastic assembly of coding exons, coupled with the imprecise nature with which the gene segments are combined, that is partly responsible for the generation of a huge repertoire of antibodies and T cell receptors that are capable of recognising a vast range of pathogens. In B cells, after binding to an antigen the antibody-antigen interaction is fine-tuned by highly localised point mutations via the process of somatic hypermutation (SHM; Section 1.21). SHM does not take place in T cells, therefore V(D)J recombination is responsible for the entirety of TCR variation,

and also explains why TCR-antigen interactions are of a lower affinity than antibody-antigen interactions.

For example, the murine *IgH* locus contains several hundred V gene segments, 13 D gene segments and four J gene segments (Fig. 1.3A; Ichihara et al., 1988). The two light chain loci, Igλ and Igκ, only contain V and J gene segments, and as a result they are less diverse than the heavy chains. In addition to the different V, D and J gene segments, the *IgH* locus contains exons for the five different constant regions. D to J rearrangement always takes place first, followed by V to DJ rearrangement, then the recombined variable domain exon is joined with the constant region exons by RNA splicing (Fig. 1.3A).

Each V(D)J coding segment is flanked by one or two recombination signal sequences (RSSs), which are each made up of a conserved heptamer (5'-CACAGTG-3') and a conserved nonamer (5'-ACAAAAACC-3'; Fig. 1.3B). In a 12RSS, the two conserved elements are separated by a non-conserved spacer of 12 base-pairs, and in a 23RSS the spacer contains 23 non-conserved base-pairs. Efficient recombination only occurs between gene segments that are flanked by different length RSSs, a condition referred to as the 12/23 rule (Tonegawa, 1983). 12 and 23RSSs are arranged at the antigen receptor loci so that only productive combinations (such as D to J) are permitted, and non-productive rearrangements are prevented (such as joining a V to another V; Fig. 1.3A).

A

Germline Configuration

B
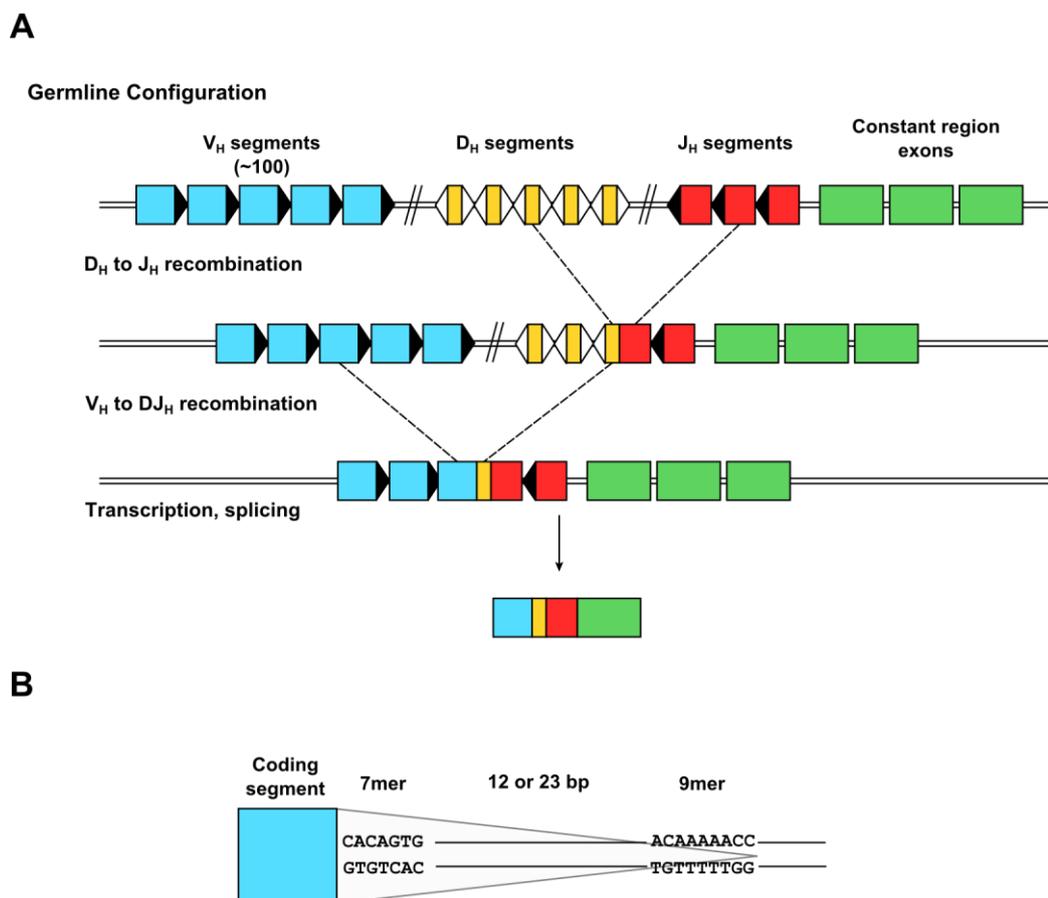
Figure 1.3 – Recombination of the *IgH* locus. (A) Schematic representation of an antigen receptor locus and the process of V(D)J recombination. Note that a real locus has many more copies of each gene segment than shown here. The germline configuration of a hypothetical heavy chain locus is shown at the top. D to J recombination always occurs first, followed by V to DJ recombination to generate a complete variable region coding exon. Splicing of the mRNA transcript then joins the variable and constant exons to create the full heavy chain coding sequence. Light chain loci (Igκ and Igλ) do not contain D segments. Open triangle = 12RSS; black triangle = 23RSS. **(B)** Recombination signal sequences (RSSs) are composed of a relatively conserved heptamer and nonamer sequence, separated by a non-conserved 12 or 23 base pair spacer. Efficient recombination only takes place between a 12 and 23 RSS.

1.5 Overview of V(D)J recombination

V(D)J recombination can be divided into two distinct phases: a DNA cleavage step and a joining step (Fig. 1.4A). The cleavage step is initiated by the lymphoid-specific recombinase activating gene (RAG) proteins 1 and 2 (McBlane et al., 1995). The RAG proteins bring a 12RSS and a 23RSS together in a synaptic complex and create a single-strand nick precisely at the heptamer/coding sequence boundary at each RSS, creating a free 3' hydroxyl group which then attacks the opposite DNA strand in a direct transesterification reaction {Fig. 1.4B`; \McBlane, 1995 #20}. This forms a double-strand break (DSB) at each RSS, where the coding end is a covalently sealed hairpin and the signal end terminates in a blunt 5'-phosphorylated end (Fugmann et al., 2000a;

Bassing et al., 2002; Gellert, 2002). Following cleavage, the RAG proteins continue to hold the coding and signal ends together in a stable postcleavage complex (Agrawal and Schatz, 1997; Hiom and Gellert, 1998; Grundy et al., 2009). Proteins of the non-homologous end joining (NHEJ) pathway are then recruited to repair the broken DNA ends.
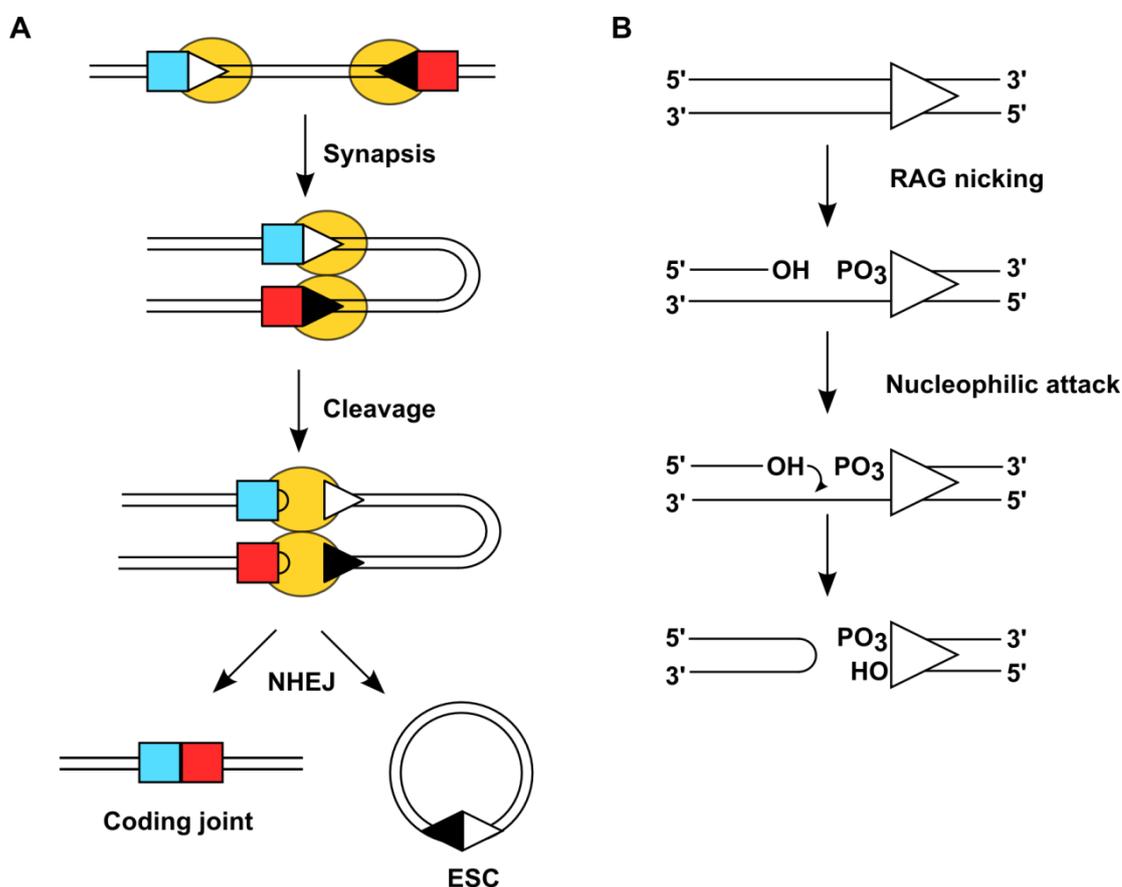
Figure 1.4 – The process of V(D)J recombination. (A) Schematic of V(D)J recombination. RAG proteins bind to a pair of RSSs, and bring them together in a synaptic complex. Double-strand breaks are made at RSS/coding segment boundary, which are then resolved via the NHEJ pathway to form a coding joint and signal joint. Blue and red squares are coding segments, triangles are RSSs, and circles are RAG proteins. **(B)** RAG proteins create a single-stranded nick precisely at the boundary of the RSS and the coding segment. This leaves a free 3'-OH group, which then attacks the opposing DNA strand via transesterification, creating a closed hairpin coding end and a blunt 5' phosphorylated signal end.

The repair reaction begins with a Ku70-Ku80 dimer binding to a coding or signal end (Fig. 1.5; Gu et al., 1997). Ku70-Ku80 then recruits the DNA-dependent protein kinase catalytic subunit (DNA-PKcs; Gottlieb and Jackson, 1993), which in turn recruits and activates Artemis. Artemis then re-opens the coding end hairpins, predominantly two nucleotides 3' of the hairpin tip, thus creating a four nucleotide overhang of palindromic (P) nucleotides (Fig. 1.6; Ma et al., 2002). Further variation is generated by the addition

of non-templated (N) nucleotides by terminal deoxynucleotidyl transferase (TdT), a lymphoid-specific polymerase, and any gaps are filled in by DNA Pol μ or DNA Pol λ (Fig. 1.6). Finally, the two modified coding ends are joined by DNA ligase IV:XRCC4 to form a coding joint. Unlike the coding ends, the two signal ends are joined together precisely, to form a signal joint (Fig. 1.5; Lieber, 2010). This creates a circular extrachromosomal piece of DNA called an excised signal circle (ESC), which is the by-product of V(D)J recombination. This by-product was previously thought to be inert, but as will be discussed below, the ESC can potentially contribute to genomic instability in developing lymphocytes by reintegrating back into the genome (Section 1.28). At some antigen receptor loci, the orientation of RSSs is such that rearrangement requires an inversional recombination, as opposed to deletional recombination. For example, half of the V gene segments at the human Igκ locus are arranged in the reverse orientation to the J gene segments (Zachau, 1993). This results in both the coding joint and the signal joint being retained in the genome, and therefore does not produce an ESC.
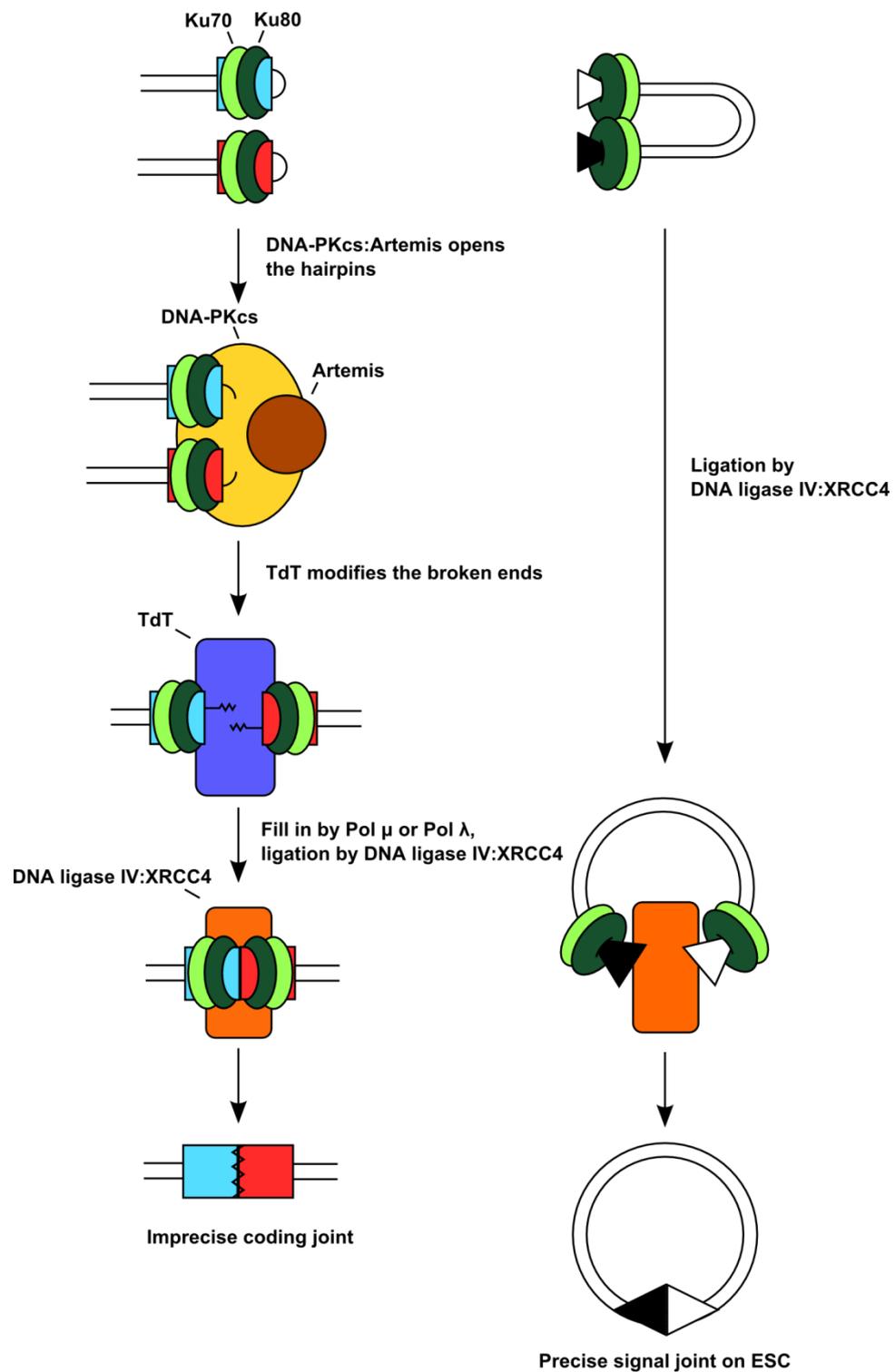
Figure 1.5 – Non-homologous end joining. A Ku70-Ku80 dimer binds each of the broken DNA ends. At the coding ends, Ku70-Ku80 recruits DNA-PKcs:Artemis, which opens the hairpinned ends. TdT, a lymphoid-specific polymerase, catalyses the addition of non-templated nucleotides to create further variation. Pol μ or Pol λ fill in any remaining gaps, and a complex of DNA ligase IV and XRCC4 ligate the modified ends to create a coding joint. Signal ends generally do not undergo further modification; instead, they are joined precisely by DNA ligase IV:XRCC4 to form a signal joint on an excised signal circle (ESC; adapted from Murphy et al., 2007).
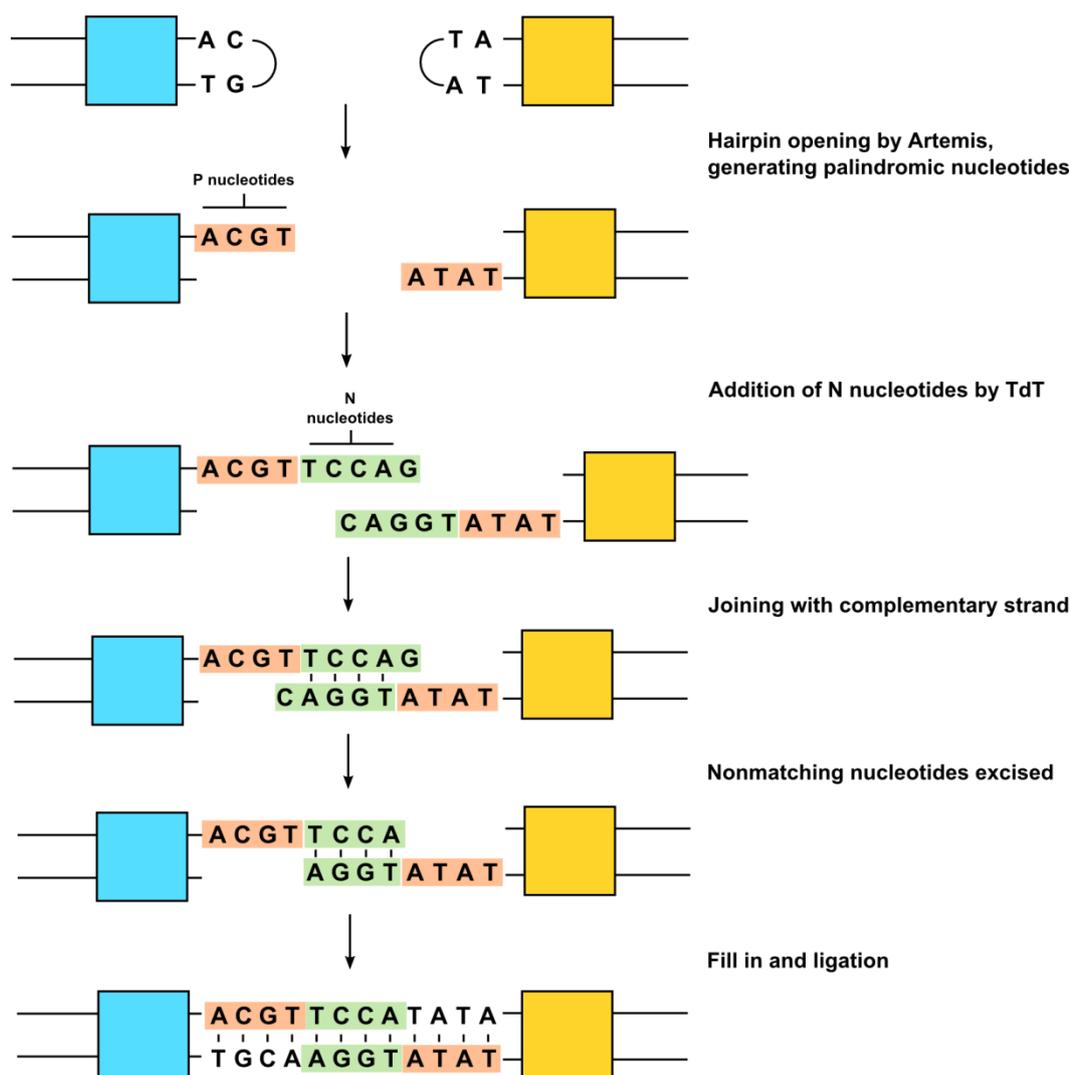
Figure 1.6 – Generation of junctional diversity. When DNA-PKcs:Artemis opens a hairpinned coding end at a site away from the apex, a short stretch of palindromic (P) nucleotides is created (indicated by orange boxes). Non-templated (N) nucleotides are added at random by the lymphoid-specific polymerase TdT (indicated by green boxes). The two strands then anneal by complementary base pairing, and any N nucleotides that are not complementary are excised by DNA-PKcs:Artemis. Note that the process of NHEJ is iterative and very flexible, and the addition and excision of nucleotides can occur in a different order than outlined above, or may not occur at all. If the P nucleotides share microhomology, the break can be repaired independently of base addition/resection. Therefore, the above scheme is only one example of how a break could be repaired.

1.6   The RAG proteins

The RAG1 and RAG2 proteins are the only proteins required to cleave RSSs and initiate recombination (McBlane et al., 1995), and are highly conserved in all jawed vertebrates. Moreover, they catalyse the only site-specific recombination reaction found in higher eukaryotes (Gellert, 2002). Mutations in either RAG1 or RAG2 are associated with a spectrum of immunodeficiencies, ranging from severe combined immunodeficiency (SCID) to the slightly less severe Omenn syndromes (Jones and Simkus, 2009). The *RAG1* and *RAG2* genes are located next to each other in the genome ~20 kb apart in opposite transcriptional directions, and are each encoded by a single exon (Oettinger et al., 1990). This close proximity led to the hypothesis that the RAG proteins evolved from a transposable element that integrated into the vertebrate genome. Indeed, it was later discovered that *in vitro* the RAG proteins can actually catalyse the transposition of DNA ends with a free 3'-OH group in a nonspecific manner (Section 1.28; Agrawal et al., 1998). Furthermore, the biochemical mechanism of the RAG proteins is very similar to transposases such as the retroviral integrases (van Gent et al., 1996b) and the bacterial Tn5 and Tn10 transposases, as will be highlighted below. The relation to retroviral integrases is further exemplified by the fact that synthetic inhibitors of the HIV integrase can also inhibit RAG activity (Melek et al., 2002).

Since the full-length RAG proteins are largely insoluble *in vitro*, the majority of work on RAG biochemistry has been done using truncated "core" versions of RAG1 (cRAG1, residues 384-1008 of 1040) and RAG2 (cRAG2, residues 1-384 of 527) that are more amenable to purification, and are the minimum regions of both proteins that are required for RSS cleavage *in vitro* and *in vivo* (Fig. 1.7; Sadofsky et al., 1993; Sadofsky et al., 1994). RAG1 is the larger of the two RAG proteins, and the core region of RAG1 contains three structurally distinct domains: a) an N-terminal nonamer binding domain that binds to the nonamer of 12 and 23RSSs (Difilippantonio et al., 1996; Spanopoulou et al., 1996), b) a central domain that displays nicking activity on single-stranded DNA, contacts the heptamer (Peak et al., 2003), and contains a zinc finger domain that mediates interactions with RAG2 (Aidinis et al., 2000), and c) a C-terminal domain that binds double-stranded DNA non-specifically and cooperatively (Arbuckle et al., 2001). The core region of RAG1 also contains a highly conserved catalytic triad of acidic residues called a DDE motif (Kim et al., 1999; Landree et al., 1999; Fugmann et al., 2000b), which co-ordinates a divalent metal cofactor in the RAG1 active site that is essential for cleavage. *In vivo*, this metal cation is most likely Mg^{2+}, since RAG activity in the presence of Mg^{2+} adheres to the 12/23 rule (van Gent et al., 1996c).

RAG1
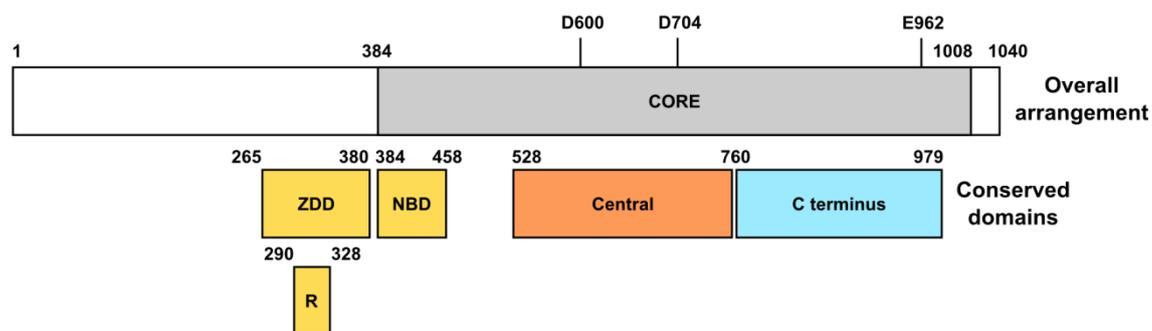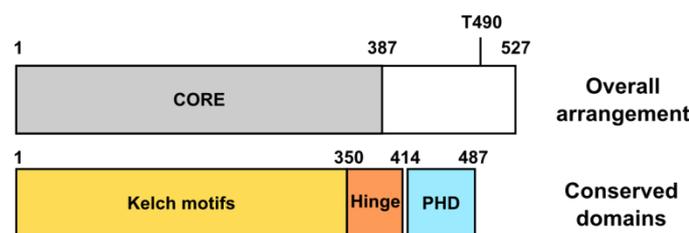
Figure 1.7 – The RAG proteins. The RAG proteins can be functionally divided into core (grey) and non-core domains (white). The conserved domains of both RAG1 and RAG2 are shown underneath the core/non-core division. RAG1 contains the major DNA binding domains (NBD, central and C-terminus domains) and the catalytic DDE motif, whilst RAG2 mainly extends RAG1 contacts with DNA. The C-terminus of RAG2 contains several highly conserved regions (shown in the lower half) responsible for regulating V(D)J recombination, including the inhibition of reintegration. All amino acid numberings are for the murine proteins. R = RING (Really Interesting New Gene) domain, ZDD = Zinc-binding Dimerisation Domain, NBD = Nonamer Binding Domain, PHD = Plant Homeodomain (Adapted from Swanson and Schatz, 2011).

Core RAG2 is comprised of six Kelch-like motifs that are thought to adopt a six-bladed propeller conformation, the sixth of which mediates contacts with RAG1 (Callebaut and Mornon, 1998; Aidinis et al., 2000), although no formal structure has been solved. In isolation, core RAG2 has no inherent DNA-binding activity (Difilippantonio et al., 1996; Spanopoulou et al., 1996; Swanson and Desiderio, 1998), so its role in the basic cleavage reaction is mainly to increase the affinity and specificity of RAG1 interactions with the RSS heptamer (Swanson, 2004).

1.7   RAG non-core regions

Although the core regions of RAG1 and RAG2 are necessary and sufficient for cleavage of RSSs *in vitro* and *in vivo*, it is clear that the non-core regions of both proteins play major roles in the efficiency, regulation and fidelity of V(D)J recombination *in vivo* (Jones and Simkus, 2009). The non-core N-terminus of RAG1 contains a basic motif that mediates interactions with karyopherin alpha 1 (KPNA1), a protein that

facilitates nuclear localisation of RAG1 (Spanopoulou et al., 1995), and a structurally characterised zinc-binding dimerisation domain (ZDD, residues 265-380), which contains a RING (Really Interesting New Gene) finger domain (residues 290-328) and a classical C_2H_2 zinc finger domain (355-376) that mediates dimerisation of RAG1 monomers (Rodgers et al., 1996; Bellon et al., 1997).

The RING domain is the prototypical member of a family of E3 ubiquitin (Ub) ligases, and its importance in V(D)J recombination is highlighted by a mutation (C328G) often found in Omenn syndrome patients (Simkus et al., 2007). In RAG1, the RING domain catalyses autoubiquitination of K233 (Jones and Gellert, 2003) and potentially ubiquitinates KPNA1, a protein involved in nuclear localisation (Simkus et al., 2009). Moreover, RAG1 has been shown to ubiquitinate histone H3 (Grazini et al., 2010), specifically the H3.3 variant (Jones et al., 2011). Recently, the RING domain of RAG1 was found to recruit a multisubunit E3 Ub ligase complex via Vpr binding protein (VprBP), and disruption of VprBP caused developmental arrest of B cells at the pro-B-to-pre-B transition. Analysis of recombination products in these cells showed a greater mutation frequency and an increased number of non-templated nucleotides (Kassmeier et al., 2012). Thus, whilst the function of RAG1-mediated ubiquitination remains unknown, it may play a role in regulating the joining of RAG cleavage products.

The non-core C-terminus of RAG2 (residues 388-527, Fig. 1.7) is also crucial for efficient, correct V(D)J recombination, as evidenced by RAG2 knock-in mice that are defective for V to DJ recombination (Akamatsu et al., 2003). The RAG2 C-terminus contains several highly conserved features: a flexible acidic hinge that has been implicated in maintaining the stability of the RAG postcleavage complex (West et al., 2005; Coussens et al., 2013), a plant homeodomain (PHD) finger that binds H3K4me3 and directs the RAG recombinase to active chromatin (Elkin et al., 2005; Matthews et al., 2007; Ramon-Maiques et al., 2007), and a threonine residue (T490) that is important for the cell cycle-regulated degradation of RAG2 at the G_1/S transition (Jiang et al., 2005). The far C-terminus of RAG2 (500-527) contains a string of conserved residues that potentially binds to phosphoinositides, although the function of this remains unclear (Elkin et al., 2005). The acidic hinge, PHD finger and T490 of the RAG2 C-terminus residue play crucial, distinct roles in maintaining accurate V(D)J recombination, as discussed in greater detail below.

1.8   HMG proteins in V(D)J recombination

The high mobility group (HMG) box proteins are a ubiquitous family of nonspecific DNA binding proteins with a wide variety of functions. First discovered over 40 years ago, their functions include acting as nonhistone chromosomal proteins that facilitate nucleoprotein complex assembly and extracellular cytokines in the inflammatory response (Malarkey and Churchill, 2012). The HMGB1/2 proteins contain two HMG box domains attached to a basic linker and an acidic tail, and can bind in the minor groove of DNA to induce a bend towards the major groove (Weir et al., 1993). Whilst the RAG proteins alone are capable of binding and cleaving RSSs, HMGB1 and HMGB2 function redundantly to stimulate cleavage at a 23RSS and promote formation of a 12/23 complex *in vitro* (Sawchuk et al., 1997; van Gent et al., 1997). HMGB1 is recruited cooperatively through interactions with RAG1 and DNA (Little et al., 2013), and has been shown to induce a large bend in the conserved spacer of the 23RSS (Ciubotaru et al., 2013). Since the conserved heptamer and nonamer are further apart in a 23RSS, the function of HMGB1 in V(D)J recombination is most likely to be bringing the heptamer and nonamer of a 23RSS closer together so that they can be bound more efficiently by the RAG proteins. Indeed, footprinting studies found that RAG-23RSS contacts at the heptamer more closely resemble RAG-12RSS contacts in the presence of HMGB1 (Yoshida et al., 2000; Swanson, 2002a). Removal of the acidic tail of HMGB1 promotes aggregation of the RAG-RSS complexes, indicating that HMGB1 might help to maintain the correct oligomeric state of these complexes (Bergeron et al., 2005). The HMGB proteins do not appear to be critical for V(D)J recombination *in vivo*, since HMGB1 knockout mice perform normal V(D)J recombination (Calogero et al., 1999). However, a more thorough investigation of the influence of HMGB1 in recombination *in vivo* is precluded by the fact that HMGB1 knockout mice perish from unrelated hypoglycaemia within 24 hours of birth (Calogero et al., 1999).

1.9   Assembly of RAG-RSS complexes

In the absence of RAG2, RAG1 forms a stable dimer (Swanson and Desiderio, 1999) that can bind an isolated RSS with a K_d of 29-41 nM (as measured by fluorescence anisotropy, Ciubotaru et al., 2003) or 92-114 nM, as measured by electrophoretic mobility shift assay (EMSA; Rodgers et al., 1999; Zhao et al., 2009). RAG1-RSS contacts are a combination of base-specific and nonspecific interactions with the DNA phosphate backbone, which centre on the nonamer region of the RSS and extend several base pairs into the spacer region (Swanson and Desiderio, 1998).

Together, RAG1 and RAG2 bind a single RSS with greater specificity and affinity, with a K_d of 12-25 nM, as measured by EMSA and fluorescence anisotropy (Zhao et al.,

2009; Wang et al., 2012). *In vitro*, the RAG proteins form two major complexes at a single RSS, called SC1 and SC2 (single RSS containing complex 1 and 2), as observed by EMSA (Swanson, 2002b). Both have been shown to contain two molecules of RAG1, and either one (SC1) or two (SC2) molecules of RAG2 that can freely re-assort (i.e. RAG2 molecules can exchange without the complex dissociating; Swanson, 2002b). Nicking of a 12 or 23RSS can occur in both SC1 and SC2, independent of synapsis with a partner RSS (Eastman and Schatz, 1997; Swanson, 2002b; Swanson, 2002a). When a second RSS is present, it is incorporated into SC2 to form a paired complex (PC) without the recruitment of any additional RAG molecules (Swanson, 2002b). Furthermore, the RAG1 active site is contributed *in trans*; that is, each RAG1 molecule binds to one RSS via the nonamer, but cleaves the heptamer of the other RSS (Swanson, 2001; Yin et al., 2009). In the presence of RAG2, RAG-DNA contacts are extended to include the entirety of the spacer region and the heptamer-coding boundary (Swanson and Desiderio, 1998). Studies from other groups indicate that a synaptic complex contains three or more RAG1 molecules (Mo et al., 1999; Mundy et al., 2002), but this is inconsistent with the studies noted above which show that a complex containing just two molecules of RAG1 can cleave a 12/23 RSS pair. Furthermore, complexes of this stoichiometry are unique to RAG proteins expressed with a GST tag, which is known to promote oligomerisation of its protein conjugates (Niedziela-Majka et al., 1998). Therefore, the general consensus in the literature, and the conclusion most consistent with the available data, is that the RAG-RSS PC contains two molecules of RAG1, two molecules of RAG2, and a 12/23RSS pair (Swanson, 2004).

Formation of a paired complex (PC) containing two RSSs could occur via two possible pathways: RSS capture, or RAG association. In the RSS capture model, RAG proteins bind a single RSS to form SC2, and subsequently bind a second RSS, without the addition of any further RAG molecules. In RAG association, two separate SCs form on separate RSSs, and then associate via protein-protein interactions (Swanson, 2004). Paired complex formation mostly likely occurs via the RSS capture model, for two reasons. Firstly, SC2 and PC both contain two molecules each of RAG1 and RAG2 (Mundy et al., 2002; Swanson, 2002b), and SC2 assembled on a single RSS can be converted to PC by addition of a suitable partner RSS (Mundy et al., 2002). Secondly, cleavage efficiency is greater when 12 and 23RSS substrates are added in a stepwise fashion, than when 12 and 23RSS complexes are formed separately and then mixed (Jones and Gellert, 2002).

Based on the observations that a) PC formation is more efficient when the first RSS bound is a 12RSS (Jones and Gellert, 2002) and b) nicking of endogenous 12RSSs, but not 23RSSs, can be detected at the *IgH*, *Igκ* and *Tcra* loci in lymphoid cells (Curry et al., 2005), it was previously thought that the RAG proteins preferentially bind a 12RSS first, then capture a 23RSS. However, it is now considered unlikely that the RAG proteins show preference for 12RSSs over 23RSSs. In the first study, truncated HMGB1 lacking the acidic C terminus was used (Jones and Gellert, 2002), but the acidic tail is known to enhance the DNA-bending activity of HMGB1 (Stros, 1998). Since HMGB1 is not needed for efficient 12RSS binding, the lack of acidic tail is unlikely to affect subsequent binding to a 23RSS. However, the additional influence of the HMGB1 acidic tail might be required to promote a RAG conformation that favours capture of a 12RSS partner (Swanson, 2004). In addition, nicking is probably not a reliable measure of RAG binding, since the sequence of the RSS and also the coding flank sequence can strongly influence the rate of nicking (Yu and Lieber, 1999).

Two recent studies have cast further doubt on the "12 first" model of RSS capture. The first study looked at endogenous RSS nicking at the *Tcrb* locus in developing T cells, but also looked at nicking of the same gene segments *in vitro*. In both cases, 23RSSs adjacent to $D_\beta 1$ gene segments were nicked more frequently than 12RSSs adjacent to J_β gene segments (Section 1.12; Franchini et al., 2009). This indicates that at the *Tcrb* locus, the RAG proteins bind to the 23RSSs first, before capturing a 12RSS. Whilst this study contradicts the findings of Curry et al. (2005) it suffers from the same problem that nicking might not be a reliable indicator of RAG binding.

In a more convincing study, chromatin immunoprecipitation sequencing (ChIP-seq) was used to directly assess RAG1 and RAG2 binding at the antigen receptor loci in developing primary lymphocytes, and it was found that the RAG proteins display no preference for either type of RSS (Ji et al., 2010). Instead, the main factor driving RAG binding was whether the RSS was found in active chromatin. In these localised regions, dubbed recombination centres, RAG binding correlates strongly with high levels of RNA polymerase II and H3K4me3, both markers of active chromatin that are crucial in regulating V(D)J recombination (Section 1.17). Importantly, this is the only study to date that has directly assessed the pattern of RAG-RSS binding *in vivo*. Therefore, the current model of RAG-RSS binding *in vivo* is that it occurs via a capture mechanism and not RAG association, as previously speculated, but the order of binding is a combination of the chromatin environment of the RSS, how close the heptamer and nonamer sequences are to the ideal consensus, and the sequence of the coding flank (Ji et al., 2010; Schatz and Ji, 2011).

1.10   RAG cleavage of RSSs

Following formation of a PC, RAG interactions with the heptamer and adjacent coding sequence are more intimate than in the SC, and these interactions are extended even further if the substrate is nicked (Nagawa et al., 2002; Nagawa et al., 2004). Furthermore, formation of a PC is accompanied by DNA distortion and unwinding at the cleavage site (Akamatsu and Oettinger, 1998; Nagawa et al., 1998a). Since the central domain of RAG1 preferentially recognises single-stranded heptamer sequence, it is thought that the RAG proteins bind to an RSS, create a nick at the heptamer/coding sequence boundary and induce localised base-pair melting at the heptamer/coding sequence region, although not necessarily in that order (Schatz and Swanson, 2011). This is supported by observations that RAG cleavage is stimulated by base-pair mismatches or abasic residues in the coding flank (Cuomo et al., 1996; Ramsden et al., 1996; Grundy et al., 2007).

A key step in hairpin formation by the related Tn5 transposase is the stabilisation of a "flipped" thymidine residue that extrudes from the DNA helix, via stacking interactions with a conserved tryptophan residue (Davies et al., 2000). This alignment enables nucleophilic attack by the opposing DNA strand to form a closed hairpin loop. A similarly flipped thymidine residue has been found in the RAG PC (Nishihara et al., 2008; Bischerour et al., 2009), and mutational studies have identified four candidate aromatic residues capable of stabilising a flipped thymidine: W893, W956, Y935 and F971 (Lu et al., 2006; Grundy et al., 2007; Bischerour et al., 2009). However, there is no single residue that completely fulfils all the requirements of base flipping, so it is suspected that all four residues play an indirect role in base stacking (Bischerour et al., 2009).

1.11   Basis of the 12/23 rule

In vivo, recombination between two different length RSSs is 30 times more efficient than recombination between either a 12/12 or 23/23 pair (Tonegawa, 1983), but the exact mechanism of the 12/23 rule remains unknown. What is known is that the 12/23 rule is established by the RAG proteins, possibly with assistance from HMGB1, at a very early stage of the V(D)J recombination, most likely during synapsis and again at double hairpin formation (Fugmann et al., 2000a). As noted above, the RAG proteins display a strong preference for binding a 12/23 pair *in vitro* over a 12/12 pair or a 23/23 pair (Jones and Gellert, 2002). Furthermore, hairpin formation at an RSS is greatly stimulated by the presence of an appropriate partner RSS (Steen et al., 1996; van Gent et al., 1996c; Hiom and Gellert, 1998; West and Lieber, 1998). The fact that the RAG1 active site is contributed *in trans* (Swanson, 2001) partly explains why hairpin formation

can only occur in the context of a paired complex, since each RAG1 subunit cannot cleave the RSS bound by its NBD, but it does not explain why the two RSSs must be a 12/23 pair.

One possible scenario that was proposed by the Gellert group is that upon capture of the first RSS, the RAG1-RAG2 dimer bound to the RSS is locked into an isomeric conformation, either a 12RSS-like or 23RSS-like conformation, depending on which RSS is bound first (Jones and Gellert, 2002). If a RAG heterotetramer binds to a 12RSS and creates a nick (as in SC2), the free RAG1 NBD has to seek out an accessible 23RSS partner against a background of other 12RSSs and nonspecific DNA. Since the major RAG-RSS interactions occur at the nonamer, it is possible that the free RAG1 subunit in the 12RSS SC2 samples incoming DNA for a nonamer sequence. Once a second nonamer is bound, the adjacent sequence is scanned for a heptamer. Since this RAG1/RAG2 heterodimer is locked in a 23RSS-binding conformation by virtue of the bound 12RSS, the heptamer sequence must be 23 base-pairs from the nonamer in order for the RAG1 active site to tightly associate. Thus, a PC containing two nicked substrates is only possible if it contains a 12/23 pair (Swanson et al., 2009). The decision to then convert a nick to a hairpin is likely a critical checkpoint in the 12/23 rule, since a flipped-out base is only detected at a pre-nicked 12RSS in the presence of 23RSS (Nishihara et al., 2008).

Binding of the second nicked RSS must then be transmitted to the active site of the RAG1 subunit bound at the heptamer of the first RSS, by an as-yet unknown conformational change. This could involve either the correct alignment of one or more of the aromatic residues suspected of stabilising a flipped base, or the alignment of one or more residues in the DDE, or both. Two gain-of-function RAG1 mutants have been described that can catalyse efficient hairpin formation in Mg^{2+} in the absence of a partner (Kriatchko et al., 2006; Nishihara et al., 2008). Furthermore, the presence of a partner does not increase cleavage efficiency, indicating that they are defective in sensing synapsis of a 12/23 pair. Notably, both mutations are proximal to the DDE motif, indicating that the conformational change that signals synaptic complex formation is likely to involve these residues.

1.12 The "beyond 12/23" rule

Whilst it is true that only different length RSSs are efficiently recombined *in vivo*, this does not mean that all 12 and 23RSSs are recombined with equal efficiency. At the *Tcrb* locus, the 23RSSs of the V_β gene segments and the J_β 12RSSs are arranged in a way that permits recombination. Despite this, V_β-to-D_β recombination almost never occurs, a phenomenon known as the "beyond 12/23 rule". A straightforward explanation would be that the chromatin structure of different regions of the *Tcrb* locus is activated at different development stages, thus enabling RAG access to the D_β gene segments before the V_β segments (see RSS accessibility, Section 1.15). However, this prediction was not borne out by experiments which showed that the V_β and DJ_β regions are both active at the same developmental stage (Mathieu et al., 2000; Jackson and Krangel, 2006). Furthermore, the beyond 12/23 rule has been recapitulated using extrachromosomal substrates in non-lymphoid cells. Therefore, restriction is not imposed either by lymphoid-specific factors or the nuclear environment of the *Tcrb* locus. The beyond 12/23 rule could be a result of the RAG proteins binding D_β 12RSSs more efficiently than J_β 12RSSs, but EMSA studies have shown that they are bound with roughly the same efficiency (Drejer-Teel et al., 2007), and the ChIP experiments discussed above (Section 1.9) did not indicate a RAG binding preference (Ji et al., 2010).

A possible explanation lies in the observation that the D_β 23RSSs contain a binding site for the AP1 transcription factor (Wang et al., 2008). c-Fos, a component of AP1, binds the D_β 23RSSs and promotes RAG association with these RSSs, thus enhancing D_β-to-J_β rearrangement. Mutation of the AP1 site abolished this effect (Wang et al., 2008). In addition, a separate *in vitro* study carried out in the absence of c-Fos found that synapsis and nicking of J_β 12RSSs is more efficient with D_β 23RSSs than V_β 23RSSs (Drejer-Teel et al., 2007). Therefore, the beyond 12/23 rule is most likely enforced by a combination of RAG interactions with c-Fos/AP1, and by intrinsic features of the *Tcrb* RSSs.

C)   Regulation of V(D)J recombination

Despite the huge payoff of having a diverse adaptive immune system, V(D)J recombination is an inherently dangerous process in that it creates double-strand breaks (DSBs) in the genome of developing B and T cells, which can potentially undergo translocation (Section 1.24). Therefore, recombination is a tightly controlled process with strict regulations imposed at a number of different levels. Firstly, it is regulated in a lineage-dependent manner, such that the RAG proteins are only expressed in B and T cells, and only the *Ig* loci are rearranged in B cells and only the *Tcr* loci are rearranged in T cells. Secondly, recombination is regulated at the cell development level, so that certain loci are arranged at specific cell development stages. Finally, rearrangement of the gene segments at each locus always occurs in the same order.

1.13   B and T cell development

Like all blood cells, B and T cells arise from haematopoietic stem cells (HSCs) in the bone marrow. Commitment of an HSC to either the B or T lineages is regulated by a complex transcription factor network. Pax5 is the guardian of B cell identity and promotes commitment to the B cell lineage (Nutt et al., 1999); the equivalent transcription factor in T cell precursors is Notch1 (Pui et al., 1999; Radtke et al., 1999), although this is a greatly simplified overview.

Commitment to the B cell lineage gives rise to the progenitor (pro-B) cell population. Further development at this stage is dependent on V(D)J recombination of the *IgH* locus, firstly by D_H to J_H gene recombination, then V_H to DJ_H joining. Once a productive VDJ_H rearrangement has occurred, the resulting μ chain is expressed on the cell surface, in a complex with the surrogate light chains, λ5 and VpreB (Bassing et al., 2002). Expression of this complex, termed the pre-B cell receptor (pre-BCR), causes a cascade of signalling events that triggers proliferation of the large pre-B cell population, followed by development to the small pre-B cell stage (Geier and Schlissel, 2006). In these cells, light chain recombination is initiated at one of the two light chain isotypes, Igκ or Igλ (Gorman and Alt, 1998) where recombination occurs between V and J gene segments only. The rearranged light chain product is then expressed with the rearranged heavy chain to form the B cell receptor molecule, IgM. Once the IgM binds to a cognate antigen in the presence of T cell help, differentiation to the mature B cell stage occurs, accompanied by somatic hypermutation (Section 1.21) and class switch recombination (Section 1.22).

Rearrangement of TCR loci also occurs at distinct developmental stages, in a fashion that mirrors B cell development. In TCRαβ cells, the TCRβ locus is rearranged at the CD4$^-$/CD8$^-$ double negative stage (equivalent to pro-B stage), with D_β to J_β rearrangement preceding V_β to DJ_β. Once a productive VDJ_β chain has been made, it is expressed on the cell surface as a pre-T cell receptor with CD3 and the surrogate pre-Tα chain. Pre-TCR signalling triggers differentiation to the CD4$^+$/CD8$^+$ double positive stage where TCRα rearrangement occurs.

1.14   Regulation of RAG expression

The first level of V(D)J recombination regulation is the expression of the RAG proteins themselves. RAG expression is restricted to the B and T cell lineages, and is also restricted to specific B and T cell developmental stages (Wilson et al., 1994; Grawunder et al., 1995). Under the control of a combination of lymphocyte-specific transcription factors including Pax5, Ikaros and E2A, RAG1 and RAG2 expression is upregulated at the pro-B and CD4$^-$/CD8$^-$ double negative stages, to enable *IgH* and TCRβ chain recombination. Following successful expression of a pre-BCR or pre-TCR, RAG expression is downregulated to allow for rapid proliferation (Grawunder et al., 1995). RAG expression is then upregulated once more for recombination of the *IgL* and TCRα loci at the pre-B and CD4$^+$/CD8$^+$ double positive stages (Wilson et al., 1994). Once a complete IgM molecule is expressed on the surface of a B cell, or a TCR undergoes positive selection by binding to an MHC molecule, expression of the RAG proteins is then sharply downregulated to prevent further rearrangements (Brandle et al., 1992). In addition to lineage- and stage-specific regulation, the expression of RAG2 is restricted to the G_1 phase of the cell cycle to prevent recombination during DNA replication and cell division (Section 1.27).

The RAG proteins can also be re-expressed in mature B cells in the lymph nodes. The purpose of this for receptor editing: if the BCR recognises a self-antigen, the RAG proteins are upregulated to allow for continuing rearrangement of the light chain loci. Thus, receptor editing has the potential to rescue a self-reactive B cell by producing a BCR with a slightly altered specificity (Gay et al., 1993; Tiegs et al., 1993).

1.15   RSS accessibility

As noted above, recombination of the *Ig* and *Tcr* loci is restricted to B and T cells, respectively, despite being catalysed by the same RAG proteins and using the same signal sequences. Furthermore, rearrangement of gene segments always occurs in a specific order (D to J followed by V to DJ at the heavy chain loci). This strict regulation can be explained by specific changes in chromatin structure that increase accessibility

of the antigen receptor loci to allow the RAG proteins to initiate rearrangement only in the correct cell type and at the correct stage of development (Alt et al., 1984; Yancopoulos and Alt, 1985). Packaging of RSSs into nucleosomes renders them inaccessible to RAG binding and consequently inhibits V(D)J recombination (Kwon et al., 1998; Golding et al., 1999; McBlane and Boyes, 2000). Furthermore, nucleosomes are preferentially positioned over RSSs (Baumann et al., 2003), implying that they must be remodelled before V(D)J recombination can be initiated. In addition, a number of studies have highlighted specific histone modifications and other changes in chromatin that correlate with increased V(D)J recombination.

1.16   Histone acetylation

Acetylation of histone H3 and H4 is highly correlated with the developmental regulation of V(D)J recombination (Roth and Roth, 2000; Huang and Muegge, 2001; Maes et al., 2001; Ye et al., 2001; Espinoza and Feeney, 2005). Indeed, McMurry and Krangel provided a striking association between histone acetylation, accessibility and V(D)J recombination at the TCRα/δ loci (McMurry and Krangel, 2000). Further analyses spanning the entire *IgH* locus demonstrated that histone acetylation was regulated both globally and locally. Large active domains were acetylated, but within these regions peaks of acetylation were localised to the recombination signal sequences (Johnson et al., 2003; Morshead et al., 2003). Histone acetylation neutralises positive charges on the nucleosome, thus reducing inter-nucleosomal contacts and increasing RSS accessibility by RAG proteins (Nightingale et al., 2007). Despite the strong correlation between histone acetylation and the initiation of V(D)J recombination, this mark alone is insufficient for full locus activation (Sikes et al., 2002; Tripathi et al., 2002; Hesslein et al., 2003).

1.17   Histone methylation

Histone H3 trimethlyated at lysine 4 (H3K4me3) is critical for efficient V(D)J recombination. Increased H3K4me3 was first observed at the *IgH* and *Tcrb* loci in peaks at the ends of regions undergoing recombination (Morshead et al., 2003). More comprehensive studies at the Igκ locus, performed in inducible cell lines, primary B cells and in cell lines representing distinct developmental stages, uncovered a correlation between increased levels of this modification and the initiation of V(D)J recombination (Perkins et al., 2004; Goldmit et al., 2005; Fitzsimmons et al., 2007; Xu and Feeney, 2009). Discovery of the RAG2 PHD finger explained the molecular basis for this correlation (Fig. 1.7; Matthews et al., 2007; Ramon-Maiques et al., 2007), and as discussed above, *in vivo* RAG2 binding correlates highly with H3K4me3 in recombination centres (Ji et al., 2010).

Mutation of crucial aromatic residues in the PHD finger (Y415, M443 and W453) eliminates RAG2 binding to an H3K4me3 peptide, and drastically reduces recombination in cell-based assays (Liu et al., 2007b; Matthews et al., 2007). Furthermore, several PHD finger mutations are implicated in Omenn syndrome (Sobacchi et al., 2006). The possible explanations for how the RAG2 PHD finger regulates efficient recombination are two-fold. First, binding of H3K4me3 at transcriptionally active Ig loci by the RAG2 PHD finger tethers the RAG recombinase to "active" RSSs (Liu et al., 2007b; Matthews et al., 2007). Second, H3K4me3 binding was found to enhance RAG cleavage efficiency (Matthews et al., 2007; Ramon-Maiques et al., 2007). A possible explanation for this is that in the absence of H3K4me3 binding, the PHD domain binds a self-peptide of RAG2, which forces RAG2 into a repressed configuration. Binding of H3K4me3 would then relieve this inhibition (Ramon-Maiques et al., 2007). This model is consistent with recent kinetic data showing that binding of H3K4me3 stimulates RAG-RSS complex formation, nicking and hairpin formation (Shimazaki et al., 2009; Grundy et al., 2010).

1.18   Non-coding transcription

Notably, Ji et al. (2010) demonstrated that the best correlation between RAG1 binding and the initiation of recombination occurs at transcriptionally active antigen receptor. Sterile, or non-coding, transcripts of the unrearranged antigen receptor loci were first identified in the mid 1980s (Lennon and Perry, 1985; Yancopoulos and Alt, 1985; Blackwell et al., 1986; Schlissel and Baltimore, 1989). Subsequently, these transcripts were found to be up-regulated at the same stage at which V(D)J recombination was initiated (Engel et al., 1999; Duber et al., 2003) suggesting that transcription could play a role in regulating V(D)J recombination. More recently, the Krangel laboratory demonstrated that transcription plays a critical role in regulating V(D)J recombination of TCR genes. Transcription through specific RSSs at the *Tcra* locus was blocked by the insertion of a transcriptional terminator downstream of the T early α (TEA) promoter (Abarrategui and Krangel, 2006; Abarrategui and Krangel, 2007); this suppressed Vα to Jα recombination and greatly diminished levels of H3K4me2, H3K4me3, H3K36me3 and H3 acetylation at the downstream Jα gene segments. Thus, these studies imply that non-coding transcription controls chromatin structure by activating immediately downstream RSSs and regulating the addition of histone modifications. Consistent with this, transcription is known to regulate chromatin changes associated with gene activation, such as histone acetylation, H2B ubiquitination (Workman, 2006) and notably H3K4me3. The latter occurs via association of the histone methyltransferase, Set1, with RNA polymerase II; this results in deposition of H3K4me3 during

transcription initiation (Ng et al., 2003; Bernstein et al., 2005; Pokholok et al., 2005), thereby helping to recruit RAG2.

As is the case for RAG binding to RSSs, the nucleosome poses a considerable barrier to passage of the polymerase. However, a number of studies have established that during the transcription of chromatin templates, H2A/H2B dimers are transiently evicted from nucleosomes to enable the passage of the RNA polymerase (Orphanides et al., 1999; Belotserkovskaya et al., 2003). Notably, a recent study found that even when all histone modifications previously associated with recombination are present, recombination is not fully activated. Instead, a strong correlation was found between the level of recombination and RSS accessibility; this accessibility was found to rely on the transcription-mediated eviction of H2A-H2B dimers at RSSs (Bevington and Boyes, 2013). This therefore answered the long-standing question of how nucleosomes are remodelled for the initiation of recombination. Moreover, the authors argue that by making the RSSs available only transiently, this may play an important role in reducing the number of RAG-mediated double stranded DNA breaks, thereby helping to maintain genomic stability.

1.19   Enhancers and transcription factors at the immunoglobulin loci

The above studies build a picture of non-coding transcription leading to activation of Ig and TCR loci in a cell- and stage-specific manner, modulating the necessary chromatin changes to allow RAG access to RSSs. However, a crucial question is how non-coding transcription itself is regulated. Notably, each antigen receptor locus contains at least one recombination enhancer and in some cases these elements share redundant functions, maintaining efficient V(D)J recombination when another element is removed or compromised, as demonstrated in mouse knockout studies (Takeda et al., 1993; Gorman et al., 1996; Xu et al., 1996; Inlay et al., 2002; Chowdhury and Sen, 2004; Schlissel, 2004). The function of the enhancers is to regulate the key changes required for V(D)J recombination, including stimulating non-coding transcription, increasing the level of activating histone modifications and opening up the chromatin to increase accessibility. Moreover, the enhancers of the antigen receptor loci are able to stimulate these changes over large distances. Indeed, some of the immunoglobulin enhancers are located hundreds of kilobases from the gene segments that are targeted for rearrangement. Analysis of these enhancers identified several transcription factors that appear to play a key role in the regulation of antigen receptor rearrangement, including E2A (Sakamoto et al., 2012), interferon regulatory factor 4 (IRF4; Johnson et al., 2008; Bevington and Boyes, 2013), and PU.1 (Nelsen et al., 1993; Rivera et al., 1993). The

cell- and stage-specific regulation of these transcription factors is thus central to the correct regulation of V(D)J recombination.

1.20   Long-range interactions between regulatory elements

Rearrangements at the antigen receptor loci occur over huge distances of up to 3 Mb. In order for this to occur, the antigen receptor loci must undergo substantial large-scale genomic reorganisation during lymphocyte development. Fluorescence *in situ* hybridisation (FISH) studies demonstrated that the *IgH* and *Igκ* loci are in an extended configuration in T cells but they become contracted in pro-B cells and pre-B cells, respectively (Kosak et al., 2002; Roldan et al., 2005; Sayegh et al., 2005; Skok et al., 2007). *IgH* locus contraction and distal V_H to DJ_H rearrangements are facilitated by the transcription factors, Pax5 and YY1, and loss of Pax5 increases the distance between the V_H and C_H regions, thus reducing the use of distal V_H gene segments (Hesslein et al., 2003; Fuxa et al., 2004; Liu et al., 2007a). In YY1-deficient mice, development is blocked prior to the onset of *IgH* recombination (Gordon et al., 2003; Liu et al., 2007a). Similar to the absence of Pax5, loss of YY1 mainly affects recombination of the distal V_H gene segments and increases the distance between V_H and C_H gene segments.

More recently, a key role for CTCF in facilitating long range interactions and chromosomal loops at the immunoglobulin loci has been uncovered. Developmentally regulated binding of CTCF and the cofactor cohesin was demonstrated at sites within the *IgH* and *Igκ* loci (Degner et al., 2011), which provide an anchor for long range interactions and the formation of chromosomal loops. Notably, YY1 binding sites were also found at the base of these looped structures (Guo et al., 2011a). Deletion of two conserved CTCF binding sites at the *IgH* locus enabled aberrant V_H to D_H recombination before D_H to J_H, and also enabled V_H to DJ_H recombination in developing T cells (Featherstone et al., 2010; Guo et al., 2011b). These observations highlight the importance of CTCF in regulating the stage and lineage specificity of recombination, and also indicate that CTCF sites act as a physical barrier that dictates the order of rearrangements.

D) Somatic hypermutation and class switch recombination in B cell maturation

After successfully expressing a functional cell-surface IgM, B cells circulate the body until they encounter a cognate antigen. These naïve B cells internalise the antibody-antigen complex, and in the presence of T cell help, form germinal centres in secondary lymphoid organs such the lymph nodes (Kelsoe, 1996; MacLennan, 2005). Here, B cells undergo rapid clonal expansion, producing a large number of cells that express the same immunoglobulin. The affinity of the B cell receptor-antigen interaction is then fine-tuned by the process of somatic hypermutation (SHM), where an extremely high rate of localised point mutation occurs at *IgH* and *IgL* variable exons (Peled et al., 2008b). These mutations are clustered around specific nucleotides corresponding to the CDR1 and CDR2 of the receptor, thus creating a huge number of B cells that express subtle variations of the same immunoglobulin. Some of these antibodies bind to the antigen with greater affinity and are subjected to positive selection and proliferate further, whereas those cells containing mutations that weaken the interaction will be lost (Li et al., 2004; Casali et al., 2006; Di Noia and Neuberger, 2007; Teng and Papavasiliou, 2007).

1.21 Somatic hypermutation

Activation of B cells by an antigen triggers the upregulation of activation-induced cytidine deaminase (AID) expression by Pax5 and E2A. AID initiates SHM by deaminating deoxycytidine (dC) residues in single-stranded DNA (ssDNA) to deoxyuridine (dU), thus creating a uracil:guanine (U:G) mismatch (Fig. 1.8). AID can only act on ssDNA, therefore transcription through the *Ig* locus is essential to create single-stranded transcription bubbles before somatic hypermutation can occur (Peled et al., 2008b).
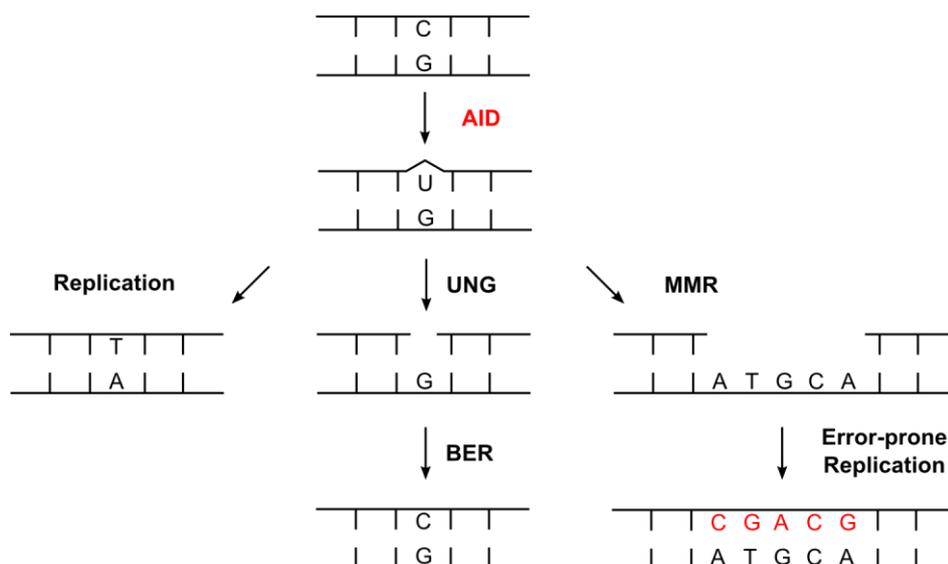
Figure 1.8 – Somatic hypermutation. AID deaminates cytidine to uridine, which is then processed in one of three ways: replication, base excision repair (BER) or mismatch repair (MMR). During replication, U is simply recognised as a T, resulting in a C to T point mutation. In BER, uracil DNA glycosylase (UNG) removes the uracil residue, leaving an abasic site, which is then repaired incorrectly by an error-prone polymerase. In BER only one residue is mutated, whereas in MMR, bases surrounding the lesion are also mutated. The BER and MMR pathways usually repair breaks correctly, but during SHM, error-prone polymerases are recruited, such as REV1 or Pol μ, resulting in mutations at sites of AID activity.

Since dU is not normally found in DNA, these residues are repaired in one of three ways: DNA replication, base-excision repair, or mismatch repair. Repair by DNA replication simply involves interpretation of the uracil residue as deoxythymidine, resulting in a C-to-T mutation, or a G-to-A mutation in the opposite strand. In base excision repair (BER), uracil DNA glycosylase (UNG) excises the uracil residue to create an abasic site. The abasic site is converted to a single-strand DNA break by apurinic/apyrimidinic endonuclease 1 (APE1), which can then be filled in with any of the four bases by a translesional error-prone DNA polymerases such as REV1 or Pol μ (Peled et al., 2008b). In mismatch repair, the U:G mismatch is recognised by a different set of proteins, but also results in the mutation of residues surrounding the U:G mismatch. Either the Msh2-Msh3 or Msh2-Msh6 heterodimer recognizes the U:G mismatch, which then recruit (amongst others) MLH1 and PMS2 leading to a single-strand nick near the mismatch. Exonuclease 1 excises a stretch of the surrounding DNA strand, and PCNA recruits error-prone polymerases to repair the gap, resulting in several new mutations surrounding the site of the original U:G mismatch (Peled et al., 2008a). Intriguingly, mismatch repair and base excision repair are normally benevolent repair pathways, but in somatic hypermutation they are hijacked and made to introduce mutations in the V region. In both cases it is thought that mono-ubiquitinylated PCNA is responsible for recruiting the error-prone polymerases (Poltoratsky et al., 2000).

1.22 Class switch recombination

Before stimulation by a complementary antigen, naïve B cells express either the IgM or IgD isotypes by alternative splicing of the D or M constant exons onto the variable gene exon. Changing to any of the other isotypes (IgG, IgA or IgE) is mediated by class switch recombination (CSR, also known as isotype switching). CSR is an intrachromosomal deletion recombination reaction that occurs between G-rich tandem-repeated DNA sequences called switch (S) regions that are located upstream of all heavy chain constant C_H regions except IgD (Fig. 1.9). Each naïve B cell has the potential to switch to any isotype, although the switch is directed by various cytokines that activate non-coding transcription through specific switch regions (Stavnezer, 1996).

The initial steps of CSR are similar to SHM. However, instead of generating a mutation, a double-stranded DNA break is made, which is subsequently repaired by the NHEJ pathway. First, non-coding transcription through the C_H locus creates localised regions of ssDNA at S regions, and AID introduces several dU residues in the S regions on both strands of DNA. UNG recognises the U:G mismatch and removes the deoxyuracil to create an apyrimidinic ribose residue, which is excised by APE1 or APE2, thus forming a single-stranded nick. These single-stranded breaks can spontaneously form a DSB if two are created close to each other on opposite strands. If not, it is possible that proteins from the MMR pathway can convert single-strand breaks that are not near each other into DSBs. The NHEJ pathway then joins DSBs at donor and acceptor switch regions via the same mechanism as in V(D)J recombination, and the intervening DNA is excised (Stavnezer et al., 2008).
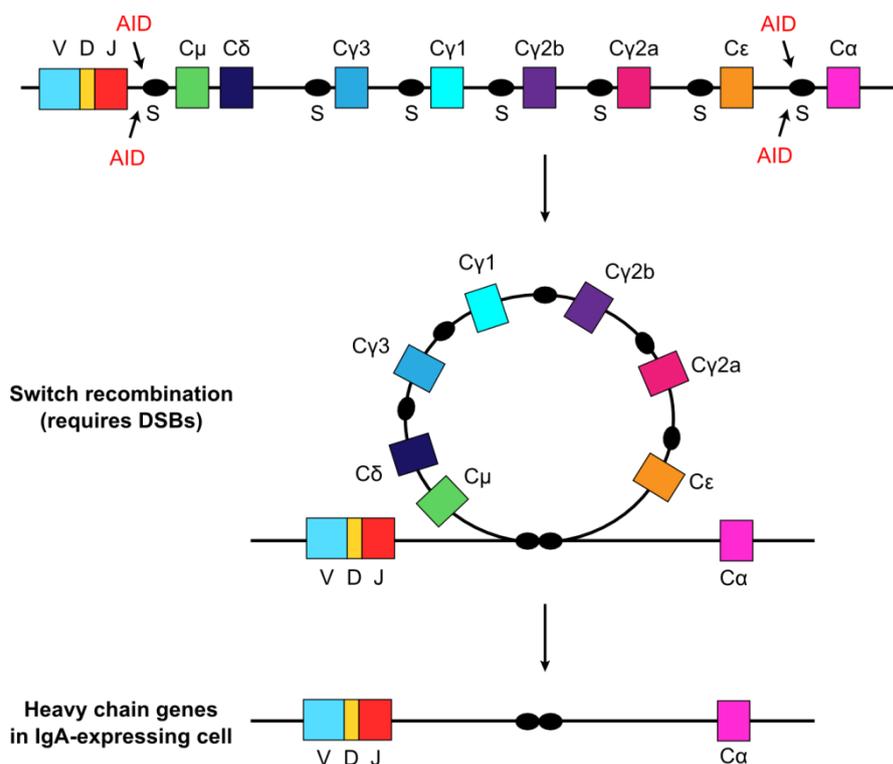
Figure 1.9 – Class switch recombination. AID creates double strand breaks at switch regions (black ovals labelled "s"). Recombination occurs between the two double strand breaks, bringing the alternative constant region adjacent to the rearranged variable region. A complete heavy chain transcript is made by mRNA splicing.

1.23 Terminal differentiation to plasma cells

Terminal differentiation of B cells into antibody-secreting plasma cells is regulated by BLIMP1 (B lymphocyte induced maturation protein 1), a zinc-finger transcription factor, considered to be the primary trigger for plasma cell differentiation. BLIMP1 orchestrates this transition by repression of Pax5, which has two major effects. The first is the repression of the germinal centre programme (i.e. the network of transcriptional regulation that drives the proliferation of B cells as well as SHM and CSR) by removing the positive regulation of *Aicda* and *Bcl6* by Pax5 (Shaffer et al., 2002; Sciammas and Davis, 2004), and the second is the derepression of *Xbp1*. *Xbp1* expression results in a dramatic expansion of the endoplasmic reticulum and activation of genes involved in the protein secretion pathway, thus allowing the production and secretion of large amounts of antibody (Shaffer et al., 2004). In addition, BLIMP1 regulates the differential 3' end processing of the immunoglobulin heavy chain pre-mRNA, leading to increased levels of antibody secretion (Shapiro-Shelef et al., 2003; Sciammas and Davis, 2004). Together, these processes produce plasma cells that are highly efficient, immunoglobulin-secreting machines.

In addition to its role in regulating recombination (Section 1.19), IRF4 plays a central role in coordinating exit from the GC programme. Whilst downregulated throughout the GC reaction, IRF4 is upregulated during exit from the germinal centres, and coordinates plasma cell development by terminating the GC programme (De Silva et al., 2012). This is achieved by derepression of BLIMP1 expression and downregulation of Bcl6 expression, which is the master regulator of the GC programme (De Silva et al., 2012). The central role of IRF4 in regulating plasma cell development is exemplified by the fact that deregulation of IRF4 alone can trigger multiple myeloma in plasma cells (Shaffer et al., 2008).

E)   Mistakes in antigen receptor production cause leukaemia and lymphoma

Normal lymphocyte development involves the programmed formation of multiple genomic DSBs and mutations, through the mechanism of V(D)J recombination, and in the case of B cells, somatic hypermutation and class switch recombination. Whilst these processes are fundamental to the generation of a diverse immune system, the formation of DSBs presents a major threat to the genomic integrity of developing lymphocytes. Indeed, chromosome translocations arising from aberrant recombination are a hallmark of lymphoid malignancies, although these aberrations alone are generally not enough to fully trigger carcinogenesis (Kuppers, 2005). For instance, the *BCL2-IgH* translocation can be found in circulating pre-leukaemic blood cells of healthy individuals (Limpens et al., 1991; Janz et al., 2003). However, they are a defining feature of many lymphoid malignancies, and are thought to be the critical initiating event (Kuppers, 2005). Full progression of these malignancies requires additional genetic mutations that further deregulate the transcriptional programmes that direct B and T cell development. The following section discusses the sources of aberrant lymphocyte development and how they trigger malignancy, with a focus on B cells.

1.24   Chromosome translocations

There are three general mechanisms by which the recurrent translocations found in lymphoid tumours are thought to promote malignancy (Nussenzweig and Nussenzweig, 2010). The first involves upregulation of a proto-oncogene or down-regulation of a tumour suppressor gene by the juxtaposition of one of these genes with a strong regulator element from one of the immunoglobulin loci (Adams et al., 1985; Vanasse et al., 1999). For example, the t(14;18) translocation, found in nearly all follicular lymphomas (Raghavan et al., 2001), brings a strong enhancer from the J_H locus on chromosome 14 within range of the *BCL2* promoter on chromosome 18. In a normal physiological setting, the role of *BCL2* is to prevent caspase 9 and 3-mediated

apoptosis by inhibiting the release of mitochondrial cytochrome c (Kridel et al., 2012). Therefore, upregulation of *BCL2* expression by the J_H enhancer promotes survival of these cells by reducing their capacity to undergo apoptosis. Another classic example is the *c-myc/IgH* translocation, which is found nearly all cases of Burkitt's lymphoma (Section 1.26; Robbiani and Nussenzweig, 2013).

The second mechanism by which translocations can promote B cell leukaemogenesis is via formation of a fusion protein. A classic example is the t(9;22)(q34;q11) Philadelphia chromosome, which encodes the BCR-ABL kinase, and whilst being a hallmark of chronic myeloid leukaemia, it is also found in 20-40% of adults with B cell acute lymphoblastic leukaemia (ALL). The BCR-ABL kinase is a constitutively-active kinase that promotes transformation by aberrantly enhancing the RAS-MAPK, JAK-STAT, and PI3K-AKT pathways (Teitell and Pandolfi, 2009). The final mechanism involves the deregulation of microRNA expression. This is a relatively recent discovery that is less well characterised than the first two mechanisms, but includes the microRNA families *miR-15 and miR-16* which act as tumour suppressors by down-regulating BCL2 (Calin et al., 2004; Calin and Croce, 2007).

The first step in a chromosomal translocation typically involves the formation of two DSBs. The majority of B cell translocations that deregulate a proto-oncogene involve a break at an immunoglobulin locus, caused by V(D)J recombinase or AID activity. There are several possible sources of the second break, including the mis-targeting of RAG activity through the recognition of cryptic RSSs (cRSSs; Papaemmanuil et al., 2014) and non-standard DNA structures, mis-targeted AID activity, or another process such as collapse of a replication fork or exposure to ionising radiation (Tsai and Lieber, 2010).

Translocations that create a fusion protein, such as *BCR-ABL* and t(12;21) *ETV6-RUNX1* (previously known as *TEL-AML1*) do not involve antigen receptor loci, therefore they are not the result of aberrant RAG activity. The sources of these breaks are less well-defined and potentially more diverse, with similar mechanisms operating widely in non-lymphoid cancers. There is some evidence that the cause of translocations involving the *Bcr* and *Abl1* breakpoint cluster regions involves Alu elements (Elliott et al., 2005), which are short repetitive retrotransposon sequences characterised by cleavage by the AluI (*Athrobacter luteus*) restriction endonuclease (Kramerov and Vassetzky, 2005). Other possible causes of translocations that do not display evidence of RAG activity include palindromic-mediated genomic instability

(Kurahashi et al., 2006), and an imbalance of metabolic pathways, based on epidemiological evidence in infant leukaemias (Rabkin and Janz, 2008).

The *ETV6-RUNX1* translocation is present in 25% of children with ALL, making it the most common genetic lesion in childhood ALL (Shurtleff et al., 1995). Expression of this fusion protein is thought to promote transformation by overexpression of the erythropoietin receptor and downstream activation of the JAK-STAT signalling pathway (Torrano et al., 2011). Furthermore, B cell development is stalled at a stage where the RAG proteins are constitutively expressed, leading to an accumulation of translocations that can eventually trigger malignant transformation (Papaemmanuil et al., 2014).

1.25   Mistargeted AID activity

Many human lymphomas involve mature B cells, which have already undergone V(D)J recombination and therefore no longer express RAG proteins. Nevertheless, these cancers are usually associated with translocations involving immunoglobulin genes, such as the *Bcl2/IgH* translocation in follicular lymphoma and the *c-myc/IgH* translocation in sporadic Burkitt's lymphoma (Kuppers, 2005). In contrast, mature T cell lymphomas are very rare, probably because these cells do not undergo further genetic alteration or double-strand break generation. Analysis of the translocations in mature B cells indicates that most of the immunoglobulin locus breakpoints are in the V (indicating SHM error) or switch (indicating CSR error) regions. Thus, it appears that AID activity plays a central role in many of these translocations (Robbiani and Nussenzweig, 2013).

Two mouse plasmacytoma models have been used to test the possibility that translocations involving switch regions are the result of mis-targeted AID activity. Plasmacytomas with *c-myc-IgH* translocations can be induced by injection of mineral oil or by overexpression of interleukin-6 (Potter and Wiener, 1992; Kovalchuk et al., 2002), and most of the translocations are between the 5' UTR of the *c-myc* gene and switch regions at the *IgH* locus. AID-deficient mice do not acquire c-myc-IgH translocations, suggesting that AID is essential to initiate the process (Unniraman et al., 2004; Ramiro et al., 2006). Uracil DNA glycosylase (UNG) is also necessary to create the DSB for these translocations (Ramiro et al., 2006), and since UNG creates DSBs in the initial steps of CSR, it appears that the source of the *c-myc/IgH* translocation is aberrant CSR and not SHM.

The cause of the *c-myc* locus break remained elusive for some time with the favoured hypothesis being that AID activity was responsible for this break as well, based on the

observations that cleavage at both loci occurs at the same stage of B cell development and that transcription through the *c-myc* locus generates G-loops that can be bound by AID (Duquette et al., 2005). This prediction was borne out by experiments by Robbiani et al. (2008), who showed that when an artificial DSB is made at the *IgH* locus, the *c-myc/IgH* translocation is detected only in AID-proficient cells, thus demonstrating the necessity of AID in creating the break at *c-myc* (Robbiani et al., 2008).

In addition to initiating translocations, AID can introduce mutations at a number of non-immunoglobulin loci in a hypermutation-like manner, including the proto-oncogenes *BCL6, PIM1, myc, RhoH/TTF* (*ARHH*) and *Pax5* (Pasqualucci et al., 1998; Pasqualucci et al., 2001). Mutations in these four oncogenes can be found in 50% of diffuse large B cell lymphomas, but not in normal germinal centre B cells, indicating that mistargeted AID activity can promote malignancy development not only by translocation-mediated mechanisms, but also by a hypermutation-mediated mechanism (Pasqualucci et al., 2001).

1.26   Mistargeted RAG activity

i)   Substrate selection errors

The first indication that V(D)J recombination might be involved in chromosome translocations arose through analysis of translocation breakpoints in B and T cell malignancies, which revealed that many of them contain cryptic RSSs, which are sequences that bear homology to true RSSs (Dalla-Favera et al., 1982; Kirsch et al., 1982). This led to the idea that the majority of translocations in developing lymphocytes take place because the RAG complex mistakenly identifies these cryptic RSSs as suitable recombination partners, but otherwise carries out normal V(D)J recombination (Fig. 1.10). There are predicted to be over 10 million cryptic RSSs scattered throughout the human genome (roughly one every 1-2 kb), thus creating huge scope for translocations by this process (Lewis et al., 1997).

However, accumulating evidence suggests that errors in substrate selection are not the major cause of translocations in developing lymphocytes (Brandt and Roth, 2009). Firstly, *in vivo* recombination experiments using extra-chromosomal plasmid substrates indicated that whilst some cryptic RSSs are utilised by the RAG proteins very efficiently, such as the *LMO2*, *TAL1* and *TAL2* cryptic RSSs, others are used very inefficiently, if at all, including those found at common translocation breakpoints such as the *BCL1* and *BCL2* loci (Raghavan et al., 2001; Marculescu et al., 2002). Secondly, many translocation breakpoints do not occur at the candidate cryptic RSS heptamer

boundary of the candidate cryptic RSS, which is inconsistent with RAG cleavage activity. Finally, some of the breakpoints contain short direct repeats, indicating that cleavage produced a single-stranded overhang, which again is inconsistent with RAG activity (Bakhshi et al., 1987; Kuppers and Dalla-Favera, 2001).
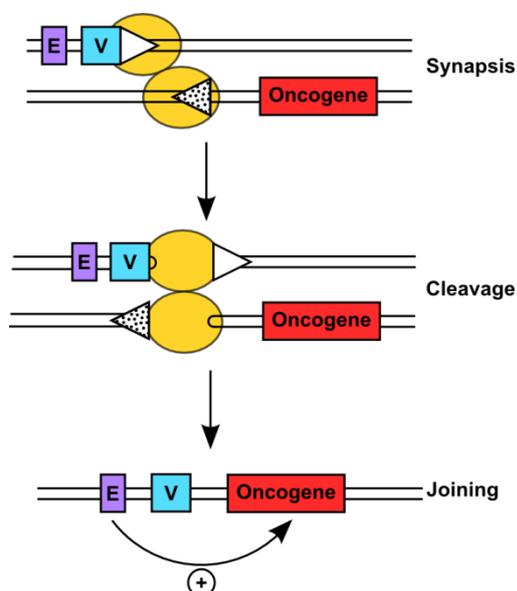
Figure 1.10 – Substrate selection error. RAG proteins carry out normal recombination but mistakenly recognise a cryptic RSS or non-standard DNA structure. If the cryptic RSS is adjacent to an oncogene, enhancers from the antigen receptor locus can upregulate the expression of the oncogene and contribute to leukaemogenesis. End donation occurs in the same way, except that the non-antigen receptor break is formed by an unrelated process, such as ionising radiation or a collapsed replication fork. Yellow circle = RAG proteins, spotted triangle = cryptic RSS, E = immunoglobulin locus enhancer, V = variable gene segment.

ii)   Structure-specific RAG binding

If the *BCL2* major breakpoint region (Mbr) does not contain a suitable cryptic RSS, then why is the t(14;18) translocation so prevalent? One possible explanation lies in the ability of the RAG proteins to bind and cleave non-B form DNA. Analysis of the Mbr revealed that it reacts with bisulphite, a chemical probe for ssDNA, suggesting that the Mbr forms a stable region of single-stranded, non-B form DNA, which is consistent with the preference of the RAG central domain to bind single-stranded DNA. Furthermore, the *BCL2* Mbr is efficiently bound and cleaved by the RAG complex in transfected cell lines, and importantly, the *BCL2-IgH* translocation can be recapitulated in transfected lymphoid cells (Raghavan et al., 2004a; Raghavan et al., 2004b; Raghavan et al., 2005; Nambiar and Raghavan, 2012), Intriguingly, this occurs independently of a discernible nonamer sequence.

iii)   End Donation

In some translocations, the non-immunoglobulin partner does not bear the hallmarks of either RAG or AID activity, and is thought to arise via another unrelated process, such as the collapse of a replication fork, or a random DNA break caused by ionising radiation or oxidative free radicals (Tycko and Sklar, 1990; Lewis, 1994; Kuppers and Dalla-Favera, 2001). In this "end donation" model of chromosome translocation, a break from one of the *Ig* loci is aberrantly joined with a break formed by one of these different processes (Fig. 1.10). End donation-style translocations are estimated to be responsible for 30-40% of the translocations found in follicular and mantle cell lymphomas (Jager et al., 2000; Welzel et al., 2001). However, it is unclear how RAG-mediated breaks become available for translocation, as they should be held together in a post-cleavage complex until they are resolved by the NHEJ pathway. One possibility put forward in a recent study is that certain non-consensus heptamer sequences destabilise the RAG postcleavage complex (Arnal et al., 2010), thus making the ends available for repair by the error-prone alternative NHEJ pathway.

1.27   RAG-mediated protection against translocations

Despite the fact that many translocations can arise from aberrant RAG activity, it is becoming clear that many aspects of RAG proteins, especially the non-core RAG2 C-terminus, actively protect against these deleterious reactions.

The cell has two general mechanisms by which DSBs can be repaired: homologous recombination and the classical NHEJ (cNHEJ) pathway. More recently, an error-prone form of the NHEJ pathway has been discovered, called alternative NHEJ (aNHEJ; Deriano and Roth, 2013). As described above, RAG cleavage products are repaired via the cNHEJ pathway (Section 1.5). Moreover, use of the homologous recombination and aNHEJ pathways is completely blocked by wild-type RAG proteins, indicating that the RAG proteins themselves guide the cleavage products to the classical NHEJ pathway (Lee et al., 2004). Why is it important that the broken ends are shepherded away from the aNHEJ machinery? With the exception of nontemplated nucleotide addition by TdT in lymphocytes, repair of DSBs in general by the cNHEJ pathway is highly accurate and efficient (Lieber, 2010). Whilst relatively little is known about the aNHEJ pathway at the molecular, it is clear that it is more error-prone than classical NHEJ, leading to large deletions and short repetitive stretches (Deriano and Roth, 2013). More importantly, the aNHEJ machinery has a propensity to catalyse chromosome translocations (Brandt and Roth, 2009). Therefore, it is crucial that repair of RAG-mediated DSBs occurs via the cNHEJ pathway.

What are the factors that guide the broken ends to the cNHEJ pathway, and simultaneously prevent repair by homologous recombination and aNHEJ? Firstly, certain RAG1 mutants (K980A and R838A/K839A/R840A) enable repair of coding and signal ends by homologous recombination and alternative NHEJ, (Huye et al., 2002; Lee et al., 2004). Secondly, a frameshift mutation in the C-terminus of RAG2 (RAG2-361^{fs}) allows proteins from the homologous recombination and alternative NHEJ pathways to access the coding and signal ends (Corneo et al., 2007). Unlike wild-type RAG proteins, the RAG2-361^{fs} mutant enables efficient repair of coding and signal ends in cells deficient for classical NHEJ. However, these joints contained short sequence homologies and large deletions, indicative of repair by the alternative NHEJ pathway (Corneo et al., 2007). Therefore, the alternative NHEJ pathway is capable of robustly repairing RAG cleavage products, but this is prevented by the wild-type RAG proteins. Remarkably, fusing RAG2 to an unrelated endonuclease (I-Sce1) prevents repair of I-Sce1 breaks by any other repair pathway than classical NHEJ. This effect is only seen with a pair of breaks containing four ends (as in V(D)J recombination), but not a single break (Cui and Meek, 2007).

Mice expressing core RAG2 display a highly increased rate of lymphoma formation on a p53-deficient background, caused by translocations at the *Tcra/Tcrd* and *IgH* loci (Deriano et al., 2011). The authors also show that the post-cleavage complex is less stable with core RAG2, suggesting that core RAG2 destabilises the postcleavage complex, allowing for repair by the translocation-prone alternative NHEJ pathway, consistent with the study discussed above (Corneo et al., 2007). The translocations seen in core RAG2/p35$^{-/-}$ are very similar to those seen in mice deficient in ATM (ataxia telangiectasia mutated). Since ATM is known to be important in the stability of RAG postcleavage complexes *in vivo* (Bredemeyer et al., 2006), it is possible that a destabilised postcleavage complex is responsible for the translocations seen in both core RAG2/p53$^{-/-}$ and ATM$^{-/-}$ mice. More recently, the region of the RAG2 C-terminus responsible for directing broken ends away from the alternative NHEJ pathway has been narrowed down to the acidic hinge region (Fig. 1.7; Chaumeil et al., 2013). Mutations in the acidic hinge which neutralised its negative charge reduced the stability of the postcleavage complex, promoted repair by alternative NHEJ, and also reduced the flexibility of the acidic hinge, implying that the acidic hinge region plays an important role in maintaining postcleavage complex stability (Chaumeil et al., 2013). Overall then, it is clear that the RAG proteins themselves play a key role in shepherding cleavage products towards the classical NHEJ pathway. However, the exact mechanism of this transfer remains unknown.

The RAG2 C-terminus contains a conserved threonine residue T490 (Fig. 1.7) that plays a key role in the cell cycle-dependent degradation of RAG2 (Li et al., 1996; Jiang et al., 2005). Cyclin-dependent kinase cyclin A/Cdk2 phosphorylates the T490 residue at G_0/S phase transition, which leads to ubiquitination of RAG2 by Skp2-SCF and degradation via the ubiquitin-proteasomal pathway (Jiang et al., 2005). Thus, RAG2 expression is tightly limited to the G_0/G_1 phases of the cell cycle. The degradation of RAG2 on entry into the S phase ensures that RAG-mediated DSBs are not generated during DNA synthesis, and also limits recombination to a phase where the expression of the NHEJ proteins is highest, thus preventing repair by alternative pathways. Consistent with this, enforced expression of RAG2 throughout the cell cycle leads to the accumulation of aberrant recombination products (Jiang et al., 2004). Since removal of the RAG2 C-terminus would also disrupt cell cycle regulation of RAG2, this could partly explain the accelerated rate of lymphoma in core RAG2 mice in the cRAG2 mice described above (Deriano et al., 2011). Consistent with this, mice expressing the RAG2 T490A mutant develop lymphomas with a similar genotype to the core RAG2 mice (Zhang et al., 2011). However, RAG2 T490A/p53$^{-/-}$ mice do not die more rapidly than p53$^{-/-}$ mice, indicating that the cell cycle-dependent regulation of RAG2 expression cannot fully account for the phenotype of core RAG2/p53$^{-/-}$ mice (Deriano et al., 2011; Zhang et al., 2011).

In addition to stabilising of the postcleavage complex and limiting the expression of RAG2, a recent study showed that the RAG2 C-terminus actively discriminates between optimal and suboptimal cryptic RSSs by increasing the rate of RAG dissociation, so that dissociation occurs before nicking can take place (Shimazaki et al., 2012). Furthermore, the H3K4me3 binding activity of the RAG2 PHD finger might prevent selection of potentially efficient cryptic RSS sites, substrates that do contain sites of H3K4me3 modification. Therefore, the RAG proteins themselves safeguard genome stability through a combination of stabilising the postcleavage complex (thus protecting the broken ends from translocation-prone repair pathways), restricting the expression of RAG2 expression to the G_1 phase of the cell cycle, and reducing the efficiency of cryptic RSS cleavage.

1.28 Reintegration of V(D)J recombination by-products

In addition to gross chromosome translocations, RAG proteins can also mediate the reinsertion of the recombination by-product, the excised signal circle (ESC), via two possible reactions, namely transposition and reintegration.

In transposition, the reactive 3' hydroxyl group from a signal end attacks the phosphodiester backbone at any location, leading to random insertion of the ESC in a joining reaction that is independent of NHEJ factors (Fig. 1.11A). Many studies have highlighted the similarity between V(D)J recombination and viral transposition. Indeed, the RAG complex catalyses transposition *in vitro* with a high efficiency (Agrawal et al., 1998; Hiom et al., 1998). However, there have been surprisingly few cases reported *in vivo*. One study in a human T cell line found an excised signal end fragment from the T cell receptor locus which had integrated into an intron of the HPRT gene (Messier et al., 2006), but importantly, there have been no known cases of leukaemia or lymphoma where RAG-mediated transposition has been identified as a causal factor. A likely explanation for this is that the RAG2 C-terminus almost entirely inhibits the transposition reaction (Elkin et al., 2003), thus making transposition an unlikely route for reinsertion of ESC by-products.

More recently, two studies have discovered a new way in which the ESC could be reinserted back into the genome, in a reaction termed reintegration (Curry et al., 2007; Vanura et al., 2007). Reintegration is distinct from transposition in that a transposition reaction shows no target specificity, whereas reintegration recombines ESCs with a genomic RSS through the same biochemistry as standard V(D)J recombination (Fig. 1.11B). This could be an authentic RSS or one of the 10 million cryptic RSSs littered throughout the human genome.

In the first study, a genome-wide assessment of signal end integrations in developing thymocytes (T cell precursors) found that half of the insertions out of a total of 43 observations were caused by reintegration of an ESC (Curry et al., 2007). Furthermore, no obvious translocations or transpositions were detected. Importantly, the number of reintegration events was 7-fold higher in mice expressing core RAG2, indicating that the RAG2 C-terminus inhibits reintegration of ESCs, possibly via one of the mechanisms described above. In the second study, mouse fibroblast cells were transfected with extrachromosomal RSS and ESC substrates and RAG expression vectors, and it was found that the ESC could efficiently recombine with the RSS-containing plasmid (Vanura et al., 2007). Furthermore, ESCs could efficiently reintegrate with the cryptic RSSs adjacent to the *TAL2* and *LMO2* oncogenes (Vanura et al., 2007).
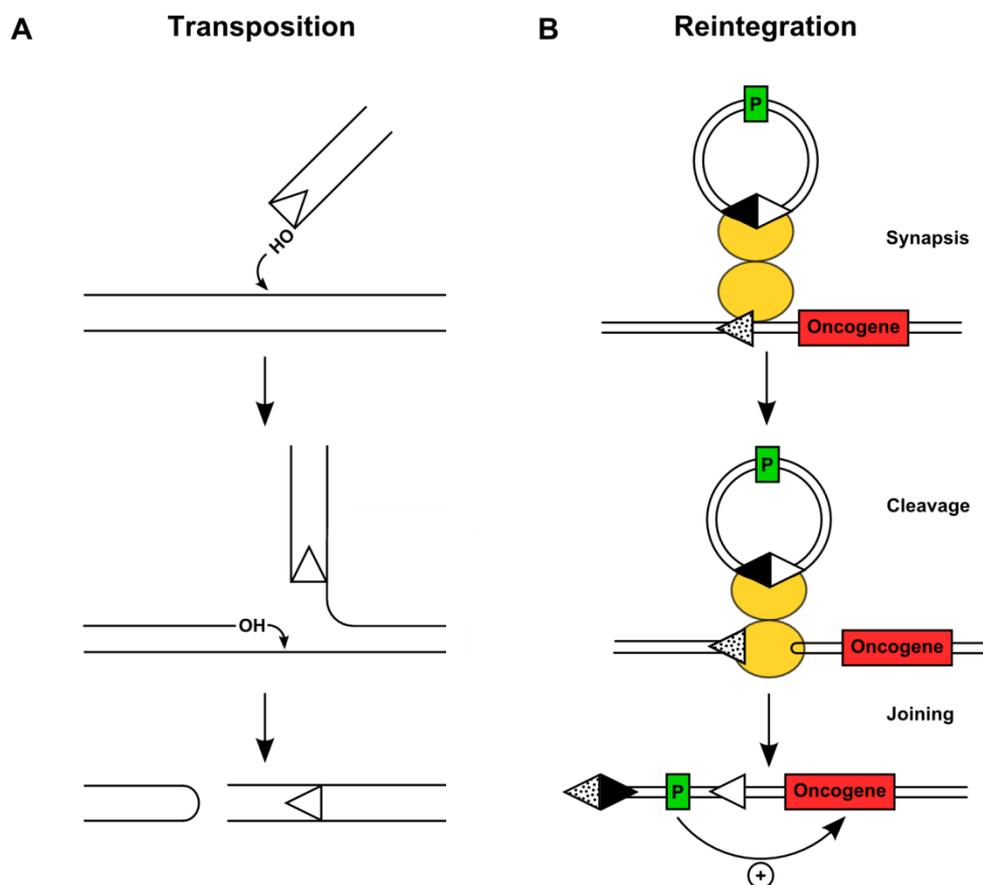
Figure 1.11 – Transposition and reintegration. (A) In transposition, the 3'-OH group of a cleaved signal end attacks the phosphodiester backbone at a random genomic locus. **(B)** In reintegration, RAG proteins catalyse normal recombination, albeit between an excised signal circle and a genomic RSS. If this occurs at a cryptic RSS adjacent to an oncogene, promoters from the antigen receptor locus can upregulate oncogene expression. P = antigen receptor locus promoter. Spotted triangle = cryptic RSS.

Gene profiling studies have shown that a substantial number of leukaemia and lymphoma cases display oncogene activation in the absence of any detectable cytogenetic abnormalities (Ferrando et al., 2002). The relatively small size of ESCs means that their reinsertion could go undetected using traditional cytogenetic methods. The occurrence of ESC reintegration *in vivo* is reinforced by a study which found an ESC that had reintegrated at the cryptic RSS in *HPRT* exon 1, a region that is commonly involved in V(D)J recombination errors (Messier et al., 2006). Together with estimates that reintegration of ESCs could represent 0.01-1% of V(D)J-mediated recombination reactions, with around 5000 reintegration events per human per day (Vanura et al., 2007), these studies indicate that reintegration of ESCs is a significant threat to genomic stability.

F)   Aims

The newly discovered reintegration reaction is a potentially significant cause of lymphoid cancers, and initial evidence suggests that the RAG2 C-terminus can inhibit reintegration by 7-fold (Curry et al., 2007). Whilst the mechanism of inhibition is unclear, it is also not sufficient to prevent reintegration entirely, as reintegration products are still detected with full-length RAG proteins (Curry et al., 2007). Complete prevention of this reaction is of paramount importance. If a mechanistic difference between recombination and reintegration exists, a small molecule inhibitor could be developed to exploit this difference, which could be used to bolster the natural inhibition of reintegration by RAG2 in people at risk, such as infants and the elderly. Therefore, the original questions I aimed to address in this project were:

1.   Do RAG-ESC interactions differ from RAG-RSS interactions? If so, how?
2.   Which part of the RAG2 C-terminus inhibits reintegration?
3.   By what mechanism does the RAG2 C-terminus inhibit reintegration?

However, whilst optimising the *in vitro* RAG assays intended to answer these questions, I discovered that when an RSS is paired with an ESC, cleavage of the RSS is more efficient than cleavage of the ESC, a phenomenon I refer to as asymmetric cleavage. Since a low level of ESC cleavage was detected, this does not completely preclude reintegration. However, the asymmetric cleavage within an RSS-ESC complex indicates that the majority of ESC-RSS pairings do not result in the cleavage of the ESC, which is a prerequisite for reintegration. Instead, I put forward an alternative outcome, named cut-and-run. In cut-and-run, an ESC stimulates the cleavage of an RSS, but the ESC remains intact. The ESC could then repeat this process, potentially creating multiple double-strand breaks in the genome of a developing B or T cell. Therefore, the aims of the project became:

1.   To verify that cut-and-run occurs *in vitro* with core and full-length RAG proteins
2.   To investigate the molecular basis of asymmetric cleavage
3.   To assess whether asymmetric cleavage occurs *in vivo* and whether cut-and-run is likely to occur in B and T cells.

These aims are addressed in turn in the following results chapters. In the first results chapter, further optimisation of *in vitro* cleavage conditions was carried out to ensure that asymmetric cleavage of an RSS-ESC pair was not the result of unsuitable or unphysiological conditions. In the second results chapter, these *in vitro* analyses were

extended to determine the molecular basis of asymmetric cleavage. Finally, I tested whether cleavage of extrachromosomal substrates is asymmetric *in vivo*, and thus whether cut-and-run is likely to occur in B and T cells.

Chapter 2

Materials and Methods

**(A)	Solutions**

### 2.1	Common Buffers

Unless stated otherwise, all reagents were from Sigma-Aldrich (Poole, UK), and all recipes are from Sambrook et al. (1989).

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	Acetic acid

DNA loading buffer (6x)

15% (w/v)	Ficoll
10 mM	Tris (pH 8.0)
1 mM	EDTA (pH 8.0)
0.01% (w/v)	Bromophenol blue
0.01% (w/v)	Xylene cyanol

Formamide loading buffer

95% (v/v)	Formamide
1 mM	EDTA (pH 8.0)
0.01% (w/v)	Bromophenol blue
0.01% (w/v)	Xylene cyanol

Lysogeny Broth (LB)

1% (w/v)	Tryptone
0.5% (w/v)	NaCl
0.5% (w/v)	Yeast Extract

LB-Agar

1% (w/v)	Tryptone
0.5% (w/v)	NaCl
0.5% (w/v)	Yeast Extract
1.5% (w/v)	Agar

Phosphate-buffered saline (PBS)

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

Adjusted pH to 7.4 with HCl

Protein loading buffer (2x)

20%	Glycerol
100 mM	Tris-HCl (pH 6.8)
284 mM	βMe
2%	SDS
0.2% (w/v)	Bromophenol Blue

SDS-PAGE running buffer

25 mM	Tris-HCl (pH 8.5)
192 mM	Glycine
0.1% (w/v)	SDS

Sodium phosphate buffer (NaP$_i$) pH 7.2

68.4 ml	1 M Na_2HPO_4
31.6 ml	1 M NaH_2PO_4

Final adjustments to pH were made using NaH_2PO_4 (acid) and Na_2HPO_4 (base).

Super Optimal Broth (SOB)

2% (w/v)	Tryptone
0.5% (w/v)	Yeast extract
10 mM	NaCl
2.5 mM	KCl
10 mM	$MgCl_2$

TAE (Tris-acetate-EDTA) buffer

40 mM	Tris-acetate
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)

TE (Tris-EDTA) buffer

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

TBS (Tris-buffered saline) buffer

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

2.2 Tissue culture media and solutions

DMEM-complete

500 ml	Dulbecco's modified Eagle's medium (DMEM)
10% (v/v)	Foetal calf serum (FCS)
120 mM	L-glutamine
2 ml	Penicillin/streptomycin (pen/strep; 100 µg/ml final concentration for both, Sigma #P4333)

DMEM-2 or DMEM-10 was made by adding FCS to 2% or 10%, respectively, and omitting pen/strep.

MEME-complete

500 ml	Minimum essential medium Eagle's (MEME)
10% (v/v)	FCS
120 mM	L-glutamine
2 ml	P/S

MEME-2.5 or MEME-10 was made by adding FCS to 2.5% or 10%, respectively, and omitting pen/strep.

Crystal Violet stain (0.1% w/v)

2.3% (w/v) Crystal violet solution

Diluted to 0.1% (w/v) with 20% ethanol before use

Polyethyleneimine (PEI, 1 mg/ml)

Linear PEI (25 kDa, Alfa Aesar #043896.01) was dissolved in water heated to ~80°C then cooled to room temperature. The pH was adjusted to 7.0 with HCl, and the solution was filter-sterilised and stored in aliquots at -80°C. Working aliquots were kept at 4°C for up to 1 month.

(B)   Manipulation of DNA

2.3   DNA restriction digest

All restriction enzymes were purchased from New England Biolabs (NEB), and reaction conditions were as specified in the manufacturer's instructions. A typical restriction digest contained 10 units of restriction enzyme/µg DNA and was carried out for 1 h in the case of diagnostic digests, or 3-16 h when creating a fragment for cloning. Typical volumes were 10 µl for diagnostic digests and 200 µl for preparative digests.

2.4   Electrophoresis of DNA

Digested plasmid DNA was separated by agarose gel electrophoresis through an agarose gel cast in 1x TAE and 5 µg/ml ethidium bromide. 1/6 volume of 6x DNA loading buffer was added before loading. The concentration of agarose was dependent on the size of fragments being separated, in accordance with the recommendations in Sambrook et al. (1989).

2.5   Routine PCR

A typical PCR reaction contained 1x ThermoPol Reaction Buffer, 200 µM each dNTP, 0.2 µM forward and reverse primers, 1 ng plasmid DNA template, and 2.5 units Taq

DNA polymerase (NEB #M0273) in a 50 µl reaction volume. Cycling conditions were as follows:

95°C	2 min	
95°C	30 s	
T_m	30 s	x 30
72°C	1 min/kb	
72°C	10 min	

Some primers contained nontemplated restriction sites at the 5' end for cloning purposes. In these cases, thermal cycling conditions used 15 cycles at the T_m of the primer without the additional restriction site, followed by 25 cycles using the T_m of the whole primer. The annealing temperature (T_m) of oligonucleotides was calculated using the OligoCalc oligonucleotide properties calculator (Kibbe, 2007), available at: http://www.basic.northwestern.edu/biotools/OligoCalc.html

2.6   Gel purification of DNA

DNA was separated by agarose gel electrophoresis (Section 2.4) and visualised by ultraviolet light for a minimal amount of time to minimise nicking of DNA. Desired fragments were excised from the gel using a sterile scalpel blade and transferred to dialysis tubing (8 kDa molecular weight cut-off) with a minimal amount of 1 x TAE (~1 ml). Fragments were electroeluted at 100 V for 1 h. Polarity of the current was reversed for ~20 s to recover DNA that had stuck to the interior surface of the dialysis tubing. DNA-containing TAE buffer was transferred to 1.5 ml tubes, and the DNA was purified by phenol-chloroform extraction/ethanol precipitation (Section 2.7) and resuspended in TE.

2.7   Phenol-chloroform extraction and ethanol precipitation of DNA

An equal volume of phenol-chloroform (1:1) was added to DNA solutions and the mixture was vortexed vigorously for ~5 sec. The aqueous and organic phases were then separated by centrifugation at 16,000 x g for 1 minute at room temperature (RT). This process was generally repeated once more. If the downstream application of the DNA was particularly sensitive (e.g. ligation), residual phenol was removed with an equal volume of 100% chloroform followed by vortexing and centrifugation as above. The aqueous phase was then transferred to a new 1.5 ml centrifuge tube and DNA was precipitated by the adding 1/10th volume of 3 M sodium acetate (pH 5.2) and 2 volumes of 100% ethanol. If a particularly small amount of DNA was to be precipitated, 20 µg glycogen was added to aid in precipitation and visualisation of the DNA pellet. Samples

were mixed and incubated on dry ice for 10 min before centrifugation at 20,000 x g for 10 min at 4°C. The DNA pellet was washed 1-2 times by adding 1 ml 70% ethanol and centrifuging at top speed for 5 min. Pellets were then air-dried for ~5 min, or until all visible traces of ethanol had evaporated, and resuspended in a suitable volume of either dH_2O or TE (depending on the downstream application of the DNA).

2.8   Quantification of DNA

DNA concentration was calculated using a NanoDrop spectrophotometer, according to the Beer-Lambert law:

$$A = \varepsilon \times c \times l$$

Where A = absorbance at 260 nm, ε is the extinction coefficient (M^{-1} cm^{-1}), and l is the pathlength of the spectrophotometer, which is usually 1 cm.

2.9   DNA ligation

Prior to ligation, digested vector was dephosphorylated using calf intestinal phosphatase (CIP) to minimise recircularisation. A typical reaction contained 1 x NEBuffer 4 and 1 unit of CIP (NEB #P0757), and was incubated at 37°C for 1 h, then purified by phenol-chloroform extraction. A typical ligation reaction contained 1 x T4 DNA Ligase Buffer (50 mM Tris-HCl [pH 7.5], 10 mM $MgCl_2$, 10 mM DTT, 1 mM ATP; NEB #B0202S), 1 unit of T4 DNA Ligase (NEB #M0202S), ~50 ng digested and dephosphorylated vector, and a 3-fold molar excess of digested insert in a 10 µl reaction volume. Ligation of overhanging "sticky" ends was either done for 20 min at room temperature (RT) or at 16°C for 16 h. Ligation of blunt DNA ends was done at RT for 2 h or 16°C for 16 h. A vector-only control ligation was carried out alongside each ligation reaction so that the vector background could be estimated following transformation of *E. coli*. Ligation reactions were stopped by heat inactivation at 65°C for 10 min before transformation of *E. coli* (Section 2.10).

2.10   Transformation of *E. coli*

A 50 µl aliquot of chemically competent DH5α cells was thawed on ice. Immediately after the aliquot had fully thawed, DNA was added and the sample was mixed gently and incubated on ice for 30 min. When transforming *E. coli* with plasmid DNA from a purified stock, 2 ng of DNA was added to cells, and when transforming ligation reactions (Section 2.9) half of the reaction (5 µl) was added. Cells were transformed by heat shock in a 42°C water bath for 90 sec, followed by a 2 min recovery on ice. 300 µl

SOB, pre-warmed to RT, was added and the cells were incubated at 37°C for 45-60 min (generally 60 min for ligations, 45 min for stock plasmids). Cells were then plated on LB-agar plates containing the appropriate selective antibiotic. For transformation of a purified plasmid stock, 10 µl was plated, and for ligation reactions, 150 µl was spread onto each of two plates. These plates were then incubated at 37°C for 16 h.

2.11 Small scale preparation of plasmid DNA (miniprep)

Small scale preparation of plasmid DNA was carried out according to Sambrook et al. (1989). A single colony of *E. coli* was picked from an LB-agar plate with a sterile pipette tip and used to inoculate 2 ml of LB with appropriate antibiotic (working concentrations taken from Sambrook et al., 1989). Cultures were incubated for 16 h at 37°C with shaking (250 rpm). 1.5 ml of the culture was transferred to a 1.5 ml centrifuge tube, and the remaining 0.5 ml was stored at 4°C to inoculate cultures for subsequent large scale plasmid preparation. Cultures were spun at 16,000 x g for 30 sec to pellet the bacteria, and the supernatant was discarded. Pellets were thoroughly resupended in 100 µl alkaline lysis buffer I by vortexing. 200 µl alkaline lysis buffer II was added, mixed gently by inverting several times, and incubated on ice for 5 min. 150 µl alkaline lysis buffer III was added and mixed vigorously by vortexing, then tubes were incubated on ice for 5 min. Cellular debris was removed by centrifugation at 16,000 x g for 2 min, and the plasmid-containing supernatant was transferred to a new 1.5 ml tube. Samples were cleaned by phenol-chloroform extraction/ethanol precipitation (Section 2.7) without sodium acetate. Pellets were washed once with 1 ml 70% ethanol, air-dried briefly and resuspended in 50 µl TE. Typically, purified samples were tested by diagnostic restriction digest (Section 2.3) to check for successful ligation of a DNA fragment or to confirm the identity of a plasmid. RNase A (10 ng/µl) was also added to simultaneously degrade RNA during the restriction digest.

2.12 Large scale preparation of supercoiled plasmid DNA (maxiprep)

10 ml LB containing the appropriate selective antibiotic was inoculated with the 0.5 ml of culture left over from small-scale plasmid preparation (Section 2.11) or with a single colony from a streaked out glycerol stock (Section 2.13). The 10 ml culture was incubated at 37°C with shaking (250 rpm) for ~8 h, then used to inoculate 500 ml LB plus antibiotic, which was then incubated for 16 h at 37°C with shaking (250 rpm). Cells were harvested by centrifugation at 4,000 x g for 20 min (4°C). Plasmid DNA was then purified using a Qiagen Plasmid Maxi Kit (Qiagen #12162) according to the manufacturer's instructions. Purified DNA was resuspended in 0.5 ml TE and the concentration was measured using a NanoDrop spectrophotometer.

2.13   Preparation of glycerol stocks for long-term storage of plasmid DNA

For long-term storage of plasmids, glycerol stocks were made by transferring 625 µl of cells cultured for 16 h for large scale plasmid purification (Section 2.12), to a sterile 1.5 ml screw-cap tube, adding 375 µl sterile 40% glycerol and snap-freezing on dry ice. Glycerol stocks were stored at -80°C.

(B)   Purification of core RAG proteins using the baculovirus/Sf9 expression system

2.14   Maintenance of suspension Sf9 cells

Suspension Sf9 cells were cultured in Sf-900 II serum-free medium (SFM; Life Technologies #10902-096) at 28°C with moderate shaking (140 rpm). Cells were grown to a density of 1.5-2.0 x 10^6 cells/ml, at which point they were diluted to 0.5 x 10^6 cells/ml. Proliferation from 0.5 to 2.0 x 10^6 cells/ml usually took ~ 48 h.

2.15   Calculation of baculovirus titer by plaque assay

The titer of baculovirus stocks was calculated according to the protocol in the Bac-to-Bac® Baculovirus Expression System manual (Life Technologies Publication Number MAN0000414). Before beginning, plaquing medium was prepared by melting 4% low-melting point agarose (Life Technologies #18300-012) in a 70°C water bath for 30 min; this was then combined with 1.3 x Sf-900 Medium (Life Technologies #10967-032) at a 1:3 ratio (i.e. 30 ml medium and 10 ml melted 4% agarose per plaque assay). The prepared plaquing medium was kept in a 40°C water bath until use to prevent the agarose from setting.

2 ml of Sf9 cells at 0.5 x 10^6 cells/ml was added to each well of two 6-well plates and incubated at 28°C for 1 h to allow cells to attach. In the meantime, a dilution range of eight 10-fold serial dilutions of baculovirus carrying RAG1 or RAG2 cDNA was made by adding 0.5 ml of virus stock (or the previous dilution) to 4.5 ml of Sf-900 II SFM. Once cells were attached, the medium was removed and immediately replaced with 1 ml of virus dilution. The dilutions 10^{-4} to 10^{-8} were used in duplicate, and a no virus negative control in duplicate (i.e. 1 ml of SFM), for a total of 12 wells. Cells were incubated at 28°C for 1 h to allow for infection. The virus dilutions were then removed and replaced with 2 ml of plaquing medium, which was left to set at RT for 1 h. The plates were then transferred to a humid box (a sterilised Tupperware box with tissues soaked in sterile dH$_2$O in the bottom to prevent desiccation of the agarose overlay) and incubated at 28°C for 7-10 days until viral plaques were visible to the naked eye. At this point, to improve visualisation of the plaques, 0.5 ml of neutral red solution (1 mg/ml) prepared

in dH$_2$O was added to each well and incubated for 1-2 h at RT. Excess stain was then removed with a pipette and the viral plaques were counted.

The viral titer (plaque forming units (pfu)/ml) was calculated according to the formula:

$$\text{Titer (pfu/ml)} = \text{number of plaques} \times \text{dilution factor} \times \frac{1}{\text{ml inoculum/well}}$$

2.16   Plaque purification of recombinant baculoviruses

Serial passage of baculovirus stocks leads to an accumulation of defective interfering particles (DIPs; Pijlman et al., 2001; Pijlman et al., 2003). These viral particles often lack some or all of the foreign gene of interest or other genes involved in very late gene expression and compete with growth of the desired recombinant virus, thus drastically reducing recombinant protein yield. Therefore, core RAG1 and core RAG2 baculoviruses (AcD26 and AcD25, respectively; McBlane et al., 1995) were plaque-purified twice to ensure that stocks were clonally pure and free from DIPs, following the protocol outlined in the Bac-to-Bac® Baculovirus Expression System manual (Publication Number MAN0000414).

The first part of the protocol is identical to a plaque assay (Section 2.15). This is followed by an intermediate amplification of individual plaques, then a second round of plaque purification. Once viral plaques were visible to the naked eye, two well-isolated plaques of each baculovirus were picked using a sterile pipette tip and transferred to 0.5 ml SFM in a 1.5 ml tube, and mixed well by vortexing. 2 ml of 0.5x10^6/ml Sf9 cells was seeded in each well of two 6-well plates and incubated at 28°C for 1 h. Sf9 cells were then infected with 100 μl of each plaque isolate in duplicate, alongside a mock-infection of 100 μl SFM. Infected cells were incubated in a humid box at 28°C for 72 h, until cells showed signs of infection. Next, a second plaque assay was set up, as described in Section 2.15. Individual plaques were picked with a sterile pipette tip, transferred to 0.5 ml SFM, vortexed, and stored at 4°C until amplification.

This new plaque isolate (0.5 ml) was now double-plaque-purified, and required amplification. Ideally, baculovirus should be amplified by infecting Sf9 cells at a multiplicity of infection (MOI; the ratio of pfu:cells) ranging from 0.05-0.1. However, infecting at the ideal MOI was not possible since the viral titer of the plaque isolates was not known, and there was not enough stock for a plaque assay. Instead, the 0.5 ml of double-purified baculovirus plaque was split between 5 wells of a 6-well plate

seeded with 1 x 10^6 Sf9 cells/well and incubated for 4 days in a humid box at 28°C. This is the P1 stock.

Assuming a P1 titer of 1x10^6 cells/ml (as suggested in the Bac-to-Bac manual), the P1 stock was used to infect 50 ml of Sf9 cells (2 x 10^6 cells/ml), using an MOI of 0.1. Infections were incubated at 28°C with moderate shaking (140 rpm). After 48 h, cells were removed by centrifugation at 1000 x g for 5 min, and the baculovirus-containing supernatant was transferred to fresh 50 ml tubes. These P2 baculoviral stocks were stored at 4°C, protected from light. Tissue culture-grade FCS was added to 2% to minimise loss of infectivity from degradation by residual proteases. The viral titer of the P2 stocks was estimated by plaque assay (Section 2.15) to be 1.2 x 10^8 pfu/ml for the RAG 1 baculovirus, and 1x10^8 pfu/ml for RAG2. Several 1 ml aliquots of the P2 viral stocks were snap-frozen on dry ice and stored at -80°C for long-term storage.

Since the P2 amplification only yielded 50 ml, large volume high-titer P3 stocks were made by infecting large Sf9 cultures (>250 ml) at an MOI of 0.1. Infected cells were incubated at 28°C for 48 h with moderate shaking (140 rpm), and baculovirus was harvested and stored as above. If subsequent amplifications of baculovirus were required, these were done using the 1 ml aliquots of P2 stored at -80°C, and not by serial passage of the P3 stocks.

2.17 Expression and purification of core RAG proteins from Sf9 cells

Suspension Sf9 cells were grown to a density of 1.5 x 10^6 cells/ml and co-infected with recombinant baculoviruses that express His- and MBP-tagged core RAG1 and core RAG2 (AcD26 and AcD25, respectively; McBlane et al., 1995) at an MOI of 3 for both viruses. Infected cells were incubated at 28°C with shaking (140 rpm) for 60-72 h. Cells were then harvested by centrifugation at 1000 x g for 20 min, washed twice with ice-cold PBS, snap-frozen in a dry ice/ethanol bath and stored at -80°C until purification.

All subsequent steps were carried out at 4°C unless stated otherwise. Infected cell pellets were thawed on ice and resuspended in 20 ml of lysis buffer (20 mM Tris-HCl [pH 7.9], 0.5 M NaCl, 5 mM imidazole, 2 mM β-mercaptoethanol (βME), 1 mM benzamidine, 1 mM phenylmethylsulphonyl fluoride (PMSF), 2X Complete Protease Inhibitor Cocktail (Roche #11697498001). Cells were mechanically lysed by ten strokes in a glass Dounce homogeniser with a tight-fitting pestle and then subjected to 3 cycles of sonication at 10 microns for 30 sec using an MSE Soniprep 150, followed by cooling on ice for 30 sec. The lysate was then cleared by centrifugation at 10,000 x g for 30 min, and loaded onto a 1 ml Ni-NTA column (Qiagen Superflow #1018611) equilibrated

with buffer A (20 mM Tris-HCl [pH 7.9], 0.5 M NaCl, 5 mM imidazole) connected to an ÄKTA Explorer (GE Healthcare). The column was washed with 5 column volumes (CV) of buffer A, followed by 5 CV of buffer B (20 mM Tris [pH 7.9], 0.5 M NaCl, 20 mM imidazole). Core RAG proteins were eluted with 10 CV of elution buffer (20 mM Tris-HCl [pH 7.9], 0.5 M NaCl, 0.5 M imidazole).

The RAG-containing fractions (identified by a peak on the chromatogram) were pooled together and mixed with 2 volumes of amylose column buffer (NaP$_i$ [pH 7.2], 0.5 M NaCl, 0.25% Tween 20$^®$), then loaded onto a 1 ml amylose column (NEB #E8022S) equilibrated with amylose column buffer. The column was washed with 5 CV of amylose column buffer, followed by 3 CV of amylose column buffer without Tween. Purified RAG proteins were eluted with 15 CV of amylose column buffer without Tween and with 10 mM maltose.

RAG-containing fractions were analysed by SDS-PAGE (Section 2.19) and fractions containing similar concentrations of protein were pooled. Protein was then concentrated 4-fold using centrifugal filter units with a molecular weight cut off of 30 kDa (Millipore #UFC803024). This procedure is also convenient for desalting samples, as the protein samples are in a solution containing salt at a high concentration (0.5 M NaCl) after elution from the amylose column. Therefore, after each spin cycle, concentrated protein was diluted with storage buffer (25 mM Tris-HCl [pH 8.0], 150 mM KCl, 2 mM DTT, 10% glycerol). Four spin cycles were carried, concentrated protein was aliquoted, snap-frozen in a dry ice-ethanol bath and stored at -80°C.

2.18 Expression and purification of HMGB1

The full-length rat HMGB1 expression plasmid pETM11-HMGB1 was a kind gift from Prof Marco Bianchi (Vita-Salute San Raffaele University, Milan, Italy). This plasmid contains the full-length protein coding sequence of rat HMGB1 (UniProt #B4F758) with an N-terminal 6xHis tag that can be removed by Tobacco Etch Virus (TEV) protease. Purification of full-length HMGB1 was carried out according to a protocol provided by Prof Bianchi. BL21(DE3) cells were transformed with pETM11-HMGB1, plated on LB/kanamycin agar plates and grown at 37°C for 16 h. A single colony was picked and used to inoculate 10 ml of LB/kanamycin and grown at 37°C for 16 h. The following day the entire 10 ml culture was transferred to 1 litre of LB/kanamycin and incubated at 37°C (250 rpm) until the OD$_{600}$ = 0.6-0.8. Expression of HMGB1 was then induced with IPTG at 0.5 mM, and incubated at 25°C with shaking (250 rpm) for 16 h.

Cells were harvested for 20 min at 4,000 x *g*, washed once in ice-cold PBS, and the pellet was snap-frozen in a dry ice/ethanol bath and stored at -80°C until purification. The following was carried out at 4°C unless stated otherwise. The pellet was thawed on ice and resuspended in 20 ml lysis buffer (20 mM Tris-HCl [pH 8.0], 150 mM NaCl, 10 mM imidazole, 2 mM βME, 0.2% Igepal CA630, 1x EDTA-free protease inhibitor cocktail (Roche #11873580001), 1 mM benzamidine, 1 mM PMSF). The resuspended cells were sonicated for 4 cycles of 2 min at 10 microns, 2 min on ice, and the lysate was clarified by centrifugation at 16,000 x *g* for 30 min. The cleared lysate was then filtered through a 0.2 μm syringe filter unit. The filtered lysate was bound to 5 ml Ni-NTA resin (Qiagen) in an Econo-Pac® chromatography column (Biorad #732-1010) on a rotating shaker for 30 min. The flowthrough was collected and applied to 2.5 ml of Ni-NTA resin in a second Econo-Pac® column for 30 min to maximise recovery of His-tagged HMGB1. The flowthrough was collected and both columns were washed with 10 CVs of lysis buffer, 5 CVs of lysis buffer without Igepal CA 630 or protease inhibitors, 5 CVs of wash buffer 1 (20 mM Tris-HCl [pH 8.0], 1 M NaCl, 10 mM imidazole, 2 mM βMe) and 5 CVs of wash buffer 2 (20 mM Tris-HCl [pH 8.0], 150 mM NaCl, 30 mM imidazole, 2 mM βMe). HMGB1 was eluted with elution buffer (Tris-HCl [pH 8.0], 150 mM NaCl, 300 mM imidazole, 2 mM βMe).

HMGB1 from the Ni-NTA column was loaded onto a 1 ml HiTrap SP XL column (GE Healthcare) equilibrated with buffer A (20 mM Tris-HCl [pH 8.0], 1 mM DTT, 80 mM NaCl). The column was washed with 9 column volumes of buffer at 80 mM NaCl and HMGB1 was eluted with a gradient of 80-650 mM NaCl (20 mM Tris-HCl [pH 8.0], 1 mM DTT) over 25 CVs. HMGB1-containing fractions were dialysed against 100 volumes of storage buffer (25 mM Tris-HCl [pH8.0], 150 mM KCl, 2 mM DTT, 10% glycerol) and analysed by SDS-PAGE (Section 2.19).

2.19 SDS-polyacrylamide gel electrophoresis of proteins

Typically, protein samples were analysed on a 10% (37.5:1 acrylamide:bis-acrylamide) polyacrylamide gel (8 x 10 x 0.07 cm). For HMGB1, it was found that the best resolution is achieved using a 12% gel (19:1 acrylamide:bis-acrylamide). In both cases, a 3% (37.5:1 acrylamide:bis-acrylamide) stacking gel was used. Running and stacking gel solutions were prepared as per the protocol in Sambrook et al. (1989) and cast using the BioRad Protean III system. Samples were mixed with an equal volume of 2x protein loading buffer by vortexing, then heated in a boiling water bath for 3 min. Denatured samples were loaded onto the gel and separated for 1 h at 180 V.

2.20 Coomassie staining of polyacrylamide gels

Polyacrylamide gels were incubated on an orbital shaker at RT in fixing solution (25% (v/v) isopropanol, 10% (v/v) acetic acid) for 10 min. Fixing solution was replaced with rapid Coomassie staining solution (10% (v/v) acetic acid, 0.006% (w/v) Coomassie Brilliant Blue G-250) for a minimum of 1 h. Gels were destained with 10% (v/v) acetic acid until the background was clear, which usually took around 20-30 min.

2.21 Estimation of protein concentration by SDS-PAGE and densitometry

The concentration of purified proteins was estimated by running a range of BSA concentrations on an SDS-PAGE gel alongside an aliquot of purified protein. Gels were stained with Coomassie (Section 2.20) and photographed using a SynGene white light transilluminator. The band densities were measured using AIDA Image Analyser software (Raytest, Germany, v.4.14). A BSA standard curve was plotted in Excel (2007) and the equation of the line of best fit was used to estimate the amount of purified protein in each band.

(C) Expression and purification of full-length RAG proteins using a vaccinia virus/HeLa cell expression system

2.22 Maintenance of cell lines used in producing recombinant vaccinia and expressing full-length RAG proteins

Adherent Baby Hamster Kidney 21 (BHK-21) and HeLa cells were maintained in DMEM-complete. HuTK$^-$ 143B cells were maintained in DMEM-complete with 25 µg/ml bromodeoxyuridine (BrdU). In general, adherent cells were cultured in a 75 cm^2 flask until ~90% confluent, at which point they were split 1:10-1:20. This was done by aspirating medium from the cells, adding 1 ml trypsin/EDTA and leaving to detach for ~5 min at RT. Detached cells were resuspended in 10 ml DMEM-complete to inhibit trypsin, then 1-2 ml was added to 20 ml DMEM-complete in a fresh 75 cm^2 flask.

Suspension HeLa cells were cultured in MEME supplemented with 5% FCS, 4 mM L-glutamine and 20 mM HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid). HEPES was included to maintain the pH at pH 7.0, since the incubator set up for spinner flasks did not contain CO_2. The log phase of suspension HeLa cell growth is between 1 and 5 x 10^5 cells/ml. HeLa cell density was maintained between 5x10^4 cells/ml and 5 x 10^5 cells/ml. Proliferation from 5x10^4 to 5x10^5 cell/ml typically took around 72 h.

In general, complete medium (i.e. medium supplemented with 10% FCS, L-glutamine and pen/strep; Section 2.2) was used for routine cell maintenance, and medium supplemented with 2% FCS and L-glutamine, but no antibiotic was used during infections. A lower concentration of FCS is used during infections because high concentrations can inhibit viral infectivity, and antibiotics are omitted because they are potentially toxic to cells during infection.

2.23 Cloning full-length RAG proteins into vaccinia transfer vector pTM1

The pTM1 expression cassette contains (5' to 3'): a T7 RNA polymerase promoter, the encephalomyocarditis virus untranslated leader region (EMCV-UTR) to increase transgene expression levels, a multiple cloning site (MCS), and a T7 terminator. The expression cassette is flanked by segments of the vaccinia virus thymidine kinase (TK) gene. During production of recombinant vaccinia virus the expression cassette is inserted into the TK gene of wild-type vaccinia virus by homologous recombination, allowing for selection of recombinant vaccinia virus in a TK-deficient (TK^-) cell line.

Complementary synthetic oligonucleotides encoding a 9xHis tag and a FLAG tag flanked by NcoI and EcoRI sites (FLAG-His Adaptor F and R, Table 2.1) were annealed and digested with NcoI and EcoRI, gel-purified and ligated into the NcoI/EcoRI sites of pTM1 to make pTM1-FH. The pTM1 vaccinia transfer vector (Moss et al., 1990) was a kind gift from Dr Bernard Moss (National Institute of Allergy and Infectious Diseases, USA). The coding sequences of full-length RAG1 and full-length RAG2 were PCR-amplified from pJH548 and pJH549, respectively (Sadofsky et al., 1993), using the Q5® high-fidelity DNA polymerase (NEB #M0491). Primers were designed so that a SmaI site was added to the N-terminus of the RAG1 coding sequence and an EcoRI site was added to the N-terminus of the RAG2 coding sequence. RAG1 was amplified with RAG1 SmaI F and RAG1 XhoI R, and RAG2 was amplified using RAG2 EcoRI F and RAG2 XhoI R (Table 2.1). Both were amplified to include the XhoI site of pJH548/pJH549 located downstream of the RAG1/RAG2 coding sequence. RAG1 and RAG2 PCR products were gel-purified and ligated into the SmaI/XhoI or EcoRI/XhoI sites of pTM1-FH, respectively, to make pTM1-FHR1 and pTM1-FHR2. DNA sequencing by GATC Biotech (Konstanz, Germany) confirmed the identity of the clones and that no mutations had been introduced.

2.24 Amplification of wild-type and recombinant vaccinia virus

Stocks of wild-type and recombinant virus were prepared in the same way. Wild-type vaccinia virus was a kind gift from Prof Geoffrey Smith (University of Cambridge, UK), which was supplied as a crude lysate of infected cells that required trypsination before

infection. An aliquot of virus was mixed with an equal volume of trypsin/EDTA and incubated at 37°C vortexing at 5-10 min intervals for 30 min, or until no visible cell clumps remained. Trypsinised virus was diluted to 12 ml with DMEM-10. Four confluent 175 cm^2 flasks of HeLa cells were infected at an MOI of 1 by adding 3 ml to each flask. Virus was evenly distributed by gentle rocking for ~30 s then incubated at 37°C/5% CO_2 for 2 hours, with gentle shaking every 30 min. After 2 h, 30 ml DMEM-10 was added and cells were incubated for 3 days until cytopathic effects were observed. To harvest the virus, cells were detached by shaking, centrifuged at 1000 x g for 5 min, and resuspended in 2 ml MEME-2. Vaccinia virus remains cell-associated, so cells were lysed by freeze-thawing three times to release the virus. Virus stocks were stored as a crude lysate in 0.5 ml aliquots at -80°C.

2.25 Calculation of vaccinia virus titer

All incubations were at 37°C/5% CO_2. A 6-well plate was seeded with 3.3 x 10^5 BHK-21 cells/well in 2 ml MEME-complete and incubated for 8-16 h. Baby hamster kidney 21 (BHK-21) cells were used because they have a long spindle-like morphology, which rounds up dramatically upon infection, making plaques easy to identify. HeLa cells can be used, but their round morphology and much faster doubling time makes it harder to identify plaques. A 10 μl aliquot of crude virus lysate was thawed and trypsined as in Section 2.24 and used to make eight 10-fold serial dilutions of virus in MEME-2. BHK-21 cells were washed with PBS to remove residual antibiotics, and then infected with 2 ml of the 10^{-6}, 10^{-7} and 10^{-8} dilutions in duplicate. Cells were incubated for 1-2 hours with gentle shaking every 15-30 min. 4 ml MEME-complete was then added (without removing virus dilutions) and cells were incubated for 24-48 h, until plaques were visible under a light microscope.

To stain cells, medium was removed and 1 ml 0.1% crystal violet (which only stains viable cells) was added and incubated at RT for a minimum of 5 min. Stain was discarded and excess stain washed off with PBS. Viral plaques were counted and averaged for each dilution and the titer (pfu/ml) was calculated according to the formula:

$$\text{pfu/ml} = \frac{\text{number of plaques} \times \text{dilution factor}}{\text{ml of virus used}} \times 2$$

The final multiplication accounts for dilution with trypsin at the start of the protocol. Since 2 ml of virus was used, this cancels the division so the formula can be simplified to:

$$pfu/ml = number\ of\ plaques\ x\ dilution\ factor$$

The final titer estimate was calculated by averaging the titers for each dilution.

2.26   Production of recombinant vaccinia virus

To produce recombinant vaccinia virus, BHK-21 cells were simultaneously infected with wild-type vaccinia virus and transfected with vaccinia transfer vector containing the full-length RAG coding sequences (Section 2.23). Two 25 cm^2 flasks were seeded with BHK-21 cells and grown to near confluency. Wild-type vaccinia virus was trypsinised as in Section 2.22 and diluted to 1.5 x 10^5 pful/ml in MEME-2. Medium was removed from BHK-21 cells and replaced with 1 ml diluted wild-type vaccinia virus (MOI ~0.05). Cells were incubated for 2 hours at 37°C/5% CO$_2$.

pTM1-FHR1 and pTM1-FHR2 were transfected using Lipofectamine reagent according to the manufacturer's instructions (Life Technologies). 8 µg of either pTM1-FHR1 or pTM1-FHR2 was added to 625 µl Opti-MEM$^®$ I medium in separate tubes and mixed gently. 25 µl Lipofectamine reagent was added to 625 µl Opti-MEM$^®$ I medium and incubated at RT for 5 min. The plasmid and Lipofectamine solutions were then mixed by vortexing briefly and incubated at RT for 20 min. Virus inoculum was removed from the cells, and replaced with 100 µl transfection solution, which was evenly distributed by gentle rocking. 5 ml of MEME-complete was added and the cells were incubated for 48 h at 37°C/5% CO$_2$. Cells were dislodged by gently shaking the flask and vigorously pipetting medium over the cells. Virus was harvested by transferring resuspended cells to a 50 ml tube and centrifuging at 1000 x g for 3 min, then resuspending in 1 ml of MEME-2. Virus was released from cells by freeze-thawing three times and stored at -80ºC until selection (Section 2.27).

2.27   Selection of recombinant vaccinia virus

When recombinant vaccinia virus is made using the pTM1 transfer vector, the expression cassette is inserted into the non-essential vaccinia virus TK gene. Therefore, recombinant vaccinia virus was selected by infecting a TK-deficient cell line. Human TK-deficient 143B cells (HuTK$^-$ 143B, ATCC #CRL-8303) were a kind gift from Prof Geoffrey Smith (University of Cambridge, UK). HuTK$^-$ 143B cells were maintained in DMEM-complete medium containing 25 µg/ml bromodeoxyuridine (BrdU). BrdU is a synthetic nucleoside analogue of thymidine, which must be phosphorylated by TK in order to be incorporated into DNA. If HuTK$^-$ 143B cells are transfected with wild-type vaccinia virus expressing the TK gene, BrdU becomes phosphorylated and lethally

incorporated into the viral genome. Therefore, only recombinant vaccinia virus with inactivated TK will survive and produce plaques upon infection of HuTK⁻ 143B cells.

3.3×10^5 HuTK- 143B cells/well were seeded in one 6-well plate per virus and incubated for 16 h. 100 μl of recombinant viral lysate was trypsinised with an equal volume of trypsin/EDTA as in Section 2.24. Four 10-fold serial dilutions (ranging from 10^{-1} to 10^{-4}) were made in DMEM-2. Medium was removed from the cells and 1 ml virus dilution was added (10^{-2} to 10^{-4} dilutions in duplicate wells) and incubated for 1-2 h at $37°C/5\%$ CO_2. An agarose overlay was prepared by mixing equal volumes of 2% low-melting point agarose (melted and cooled to 45°C) and 2x selective plaquing medium (2x plaque medium, Life Technologies #21935028), 5% FCS, 240 mM L-glutamine, 0.5% (v/v) pen/strep, 50 μg/ml BrdU). The viral inocula were removed and cells were overlaid with 3 ml selective agarose per well. Cells were incubated for 48 h ($37°C/5\%$ CO_2). A second overlay, containing neutral red stain, was prepared by mixing equal volumes of 2% low melting point agarose and 2x selective plaquing medium + 0.1% neutral red. 2 ml of selective agarose was added to each well and incubated for 16 h. The following day, ten well-isolated plaques were picked using sterile pipette tips and the agarose plugs were transferred to 0.5 ml DMEM-2% (no pen/strep or BrdU). Samples were vortexed vigorously and freeze-thawed 3 times, then snap-frozen on dry ice and stored at -80°C.

2.28   Screening of recombinant vaccinia virus

TK⁻ plaques need to be screened by PCR to check for insertion of the transgene, since spontaneous TK⁻ virus mutants can arise at a frequency of 1:10,000. Recombinants can represent anywhere between 10 to >90% of plaques, depending on the transfection efficiency (Earl et al., 2001). Screening required infecting a small culture of HuTK⁻ 143B cells, harvesting the virus and using a small amount of infected cell lysate in PCR assays.

1.25×10^5 HuTK⁻ 143B cells were seeded in each well of a 24-well plate in 0.5 ml complete DMEM-10/BrdU and incubated ($37°C/5\%$ CO_2) for 16 h until 90% confluent. Plaque isolates (Section 2.27) were thawed and mixed, and each well was infected with 0.25 ml of one plaque isolate and incubated for 2 h ($37°C/5\%$ CO_2) with gentle shaking at 15 min intervals. 1 ml DMEM-2/BrdU was added and cells were incubated for 48 h ($37°C/5\%$ CO_2). Cells were scraped using the rubber end of a 1 ml syringe plunger, transferred to 1.5 ml tubes, and centrifuged for 30 s at top speed to pellet infected cells. Cell pellets were washed once with 1 ml PBS (30 s, 16,000 x g) and resuspended in 0.5 ml PBS.

DNA was extracted from plaque isolates by freeze-thawing three times to lyse the cells. 100 µl of lysate was transferred to a 1.5 ml tube with 10 µl DNA extraction buffer (20 mM Tris-HCl [pH 7.5], 20 mM EDTA, 0.5% SDS, 1 mg/ml Proteinase K), mixed gently by inversion and incubated at 37°C for 3.5 h. DNA extracted from each isolate was purified by phenol-chloroform extraction/ethanol precipitation and resuspended in 20 µl TE. 10 µl of each was then tested for insertion of the RAG transgenes by PCR using the VacV RAG F and VacV RAG R primers (Table 2.1), followed by agarose gel electrophoresis (Section 2.4). To ensure that the recombinant vaccinia virus was clonally pure, the processes of selection and screening (Section 2.27 and 2.28) were repeated twice more, taking four plaque isolates forward each time. Once recombinant viruses had been through three rounds of selection and screening, they were amplified by infection of successively larger numbers of cells according to the amplification protocol in Section 2.24. The first two stages of amplification (12-well plate then 25 cm^2 flask) were done using selective conditions (i.e. HuTK- 143B cell grown in BrdU). Large stocks of full-length RAG1 and full-length RAG2-expressing vaccinia viruses were then produced by infection of HeLa cells in 175 cm^2 flasks. The viral titer was then estimated as in Section 2.25.

2.29   Purification of full-length RAG proteins from vaccinia-infected HeLa suspension cells

Suspension HeLa cells were infected with three viruses: both viruses expressing full-length RAG1 or full-length RAG2 and a recombinant vaccinia virus that expresses T7 RNA polymerase (vTF7-3, a kind gift from Dr John Barr, University of Leeds, UK) at an MOI of 10 for each virus. All three vaccinia viruses were trypsinised as in Section 2.24, added to cells and incubated for 24 h (37°C/5% CO$_2$). Infected cells were harvested by centrifugation at 1000 x g for 5 min at 4°C, washed twice with ice-cold PBS, snap-frozen in a dry ice/ethanol bath and stored at -80°C until purification.

Full-length RAG proteins were purified according to a protocol based on McBlane et al. (1995). All steps were carried out at 4°C. An infected cell pellet was thawed on ice and resuspended in 10 ml buffer C0 (20 mM Tris-HCl [pH7.5], 500 mM NaCl, 1x complete protease inhibitors (Roche), 10% glycerol, 0.1% Triton X-100). Cells were lysed by 20 strokes in a Dounce homogeniser with a tight-fitting pestle and the lysate was cleared by centrifugation for 30 min at 11,000 x g (30,000 rpm), 4°C, using a Beckman SW55 Ti rotor. 1 ml of anti-FLAG M2 resin in an Econo-Pac® column (BioRad) was washed with three column volumes (CV) 0.1 M glycine [pH3.5] and equilibrated with 5 CV of buffer C0 without protease inhibitors. Cleared lysate was applied to the column by gravity flow, and then re-applied to maximise binding. The column was washed with 10

CV Buffer C0 without protease inhibitors, and full-length RAG proteins were eluted with 10 CV of buffer C0 + 100 µg/ml FLAG peptide.

βMe and imidazole were added to the eluate at 2 mM and 20 mM, respectively, and applied to a 1 ml Ni-NTA column attached to an ÄKTA Explorer and equilibrated with buffer A (10 mM Tris-HCl [pH 8.0], 500 mM NaCl, 20 mM imidazole, 2 mM βMe). The column was washed with 10 CV buffer A, and His-tagged RAG proteins were eluted using a gradient of 20 to 500 mM imidazole. Samples were analysed by SDS-PAGE and western-blot (Section 2.30), which indicated that the full-length RAG proteins were largely insoluble, since they were detected in the pellet after centrifugation but not in any subsequent fraction.

2.30 Western blot of full-length RAG proteins

Samples were taken at each step of the full-length RAG purification for detection of full-length RAG proteins by western blot. First, samples were separated by SDS-PAGE (Section 2.19). Prior to transfer, polyvinylidene fluoride (PVDF) membrane (Hybond P, GE Healthcare #10600023) was soaked in 100% methanol, washed four times with dH_2O, then washed in transfer buffer for 10 min (20 mM Tris-HCl [pH 8.0], 150 mM glycine, 20% (v/v) methanol, 0.038% (w/v) SDS). Proteins were transferred from the gel to PVDF membrane using a semi-dry blotter for 1 h at 0.68 mA/cm^2. Non-specific binding sites were blocked for 1 h in 20% horse serum (HS)/PBS on a rocking platform. The blocked membrane was then hybridised in 3 ml 1:2000 monoclonal mouse anti-FLAG antibody (Sigma #F3165) in HS/PBS/1% TWEEN-20 (PBST) solution for 1 h at RT on a roller mixer. Unbound primary antibody was removed by washing with PBST, changing the PBST every 10 min. The secondary hybridization was with 1:10,000 horse radish peroxidase (HRP)-conjugated polyclonal anti-mouse antibody in HS/PBST for 1 h at RT, followed by washing in PBST for 45 min, replacing the PBST every 10 min. Chemiluminescent HRP substrate was prepared according to the manufacturer's instructions (ThermoPierce #34077) and added to the membrane for ~3 min. Blots were then exposed to photosensitive film and developed in a film processor.

(D) **Expression and purification of core and full-length RAG proteins from HEK-293T cells**

2.31 **Cloning MBP-tagged core and full-length RAG coding sequences into a mammalian expression vector**

The full-length coding sequences of RAG1 and RAG2 were PCR amplified from pJH548 and pJH549, respectively, using the Q5 high-fidelity polymerase (NEB #M0491) and primers that add NdeI and BamHI sites. Full-length RAG1 was amplified using primers RAG1 NdeI F and RAG1 BamHI R, and full-length RAG2 was amplified using primers RAG2 NdeI F and RAG2 BamHI R (Table 2.1). Core RAG1 was amplified using primers cRAG1 NdeI F and cRAG1 R. The PCR products were blunt-ligated into pBluescript cut with EcoRV, then subcloned into the MBP-tagging vector pET-MBPL (a kind gift from Dr Chris Thomas, University of Leeds) via the NdeI and BamHI sites, such that the RAG coding sequences were in-frame and downstream of the MBP coding sequence. MBP-tagged RAG sequences were subcloned into pEF-XC (a modified version of pEF-BOS; Mizushima and Nagata, 1990) via the XbaI and BamHI sites, to make pEF-MfR1, pEF-MfR2 and pEF-McR1 (Fig. 2.1-2.3). MBP-tagged core RAG2 in pEF-XC was made by amplifying the MBP-core RAG2 sequence from pEF-MfR2 and adding XbaI and BamHI sites, using primers MBP flRAG F and MBP cRAG2 (Table 2.1). The MBP-cRAG2 fragment was then inserted into the XbaI/BamHI sites of pEF-XC to make pEF-McR2 (Fig. 2.4). DNA sequencing by GATC Biotech (Konstanz, Germany) confirmed the identity of the clones and that no mutations were introduced during PCR.

2.32 **Transfection of HEK-293T cells with full-length RAG expression vectors**

10 cm^2 dishes (typically 15 per transfection) were seeded with 2 x 10^6 HEK-293T cells/dish in complete DMEM-10 and incubated at 37°C/5% CO$_2$ for 16 h. Medium was replaced with fresh medium and incubated for 3 h at 37°C/5% CO$_2$. HEK-293T cells were then co-transfected with pEF-MfRAG1 and pEF-MfRAG2 using polyethylenimine (PEI). Transfection solution was prepared in serum- and antibiotic-free DMEM using 5 μg of each plasmid and 30 μg of PEI per dish in a total volume of 1 ml/dish. The DNA/PEI mixture was vortexed briefly and incubated at RT for 10 min. 1 ml of DNA/PEI mix was added drop-wise to each dish and distributed evenly by gentle rocking, then incubated for 48 h at 37°C/5% CO$_2$. Cells were harvested by scraping and transferring to 50 ml tubes, and centrifuging at 1000 x g for 3 min at 4°C. Cells were washed twice with ice-cold PBS, snap-frozen in dry ice/ethanol and stored at -80°C.

2.33   Purification of full-length RAG proteins from HEK-293T cells

Core and full-length RAG proteins, expressed in HEK-293T cells, were purified in the same way. All steps were performed at 4°C. Cell pellets (typically from two transfections per purification i.e. 30 dishes) were thawed on ice and resuspended in a total of 10 ml of lysis buffer (10 mM NaP$_i$ [pH 7.2], 0.5 M NaCl, 1 mM DTT, 0.25% TWEEN-20, 1x protease inhibitors with EDTA (Roche), 1 mM PMSF, 1 mM benzamidine). Cells were lysed using a Dounce homogeniser by 20 strokes with a tight-fitting pestle, then centrifuged for 40 min at 11,000 x g (30,000 rpm), 4°C, with a SW55 Ti rotor (Beckman). The supernatant was loaded (0.5 ml/min) onto a 1 ml amylose column (NEB # E8022S) equilibrated with buffer A (10 mM NaP$_i$ [pH 7.2], 0.5 M NaCl, 1 mM DTT, 0.25% TWEEN-20). The column was washed with 10 CV of buffer A, followed by 10 CV of buffer B (buffer A without TWEEN-20) and protein was eluted with 10 CV buffer C (Buffer A without TWEEN-20, with 10 mM maltose). Fractions were analysed by SDS-PAGE, and RAG-containing fractions were pooled and dialysed twice for 3 h against dialysis buffer (25 mM Tris-HCl [pH8.0], 150 mM KCl, 2 mM DTT, 10% glycerol). Dialysed protein was aliquoted, snap-frozen in a dry ice/ethanol bath and stored at -80°C.

(E)   *In vitro* RAG assays

2.34   Gel purification of DNA oligonucleotides

All synthetic oligonucleotides, purified by desalting, were supplied by Sigma-Aldrich (Poole, UK). DNA oligonucleotides that were used as *in vitro* RAG substrates (Table 2.2) were gel-purified using the "crush-and-soak" method (Ellington and Pollard, 2001). Oligonucleotides were run on a 12% (19:1) polyacrylamide gel containing 7M urea and 1x TBE for 1 h at 500 V. Gels were wrapped in cling film and oligonucleotides were visualised by UV shadowing using a hand-held UV monitor at 254 nm and the intensifying screen of a western blot cassette, and excised using a sterile scalpel blade. The gel slice was finely chopped on a glass gel plate and transferred to a 1.5 ml tube in 1 ml TE, 100 mM NaOAc, and 0.5% SDS to prevent nuclease activity during elution. Tubes were heated in a boiling water bath for 3 min, then freeze-thawed 3 times in dry ice and a 42°C water bath. The oligonucleotide was left to diffuse from the gel matrix on a rotating wheel at RT overnight. Polyacrylamide was removed by centrifugation at top speed for 2 min, and supernatant was transferred to a fresh 1.5 ml tube. Residual polyacrylamide was removed by filtering through a 0.2 µm syringe filter unit. The volume of oligonucleotide solution was reduced to 400 µl by butanol extraction, prior to purification by phenol-chloroform extraction/ethanol precipitation. DNA pellets were

resuspended in TE and the concentration of purified oligonucleotide was measured using a NanoDrop spectrophotometer (Section 2.8).

2.35 Preparation of radiolabelled oligonucleotides

The "top" strand of each RSS-containing oligonucleotide (Table 2.1) was 5'-end labelled with $[\gamma-^{32}P]$-ATP using T4 polynucleotide kinase (T4 PNK, NEB #M0201S). Labelling reactions containing 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 $[\gamma-^{32}P]$-ATP (Perkin-Elmer Life Sciences) in a total volume of 15 µl were incubated at 37°C for 30 min. The volume was increased to 50 µl with TE and unincorporated $[\gamma-^{32}P]$-ATP was removed using MicroSpin^{TM} G-25 columns (GE Healthcare) for 2 min at 700 x g. Purified radiolabelled oligonucleotide was then annealed to its unlabelled reverse complement (Section 2.36).

2.36 Annealing oligonucleotides

Oligonucleotide annealing reactions contained both complementary strands at 100 nM, 100 mM NaCl, and 0.5x TE in a total volume of 100 µl. Annealing reactions were then heated in screw-cap tubes for 3 min in a boiling water bath and left to cool gradually overnight. The final concentration of annealed oligonucleotide was 100 nM. Radiolabelled and unlabelled oligonucleotides were annealed in the same way.

2.37 *In vitro* RAG cleavage assays

A typical RAG cleavage reaction contained 25 mM MOPS pH 7.0, 50 mM potassium acetate, 1 mM DTT, 100 µg/ml BSA and 1 mM divalent metal cation chloride salt (either $CaCl_2$, $MnCl_2$ or $MgCl_2$) prepared as a 10x concentrated mix, 2 nM radiolabelled oligonucleotide, 30 ng (120 nM) HMGB1, and 4 µl of purified RAG proteins (usually ~ 80 nM) in a total volume of 10 µl. Where indicated, a second unlabelled partner oligonucleotide was included. Reactions were incubated at 37°C, usually for 15 min (or as indicated in the figure legend), and stopped by addition of 3.3 µl of 4x stop buffer (200 mM Tris-HCl [pH 8.0], 0.4% SDS, 20 mM EDTA, 0.7 mg/ml proteinase K, 4x DNA loading buffer), followed by incubation at 37°C for 30 min. SDS and proteinase K are included in the stop buffer to degrade the RAG proteins. If this step is omitted, a portion of the radiolabelled oligonucleotide remains bound by RAG proteins and HMGB1, and becomes stuck in the wells of the gel during electrophoresis, thus preventing accurate quantification of the level of cleavage. The whole reaction was loaded onto a 12% (19:1) polyacrylamide gel in 1x TAE (17 cm x 15 cm x 1.6 mm), and separated for 1.5 h at 170 V (constant). Gels were soaked in 70% methanol on an orbital shaker for 20 min at RT to prevent them from cracking when drying due to the high percentage of

polyacrylamide. Gels were dried on to Whatman® 3MM chromatography paper using a Bio-Rad gel drier with an 80°C heated lid for 1 h. The dried gel was exposed to a phosphorimager screen and visualised using an FLA-1500 phosphorimager (Fujifilm). The percentage of substrate cleaved was quantified using AIDA Image Analyser software (Raytest, Germany, v.4.14).

2.38 Analysis of RAG cleavage activity by denaturing gel electrophoresis

To analyse the nicking and hairpinning activity of RAG proteins, cleavage assays were set up as described in Section 2.37. Then, half of the reaction was added to 4x stop buffer and run on a native polyacrylamide gel as described in Section 2.37. The remaining half of the reaction was purified by phenol-chloroform extraction/ethanol precipitation, resuspended in 5 µl formamide loading buffer, and run on a 12% (19:1) polyacrylamide gel containing 7 M urea. DNA was electrophoresed for 1 h at 500 V and the gel was transferred to Whatman 3 MM paper. Gels were dried and processed as in Section 2.37.

2.39 Electrophoretic mobility shift assay (EMSA)

RAG-RSS binding assays were identical to cleavage reactions, except that they were incubated at 37°C for 10 min only in the presence of either $MgCl_2$ or $CaCl_2$, as indicated. 1 µg of monomeric streptavidin (Thermo #21125) per reaction was included in the streptavidin supershift assay (Section 4.1). After incubation, reactions were chilled on ice for 5 min and loaded onto a 4% (19:1) polyacrylamide gel (17 cm x 15 cm x 1.6 mm) in 0.5x TBE, pre-chilled to 4°C in a cold room. The gels were run and dried as for cleavage assays (Section 2.37), except the 70% methanol soak was omitted. Dried gels were exposed to a phosphorimager screen and visualised using an FLA-1500 phosphorimager (Fujifilm).

2.40 Estimating RAG-RSS and RAG-ESC dissociation constant (K_d) by EMSA

The affinity of RAG-ESC binding was estimated by EMSA. Binding reactions were set up as in Section 2.39, with core RAG concentration ranging from 0-80 nM. The amount of dialysis buffer was kept constant by keeping the combined volume of RAG and dialysis buffer at 4 µl/reaction. This was to ensure that the concentration of monovalent cations and glycerol, which might affect binding efficiency, was the same for each RAG concentration. Gels were run and processed as in Section 2.39, and the fraction of bound and unbound oligonucleotide was measured by densitometry. The dissociation constants (K_d) were estimated by plotting RAG concentration against % bound probe in Excel, and calculating the concentration of RAG protein at which 50% of probe was bound.

2.41   DNase I footprint analysis of RAG-RSS binding

The 12RSS footprint probe was prepared by digesting 100 µg pJβ2.7+, a 12RSS-containing plasmid, with EcoRV and NdeI. The 12RSS containing fragment was gel-purified, resuspended in dH$_2$O, and radiolabelled by filling in the recessed NdeI end with α-^{32}P-dCTP and Klenow polymerase. Labelling reactions, containing the EcoRV/NdeI fragment from pJβ2.7+, 5 µl α-^{32}P-dCTP, 1x NEBuffer 2, and 10 units of Klenow polymerase, were incubated for 30 min at RT, then supplemented with 5 mM dNTP mix to complete the fill-in reaction. The labelled fragment was then purified by phenol-chloroform extraction/ethanol precipitation and resuspended in 50 µl TE. The ESC footprint probe was prepared in the same way, except that the NheI/ApaI fragment from pJδ1+ was used.

Binding reactions containing 1x buffer (25 mM MOPS [pH 7.0], 0.1 mg/ml BSA, 42 mM KOAc, 2 mM DTT, 1 mM CaCl$_2$), 12RSS probe or ESC probe (~30 counts per second), 20% DMSO, 0.2 µg polydI.dC, and 100 nM cRAG proteins, were incubated at 37°C for 10 min. MgCl$_2$ (5 mM final concentration) and 2 units of DNase I were added and the mixture was incubated at RT for 1 min. DNase I reactions were quenched by the addition of EDTA at 15 mM, and then purified by phenol-chloroform/ethanol precipitation, using GlycoBlue (Life Technologies #AM9516) to aid in the visualisation of the small radioactive DNA pellet. DNA was resuspended in 3 µl formamide loading buffer, boiled for 3 min and run alongside the appropriate GA marker on a pre-warmed 6% (19:1) denaturing polyacrylamide gel for 45 min at 38 mA. Gels were transferred to Whatman 3MM paper and dried in a BioRad vacuum drier for 45 min with an 80°C heated lid. The dried gel was exposed to a phosphorImager screen and scanned using an FLA-1500 PhosphorImager (FujiFilm). The relative density of each band was measured using ImageJ.

2.42   Preparation of GA ladder

Preparation of a GA ladder was required for determining the identity of each band in the DNase I footprint experiments. The same radiolabelled probe that is used in the footprint is partially chemically degraded at only the G and A residues, thus creating a radiolabelled ladder where each G and A residue of the probe is seen on a gel. The GA ladder was prepared by precipitating a fifth of the labelled probe stock (Section 2.41) and resuspending it in 10 µl dH$_2$O. 25 µl 100% formic acid was added and the mixture was incubated for 2 min at RT, and quenched by addition of 200 µl stop buffer (300 mM NaOAc [pH 5.2], 0.1 mM EDTA [pH 8.0], 0.1 mg/ml yeast tRNA) and 150 µl 100% ethanol chilled to -20°C. DNA was precipitated by centrifugation at 14,000 rpm for 10 min and washed twice with 70% ethanol. The pellet was resuspended in 200 µl dH$_2$O

and precipitated once more, washed with 70% ethanol, and dried in a SpeedVac. The dried pellet was resuspended in 70 µl 10% (v/v) piperidine, transferred to a screw-cap tube and incubated in a 90°C heat block for 30 min. Piperidine was evaporated in a SpeedVac, and the pellet was resuspended in 30 µl of dH$_2$O. The DNA was then transferred to a fresh tube to minimise piperidine contamination, and dried again in a SpeedVac. This was repeated twice more, resuspending in 20 µl dH$_2$O the first time, and 10 µl formamide loading buffer the second time.

(F)   *In vivo* RAG activity assays

2.43   Cloning of 23RSS-containing plasmid substrate

The 23RSS-containing plasmid used in extrachromosomal RAG activity assays (Section 2.44-2.47) was made by replacing the 12RSS in pJβ2.7+ (Fig. 2.5). Briefly, oligonucleotides containing a consensus 23RSS oligonucleotide with terminal BamHI and HindIII sites (23RSSF and 23RSSR, Table 2.1) were annealed and phosphorylated using T4 PNK as in Section 2.35, except non-radioactive dATP was used instead of [γ-^{32}P]-ATP. The phosphorylated 23RSS oligonucleotide was then ligated (Section 2.9) with pJβ2.7+ cut with BamHI and HindIII to make p23+ (Fig. 2.6).

2.44   Transfection of NIH/3T3 cells

NIH/3T3 cells were maintained in DMEM-complete, splitting 1:10 when a confluency of ~90% was reached. NIH/3T3 cells were plated at 1 x 10^6 cells/10 cm dish in 10 ml DMEM-complete the day before transfection. Cells were transfected with 2 µg of RSS-containing plasmid and/or an ESC plasmid, and 1 µg each of RAG1 and RAG2 expression vectors using PEI as in Section 2.32. Cells were harvested by scraping 48 h after transfection. Plasmid DNA was extracted using the method described by Hirt (1987; Section 2.42), and resuspended in 20 µl of dH$_2$O. 1 µl of each sample was analysed by qPCR (Section 2.46). The amount of intact 12RSS plasmid (pJβ2.7+) was quantified using primer pair re-int1 and re-int2, and the amount of intact 23RSS plasmid (p23+) using 23RSS fwd and 23RSS rev. The total amount of 12RSS (pJβ2.7+) or 23RSS plasmid (p23+) was quantified using primers Neo3 fwd and Neo3 rev. The amount of intact ESC plasmid (pJδ1+) was quantified using primer pair Jdcut fwd and Jdcut rev. The total amount of pJδ1+ was quantified using primers Jdnorm fwd and Jd norm rev (all primer sequences can be found in Table 2.3). Thermal cycling conditions were as follows:

```
95°C   6 min
95°C   10 s  ⎫
Tm     15 s  ⎬  40 cycles
72°C   10 s  ⎭
```

2.45   Extraction of plasmid DNA from transfected mammalian cells

Plasmid DNA was extracted from transfected NIH/3T3 cells using the method described by Hirt (1967). Transfected cells were washed twice with 3 ml of PBS, and lysed directly on the plate by addition of 0.8 ml of buffer 1 (0.6% (w/v) SDS, 0.01 M EDTA [pH 7.5]), and incubation at RT for 10 min. 0.2 ml of buffer 2 (5 M NaCl) was added and mixed by swirling to neutralize, and left for 2 min, after which the cells were scraped into a 1.5 ml tube and stored at 4°C overnight. SDS and protein was removed by centrifugation at 14,000 rpm, 4°C, for 40 min. The supernatant was transferred to a fresh tube and DNA was cleaned by two rounds of phenol-chloroform extraction/ethanol precipitation. 1 ml 100% butanol was added, mixed vigorously, and spun at top-speed for 1 min. The butanol layer was removed and DNA was precipitated in 100% ethanol and NaOAc, then washed in 1 ml 70% ethanol (Hirt, 1967).

2.46   Quantitative PCR

A typical quantitative PCR contained 5 µl 2x SensiFAST SYBR® No-ROX mix (Bioline #BIO-98080), 400 nM of forward and reverse primers (Table 2.1) and 1 µl of sample in a total volume of 10 µl. PCRs were carried out in a Corbett RotorGene 6000 qPCR machine, and analysed using the Corbett Rotor-Gene 6000 Series Software (v.1.7, build 87). In each case, a standard curve of samples containing the amplified region was analysed alongside in order to calculate the relative amount of product in each of the unknown samples. A no-template control was included in each run to test for cross-contamination.

2.47   Ligation-mediated PCR (LM-PCR) of extrachromosomal substrates

Plasmid DNA extracted from transfected NIH/3T3 cells (Section 2.44) was used in an LM-PCR assay to determine whether a decrease in qPCR product (Section 2.46) was caused by double-strand breaks at the plasmid RSS, or by nicking of the plasmid RSS, since qPCR cannot distinguish between these two possibilities. Only double-strand breaks will ligate with a linker oligonucleotide and give an LM-PCR product. 5 ng of extracted plasmid DNA digested with EcoRV was used as a positive control and processed in the same way as undigested. FM25 and FM11 oligonucleotides (Table 2.3) were annealed and used in a ligation reaction containing 5 ng extracted plasmid

DNA (from Section 2.45), 3 µl 10x T4 DNA ligase buffer and 5 units T4 DNA ligase in a final volume of 30 µl. Ligation reactions were incubated at 16°C for 16 h, then purified by phenol-chloroform extraction/ethanol precipitation and resuspended in 10 µl dH$_2$O. In the first round of PCR, 2 µl resuspended DNA was used in a reaction containing 3 µl 10x ThermolPol buffer, 1 µl Vent polymerase (NEB #M0254), 0.6 µM each of CMVfwd and FM25 primers (Table 2.3), and 300 nM each dNTP in a total of 30 µl. Cycling conditions were as follows:

95°C 5 min

95°C 30 s ⎫
57°C 20 s ⎬ 25 cycles
72°C 30 s ⎭

72°C 7 min

In the second round of PCR, 1 µl first round PCR product was used in a reaction identical to the first round of PCR, except using the 12RSS fwd and FM25 primers (Table 2.3). Cycling conditions were also identical, except the temperature of the annealing step was 60°C. Reaction products were then separated by agarose gel electrophoresis, dried in a vacuum drier with a heated lid, and visualised using an FLA-1500 phosphorimager (FujiFilm).

2.48   Quantitative PCR of Vλ1-JCλ1 recombination in mouse pro-B cells

Purified samples of transgenic (PIP3; Bevington and Boyes, 2013) and non-transgenic mouse pro-B cell DNA were a kind gift from James Scott (Boyes laboratory, University of Leeds, UK). Amplification of recombined Vλ1-JCλ1 recombined gene segments required an initial round of standard PCR (Section 2.5) to improve specificity of the quantitative step, using primers V1SacIF and J1StyIR (Table 2.3). The following thermal profile was used:

95°C 5 min
95°C 30 s ⎫
55°C 20 s ⎬ 15 cycles
68°C 30s   ⎭
68°C 5 min

Products from the first round of PCR were then diluted 1:10 with dH$_2$O, and 1.5 µl was used in a nested qPCR assays, using primers V1RSSF and J1RSSR (Section 2.46).

Vλ1-Jλ1 ESC was quantified using 1 µl of each DNA sample and primers ESC F and ESC V1 R (Table 2.3). To test the specificity of ESC amplification, a portion of each sample was digested with ApaLI restriction enzyme (NEB #R0507), which cleaves the ESC at the head-to-head RSS junction, for 16 h at 37°C. Digestions were quenched by SDS/Proteinase K digestion (0.5% SDS and 0.2 µg/µl proteinase K) at 56°C for 1 h, and purified by phenol-chloroform extraction/ethanol precipitation. In all cases, the relative amount of DNA in each sample was quantified using primers that amplify the GAPDH gene (GAPDH F and GAPDH R2, Table 2.3). Data were then normalised by dividing the amount of Vλ1-Jλ1 recombination or Vλ1-Jλ1 ESC product by the amount of GAPDH product.

Table 2.1 – Oligonucleotides used for cloning. All oligonucleotides were synthesised then purified by desalting by Sigma (Poole, UK).

Oligonucleotide	Sequence (5'-3')
FLAG-His adaptor F	GGATCCCATGGCACATCATCACCATCACCACCATCACCATCCAGACTACAAGGACGACGACGACAAACCAGAATTCCTA
FLAG-His adaptor R	TAGGAATTCTGGTTTGTCGTCGTCGTCCTTGTAGTCTGGATGGTGATGGTGGTGATGGTGATGATGTGCCATGGGATCC
RAG1 SmaI F	CAGACCCGGGATGGCTGCCTCCTTGCCGTCT
RAG1 XhoI R	ACTCCTCGAGGGAGAAGACCTCACTGCAACCC
RAG2 EcoRI F	CCGAATTCATGTCCCTGCAGATGGTAACAGTG
RAG2 XhoI R	GTCCTCGAGGTCCGACGGTATCGATAAGCTTG
VacV RAG F	TCAACAAGGGGCTGAAGGATGC
VacV RAG R	TTTCGGGCTTTGTTAGCAGCCG
RAG1 NdeI F	GACTACCATATGATGGCTGCCTCCTTGCCG
RAG1 BamHI R	TCAGTCGGATCCAAATACCAACTTCTATGTGG
RAG2 NdeI F	GACTACCATATGATGTCCCTGCAGATGGTAAC
RAG2 BamHI R	TCAGTCGGATCCCAATATACCTGAGTCTGAGG
MBP flRAG F	TCGATGTCTAGACCACCATGGAAATCGAAGAAGGTAAAC
MBP cRAG2	TAGCATGGATCCTTAGTCGACACTGAAACAAAATTCCTCTGAGTC
cRAG1 F NdeI	ATGCTACATATGGTGCATATCAATAAAGGGGGACG
cRAG R	TAGCATGGATCCTTACGCGTTATGAGCATTCATA
23RSSF	AGCTTCCAAAAACATGTCGGTCTGTCACCTCATGATGGTGACACG
23RSSR	GATCCGTGTCACCATCATGAGGTGACAGACCGACATGTTTTTGGA

Table 2.2 – Oligonucleotides used as *in vitro* RAG substrates. Heptamer and nonamer sequences are underlined. All oligonucleotides were synthesised and desalted by Sigma, then purified by denaturing PAGE (Section 2.34).

Oligonucleotide	Sequence (5'-3')	Reference
DAR39 (12RSS top)	GATCTGGCCTGTCTTA<u>CACAGTG</u>CTACAGACTGGA<u>ACAAA</u><u>AACC</u>CTGCAG	McBlane et al. (1995)
DAR40 (12RSS bottom)	CTGCAG<u>GGTTTTTGT</u>TCCAGTCTGTAG<u>CACTGTG</u>TAAGAC AGGCCAGATC	McBlane et al. (1995)
DG61 (23RSS top)	GATCTGGCCTGTCTTA<u>CACAGTG</u>GTAGTACTCCACTGTCT GGCTGT<u>ACAAAAACC</u>CTGCAG	McBlane et al. (1995)
DG62 (23RSS bottom)	CTGCAG<u>GGTTTTTGT</u>ACAGCCAGACAGTGGAGTACTAC<u>C</u> <u>ACTGTG</u>TAAGACAGGCCAGATC	McBlane et al. (1995)
MBNSJ top (ESC top)	CTGCAG<u>GGTTTTTGT</u>TCCAGTCTGTAG<u>CACTGTG</u>CACAGT GGTAGTACTCCACTGTCTGGCTGT<u>ACAAAAACC</u>CTGCAG	Neiditch et al (2002)
MBNSJ bottom (ESC bottom)	CTGCAG<u>GGTTTTTGT</u>ACAGCCAGACAGTGGAGTACTAC<u>C</u> <u>ACTGTG</u>CACAGTGCTACAGACTGGA<u>ACAAAAACC</u>CTGCA G	Neiditch et al. (2002)
DAR81 (nonspecific top)	GATCTCGCCTCTCTTAGGTTAATCCTATAGAACTCGTCCC CGTACCTCGAG	Ramsden et al., 1996
DAR82 (nonspecific bottom)	CTAGAGCGGAGAGAATCCAATTAGGATATCTTGAGCAGG GGCATGGAGCTC	Ramsden et al., 1996
ESC 12d7 top	CTGCAG<u>GGTTTTTGT</u>TCCAGTCTGTAG<u>CACAGTG</u>GTAGTA CTCCACTGTCTGGCTGT<u>ACAAAAACC</u>CTGCAG	-
ESC 12d7 bottom	GACGTC<u>CCAAAAACA</u>AGGTCAGACATC<u>GTGTCAC</u>CATCAT GAGGTGACAGACCGACA<u>TGTTTTT</u>GGGACGTC	-
ESC 23d7 top	CTGCAG<u>GGTTTTTGT</u>TCCAGTCTGTAG<u>CACTGTG</u>GTAGTA CTCCACTGTCTGGCTGT<u>ACAAAAACC</u>CTGCAG	-
ESC 23d7 bottom	GACGTC<u>CCAAAAACA</u>AGGTCAGACATC<u>GTGACAC</u>CATCA TGAGGTGACAGACCGACA<u>TGTTTTT</u>GGGACGTC	-
ESC 23d9 top	CTGCAG<u>GGTTTTTGT</u>TCCAGTCTGTCG<u>CACTGTG</u>CACAGT GGTAGTACTCCACTGTCTGGCTGTCTGCAG	-
ESC 23d9 bottom	CTGCAGACAGCCAGACAGTGGAGTACTAC<u>CACTGTG</u>CAC <u>AGTG</u>CGACAGACTGGA<u>ACAAAAACC</u>CTGCAG	-
12d7 top	GATCTGGCCTGTCTTACTACAGACTGGA<u>ACAAAAACC</u>CTG CAG	-
12d7 bottom	CTGCAG<u>GGTTTTTGT</u>TCCAGTCTGTAGCACTGTGTAAGAC AGGCCAGATC	-
23RSSm9 top	GATCTGGCCTGTCTTA<u>CACAGTG</u>GTAGTACTCCACTGTCT GGCTGTGCGCTTATTCTGCAG	-
23RSSm9 bottom	CTGCAG<u>AATAAGCGC</u>ACAGCCAGACAGTGGAGTACTACC ACTGTGTAAGACAGGCCAGATC	-

Table 2.3 – Primers used for LM-PCR and qPCR assays. All oligonucleotides were synthesised and purified by desalting by Sigma (Poole, UK). The T_m used for each primer pair is shown, as calculated using the online OligoCalc software (Kibbe, 2007), using the salt-adjusted formula.

Oligonucleotide	Sequence (5'-3')	T_m (°C)
Re-int 1	GAGCTCGGGGAGCCTTAG	60
Re-int 2	AATCTCACCTGTGACCGTGAG	
23RSS fwd	AAACATGTCGGTCTGTCACC	60
23RSS rev	TCTGCAGAATTCCACCACAC	
Neo3 fwd	TGCTCCTGCCGAGAAAGTATC	60
Neo3 rev	TTTCGCTTGGTGGTCGAATG	
Jdcut fwd	CGGTTTTTGGAACGTCCTCAAG	60
Jdcut rev	TTTTCCAAAGCGGTCGACAG	
Jdnorm fwd	GTCGCCTTCTTGCACATTACTC	60
Jdnorm rev	AACAACAGATGGCTGGCAAC	
FM25	GCGGTGACTCGGGAGATCTGAAGTG	-
FM11	CACTTCAGATC	-
CMVfwd	CGCAAATGGGCGGTAGGCGTG	-
12RSS fwd	CTTCCTGTGCTGGGAGACCT	-
V1SacIF	GCCAACTGGGTCCAAGAAAAAC	58
J1StyIR	GCACCTCAAGTCTTGGAGAG	
V1RSSF	GGCACAGACTGAGGATGAG	55
J1RSSR	CAGTCAGTTTGGTTCCTCCA	
ESC F	GATGTCACCACCTTGGAAGAA	58
ESC V1R	GAAGATGGTAGTTATGAGACTGTACCAG	
GAPDH F	ACTTTCTTGTGCAGTGCCAGC	56
GAPDH R2	GCACACTTCGCACCAGCATC	

2.49   Plasmid Maps

Maps for all plasmids used in this thesis are provided below. Only the final clones are given; cloning intermediates and source vectors are not shown. The complete sequence of pETM11-HMGB1 was not provided by Prof Bianchi so a map could not be made. Plasmid maps were generated using SnapGene software (v2.5).
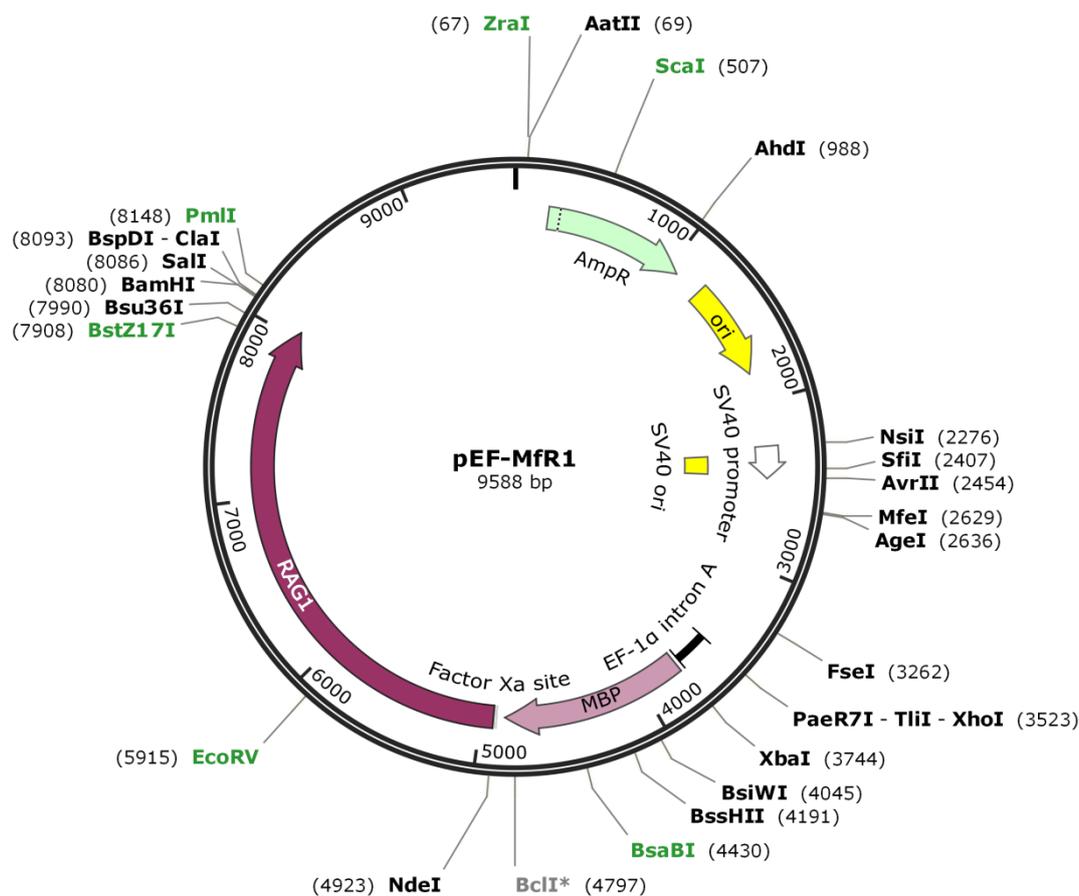
Figure 2.1 – Map of pEF-MfR1. All unique restriction sites with a recognition sequence of 6 bp or more are shown. Blunt cutters are shown in green. AmpR = Ampicillin resistance gene, ori = origin of replication, SV40 = simian virus 40, MBP = maltose-binding protein, EF-1α intron A = elongation factor-1α promoter.
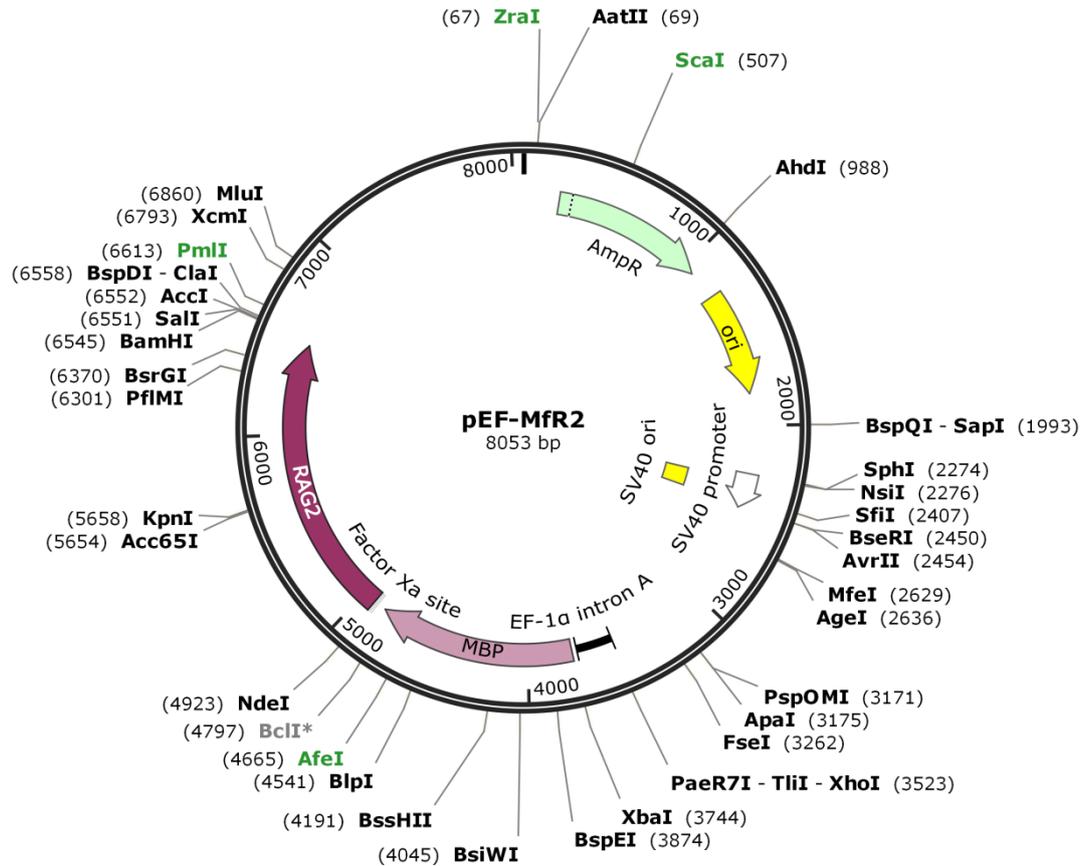
Figure 2.2 – Map of pEF-MfR2. All unique restriction sites with a recognition sequence of 6 bp or more are shown. Blunt cutters are shown in green. AmpR = Ampicillin resistance gene, ori = origin of replication, SV40 = simian virus 40, EF-1α intron A = elongation factor-1α promoter, MBP = maltose-binding protein.
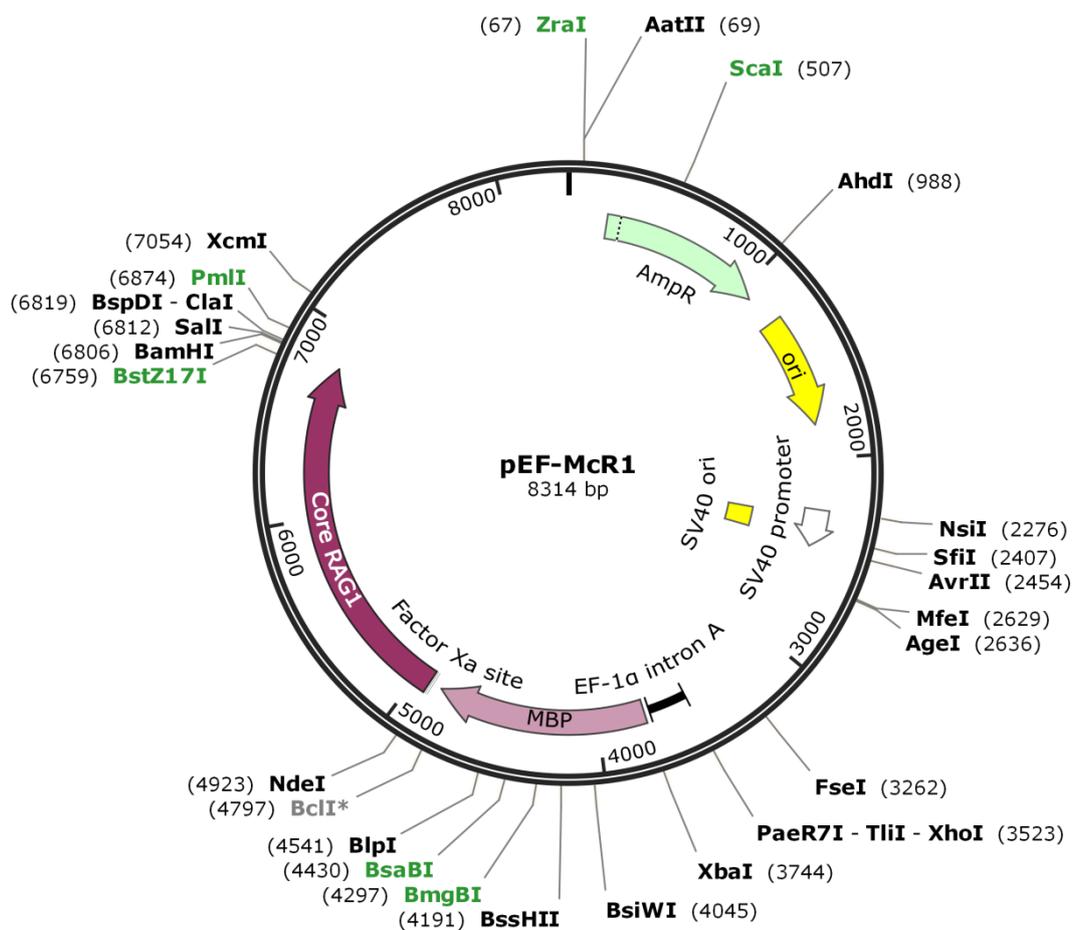
Figure 2.3 – Map of pEF-McR1. All unique restriction sites with a recognition sequence of 6 bp or more are shown. Blunt cutters are shown in green. AmpR = Ampicillin resistance gene, ori = origin of replication, SV40 = simian virus 40, EF-1α intron A = elongation factor-1α promoter, MBP = maltose-binding protein.
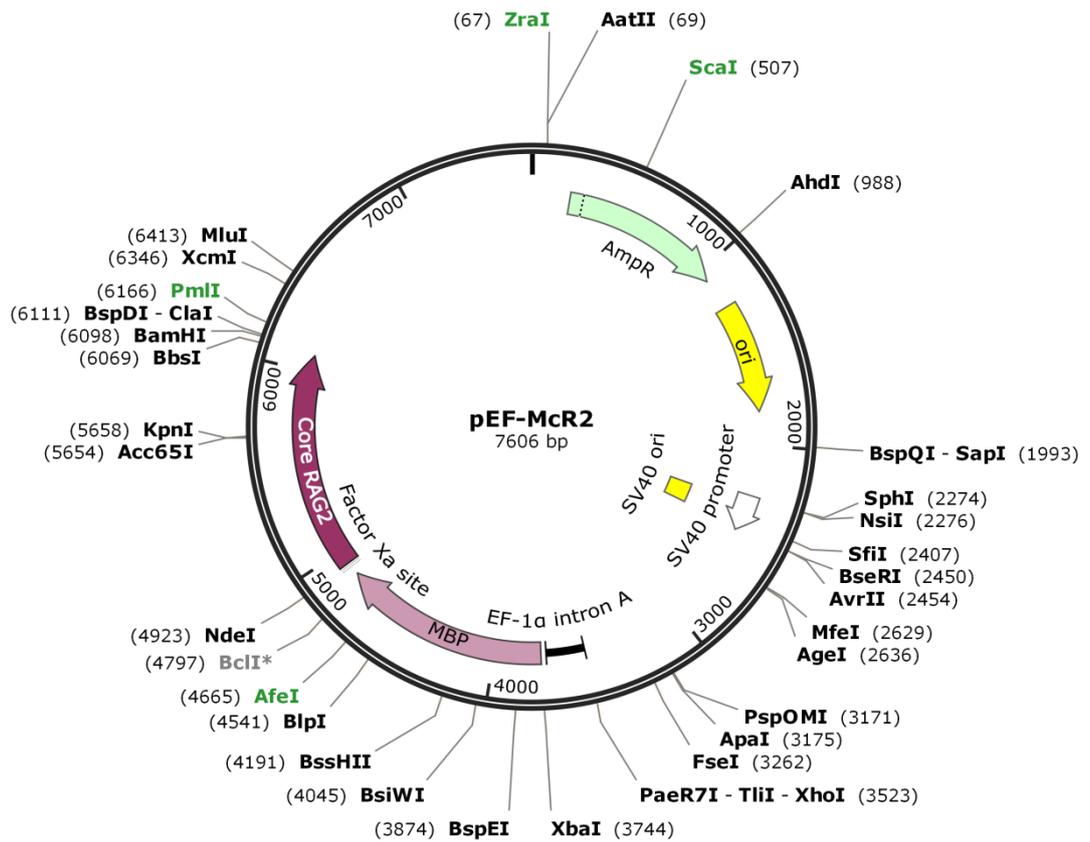
Figure 2.4 – Map of pEF-McR2. All unique restriction sites with a recognition sequence of 6 bp or more are shown. Blunt cutters are shown in green. AmpR = Ampicillin resistance gene, ori = origin of replication, SV40 = simian virus 40, MBP = maltose-binding protein coding sequence, EF-1α intron A = elongation factor-1α promoter.
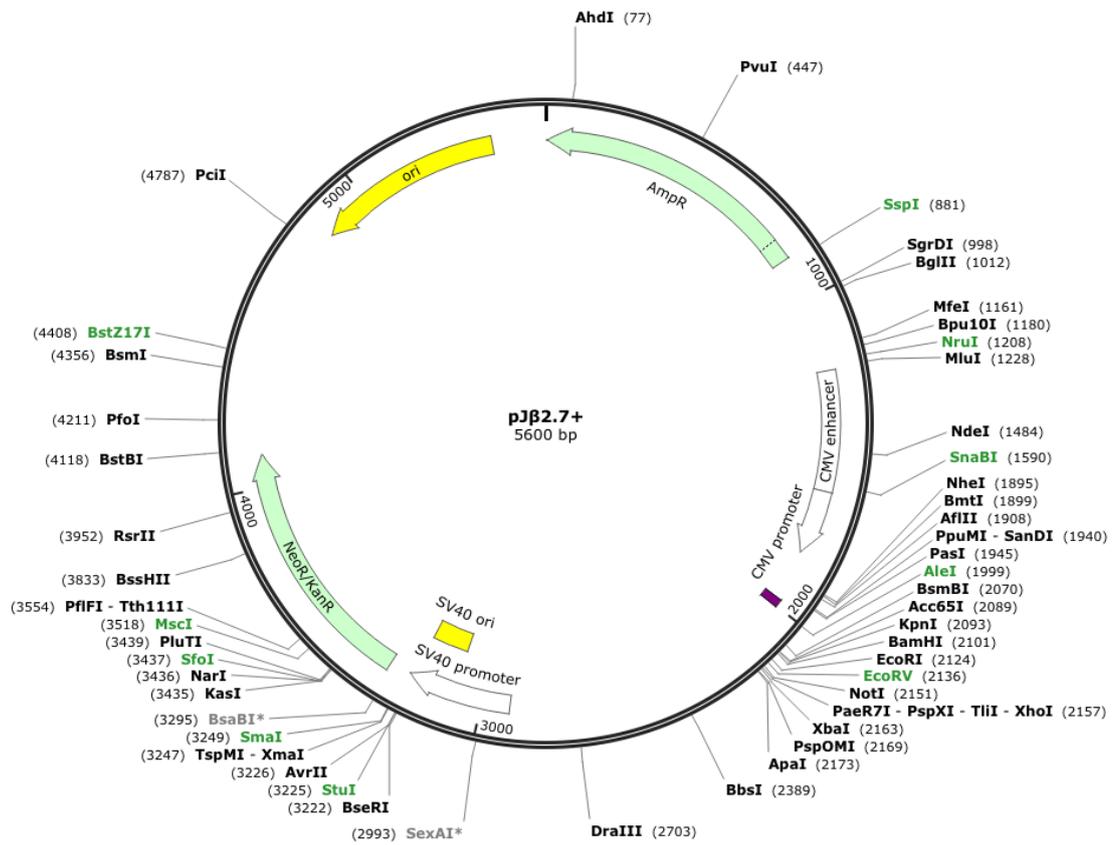
Figure 2.5 – Map of pJβ2.7+. pJβ2.7+ contains the CMV promoter upstream of a 12RSS to enable transcription through the 12RSS. All unique restriction sites with a recognition sequence of 6 bp or more are shown. Blunt cutters are shown in green. AmpR = Ampicillin resistance gene, CMV = human cytomegalovirus, ori = origin of replication, SV40 = simian virus 40, NeoR/KanR = neomycin/kanamycin resistance gene.
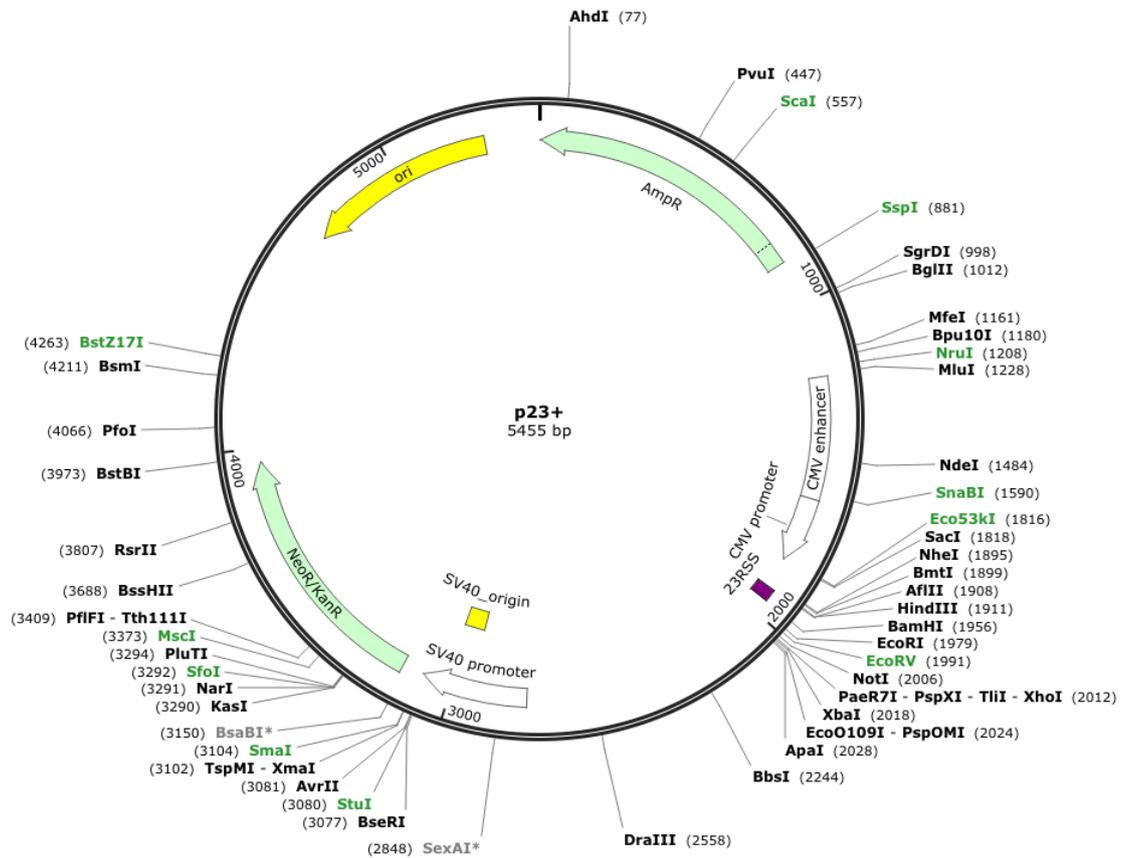
Figure 2.6 – Map of p23+. All unique restriction sites with a recognition sequence of 6 bp or more are shown. Blunt cutters are shown in green. AmpR = Ampicillin resistance gene, CMV = human cytomegalovirus, ori = origin of replication, SV40 = simian virus 40, NeoR/KanR = neomycin/kanamycin resistance gene.

AhdI (234)
FspI (456)
PvuI (604)
ScaI (714)
(4101) PciI
(3985) BspQI - SapI
(3722) BstZ17I
(3670) BsmI
ori
AmpR
SspI (1038)
SgrDI (1155)
BglII (1169)
pJδ1+
4757 bp
SV40 ori
MfeI (1318)
Bpu10I (1337)
NruI (1365)
MluI (1385)
(3409) AvrII
(3408) StuI
(3405) BseRI
SV40 promoter
SnaBI (1747)
CMV promoter
CMV enhancer
NheI (2052)
BmtI (2056)
AflIII (2065)
HindIII (2068)
Acc65I (2074)
KpnI (2078)
BamHI (2086)
EcoRI (2109)
EcoRV (2121)
EagI - NotI (2136)
BsrGI (2208)
(3176) SexAI*
Signal Joint
(2886) DraIII
AvaI - BsoBI - PaeR7I - PspXI - TliI - XhoI (2340)
BmeT110I (2341)
XbaI (2346)
EcoO109I - PspOMI (2352)
ApaI (2356)
(2572) BbsI
(2370) BclI*
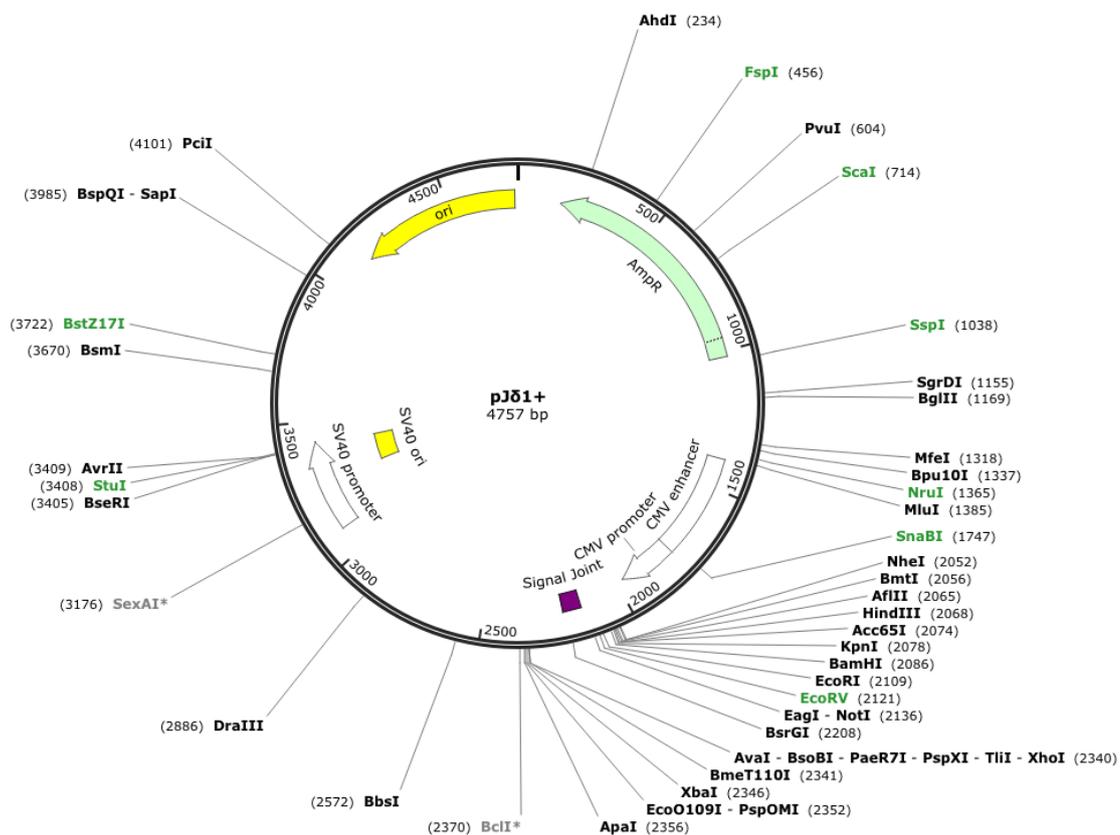
Figure 2.7 – Map of pJδ1+. All unique restriction sites with a recognition sequence of 6 bp or more are shown. Blunt cutters are shown in green. AmpR = Ampicillin resistance gene, CMV = human cytomegalovirus, Signal Joint = head-to-head 12RSS-23RSS, ori = origin of replication, SV40 = simian virus 40.

Chapter 3

Analysis of RAG cleavage in an RSS-ESC complex

A)   Introduction

It is well-documented that mistakes in V(D)J recombination trigger leukaemias and lymphomas via chromosome translocations, and these translocations seem to be the major way in which V(D)J recombination causes cancer (Kuppers, 2005). In the late 1990s, the idea emerged that the RAG recombinase could trigger leukaemia via a process called transposition (van Gent et al., 1996b; Agrawal et al., 1998; Hiom et al., 1998). In a transposition reaction, the ESC is reinserted back into the genome at any location in a non-sequence specific reaction (Hiom et al., 1998). Whilst random insertion of broken signal ends has been proven to occur *in vitro*, to date there is only limited evidence that this reaction occurs *in vivo* (Messier et al., 2003; Messier et al., 2006). Importantly, no transposition events have been detected in cancer patients.

Two papers published in the same year showed that ESCs could be reinserted into the genome in a different way, via a process called reintegration (Curry et al., 2007; Vanura et al., 2007). During reintegration, RAG proteins form a complex between an ESC and a genomic RSS, catalyse the cleavage of both substrates, and transfer the broken ends to the NHEJ proteins so that the ESC is reinserted back into the genome. Reintegration differs from transposition in that reintegration is sequence-specific, occurring only at a genomic RSS, and proceeds via the same biochemical mechanism as standard V(D)J recombination (Vanura et al., 2007). This reaction could occur at a legitimate RSS, or perhaps more dangerously, at a cryptic RSS. There are predicted to be over 15 million cryptic RSSs scattered throughout the human genome (Lewis et al., 1997), and a number of these are adjacent to oncogenes (Boehm et al., 1988; Kuppers and Dalla-Favera, 2001). Some ESC by-products, depending on their origin, will contain strong promoter elements from the antigen receptor loci (Fig. 1.11B). Therefore, it is feasible that an ESC containing a promoter could undergo reintegration adjacent to an oncogene, thereby upregulating the expression of that oncogene, and contribute to leukaemia and lymphoma formation.

One of the studies that reported reintegration of ESCs also found that mice expressing core RAG2 gave a higher level of reintegration events, suggesting that the RAG2 C-terminus can inhibit reintegration (Curry et al., 2007). In order determine whether there

are any key differences in RAG-RSS and RAG-ESC interactions that could either explain the mechanism by which the RAG2 C-terminus inhibits reintegration, or be exploited by small molecule inhibitors to completely inhibit reintegration, I carried out *in vitro* analyses of RAG activity at RSSs versus ESCs.

The first part of this chapter describes the purification of core RAG1 and core RAG2 (cRAG) and high mobility group B1 (HMGB1) proteins required for the *in vitro* analyses and the optimization of the reaction conditions required to accurately recapitulate *in vivo* recombinase activity. Once I had determined suitable reaction conditions, I found that cleavage and stimulation of 12 and 23RSSs is roughly equivalent, but that cleavage of the ESC is less efficient, indicating an asymmetry in cleavage between RSSs and ESCs. Whilst not entirely at odds with the model of reintegration, I put forward an alternative mechanism to describe the outcome of the majority of RAG-ESC-RSS interactions called cut-and-run. In cut-and-run, an ESC stimulates cleavage at an RSS, whilst remaining intact in the majority of interactions. This intact ESC could potentially stimulate cleavage at a second RSS, and repeat the process of cleavage and running to another RSS until it is lost through cell division or cleaved by RAG proteins or other cellular nucleases.

Whilst cRAG proteins are more amenable to purification, they might not accurately recapitulate full-length RAG activity, so it was necessary to purify the full-length RAG (flRAG) proteins. The second part of this chapter discusses issues surrounding successful purification of flRAG proteins and presents the results from their use in activity assays. I find that cleavage shows some asymmetry between a 12RSS and ESC, albeit with a lower overall asymmetry than with core RAG proteins, but cleavage between a 23RSS and ESC is not at all asymmetric. Inaccuracies caused by the overall low activity of the flRAG proteins could have influenced these results; nevertheless, *in vivo* data presented in Chapter 5 shows that full-length RAG cleavage of RSSs and ESCs is indeed asymmetric.

B) Results

3.1 Purification of core RAG proteins and HMGB1

Until recent years, the majority of biochemical analyses of V(D)J recombination and the RAG recombinase have been performed using truncated versions of the RAG1 and RAG2 proteins. Early work on the RAG proteins identified the minimal core regions of RAG1 and RAG2 that are necessary and sufficient to carry out recombination of extrachromosomal substrates *in vivo* (Sadofsky et al., 1993; Silver et al., 1993; Cuomo and Oettinger, 1994; Sadofsky et al., 1994; Kirch et al., 1996). For RAG1, this encompasses residues 384-1008 of 1040, which contain the nonamer binding domain (Difilippantonio et al., 1996; Spanopoulou et al., 1996) and the DDE motif of the active site (Kim et al., 1999). The core region of RAG2 includes residues 1-384 out of 527, which removes several conserved but non-essential domains including a plant homeodomain (PHD) finger which binds to H3K4me3, and an acidic "hinge" region (Matthews et al., 2007; Ramon-Maiques et al., 2007), the function of which is less clear, but which might be involved in directing the broken ends RSS ends to the nonhomologous NHEJ machinery, and away from the alternative NHEJ pathway (Chaumeil et al., 2013). Recently, the minimal region of RAG2 has been reduced to residues 1-360 (Corneo et al., 2007; Schatz and Swanson, 2011), but the vast majority of published data have been made using residues 1-387 of RAG2, and it is this version of core RAG2 that is used here.

The core versions of the RAG proteins are considerably easier to express and purify in large quantities than the full-length counterparts, and co-expression increases both the yield and activity of the RAG proteins compared to when they are expressed separately (Swanson, 2004). Therefore, core RAG1 (cRAG1) and core RAG2 (cRAG2) were purified from Sf9 cells infected with baculoviruses that express cRAG1 and cRAG2 with maltose binding protein (MBP), 9xHis and 3xmyc tags (Fig. 3.1A). To obtain optimal cRAG expression, stocks of cRAG1 and cRAG2 baculoviruses were plaque-purified and amplified to ensure clonally identical stocks (Section 2.16). Following estimation of their titers by plaque assay (Section 2.15), they were then used to infect large suspension cultures of Sf9 insect cells. cRAG proteins were co-purified from infected cells by Ni^{2+} affinity chromatography followed by amylose affinity chromatography (Section 2.17; Fig. 3.1B). One litre of infected cells typically yields ~150 µg of purified RAG proteins that are greater than 95% pure, as judged by SDS-PAGE (Fig. 3.1B). The ratio of cRAG1:cRAG2 is usually around 2:1, possibly because RAG2 is more

prone to degradation by cellular proteases, or is less efficiently expressed in Sf9 cells than RAG1.
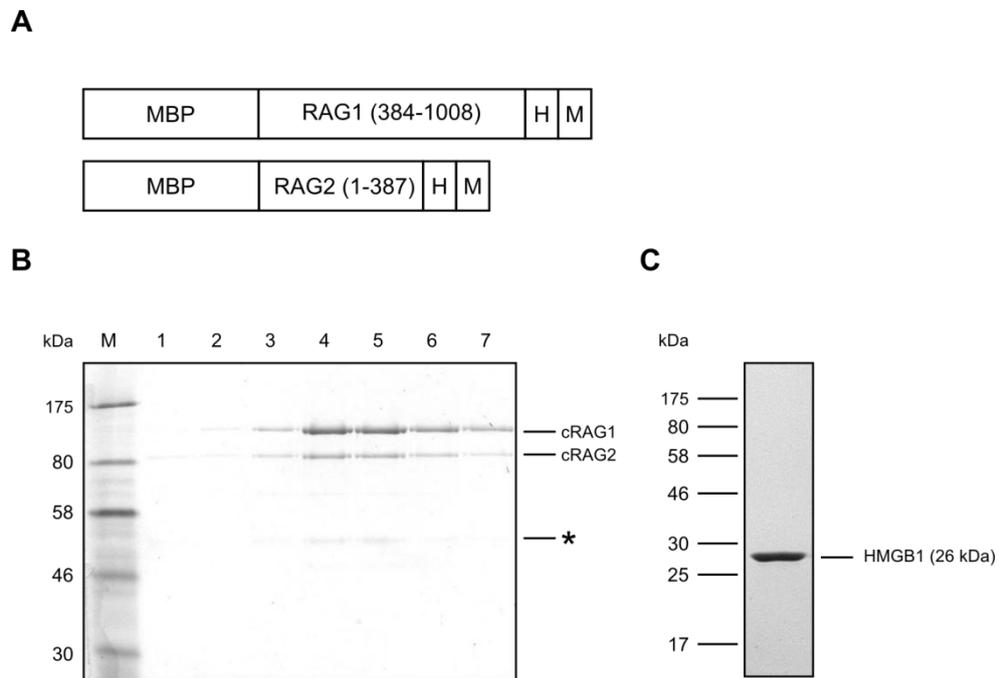
Figure 3.1 – Purified core RAG proteins and HMGB1 (A) Schematic of MBP-, 9xHis- and 3xmyc-tagged core RAG fusion proteins. MBP = Maltose binding protein, H = 9xHis tag, M = 3xmyc tag. **(B)** Purified MBP-tagged core RAG1 (118 kDa) and MBP-tagged core RAG2 (90 kDa). The different lanes show samples of the 1 ml fractions collected from the amylose column. M = marker, * = contaminant/degradation product. (C) Purified full-length HMGB1 (26 kDa)

As described in Section 1.8, HMGB1 is a nonspecific DNA-binding protein that can incorporate into the RAG-RSS complex and stimulate RAG cleavage of 23RSSs by binding in the spacer and bringing the heptamer and nonamer closer together (van Gent et al., 1997; Swanson, 2002a). As a result, HMGB1 is routinely included in *in vitro* RAG activity assays (Swanson, 2004). Therefore, 6xHis-tagged full-length rat HMGB1 was expressed in the BL21(DE3) strain of *E.coli* and purified first by Ni^{2+} affinity chromatography and then by anion exchange chromatography (Section 2.18; Fig. 3.1C). One litre of induced *E. coli* typically yields ~10 mg of highly pure full-length HMGB1 (>98% pure, Fig. 3.1C).

3.2   Establishing in vitro reaction conditions: stimulation of RAG cleavage is non-specific in the presence of Mn^{2+}

Reintegration involves pairing either the 12 or 23RSS from an ESC with an appropriate partner RSS. Before investigating this mechanism, it was important to first verify that

the purified RAG proteins were behaving in the manner expected when utilising standard 12 and 23RSSs. Therefore, well-established reaction conditions were used to assess the activity of RAG proteins on radiolabelled oligonucleotides containing the sequence of either a 12 or 23RSS (McBlane et al., 1995).

In a typical cleavage assay, a radiolabelled oligonucleotide encoding a 12RSS, 23RSS or ESC is incubated with core RAG1, core RAG2, HMGB1, and a non-labelled partner RSS. The cleavage products are separated on a native polyacrylamide gel and visualised using a PhosphorImager (Section 2.37). The oligonucleotides contain the same consensus heptamer and nonamer sequences and differ only in spacer length. Heptamer and nonamer sequences that deviate from the consensus can affect the efficiency with which RSSs are bound and cleaved (Ramsden et al., 1996); therefore identical heptamer and nonamer sequences were used for both 12 and 23RSS oligonucleotides.

In vivo, efficient recombination only occurs between two coding segments flanked by RSSs of different spacer length, i.e. a 12 and 23 RSS, a requirement known as the 12/23 rule (Tonegawa, 1983) and this specificity is reflected *in vitro* (van Gent et al., 1996c; Hiom and Gellert, 1998). When a radiolabelled 12RSS (DAR39/DAR40, Table 2.2) is incubated with cRAG proteins, HMGB1 and a less than equimolar amount of unlabelled 23RSS, the level of cleavage is low (Fig. 3.2A, lanes 2-7). However, the level of cleavage increases up to 40-fold when the unlabelled 23RSS (DG61/DG62, Table 2.2) is at equimolar concentrations or higher (Fig. 3.2A, lanes 7-11). The same stimulatory effect is also seen when the 23RSS is labelled. The efficiency of cleavage is low in the absence of a partner RSS, but addition of unlabelled 12RSS above an equimolar amount causes an increase in cleavage (Fig. 3.2B). Overall, the 23RSS is cleaved less efficiently than a 12RSS, consistent with previous studies (van Gent et al., 1996a; Ramsden et al., 1997).

Next, to mimic synaptic complex formation during reintegration, a labelled 12 or 23RSS oligonucleotide was paired with an oligonucleotide containing a signal joint where a 12 and 23RSS are joined at the heptamer sequence (MBNSJ-top/MBNSJ bottom, Table 2.2), as is the case with an ESC. The level of cleavage at both a 12 and 23RSS is stimulated to a similar level, as would be expected since an ESC contains both a 12 and a 23RSS and therefore could form a synaptic complex with either type of RSS (Fig. 3.3). However, when the ESC is labelled, the level of cleavage remains relatively low regardless of the amount or type of partner RSS present (Fig. 3.4), indicating a potential asymmetry in stimulation and cleavage at an RSS-ESC pair.
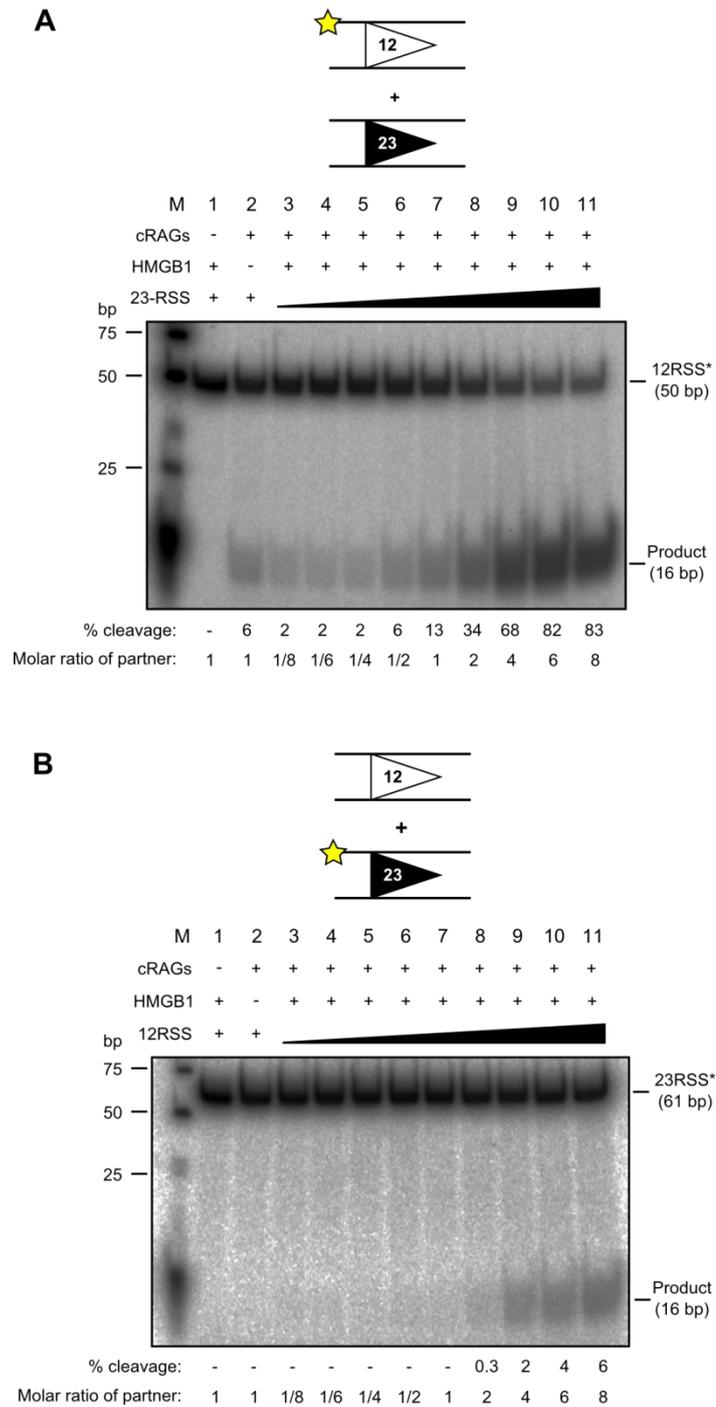
Figure 3.2 – RSSs co-stimulate RAG cleavage in the presence of Mn²⁺. (A) 2 nM radiolabelled 12RSS oligonucleotide (DAR39/DAR40) was incubated at 37°C for 1.5 hours with purified cRAG1 and cRAG2, HMGB1, and an unlabelled 23RSS oligonucleotide (DG61/DG62), ranging from 0.16- to 8-fold excess of unlabelled partner. The cleavage efficiency as a percentage of the total labelled substrate and the molar ratio of unlabelled partner is shown below each lane. The substrates used are shown above the gel, with a yellow star indicating the location of the radiolabel. **(B)** As for A, except the 23RSS was radiolabelled and an unlabelled 12RSS oligonucleotide was titrated.
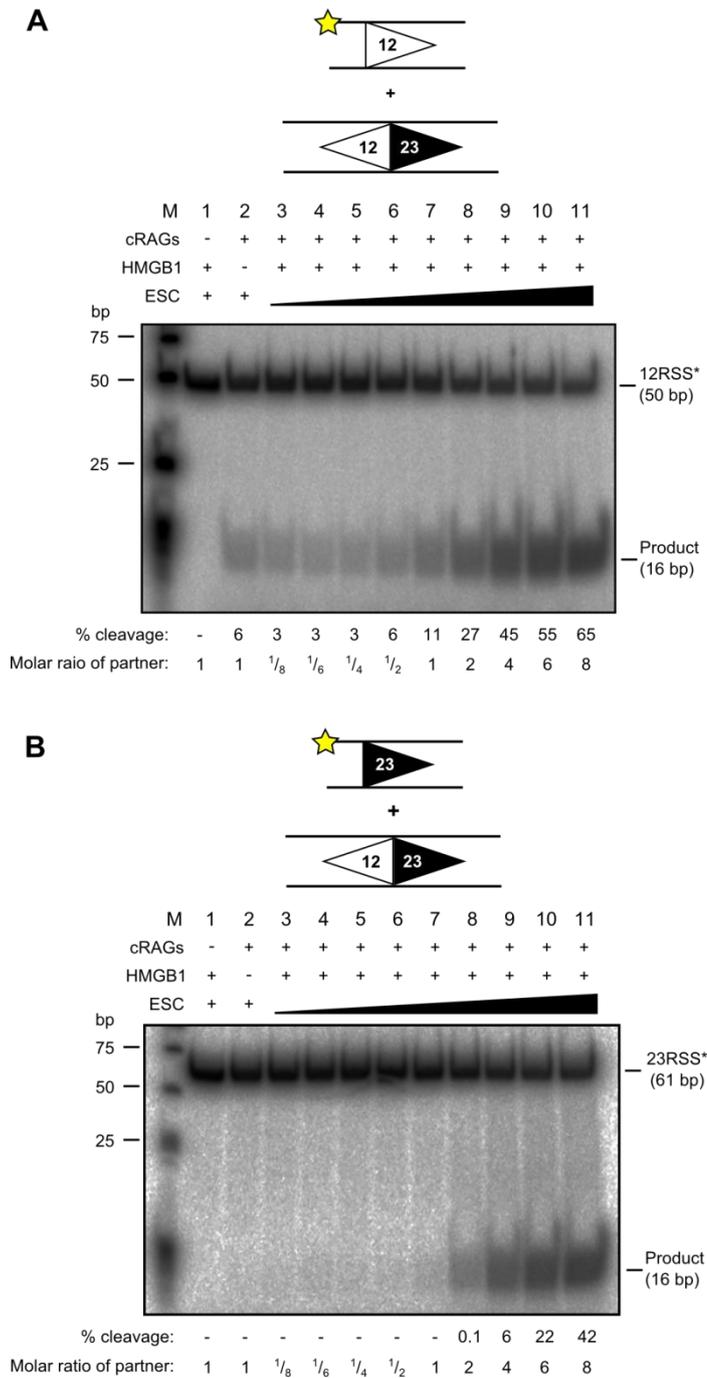
Figure 3.3 – ESCs stimulate RSS cleavage in Mn^{2+}. (A) 2 nM radiolabelled 12RSS oligonucleotide was incubated at 37°C for 1.5 hours with purified cRAG1 and cRAG2, HMGB1, and an unlabelled ESC oligonucleotide (MBNSJ top/MBNSJ bottom), ranging from 0.16- to 8-fold excess of unlabelled partner. The cleavage efficiency as a percentage of the total labelled substrate and the amount of unlabelled partner is shown below each lane. The substrates used are shown above the gel, with a yellow star indicating the location of the radiolabel. **(B)** As for A, except the 23RSS (DG61/DG62) was radiolabelled.
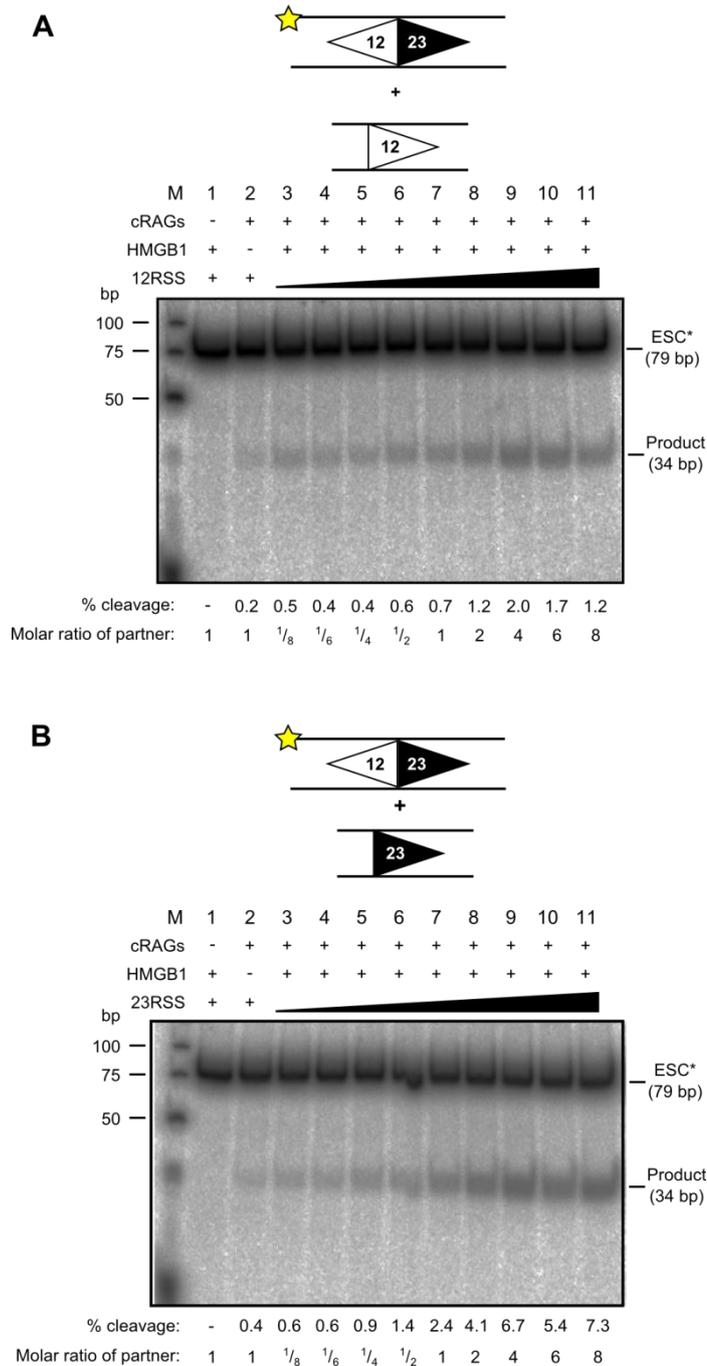
Figure 3.4 – Cleavage at RSS-ESC complexes in Mn^{2+}. (A) 2 nM radiolabelled ESC oligonucleotide was incubated at 37°C for 1.5 hours with purified cRAG1 and cRAG2, HMGB1, and an unlabelled 12RSS oligonucleotide, ranging in concentration from 0.16- to 8-fold excess. The cleavage efficiency as a percentage of the total labelled substrate and the amount of unlabelled partner is shown below each lane. The substrates used are shown above the gel, with a yellow star indicating the location of the radiolabel. **(B)** As for A, except an unlabelled 23RSS was titrated.

The ESC substrate would more accurately be called a signal joint or excised signal joint instead of an ESC, since the two head-to-head RSSs are located on a short linear oligonucleotide and not a circular piece of DNA. For the sake of clarity and consistency however, the oligonucleotide ESC substrate will be referred to as an ESC for the entirety of this thesis.

To test the sequence-specificity of this reaction, I added a non-specific oligonucleotide (DAR81/DAR82, Table 2.2) to a labelled 12 or 23RSS. Surprisingly, the non-specific oligonucleotide also stimulates cleavage at an RSS by the same amount as an RSS or ESC (Fig. 3.5). A likely cause of this non-specific stimulation is the divalent metal cation used in the reaction buffer. I initially used Mn^{2+} as the divalent metal cation because higher levels of RAG activity were generally observed with this cation, and at first, my purified cRAG proteins were not active in Mg^{2+}. As is the case with many DNA-processing proteins, including retroviral integrases (Bujacz et al., 1997), T7 DNA polymerase and *E. coli* DNA polymerase I (Tabor and Richardson, 1989), Mn^{2+} can be utilised by RAG proteins in place of Mg^{2+} (Santagata et al., 1998). In most cases, the catalytic activity of DNA processing proteins is more efficient in the presence of Mn^{2+} than Mg^{2+}, but at the expense of greater promiscuity. This is also true of the RAG proteins, since Mn^{2+} has been shown to uncouple the cleavage reaction, allowing efficient hairpin formation at a single RSS outside of the 12/23 restriction (McBlane et al., 1995; van Gent et al., 1996c; Santagata et al., 1998). A bandshift assay (Section 2.39) confirmed that the non-specific oligonucleotide is indeed bound by the RAG proteins (Fig. 3.6, top panel), as would be expected if it is to stimulate cleavage at an RSS. However, a cleavage assay in the same conditions shows that the non-specific oligonucleotide is not cut by RAG proteins (Fig. 3.6, bottom), confirming that although the RAG proteins can efficiently bind to DNA without a signal sequence, an RSS is essential for RAG cleavage of DNA. Since Mn^{2+} facilitates non-specific stimulation of RAG activity outside the 12/23 rule, it is not suitable for use as a cofactor in determining the effect of ESCs on RSS cleavage and RAG-ESC interactions. Therefore, it was important to repeat these experiments in the presence of Mg^{2+}, as the stimulation of RAG activity is expected to be sequence-specific under these conditions.
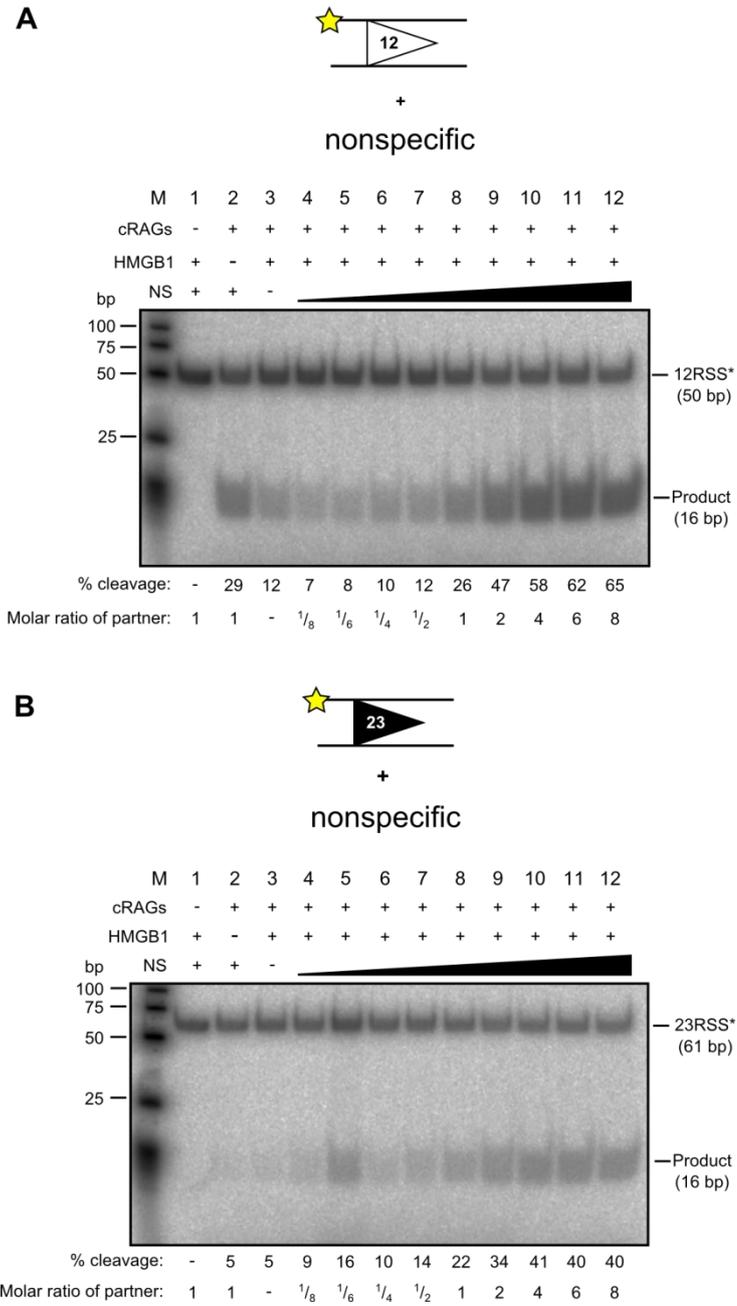
Figure 3.5 – Stimulation of core RAG activity is not sequence-specific in the presence of Mn^{2+}. (A) Addition of a non-specific 50 bp oligonucleotide to 2 nM 12RSS stimulates cRAG activity by the same amount as another RSS, indicating that stimulation is not sequence-specific in the presence of Mn^{2+}. **(B)** As for (A) except a labelled 23RSS oligonucleotide was used. NS = non-specific 50mer.
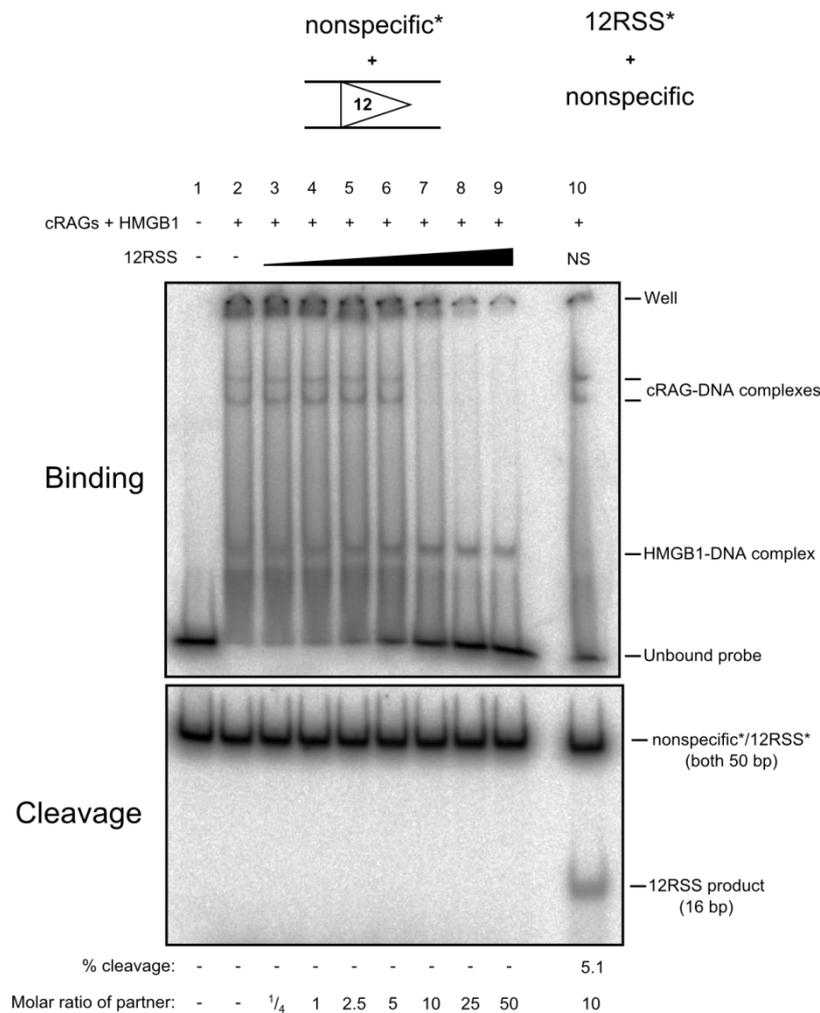
Figure 3.6 – Non-specific DNA is bound but not cleaved in the presence of Mn^{2+} by core RAG proteins. A bandshift assay using a labelled non-specific 50mer indicates that the non-specific oligonucleotide is bound by the cRAG proteins, but is not cleaved (bottom panel). The far right lane is a positive control showing binding and cleavage at a labelled 12RSS.

3.3 Establishing *in vitro* reaction conditions: RAG cleavage adheres to the 12/23 rule in the presence of Mg^{2+}

Initially, Mg^{2+} was substituted for Mn^{2+}, keeping all other factors and reaction components the same as in previous experiments. However, I did not observe efficient hairpin formation under these conditions (Fig. 3.7A, left). Although several studies have reported RSS cleavage in these conditions (van Gent et al., 1996c; Eastman and Schatz, 1997; Hiom and Gellert, 1998), the Swanson group routinely includes 20% dimethyl sulphoxide (DMSO) in their reactions, since they observe that addition of DMSO promotes the accurate production of hairpins and without DMSO, RAG cleavage in Mg^{2+} is inefficient and largely off-target (Bergeron et al., 2006). In my hands, DMSO is essential for efficient hairpin formation when Mg^{2+} is the metal

cofactor (Fig. 3.7A, right panel). Although several faint bands can be seen in the absence of DMSO, one of which is the correct size for a hairpin product (Fig. 3.7A, left panel), DMSO appears to be essential for efficient and correct hairpin formation in Mg^{2+}-containing buffer (Fig. 3.7A, right panel).

I next examined the stimulation of RAG cleavage by a partner substrate in the presence of Mg^{2+} and DMSO. Other groups have reported that cleavage of a single RSS in Mg^{2+} is inefficient, and efficient cleavage requires the presence of a suitable partner RSS. Contrary to this, I observed that after incubation for 2 hours at 37°C, addition of a partner reduces the cleavage efficiency (Fig. 3.7A, right panel, and Fig. 3.7B, top panel). However, co-stimulation is observed if the incubation time is shortened to 15 minutes (Fig. 3.7B, bottom panel).

One possible explanation is that in the presence of Mg^{2+} conditions, cRAG proteins can undergo one of two reactions: an initial "fast" coupled reaction, and a slower uncoupled reaction. In the initial coupled reaction, RAG proteins form a synaptic complex with two RSS substrates, where each substrate stimulates cleavage at the other, as seen after a 15 minute incubation (Fig. 3.7B, bottom panel). In the uncoupled reaction, RAG proteins bind to and cleave a single RSS. The rate of this reaction is slower than the coupled reaction, and after longer incubation periods the products from cleavage at a single RSS accumulate. At concentrations where the unlabelled partner is in excess, single cleavage of the unlabelled partner outcompetes cleavage at the labelled substrates, hence the apparent decrease in cleavage at higher concentrations of partner. An undesirable side-effect of a short incubation time is that the overall level of cleavage (between 1-10%) is much lower than in previous experiments performed for 1.5 hours in the presence of Mn^{2+} (e.g. up to 80% cleavage, Fig. 3.2A).
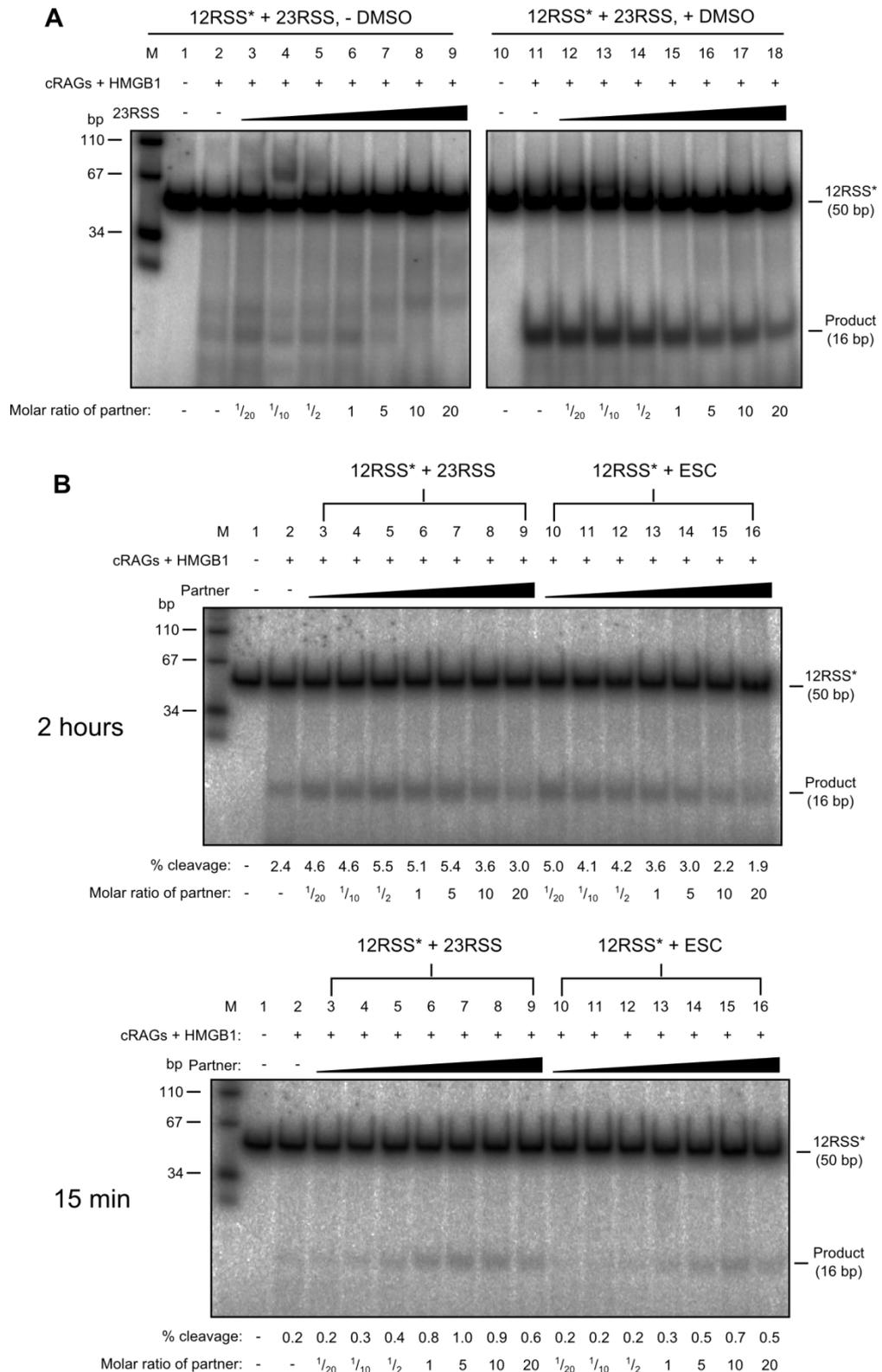
Figure 3.7 – DMSO is essential for efficient RAG cleavage in the presence of Mg^{2+} (A) 2 nM radiolabelled 12RSS was incubated for 2.5 hours at 37°C with cRAG proteins, HMGB1, and 0.01 - 4 nM unlabelled 23RSS in the presence of Mg^{2+}, either without (left) or with (right) 20% (v/v) DMSO. Efficient cleavage is only observed in the presence of DMSO **(B)** Labelled 12RSS oligo was incubated with unlabelled 23RSS (left), lanes 3-9) or ESC (lanes 10-16) for either 2 hours or 15 min. Stimulation by a partner or an ESC is only observed after a short incubation (lanes 6-9 and 13-16). After longer incubations, stimulation is less pronounced.

3.4 Core RAG proteins purified from HEK-239T cells are more active than those purified from Sf9 cells

cRAG proteins purified from Sf9 cells showed relatively low levels of activity in Mg^{2+} compared to Mn^{2+}, especially in the short incubation periods necessary to observe substrate co-stimulation (Fig. 3.7B), which complicated accurate quantification and interpretation. It is possible to purify RAG proteins from transfected human embryonic kidney 293T (HEK-293T) cells (Bergeron et al., 2006). Using this expression system, I found that the cRAG proteins showed a higher level of activity than cRAG proteins purified from Sf9 cells. Therefore, cRAG proteins purified from HEK-293T cells (Fig. 3.8) were used in all relevant experiments described below.

Using the newly optimised reaction conditions (namely, 20% DMSO and 15 min incubation time) and cRAG proteins purified from HEK-293T cells, I tested RAG activity with different substrate combinations to determine whether or not the 12/23 rule is adhered to. As I observed for cRAG proteins in the presence of Mn^{2+}, 12 and 23RSSs co-stimulate, with the greatest stimulation observed at a 10-fold excess of unlabelled partner (Fig. 3.9A). However, a non-specific oligonucleotide does not stimulate cleavage, indicating that stimulation is sequence-specific and that cRAG proteins purified from HEK-293T cells adhere to the 12/23 rule under these conditions (Fig. 3.9B).

The fact that stimulation occurs at a 10:1 ratio of unlabelled:labelled substrate is not physiologically accurate. However, a greater concentration of partner relative to labelled substrate is needed because cRAG proteins are present in excess. Therefore, when both substrates are present at a 1:1 ratio, most of the labelled and unlabelled substrates are present in separate complexes, hence the need to add an excess amount of unlabelled partner to force the formation of a synaptic complex. A 10-fold excess of unlabelled partner appears to be the ideal concentration; at concentrations greater than this the unlabelled partner begins to compete with the labelled substrate (e.g. Fig. 3.9A, lanes 7 and 8).

It should be noted that under these new conditions there is a diffuse band that runs slightly faster than the expected 16 bp product (Fig. 3.9). Titration of a suitable unlabelled partner reduces the abundance of this band. The nature of this band is unclear, but it is specific to reactions performed in buffer containing DMSO and Mg^{2+}, and is not observed in the presence of Mn^{2+} (Fig. 3.10). There are three possible
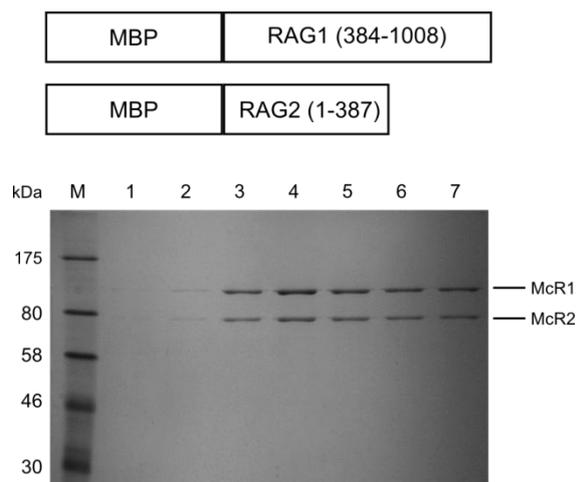
Figure 3.8 – Core RAG proteins purified from transfected HEK-293T cells. Adherent HEK-293T cells were co-transfected with MBP-tagged core RAG1 and core RAG2 expression vectors. Core RAG proteins were purified in a single step by amylose affinity chromatography. Each lane is a different fraction eluted from an amylose column. McR1 = MBP-tagged core RAG1 (112 kDa); McR2 = MBP-tagged core RAG2 (86 kDa).

for this. Firstly, small amounts of contaminating nuclease in the RAG prep, that are active in Mg^{2+} but not Mn^{2+}, degrade the labelled oligonucleotide. Addition of a partner oligonucleotide then stimulates RAG activity so that nuclease activity is out-competed. The second possibility is that in the absence of a partner substrate, RAG proteins perform off-target cleavage of the labelled substrate, and addition of a partner promotes correct hairpin formation. Finally, it is possible that the diffuse band represents the single-stranded 16 bp product caused by RAG nicking, but not hairpinning. In the presence of Mg^{2+}, RAG proteins are capable of nicking RSSs in the absence of a suitable partner, but a partner is required for efficient hairpin formation (van Gent et al., 1997). In the original conditions (i.e. in the absence of DMSO), the nicked strand stays annealed to the opposite strand, since the T_m of the 16 base nick product (~43°C) is greater than the incubation temperature (37°C). DMSO lowers the T_m of DNA by 0.75°C per 1% DMSO (von Ashen et al., 2001), so the presence of 20% DMSO in the new conditions will decrease the T_m of the 16 base nick product to ~28°C. Furthermore, the smaller diffuse band is more abundant in the presence of Mg^{2+} than in Mn^{2+} (Fig. 3.10), since a suitable partner is not required for efficient hairpin formation in Mn^{2+} (van Gent et al., 1997). Therefore, the diffuse band is most likely to be nicked substrate that has denatured at 37°C in the presence of DMSO, and runs slightly faster than the double-stranded 16 bp product caused by hairpinning. Addition of a partner RSS allows for the majority of nicks to be converted hairpins, thus reducing the abundance of denatured nicked products (Fig. 3.9A). In any case, although this diffuse

band can complicate densitometry, it does not affect the overall interpretation or conclusions of these data.
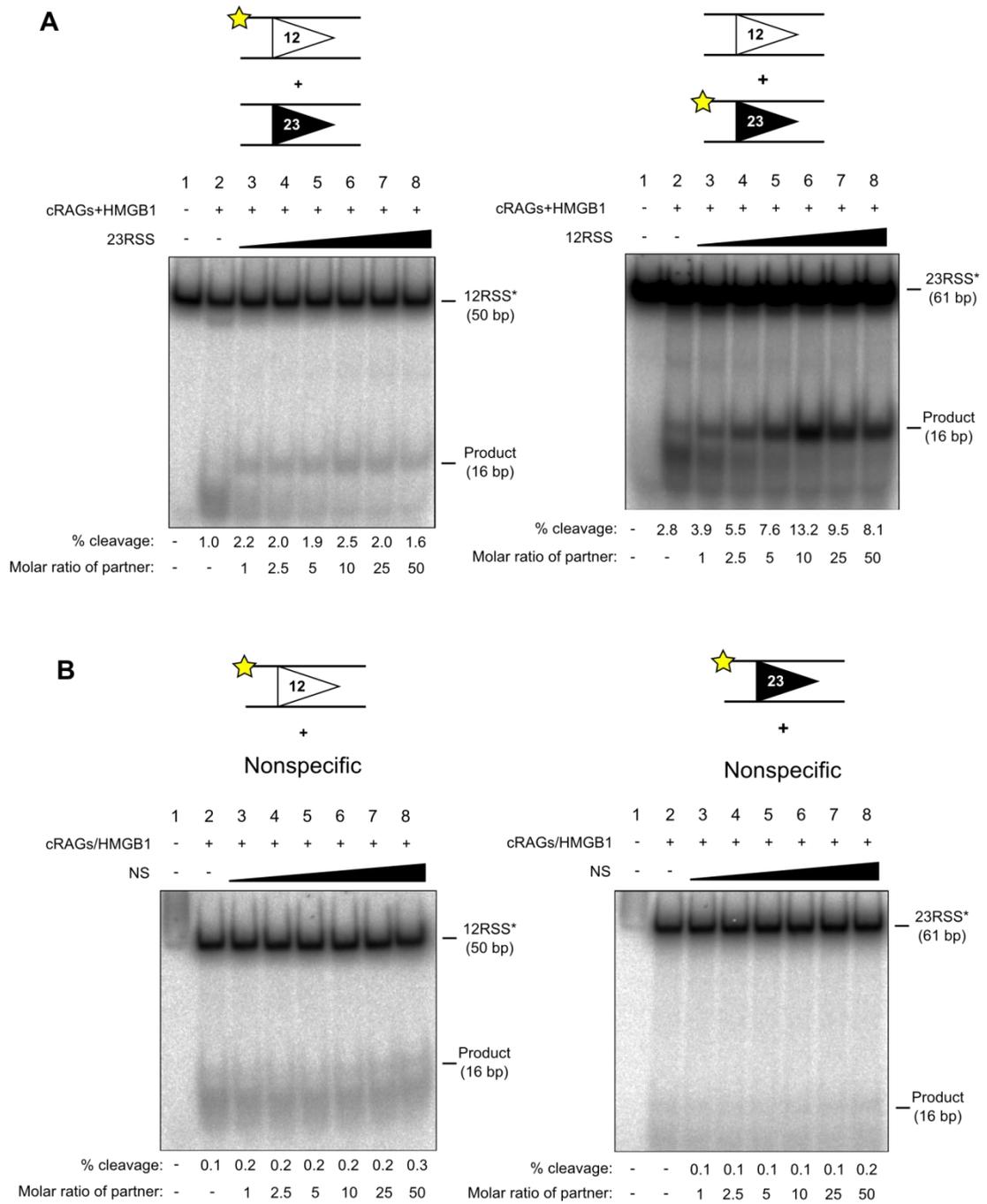
Figure 3.9 – RSSs co-stimulate in a sequence-specific manner in the presence of Mg^{2+}. (A) 2 nM labelled 12 (left) or 23RSS (right) was incubated in Mg^{2+}-containing buffer with cRAGs purified from HEK-293T cells and an RSS oligonucleotide of different spacer length. **(B)** As for (A), except unlabelled nonspecific 50mer was added instead of RSS.

	12RSS*						23RSS*			
	1	2	3	4	5	6	7	8	9	10
Mn^{2+}	-	-	+	+	-	-	-	-	+	+
Mg^{2+}	+	+	-	-	+	+	+	+	-	-
cRAGs	-	-	-	-	+	+	-	-	+	+
flRAGs	+	+	+	+	-	-	+	+	-	-
ESC	-	+	-	+	-	+	-	+	-	+
% cleavage:	1.7	2.8	23	15	-	-	1.7	1.6	1.2	5.4
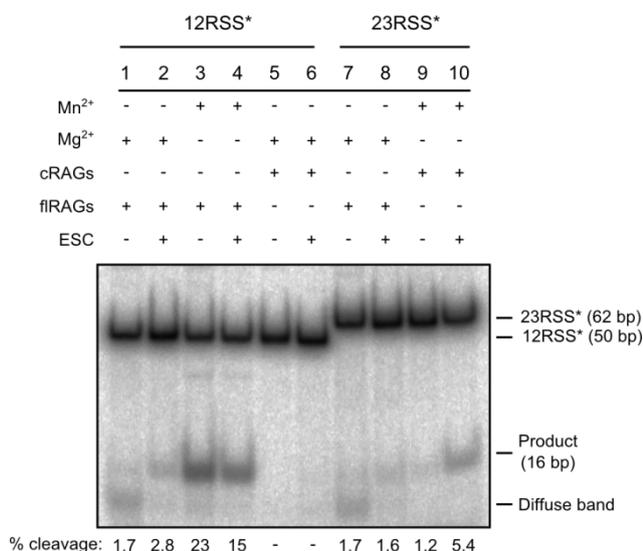
Figure 3.10 – A diffuse band is observed in the presence of Mg^{2+}. The cleavage of 12 and 23RSSs by core or full-length RAG proteins was analysed under different conditions, namely in the presence of Mg^{2+} or Mn^{2+}, and with or without an ESC partner. A diffuse band appears most strongly when in the presence of Mg^{2+} and in the absence of an ESC partner (lanes 1 and 7), suggesting that it could be caused either by contaminating nucleases or by off-target RAG cleavage.

Next, I examined the effect of an ESC in the presence of Mg^{2+}. As observed in the presence of Mn^{2+}, an ESC stimulates cleavage at both a 12RSS and a 23RSS, with the greatest level of stimulation (approximately 4-fold in each case) seen when there is a 10-fold excess of ESC (Fig. 3.11A). However, neither a 12RSS nor a 23RSS stimulates cleavage at an ESC (Fig. 3.11B). In fact, addition of a partner causes a decrease in ESC cleavage. The difference in cleavage efficiency of 12 and 23RSS (at a 10-fold excess of partner) is only 1.2-fold, suggesting that the level of RSS cleavage and stimulation is relatively symmetric. By contrast, a 12RSS and ESC have a 4-fold difference in cleavage efficiency, and with a 23RSS, the difference is 10-fold. These data therefore indicate that cleavage and stimulation within an RSS-ESC complex is asymmetric in the presence of Mg^{2+}.

To see whether asymmetric cleavage between ESCs and RSSs is due to limiting RAG concentration, labelled 12RSS or ESC oligonucleotides were incubated with a 10-fold excess of unlabelled 23RSS (since this is typically the concentration at which maximum cleavage is observed), and cRAG proteins at concentrations ranging from 20 to 80 nM (Fig. 3.12). At each concentration of RAG proteins, ESC cleavage was less efficient than RSS cleavage, on average by a factor of 12. Therefore, asymmetric cleavage is independent of cRAG concentration, at least at the concentrations tested.
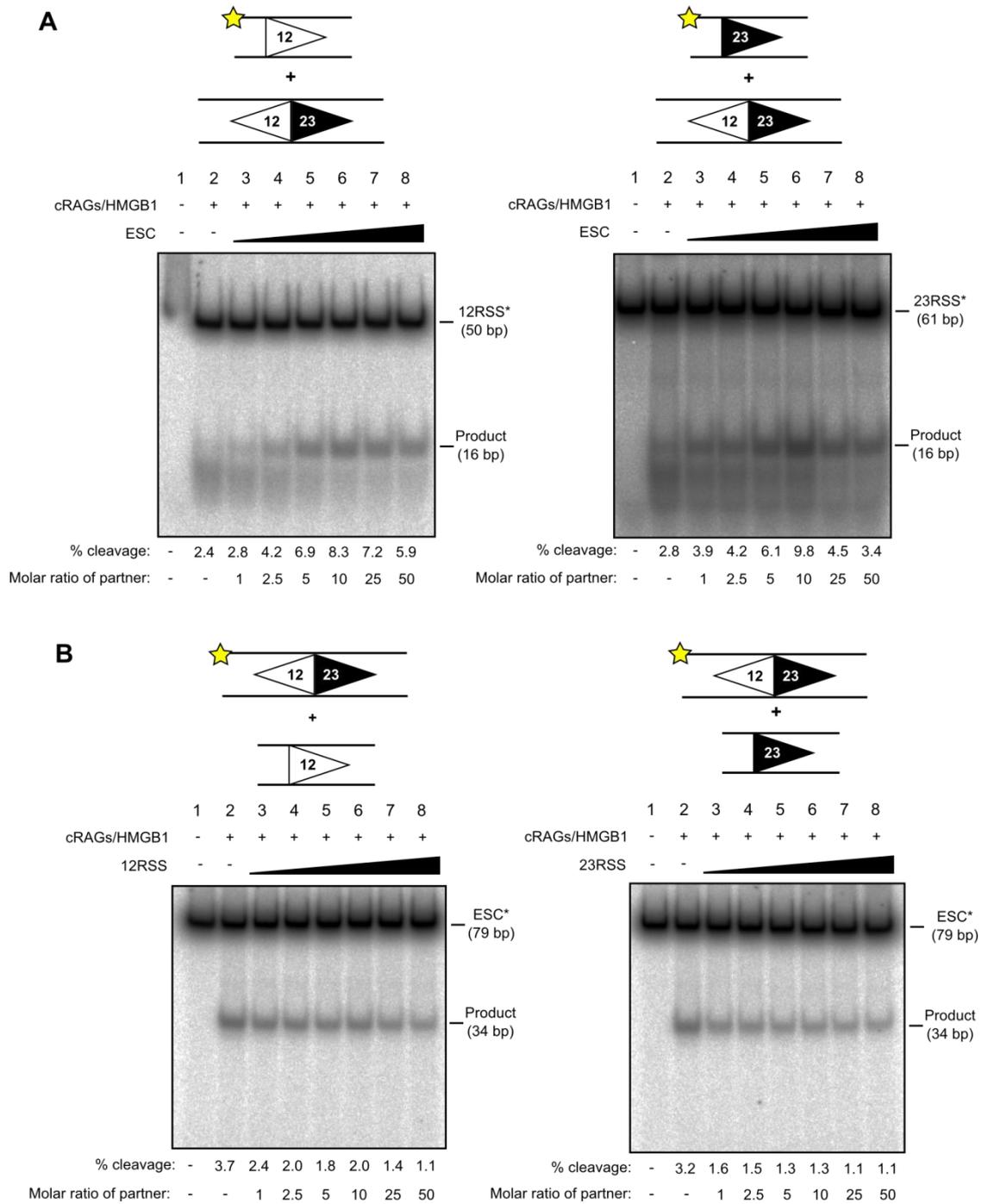
Figure 3.11 – ESCs stimulate RSS cleavage in the presence of Mg^{2+}, but RSSs do not stimulate ESC cleavage (A) 2 nM radiolabelled 12 or 23RSS-containing oligonucleotide was incubated with cRAG proteins purified from HEK-293T cells, HMGB1 and 1 to 50-fold excess of unlabelled ESC oligonucleotide. Reactions were carried out for 15 min at 37°C. **(B)** As for (A), except a radiolabelled ESC was used and an unlabelled 12 or 23RSS was added.
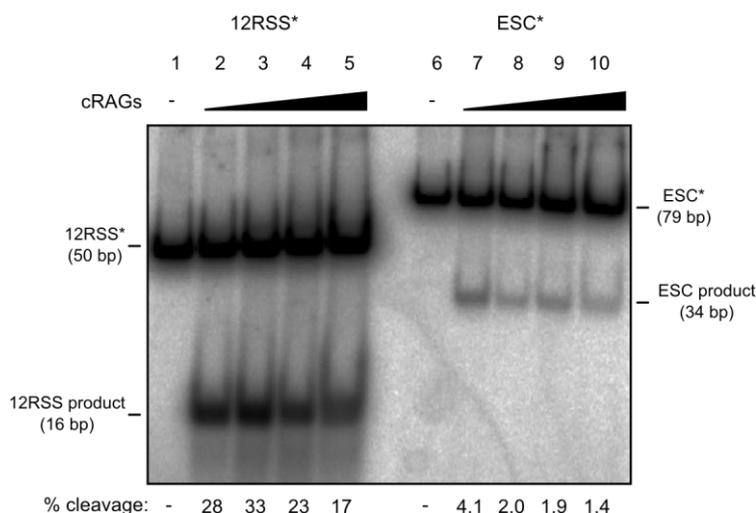
Figure 3.12 – ESC cleavage is less efficient than RSS cleavage over a range of cRAG concentrations. 2 nM radiolabelled 12RSS (left) or ESC (right) oligonucleotide was incubated for 30 min at 37°C with 20 nM unlabelled 23RSS oligonucleotide and cRAG proteins ranging from 20 to 80 nM. ESC cleavage was less efficient than RSS cleavage at each concentration of cRAG proteins.

If an RSS does not stimulate cleavage at ESC, then this is at odds with the proposed model of reintegration. In order for reintegration to occur, an ESC must first be cut by RAG proteins. Whilst reintegration could occur to an extent, since some ESC cleavage does occur (Fig. 3.11), reintegration might occur less frequently than previously predicted since the cleavage in RAG-RSS-ESC complexes appears to be asymmetric (Fig. 3.13A). Instead, I propose an alternative mechanism by which ESCs could trigger genome instability in developing B and T cells, called "cut-and-run" (Fig. 3.13B). In cut-and-run, when RAG proteins form a synaptic complex with a genomic RSS and an ESC, the ESC stimulates cleavage at the RSS, but cleavage at the ESC is not stimulated. The data in Fig. 3.11 indicate that in an RSS-ESC complex, the ESC would remain intact in three quarters of these interactions. If the ESC remains intact, depending on the fate of the post-cleavage complex in these instances, RAG proteins could potentially form a complex with this ESC and a second genomic RSS, and stimulate cleavage at the RSS. This process of cleavage and "running" to another genomic RSS (hence "cut-and-run") could be repeated until the ESC is cleaved or lost through cell division. In the remainder of this thesis, I further investigate this cut-and-run hypothesis.
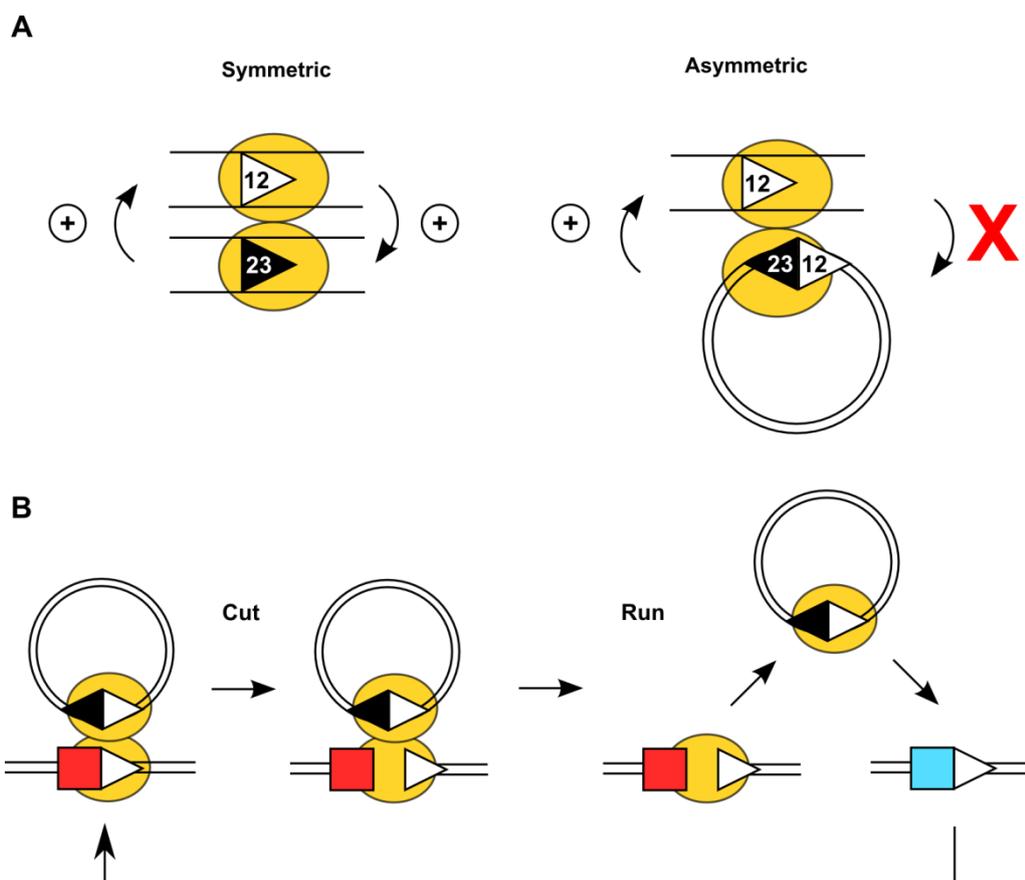
Figure 3.13 – Asymmetric cleavage and the cut-and-run hypothesis. (A) When two RSSs of different spacer length are incubated together with RAG proteins, RAG cleavage is stimulated at both substrates. Therefore, both cleavage and stimulation are "symmetric". However, when one of the RSSs is replaced with an ESC, cleavage at the RSS is stimulated, but ESC cleavage is not stimulated, therefore cleavage with an ESC-RSS pair is "asymmetric". **(B)** If asymmetric cleavage occurs *in vivo*, it is possible that instead of ESCs reintegrating into the genome, the more frequent outcome is that an ESC stimulates cleavage of a genomic RSS (either legitimate or cryptic), whilst the ESC itself remains intact. This ESC could potentially be re-bound by the RAG proteins, and "run" to another RSS to stimulate further cleavage.

3.5 Expression of full-length RAG proteins using a vaccinia-based system

The data shown up until this point have been produced using truncated versions of the RAG1 and RAG2 proteins. The non-core regions, whilst not vital for the basic biochemical activity of the RAG proteins, are known to play a number of important roles in the regulation of V(D)J recombination. For example, the RAG2 C-terminus contains a PHD finger that binds to trimethylated lysine 4 of histone H3 (H3K4me3), thus directing the RAG recombinase to actively transcribed regions of the antibody loci (Matthews et al., 2007), and also plays a role in reducing the frequency of reintegration events (Curry et al., 2007). Since reintegration and cut-and-run both involve RAG-ESC interactions, it was of critical importance to test whether the asymmetry in cleavage observed with a cRAG-RSS-ESC complex occurs with the full-length RAG proteins.

The first approach to obtain purified full-length RAG proteins involved expression in a vaccinia-based system, which has previously been shown to produce active full-length RAG proteins (McBlane et al., 1995; Elkin et al., 2003). In this expression system, vaccinia virus containing the cDNA of the protein of interest under the control of a T7 promoter is used to infect mammalian cells along with a recombinant vaccinia virus that expresses the T7 RNA polymerase (Elroy-Stein and Moss, 2001). Expression of the protein of interest is directed by T7 RNA polymerase, leading to very high yields. First, the flRAG coding sequences with N-terminal 9xHis and FLAG epitopes were cloned into the vaccinia virus transfer vector pTM1 (Section 2.23). Then, to produce recombinant vaccinia virus, BHK-21 cells were infected with wild-type vaccinia virus and co-transfected with the pTM1-flRAG transfer vectors (Section 2.26). In these vectors, the flRAG genes are flanked by sections of the vaccinia virus thymidine kinase (TK) gene, so that in infected cells they are inserted into the vaccinia genome by homologous recombination, allowing for selection of TK-deficient mutants (Sections 2.27 and 2.28). Recombinant vaccinia virus was then purified through several rounds of plaque purification to ensure a clonally identical stock. Stocks of flRAG1 and flRAG2 vaccinia viruses and a T7 RNA polymerase-expressing virus, vTF7-3, were made by amplification in adherent HeLa cells prior to large-scale infection for protein purification.

An anti-FLAG western blot of the purification process indicated that only trace amounts of flRAG proteins were present after the centrifugation step, indicating that although the flRAG proteins were expressed using this system, they were largely insoluble in the conditions used (Fig. 3.14A). Optimisation of the lysis conditions was carried out in an attempt to solubilise the full-length RAG proteins, but unfortunately this approach ultimately proved fruitless, and the vaccinia/T7 expression system was abandoned.

I therefore turned to the other main way in which full-length RAG proteins have been purified, which is based on plasmid transfection into adherent human embryonic kidney HEK-293T cells (Bergeron et al., 2006). In this system, HEK-293T cells are transfected with mammalian expression vectors containing the full-length RAG protein coding sequences with N-terminal MBP tags. Full-length RAG1 and full-length RAG2 (flRAG proteins) were then purified via one-step affinity chromatography on an amylose column (Fig. 3.14B).

A

B

1 = Crude lysate
2 = Crude lysate after low salt wash
3 = Insoluble fraction
4 = Supernatant
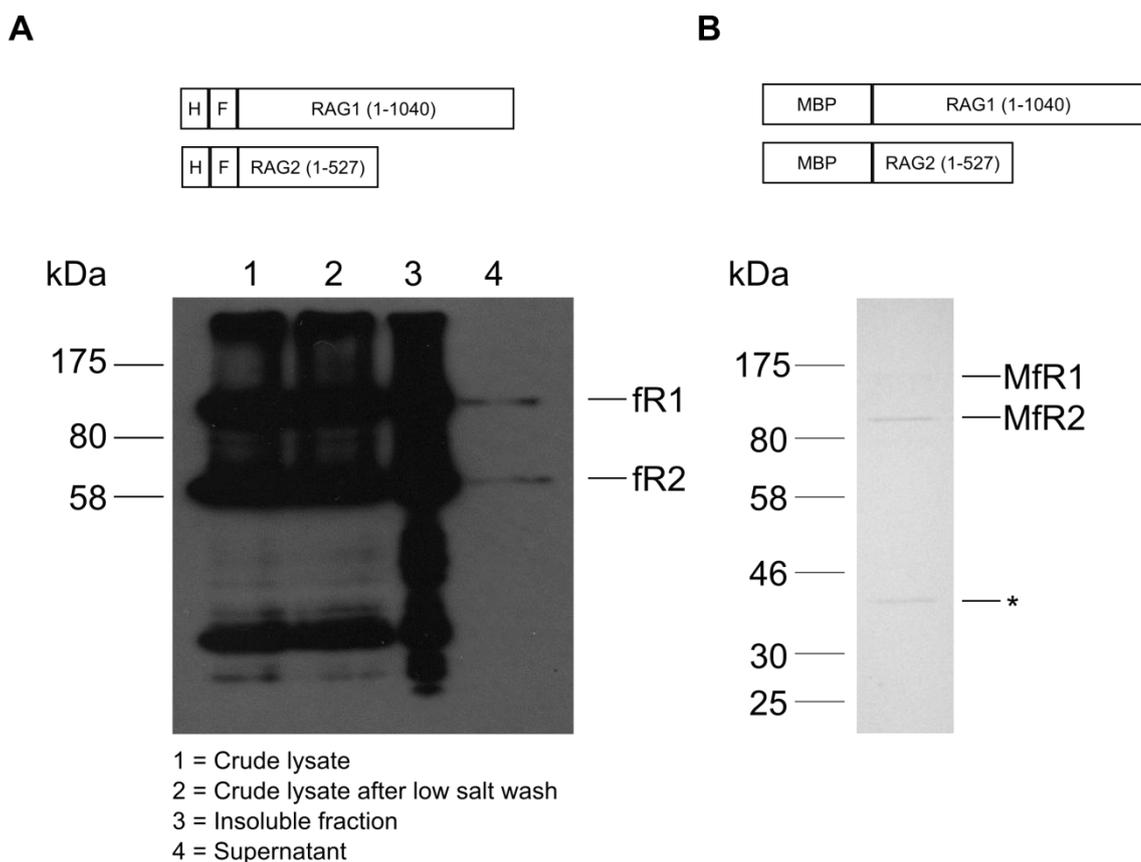
Figure 3.14 – Purification of full-length RAG proteins. (A) Expression via the vaccinia/T7 system did not yield soluble flRAG proteins. flRAG proteins can be detected up until the point of clarification by centrifugation, after which only a minimal amount can be detected in the supernatant. The schematic at the top shows the protein constructs and the tagging scheme. H = 9xHis tag, F = FLAG tag, fR1 = 9xHis/FLAG-tagged full-length RAG1 (122 kDa), fR2 = His9/FLAG-tagged full-length RAG2 (62 kDa) **(B)** MBP-tagged flRAG proteins purified from transfected HEK-293T cells. The schematic shows the protein constructs. MfR1 = MBP-full-length RAG1 (162 kDa), MfR2 = MBP-full-length RAG2 (102 kDA), * = contaminant or degradation product. The amount of MfR1 is just at or below the detection limit of Coomassie stain.

Full-length RAG activity was first tested with standard 12 and 23RSS substrates, using the same conditions determined for core RAG proteins in the presence of Mg^{2+}, to verify that the purified full-length RAG proteins adhere to the 12/23 rule. Addition of a 23RSS to a 12RSS stimulates RAG cleavage, with the greatest stimulation (four-fold) seen when the partner concentration is 10-fold greater (Fig. 3.15A, left). A 12RSS also stimulates 23RSS cleavage by a maximum of 3.1-fold when the concentration of 12RSS is 10 times greater (Fig. 3.15A, right). Addition of a non-specific oligonucleotide does not stimulate RAG cleavage by any more than 1.5-fold at either a 12 or 23RSS, indicating that stimulation of full-length RAG activity by a partner RSS is sequence-specific and symmetric, and that the purified full-length RAG proteins adhere to the 12/23 rule *in vitro* (Fig. 3.15B).
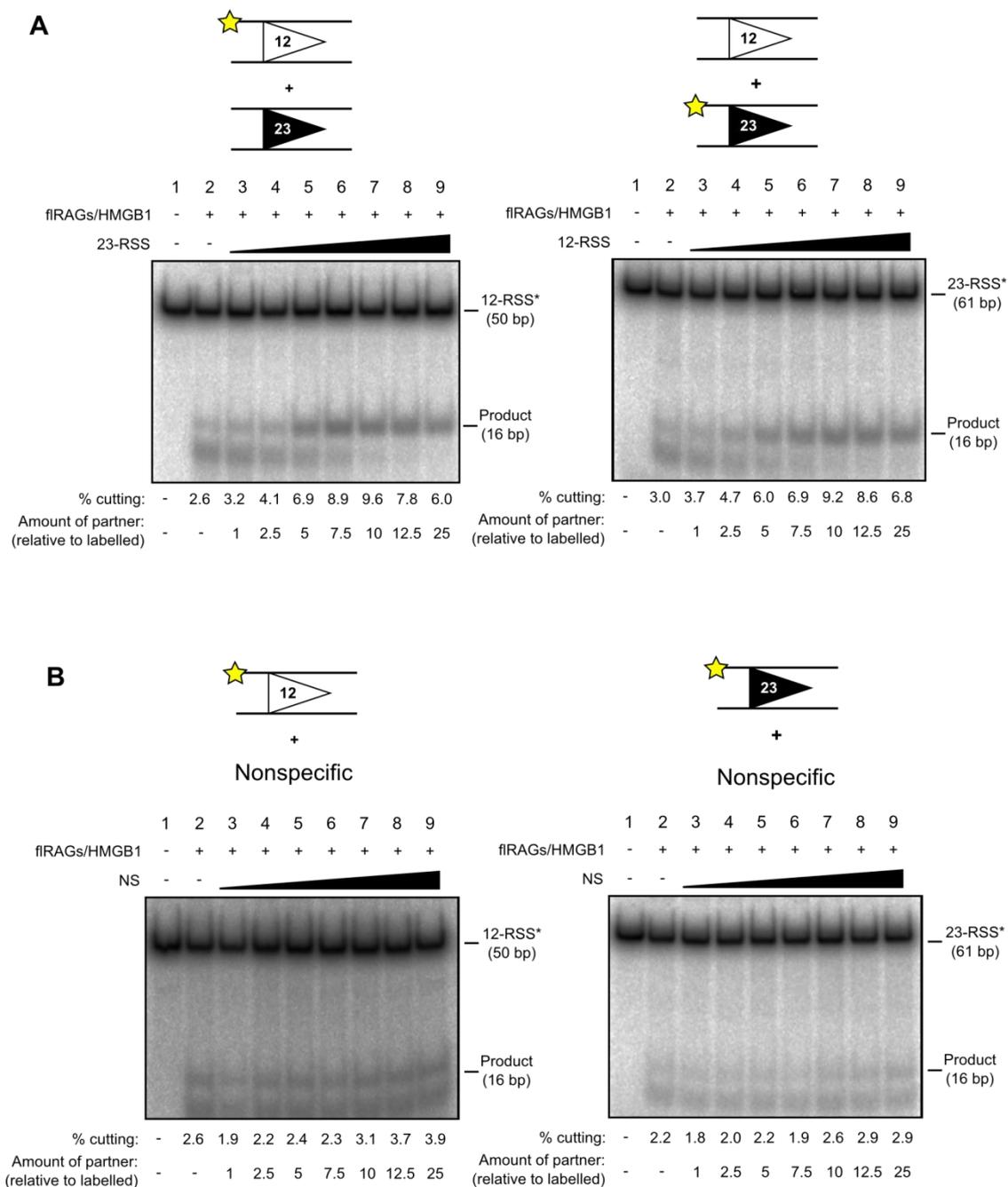
Figure 3.15 – Purified full-length RAG proteins adhere to the 12/23 rule in vitro (A) Addition of an unlabelled 23RSS oligonucleotide to a radiolabelled 12RSS oligonucleotide stimulates flRAG activity at the 12RSS. The greatest stimulation is seen when the partner substrate is 10 times more concentrated (lane 7) **(B)** A nonspecific 50mer does not significantly stimulate cleavage at either a 12 (left; 1.5-fold difference) or 23RSS (right; 1.3-fold difference). Diagrams above each gel indicate the substrates used, and a star indicates which substrate is radiolabelled. The percentage of cut substrate and the concentration of unlabelled partner is shown beneath each lane. The concentration of labelled oligonucleotide is 2 nM in each case.

I then tested whether cleavage of an RSS-ESC pair is also asymmetric when using full-length RAG proteins. Addition of an ESC to a labelled 12RSS causes a maximum increase of 1.6-fold, seen at a 5-fold excess of ESC, and addition of an ESC to a labelled 23RSS does not stimulate cleavage by a significant amount (Fig. 3.16A). Contrary to observations with core RAG proteins, addition of either a 12RSS or 23RSS partner to an ESC causes an increase in cleavage of 2.2-fold with a 12RSS and 3.1-fold with a 23RSS (Fig. 3.16B), but addition of a non-specific oligonucleotide does not stimulate cleavage (Fig. 3.17). Furthermore, cleavage and stimulation in a 12RSS-ESC complex is roughly equivalent (compare lane 7 in Fig. 3.16A left and 3.16B left), whilst an ESC is cut more efficiently in a 23RSS-ESC complex (lane 7 in Fig. 3.16A right and Fig. 3.16B right). Therefore, it seems that ESC-RSS cleavage *in vitro* is not asymmetric when using these purified full-length RAG proteins. However, results from qPCR-based experiments indicate that ESC-RSS cleavage with full-length RAG proteins is asymmetric *in vivo* (Section 5.1).
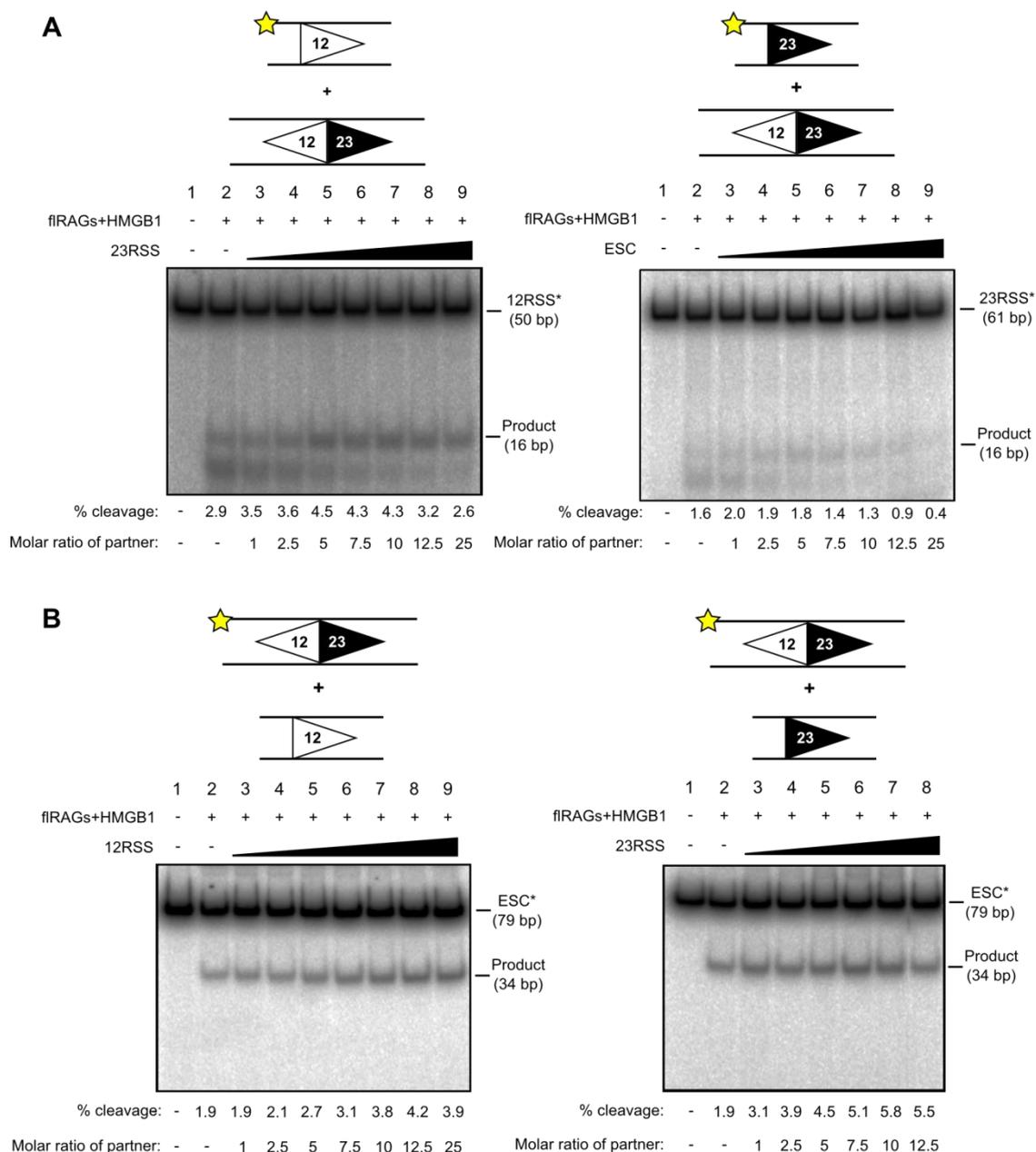
Figure 3.16 – Full-length RAG cleavage of an ESC-RSS pair is not asymmetric (A) Addition of an unlabelled ESC oligonucleotide to a radiolabelled 12RSS oligonucleotide (left) or 23RSS (right) stimulates flRAG-mediated cleavage, with the greatest stimulation seen when the ESC is 10 times more concentrated. **(B)** A 12RSS stimulates ESC cleavage by a maximum of 2.2-fold, whereas a 23RSS stimulates ESC cleavage by 3-fold. Diagrams above each gel indicate the substrates used, and the star indicates which substrate is radiolabelled. The percentage of cut substrate is shown below each lane with the concentration of unlabelled partner shown beneath that. The concentration of labelled oligonucleotide is 2 nM in each case.
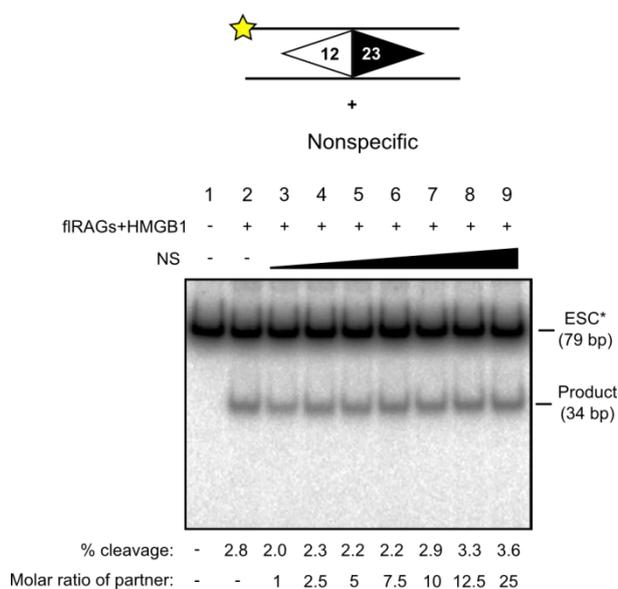
Figure 3.17 – Cleavage of an ESC is not stimulated by a nonspecific oligonucleotide. Labelled ESC oligonucleotide was incubated for 15 min at 37°C in Mg^{2+}-containing buffer with full-length RAG proteins, HMGB1 and an unlabelled nonspecific oligonucleotide (concentration shown underneath each lane). NS = non-specific oligonucleotide.

3.6   Mechanism of ESC cleavage by RAG proteins

Cleavage of standard 12 or 23RSSs occurs by nicking an RSS precisely between the conserved heptamer and the coding segment, creating a free 3' hydroxyl group which then attacks the opposite DNA strand in a direct trans-esterifcation reaction (McBlane et al, 1995). Hairpin formation requires base flipping of the nicked strand, catalysed by aromatic residues in RAG1, such that it can undergo nucleophilic attack by the opposing DNA strand (Lu et al., 2006; Bischerour et al., 2009).

The two papers that have investigated ESC cleavage disagree on the mechanism of its cleavage. One group's data indicate that cleavage of an ESC occurs via the standard nick-hairpin mechanism (Vanura et al., 2007), whereas Roth's group propose an alternative "nick-nick" mechanism, whereby the RAG proteins nick one strand of the ESC between the two heptamers, and then bind to the other RSS and nick the opposite strand, thus creating a double strand break with two blunt ends and no hairpin (Neiditch et al., 2002). The different conclusions reached by these two groups could be a result of the different experimental approaches. Vanura et al. (2007) detected reintegration products on transfected substrates by PCR, and Roth's group used ligation-mediated PCR (LM-PCR) to amplify broken ESC ends. Since the latter is a highly sensitive technique, it is possible that the contribution of nicked substrates to overall cleavage was overestimated.

Nevertheless, since the available information suggests that both scenarios are a possibility, I examined whether the cleavage products formed in my experiments are the result of the standard nick-hairpin or the alternative nick-nick mechanism. A labelled 12RSS, 23RSS or ESC oligonucleotide was incubated for 15 min at 37°C with cRAG proteins, HMGB1, and a 10-fold excess of unlabelled partner (Fig. 3.18). Half of the reaction was run on a native polyacrylamide gel (Fig. 3.18A), and the other half was purified by phenol-chloroform extraction and run on a denaturing polyacrylamide gel (Fig. 3.18B).

On a native gel, single-stranded nicks will remain annealed to the complementary strand. Therefore, only double-strand breaks generated by either hairpin formation of nick-nick cleavage are visible. Substrates that contain a single-strand nick will co-migrate with uncut substrate. When the 12RSS is labelled, a 10- to 12-fold stimulation in cleavage is observed in the presence of a 23RSS (Fig. 3.18A, lane 3) or ESC (lane 4), and when the 23RSS is labelled, there is a 35-fold stimulation in cleavage in the presence of 12RSS and a 17.5-fold stimulation in the presence of ESC (Fig. 3.18A, lanes 7 and 8). Addition of a 12RSS to an ESC causes a 7.7-fold stimulation in cleavage (Fig. 3.18A, lane 11), but a 23RSS does not greatly stimulate cleavage above the level seen at an ESC alone (Fig. 3.18A, lanes 10 and 12). Overall, the level of asymmetry in cleavage between an ESC and a 12RSS is 2.6-fold, similar to my previous observations, and the asymmetry between a 23RSS and an ESC is 12.7-fold.

On a denaturing gel, single-strand nicks and nick-nick products (if they occur) co-migrate and denatured hairpins run at a size roughly twice as large as the nick products. If the ESC is cleaved by the nick-nick mechanism, then no hairpin products will be visible. When a lone 12RSS is incubated with cRAG proteins, 70% of the oligo is nicked (Fig. 3.18B, lane 2). Only 1.2% of substrate is cut on a native gel (Fig. 3.18A), therefore the nick products seen on the denaturing gel are caused by nicking of the oligonucleotide on one strand. This high level of nicking is consistent with the idea that RAG proteins can efficiently nick a single RSS in the absence of partner (Eastman and Schatz, 1997; Yu and Lieber, 2000). The amount of 12RSS hairpin product increases 35-fold in the presence of 23RSS or ESC (Fig. 3.18B, lanes 3 and 4, respectively). The increase in the percentage of hairpin products is similar to the increase in cleavage on the native gel, and the amount of nicked product only changes by twofold, relative to RSS alone. These observations are consistent with those made by other groups that in the presence of Mg^{2+}, RAG proteins can only make a single-strand nick at a lone RSS, and a second RSS of different spacer length is required for efficient hairpin formation (Yu and Lieber, 2000). The same effect is observed with a 23RSS (Fig. 3.18B, lanes 5-

8), indicating that RAG cleavage of RSSs occurs via the standard nick-hairpin mechanism in the presence of a suitable partner RSS, as expected, but also with an ESC.
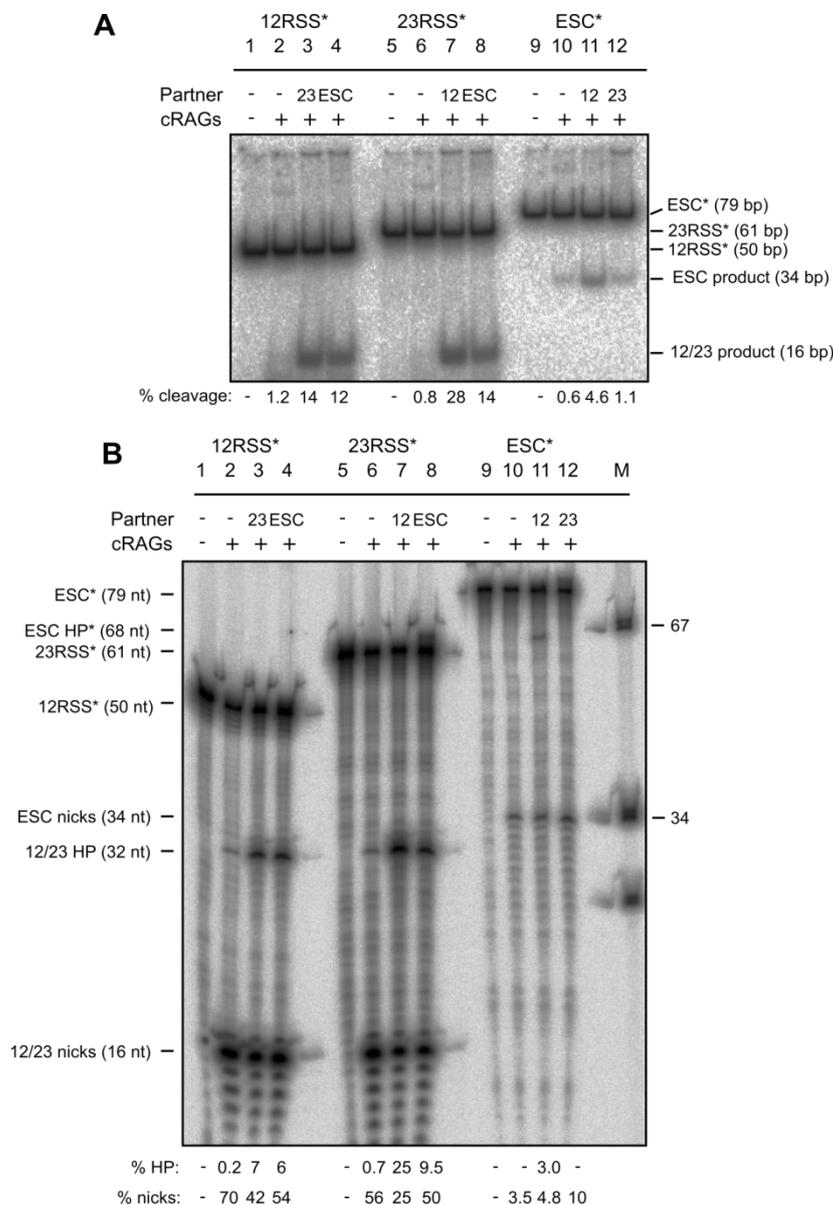
Figure 3.18 – Core RAG proteins cleave ESCs via a nick-hairpin mechanism. Core RAG cleavage reactions were performed as in previous experiments, using a labelled 12RSS, 23RSS or ESC oligonucleotide and a 10-fold excess of unlabelled partner as indicated. **(A)** Half the reaction was run on a native polyacrylamide gel, which only separates uncut and completely cut substrate. Substrates nicked on one strand co-migrate with uncut substrate, and nick-nick products (if formed) co-migrate with hairpin products. **(B)** The other half of the reaction was purified by phenol-chloroform extraction and run on a denaturing polyacrylamide gel. The denaturing conditions allow for separation of nicked products and hairpin products. HP = hairpins, nt = nucleotides.

At an ESC alone, no hairpin product is seen (Fig. 3.18B, lane 10). Either the ESC DSBs seen on the native gel (Fig. 3.18A, lane 10) are caused entirely by nick-nick cleavage, or the level of hairpinning was too low to be observed above background on the denaturing gel. In the presence of a 12RSS, there is an increase in the level of ESC hairpin formation that is similar to the level of cleavage seen on the native gel (Fig. 3.18B, lane 11), indicating that the addition of a 12RSS stimulates cleavage of ESCs via the standard nick-hairpin mechanism rather than the nick-nick mechanism. 12RSS stimulation of an ESC was not previously observed with cRAG proteins (Fig. 3.11B). The reason for this discrepancy is unclear, but it could be that the different cRAG preps used in these experiments show slightly different levels of activity. Importantly, cleavage of an RSS-ESC pair is asymmetric in both cases.

An increase in ESC hairpin products was not detected in the presence of a 23RSS (Fig. 3.18B, lane 12), consistent with the lack of stimulation by a 23RSS on the native gel (Fig. 3.18A, lane 12). However, the level of nicking increases 3.7-fold compared to ESC alone, with no corresponding increase in ESC cleavage on the native gel (Fig. 18A, lane 12). This indicates that a 23RSS can stimulate single-strand nicking of an ESC, but hairpin formation with a 23RSS partner is blocked.

When using full-length RAG proteins, the ESC did not stimulate cleavage of either RSS above an RSS alone (Fig. 3.19A). Similar to the results presented above (Fig. 3.16), the cleavage efficiency of an ESC-RSS pair is roughly equivalent, indicating that the cleavage is not asymmetric with these full-length RAG proteins. 12/23RSSs stimulate cleavage of one another, which correlates with an increase in hairpin products on the denaturing gel, as expected (Fig. 3.19B, lanes 3 and 7). There is little to no hairpin product formed at an isolated ESC (Fig. 3.19B, lane 10). However, addition of a 12RSS, but not a 23RSS, results in a slight increase in hairpin formation (Fig. 3.19B, lanes 11 and 12). In this instance, the ESC hairpin product does not run at the expected size of 68 nucleotides (Fig. 3.19B, lane 11). One possible explanation for this is that the ESC hairpin product was not fully denatured, and therefore migrated faster than expected. The fact that the increase in this band correlates well with the increase in product seen on a native gel suggests that this band does in fact represent ESC hairpin products.
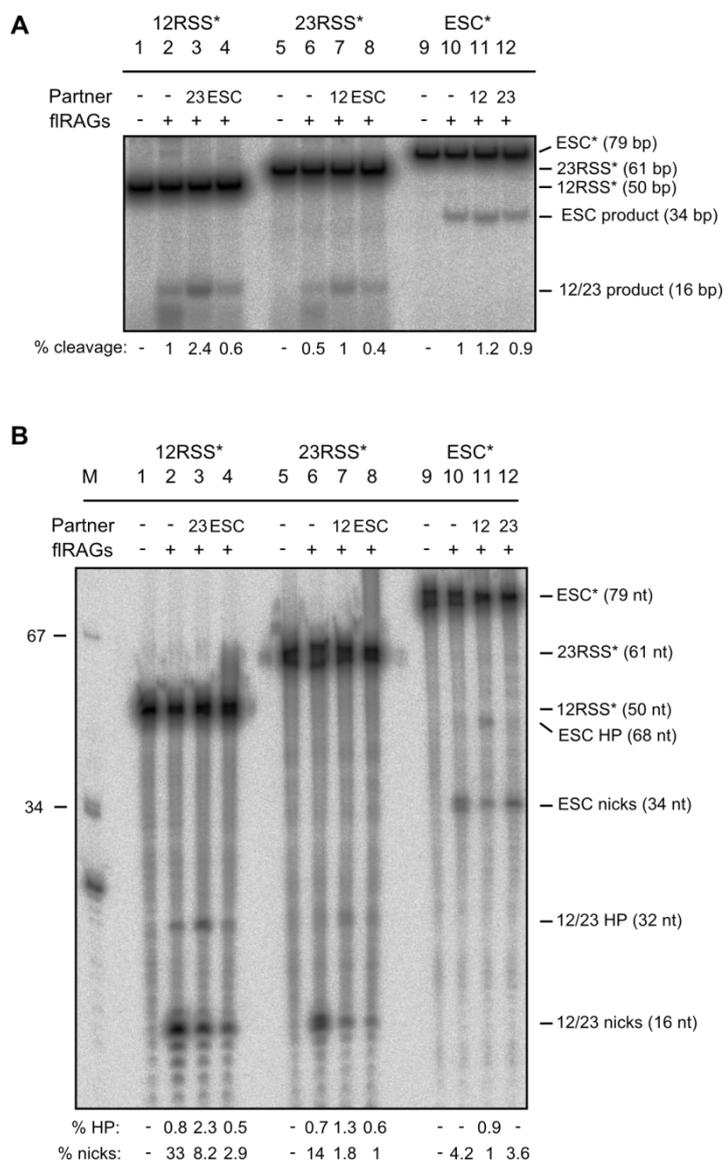
Figure 3.19 – Full-length RAG proteins cleave ESCs via a nick-hairpin mechanism. Full-length RAG cleavage reactions were performed as in previous experiments, using a labelled 12RSS, 23RSS or ESC oligonucleotide and a 10-fold excess of unlabelled partner as indicated. **(A)** Half the reaction was run on a native polyacrylamide gel, which only separates uncut and completely cut substrate. Substrates nicked on one strand co-migrate with uncut substrate, and nick-nick products (if formed) co-migrate with hairpin products. **(B)** The other half of the reaction was purified by phenol-chloroform extraction and run on a denaturing polyacrylamide gel. The denaturing conditions allow for separation of nicked products and hairpin products. HP = hairpins, nt = nucleotides.

Although these full-length RAG proteins do not display asymmetric cleavage activity, the lack of ESC hairpin formation in the presence of a 23RSS is common to both the core and full-length RAG proteins. This suggests that the conformation of a 23RSS-ESC is somehow more inhibitory to ESC cleavage than the conformation of a 12RSS-ESC complex. Overall, these results indicate that in my experimental system, ESCs are predominantly cleaved via the standard nick-hairpin mechanism when paired with a

12RSS, but a low level of nick-nick cleavage of ESCs could occur in the absence of partner.

C)   Discussion

RAG proteins cleave RSSs and ESCs in an asymmetric manner

Two groups have reported that the RAG proteins can catalyse reintegration of the by-product of V(D)J recombination, the excised signal circle, back into the genome of B and T cells (Vanura et al., 2007, Curry et al., 2007). Furthermore, Curry et al. (2007) reported that the presence of the RAG2 C-terminus reduced the level of reintegrated products by around 7-fold. Since this is potentially a very harmful reaction, I aimed to investigate the biochemistry of this reaction, specifically the inhibitory effect of the RAG2 C-terminus. Using purified core RAG proteins, HMGB1 and different combinations of labelled and unlabelled oligonucleotide substrates, I made the unexpected discovery that whilst addition of an ESC can stimulate cleavage of both 12 and 23RSSs, the level of cleavage at an ESC under identical conditions is consistently lower than at an RSS (i.e. cleavage of an RSS/ESC is asymmetric).

This observation was first made using core RAG proteins in the presence of Mn^{2+}, but the reaction conditions had to be optimized since stimulation in Mn^{2+} was found to be non-specific. After achieving cleavage and stimulation in the presence of Mg^{2+}, the asymmetry in substrate cleavage was clearly observed with core RAG proteins, but was much less pronounced with full-length RAG proteins.

Stimulation of RAG activity is non-specific in the presence of Mn^{2+}

RAG activity requires a divalent metal cation cofactor, and can utilise either Ca^{2+}, Mn^{2+} or Mg^{2+}, although each one influences RAG recombinase activity in a different way. Ca^{2+} enables RAG proteins to bind RSSs, but precludes cleavage, whereas Mn^{2+} and Mg^{2+} both allow binding and cleavage (Swanson 2004). However, the consensus in the literature is that in the presence of Mn^{2+}, RAG proteins can create a double-strand break at a lone RSS (McBlane et al., 1995), whereas in Mg^{2+} RAG proteins can create single-strand nicks at a lone RSS, but hairpin formation can only occur in the context of a paired complex between two different length RSSs (van Gent et al., 1996c). Thus, cleavage in the presence of Mn^{2+} is uncoupled, whereas cleavage in the presence of Mg^{2+} is coupled. Since V(D)J recombination follows the 12/23 rule (Tonegawa 1983),

and the nuclear concentration of Mg^{2+} is greater than Mn^{2+}, Mg^{2+} is considered to be the physiologically relevant divalent metal cation.

My initial experiments were performed in the presence of Mn^{2+}, because the level of RAG activity in the presence of Mg^{2+} was too low for detection. It was in the presence of Mn^{2+} that I first observed a very striking asymmetry in RSS-ESC cleavage (Fig. 3.3 and Fig. 3.4). Since Mn^{2+} supported stimulation outside of the 12/23 restriction, it was replaced with Mg^{2+}.

Efficient hairpin formation in the presence of Mg^{2+} was only observed when DMSO was included in the reaction (Fig. 3.7A). The function of DMSO in this context is unclear. DMSO can cause aggregration or oligomerisation of proteins, but it is unlikely that this is the case here, since the complexes, or the abundance of certain complexes seen in a bandshift assay do not change in the presence of DMSO (see Chapter 4). Therefore DMSO does not seem to be affecting the overall oligomeric state of the RAG-RSS complex. DMSO is commonly used to lower the melting point of DNA, for example when PCR-amplifying GC-rich sequences. The mechanism of RAG-catalysed DNA cleavage is thought to involve the local melting of DNA structure since the RAG1 central domain preferentially binds single-stranded DNA (Peak et al., 2003), and base pair mismatches at the heptamer-coding boundary stimulate RAG activity (Cuomo et al., 1996; Ramsden et al., 1996; Grundy et al., 2007). Therefore, a possible reason for the requirement for DMSO is that in Mg^{2+}, RAG proteins need an extra "helping-hand" from DMSO to fully denature the DNA around the cut site. The use of DMSO is, of course, not physiological, and since other groups report Mg^{2+}-catalysed cleavage without the need for DMSO, it is unclear why DMSO is needed for RAG cleavage in my experimental system. It is possible that the active fraction of RAG proteins in my preparations is low, and DMSO is required to stimulate cleavage activity to levels that can be seen on a gel.

Another factor that required optimisation was the incubation length of the reactions. When reactions containing Mg^{2+} were incubated for as long as reactions containing Mn^{2+}, a partner substrate competed for cleavage rather than stimulated cleavage (Fig. 3.7). A possible explanation is that there is an initial coupled reaction, where RAG proteins form a synaptic complex with two substrates, and a "slow" uncoupled reaction, where RAG proteins form hairpins at a lone RSS. Stimulation by another substrate is only seen at or before 15 minutes of incubation. Beyond this, products formed by cleavage of a single RSS begin to outnumber those formed by paired cleavage in a synaptic complex.

Purified full-length RAG proteins do not catalyse asymmetric cleavage

In order to assess the likelihood that asymmetric cleavage could occur *in vivo*, it was important that I repeated these observations with full-length RAG (flRAG) proteins. The first attempt to express and purify flRAG proteins used a vaccinia/HeLa-based system (McBlane et al., 1995, Elkin et al., 2005). However, I found that this system did not yield soluble flRAG proteins, so expression in HEK-293T cells was adopted instead. The major difference between the two approaches that is likely to affect solubility is the tagging scheme. flRAG proteins purified from HEK-293T cells were soluble, but unlike flRAG proteins expressed using the vaccinia system, they were fused with MBP. Although flRAG proteins without the MBP tag have reportedly been purified using the vaccinia system, the yield and activity were very low (Elkin et al., 2003). It is possible that the vaccinia expression system might have worked in my hands, and produced higher yields for other groups, if the flRAG proteins had been fused with MBP. Furthermore, recent studies have highlighted that full-length RAG2 mediates autoinhibition of RAG cleavage efficiency via the PHD finger. The presence of H3K4me3 is thought to relieve this autoinhibition (Shimazaki et al., 2009; Grundy et al., 2010). Therefore, the lack of H3K4me3 in these experiments, coupled with the low yield of flRAG proteins, could account for the low activity levels.

Cleavage was not found to be asymmetric when using full-length RAG proteins. The yield and concentration of full-length RAG proteins (Fig. 3.14B) were low compared to core RAG proteins (Fig. 3.8). Concentrating the full-length proteins using a spin concentrator did not work since a large proportion of full-length RAG proteins were seemingly lost through non-specific binding to the concentrator membrane, and pre-treating the concentrator with BSA did not help. Therefore, relatively low levels of full-length RAG activity had to be used in cleavage experiments, which could have affected the results. It is possible that the non-core regions of the full-length RAG proteins somehow affect interactions with ESCs and RSSs. However, *in vivo* experiments presented in Chapter 5 show that ESC-RSS cleavage is indeed asymmetric with full-length RAG proteins, so these *in vitro* results do not rule out the possibility of asymmetric cleavage and cut-and-run occurring in lymphocytes.

ESC cleavage is blocked at the nicking step

Analysis of the cleavage products on a denaturing gel shows that there is very little nicking of the ESC, either alone or in the presence of an RSS (Fig. 3.18B). The fact that there is very little nicking at an ESC alone, unlike RSS alone, suggests that

cleavage of ESCs is blocked at the nicking step. Since an ESC contains two RSSs joined at the heptamer, it is possible that RAG proteins simultaneously bind to both sides, thus preventing the RAG complex bound on either side from binding to heptamer boundary and catalysing a nick. The lack of nicking at an ESC is consistent with this hypothesis, which is addressed in further detail in Chapter 4.

The denaturing gels also show that a low level of hairpinning is observed at an ESC in the presence of a 12RSS partner, but not in the presence of a 23RSS, with both core and full-length RAG proteins (Fig. 3.18B and Fig. 3.19B). It is unclear why hairpin formation is less efficient when using 23RSS partner. It could be that when bound to a 12RSS and ESC pair, the conformation of RAG proteins is such that a low level of hairpin formation is possible, but when paired with a 23RSS, blocking of the heptamer-heptamer junction is greater and entirely prevents hairpin formation. Interestingly, the lack of ESC hairpins with a 23RSS partner is consistent with a study which showed that the majority of reintegration events *in vivo* occur at Vκ 12RSSs and not Jκ 23RSSs (Curry et al., 2007). The fact that ESC hairpin formation occurs less efficiently when paired with a 23RSS might explain this observation.

The cut-and-run hypothesis

If cleavage within an ESC-RSS complex truly is asymmetric in cells, as I have observed *in vitro* using core RAG proteins, it is possible that the predominant outcome of RAG-ESC-RSS interactions *in vivo* is my proposed mechanism of cut-and-run, and not reintegration. Aside from the challenges that the introduction of multiple double strand breaks pose to genomic stability, the most significant risk is that they could be a potent source of chromosome translocation partners at the antigen receptor loci. The most common chromosome translocation resulting from aberrant V(D)J recombination is end donation (Section 1.24), where a break at one of the antigen receptor loci is joined with a break formed by an unrelated process. However, the source of orphan broken RSSs is unclear. It is possible that cut-and-run is a source of these broken ends.

Chapter 4

Determining the molecular basis of asymmetric ESC-RSS cleavage

A)   Introduction

In Chapter 3, I presented data showing that cleavage of an RSS and an ESC is asymmetric when bound by core RAG proteins. Extrapolating from this observation, I put forward the cut-and-run hypothesis, which states that ESC-RAG complexes are able to stimulate multiple double-strand breaks at RSSs within the genome of developing B and T cells when these ESCs are present at the same time as RAG proteins. The aim of this chapter is to identify the molecular basis of asymmetric cleavage by examining RAG-ESC interactions and whether there are any key differences between RAG-RSS and RAG-ESC binding.

The majority of information regarding RAG-RSS interactions comes from experiments using electrophoretic mobility shift assays (EMSAs) and DNA footprinting. Core RAG1 alone forms a stable dimer in solution (Bailin et al., 1999; Swanson and Desiderio, 1999), and can bind to an isolated RSS with an apparent K_d of ~35 nM or ~100 nM, depending on the method employed (fluorescence anisotropy or EMSA, respectively; Rodgers et al., 1999; Ciubotaru et al., 2003; Zhao et al., 2009). Unlike RAG1, RAG2 alone has little to no intrinsic DNA binding activity (Difilippantonio et al., 1996; Spanopoulou et al., 1996; Hiom and Gellert, 1998; Swanson and Desiderio, 1998; Mundy et al., 2002). However, RAG1 and RAG2 can interact in the absence of DNA (forming a 1:1 heterotetramer; (Bailin et al., 1999; Swanson and Desiderio, 1999) and together, RAG1 and RAG2 bind a single RSS with greater affinity (K_d ~13.5-25 nM) and specificity than RAG1 alone (Zhao et al., 2009; Wang et al., 2012).

In the absence of RAG2, the majority of RAG1-DNA interactions occur between the RSS nonamer and the nonamer binding domain (NBD) of RAG1 (Yin et al., 2009). RAG1-nonamer interactions are a combination of several base-specific interactions and extensive nonspecific contacts with the phosphate backbone, which probably explains why RAG proteins efficiently bind sequences nonspecifically (Swanson and Desiderio, 1998). In the presence of RAG2, these contacts are extended to include the entirety of the spacer, the heptamer, and the heptamer-coding sequence boundary (Difilippantonio et al., 1996). Therefore, the primary role of RAG2 is to direct the RAG1 C-terminus (containing the catalytic DDE residues) to the heptamer and heptamer-coding sequence boundary. Furthermore, the active site of one RAG subunit is

contributed *in trans* to the partner RSS (i.e. each RAG1 subunit cleaves the RSS that is not bound by its NBD; Swanson, 2001).

There has been no detailed analysis of RAG-ESC interactions published to date. Since an ESC contains two RSSs joined at the heptamer, it is possible the RAG proteins bind to both RSSs of the ESC simultaneously, and the competing interactions at the heptamer-heptamer junction prevent efficient cleavage of the ESC, leading to asymmetric cleavage in an RSS-ESC complex. Here, this hypothesis was tested using core RAG proteins and radiolabelled oligonucleotide substrates in electrophoretic mobility shift assays (EMSAs), which show that an additional higher molecular weight complex forms at an ESC that is not seen at an RSS, consistent with RAG proteins simultaneously binding to both RSSs of the ESC. To test whether this is indeed the source of asymmetric ESC-RSS cleavage, I carried out *in vitro* cleavage assays using ESC substrates where one RSS of the ESC is mutated to prevent RAG binding. These experiments showed that mutation of one of the RSSs restores the level of cleavage in an RSS-ESC complex to symmetric levels. Therefore, RAG binding to both sides of the ESC appears to be the source of asymmetric RSS-ESC cleavage (i.e. inhibits ESC cleavage).

In addition, I tested one of the predictions of the cut-and-run hypothesis, namely that when an ESC stimulates RAG cleavage of a genomic RSS, the RSS is then released by the RAG complex, but the ESC remains bound by RAG proteins. To test this prediction, I performed a series of electrophoretic mobility shift time course assays under conditions that allow cleavage. Data from these experiments suggest that as predicted, RAG proteins release the cleaved RSS after stimulation by an ESC, but remain bound to the ESC.

B) **Results**

4.1 **Analysis of RAG-RSS and RAG-ESC complexes by EMSA**

Since an ESC contains two RSSs in a head-to-head configuration, I predicted that the RAG proteins can bind to both sides of the ESC simultaneously and that in doing so, they prevent cleavage at the heptamer-heptamer junction of the RSSs in an ESC. To compare RAG-RSS and RAG-ESC complexes, I carried out EMSAs using the same radiolabelled oligonucleotide substrates that were used in the cleavage experiments in Chapter 3. Reactions were set up in the same manner except that Ca^{2+} was used instead of Mg^{2+} to allow binding but prevent cleavage. Thus, only pre-cleavage complexes are observed.

When a labelled 12RSS is incubated with HMGB1 and no RAG proteins, a low molecular weight HMG-RSS complex is observed (Fig. 4.1A, lane 2), consistent with previous reports that HMGB1 can bind RSS-containing DNA in the absence of RAG proteins (Swanson, 2002a). When a 12RSS is bound to cRAG proteins alone without HMGB1, two major complexes are observed (Fig. 4.1A, lane 3), consistent with previously published data (Mundy et al., 2002; Swanson, 2002b). These two complexes are generally referred to in the literature as single complex (SC) 1 and SC2 (Swanson et al., 2009). Work from other laboratories has shown that SC1 contains two molecules of RAG1 and one molecule of RAG2, whereas SC2 contains two molecules each of RAG1 and RAG2 (Fig. 4.2; Bailin et al., 1999; Swanson, 2002b).

Addition of HMGB1 supershifts both of these complexes to form HSC1 and HSC2 (HMGB1-containing single RSS complex 1 and 2; Fig. 4.1A, lane 4), consistent with previous reports that HMGB1 can incorporate into the RAG-RSS complex (Swanson, 2002a). The stoichiometry of HMGB1 within this pre-cleavage complex is unclear, but a study of the post-cleavage signal end complex (SEC; RAG proteins bound to cleaved signal ends) indicates that the ratio of RAG1:RAG2:HMGB1 is 2:2:1 (Fig. 4.2; Grundy et al., 2009).
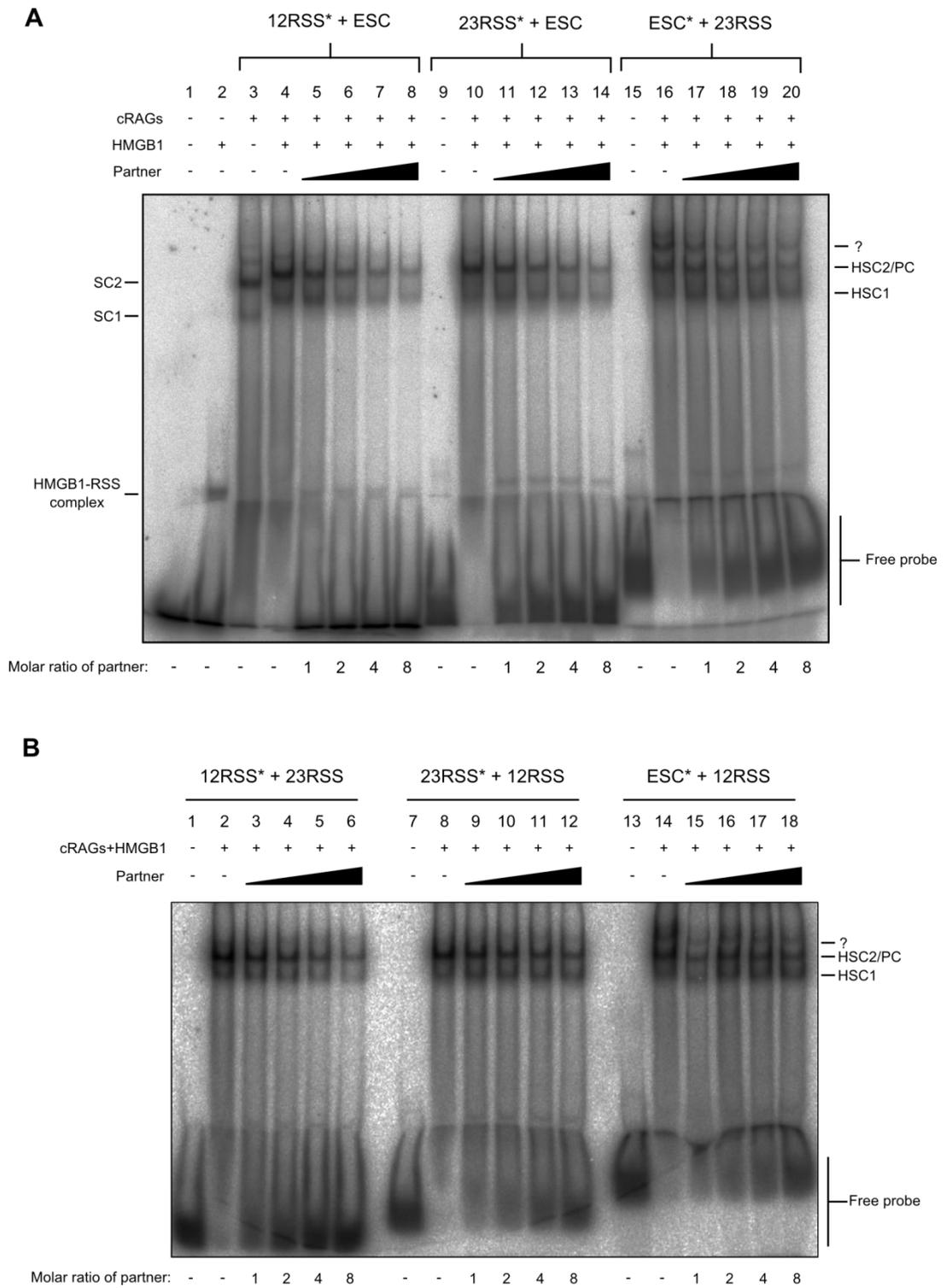
Figure 4.1 – cRAG proteins form an additional complex with an ESC. **(A)** To compare the RAG complexes that form with different substrates, a radiolabelled 12RSS, 23RSS or ESC oligonucleotide was incubated in the presence of Ca^{2+} with cRAG proteins, HMGB1 and an unlabelled partner substrate (indicated above the gel) for 10 min at 37°C, and run on a 4% polyacrylamide gel. A third complex of higher molecular weight is more visible when the labelled substrate is an ESC (indicated by "?"). **(B)** As for (A), but using different substrate combinations, as noted above the gel. SC1/2 = single complex 1/2, HSC1/2 = HMGB1 single complex 1/2, PC = paired complex
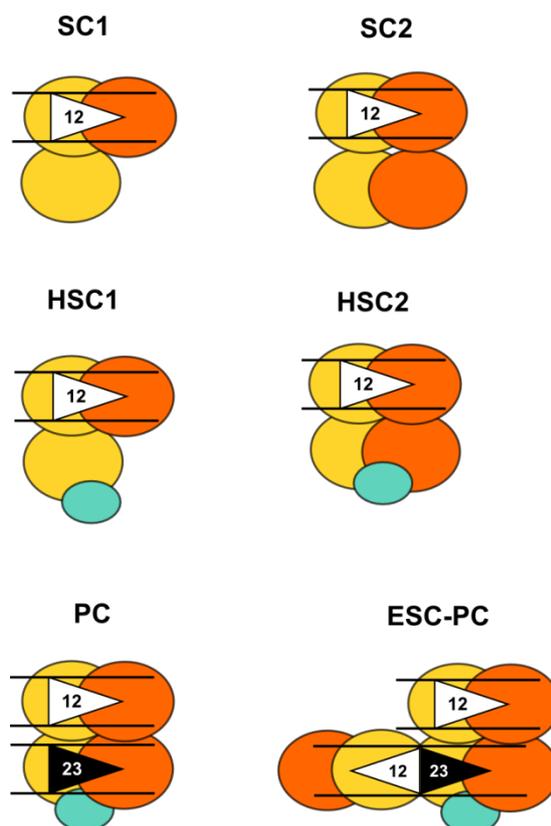
Figure 4.2 – Schematic of RAG-RSS complexes observed in an EMSA. The stoichiometry of subunits within each complex is depicted, based on results from other laboratories. In the presence of a single substrate, RAG proteins form single complex 1 (SC1) and single complex 2 (SC2). SC1 contains a RAG1 dimer and a RAG2 monomer which freely reassorts (Swanson 2004). SC2 contains two molecules of RAG2 that freely reassort (Swanson et al., 2004). Addition of HMGB1 creates HMGB1-SC1 (HSC1) and HSC2. It is thought that a RAG heterotetramer can form on a single substrate, and a second substrate incorporates into HSC2 to form a paired complex (PC), which contains exactly the same complement of proteins as HSC2. When an ESC is used, an additional complex of higher molecular weight is observed in an EMSA, indicating that one or more additional RAG molecules binds to the ESC. The schematics only depict the stoichiometry of each component within each complex, and are not intended to accurately reflect the overall 3D structure or conformation of the complexes. Yellow oval = RAG1, orange oval = RAG2, blue oval = HMGB1, triangles with either 12 or 23 depict 12 or 23RSSs.

Addition of an unlabelled 23RSS to a labelled 12RSS does not cause a detectable shift in mobility of either HSC1 or HSC2 (Fig. 4.1B, lane 6). Analysis of the cleavage activity of the different RAG-RSS complexes demonstrated that addition of a partner substrate stimulates the cleavage activity of HSC2, but not HSC1, by around 4-fold (Swanson, 2002a). Therefore, it is generally inferred that the second partner RSS incorporates into HSC2 to form a paired complex (PC) that co-migrates with HSC2, since no additional RAG molecules are recruited (Fig. 4.2). Since HSC1 and HSC2 are observed in Fig. 4.1, and addition of a partner stimulates cleavage (Fig. 3.9A and B), it is assumed that PC formation also occurs in the gels shown in Fig. 4.1. Analysis of full-length RAG-

RSS complexes was attempted, but unfortunately the concentration of flRAG proteins was too low to form distinct complexes (Fig. 4.3), which was the case with three different flRAG preparations.
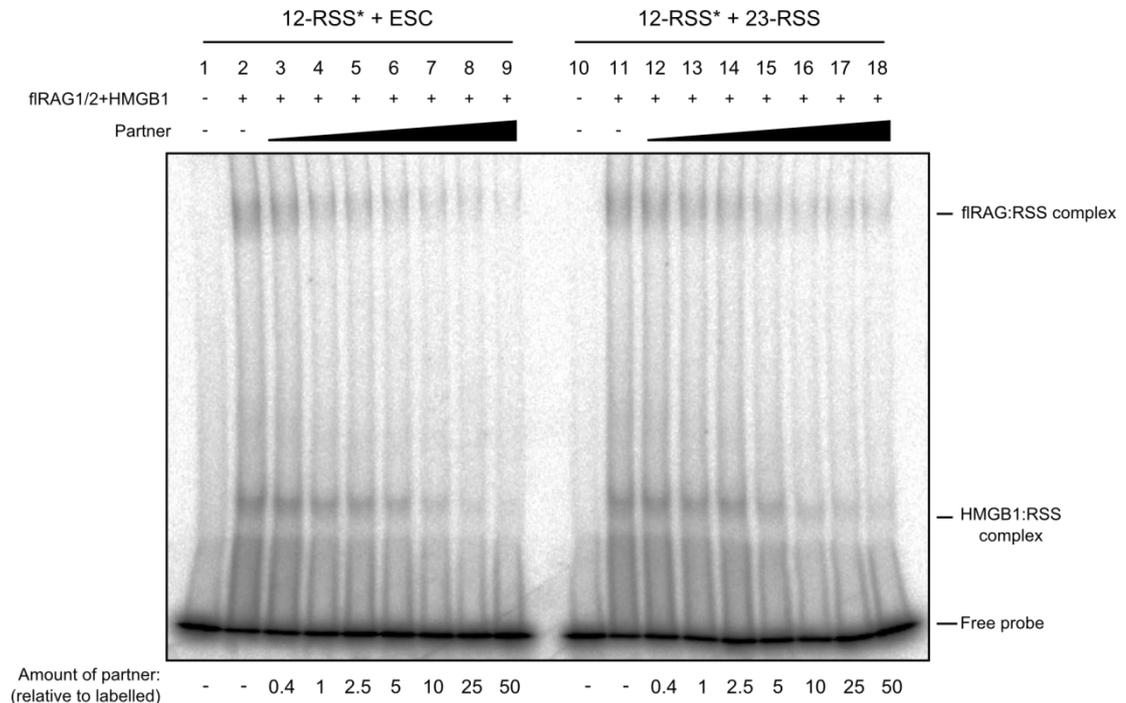
Figure 4.3 – EMSA using flRAG proteins. To investigate full-length RAG complex formation, EMSAs were carried out in an identical manner to Figure 4.1, except that purified flRAG proteins were used instead of purified cRAG proteins. A labelled 12RSS oligonucleotide was incubated with increasing amounts of unlabelled ESC or 23RSS oligonucleotide.

In an attempt to more clearly demonstrate PC formation, cRAG proteins were incubated with a 23RSS oligonucleotide containing biotin and Alexa Fluor 488 labels at opposite ends, and either a radiolabelled 12RSS or ESC (Fig. 4.4). Addition of monomeric streptavidin should supershift any complex containing biotinylated 23RSS, but a supershifted 12RSS complex will only be observed if both substrates are contained with the same complex (i.e. if they form a PC). Two supershifted complexes are seen when streptavidin is added, concomitant with a decrease in the intensity of both HSC1 and HSC2 (Fig. 4.4, bottom panel, lanes 6 and 13).

However, neither of these changes were seen with the radiolabelled 12RSS or ESC complexes (Fig. 4.4, top panel). This could indicate that the PC does not form, but since RAG cleavage is stimulated by a 12/23RSS pairing under these conditions (Fig. 3.9A), it is more likely that the supershift is not observed because the PC is relatively transient, or is unstable during electrophoresis. However, the biotinylated,

fluorescently-labelled 23RSS should be used in a cleavage assay to confirm that these additional groups do not adversely affect stimulation.
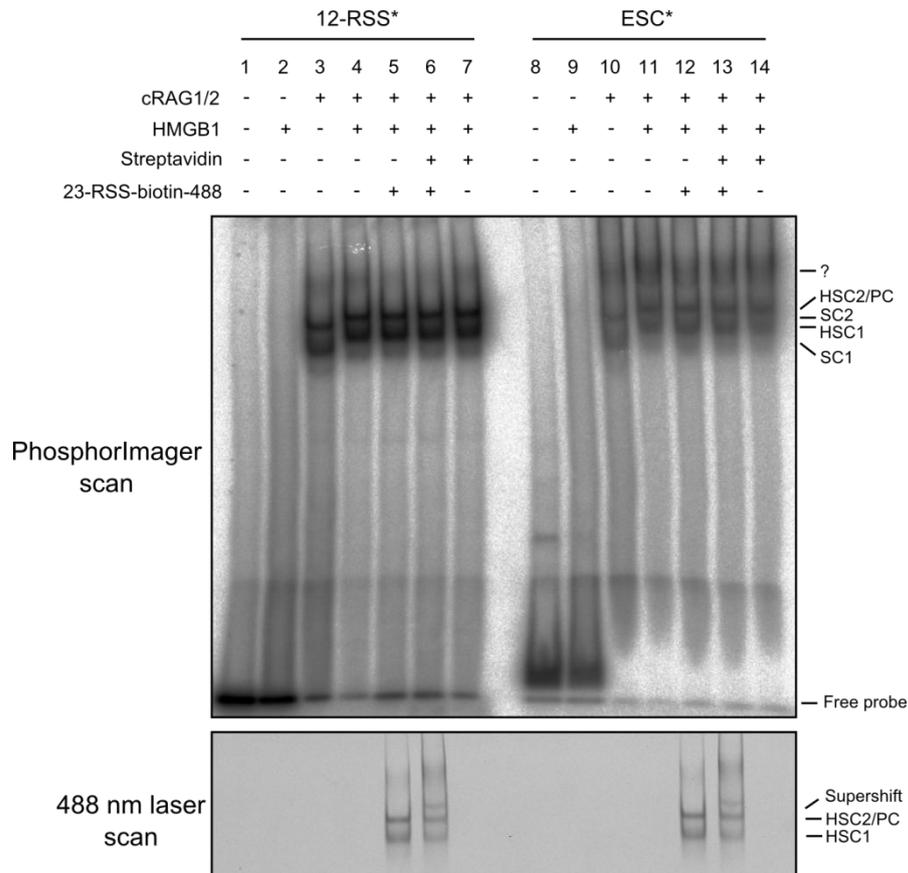
Figure 4.4 – A streptavidin supershift assay does not demonstrate paired complex formation. In an attempt to more clearly demonstrate paired complex (PC) formation, a biotinylated, Alexa Fluor 488-labelled 23RSS oligonucleotide was incubated with a radiolabelled 12RSS, cRAG proteins and HMGB1, and in lanes 6-7 and 13-14, monomeric streptavidin. Complexes were formed at 37°C for 10 min, and run on a 4% polyacrylamide gel. The gel was first scanned using a 488 nm laser (bottom panel) to visualise the 488-labelled 23RSS, then dried and exposed to a PhosphorImager screen to visualise the radiolabelled 12RSS (top panel). SC1/2 = single complex 1/2, HSC1/2 = HMGB1 single complex 1/2.

Notably, the titration of an unlabelled 23RSS substrate, ranging in concentration from equimolar to an 8-fold excess, competes with the labelled substrate for cRAG binding (Fig. 4.1B, lanes 3-6). Whilst this initially appears to be at odds with the fact that the level of RAG cleavage is greatest at a 10-fold excess of partner (Fig. 3.9), a simple explanation for this is that although the proportion of bound molecules is lower, those molecules which are bound are cleaved more efficiently because of the 23RSS partner. Alternatively, as mentioned above, the PC might be less stable than a single-substrate complex.

One or two faint complexes of a higher molecular weight than HCS2/PC can be seen in many instances (Fig. 4.1). Others have previously noted the presence of complexes above the PC in EMSAs (Schatz and Swanson, 2011), but the content or biological relevance has not been investigated. In lanes where only standard RSSs are present (e.g. Fig. 4.1A, lanes 3 and 4) it is possible that they represent higher-order oligomerisation of the RAG-RSS complexes into an octamer, as has been observed in atomic force microscopy studies (Shlyakhtenko et al., 2009). However, this complex is much more abundant in reactions containing a labelled ESC oligonucleotide (Fig. 4.1A, lanes 16-20 and Fig. 4.1B, lanes 14-18). There are two explanations for this phenomenon. The first is that RAG heterotetramers are more likely to oligomerise into higher-order complexes in the presence of an ESC. Alternatively, the higher molecular weight complex might represent RAG proteins bound to both sides of the ESC simultaneously, since such a complex would contain more RAG molecules than a standard PC (Fig. 4.2). The latter explanation is consistent with the prediction that in an ESC-RSS pairing, RAG proteins bind to the RSS and both sides of the ESC simultaneously. In this complex, an ESC could stimulate cleavage of the RSS, but the presence of RAG proteins bound to the other side of ESC might prevent efficient cleavage of the ESC, and therefore could be the basis of asymmetric cleavage. Whilst it might be expected that the same complex would be observed with a labelled RSS at a similar abundance when paired with an ESC, this is not the case. As discussed above, this might be because the ESC-RSS PC is transient or unstable upon electrophoresis.

4.2   DNase I footprint analysis of RAG binding

To further examine RAG-ESC interactions and compare these with RAG-RSS interactions, DNase I footprinting experiments were performed with either a radiolabelled 12RSS or ESC substrate (Fig. 4.5). Analysis of band intensities by densitometry suggests that there might be several protected bases in the nonamer region of both RSSs in the ESC (Fig. 4.5A, lane 5), which would be consistent with DNase I experiments by others which found no protection of the heptamer sequence, either with or without RAG2 (Nagawa et al., 1998b). This would also be consistent with RAG proteins binding to both RSSs of an ESC simultaneously. However, the strength of the footprint here is too weak to reach a definitive conclusion. Furthermore, no protection was detected at the RSS (Fig. 4.5A, lane 2). A binding assay under identical conditions shows that the majority of the labelled probe is bound by cRAG proteins, but they do not form a distinct complex (Fig. 4.5B). This, combined with the relatively high nuclease background (Fig. 4.5A, lanes 3 and 6), is likely to have contributed to the

weak footprint. Footprint analysis of full-length RAG proteins was not attempted since the concentration of purified flRAG proteins was too low to saturate the labelled probe in an EMSA (Fig. 4.3). Overall, the quality of the footprinting data precludes any accurate interpretation, but the preliminary data are consistent with the hypothesis that RAG proteins bind to both RSSs of an ESC simultaneously.
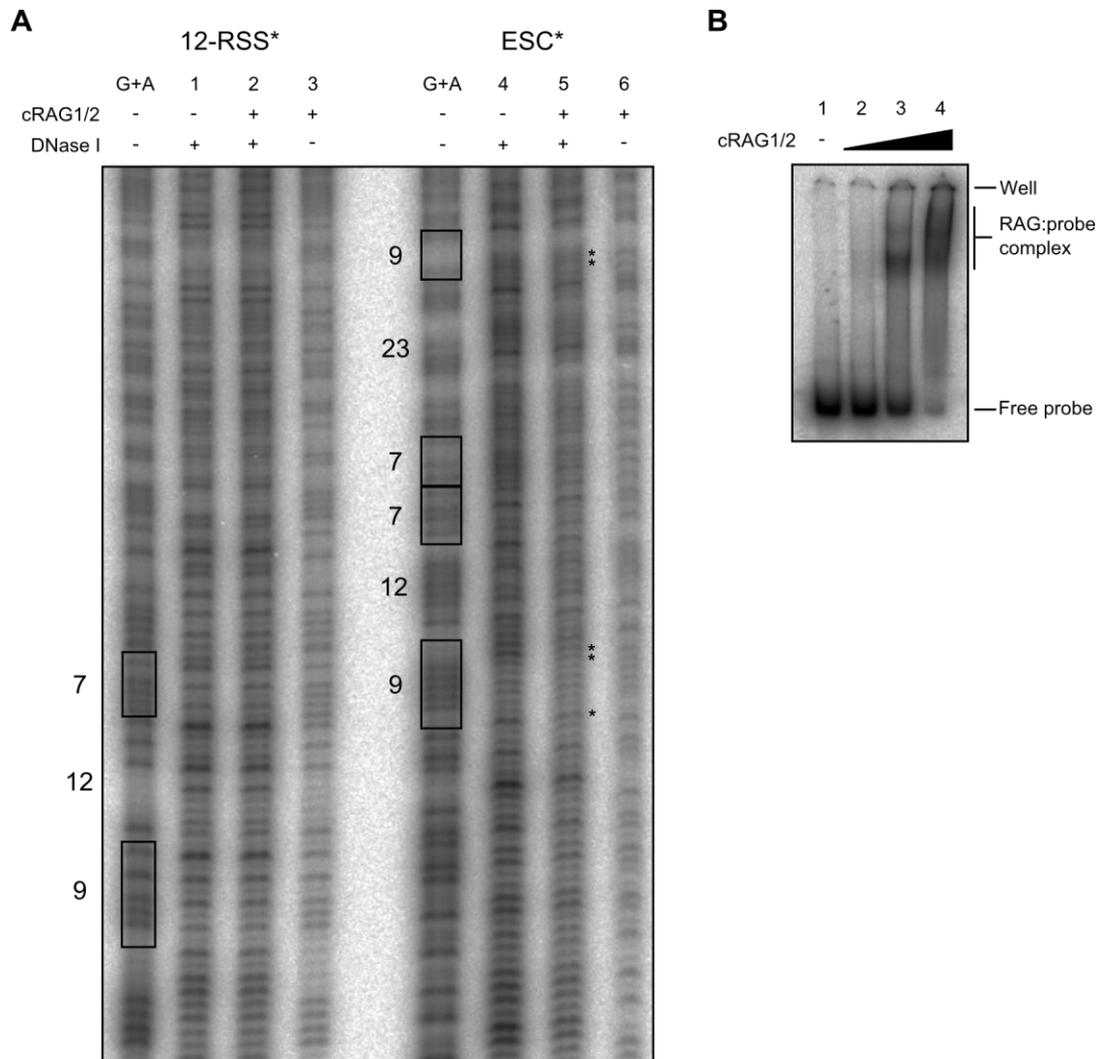
Figure 4.5 – DNase I footprinting assay of 12RSS and ESC. (A) DNase I footprinting experiments were performed to compare RAG-RSS and RAG-ESC contacts, and to further test the hypothesis that RAG proteins bind to both sides of the ESC simultaneously. Binding reactions were set up as for the EMSAs, and DNase I was added for 1 min. Reactions were purified and separated on a 6% denaturing polyacrylamide gel. Bands that showed a difference in band intensity +/- RAG protein, as judged by densitometry, are indicated by an asterisk **(B)** An EMSA using the footprinting probes, to test RAG binding under the conditions using for the DNase I footprint. RAG concentration ranges from 20 to 150 nM.

4.3   cRAG proteins bind to RSSs and ESCs with similar affinity

It is possible that the RAG proteins cleave ESCs less efficiently than RSSs because the affinity of RAG-ESC interactions is lower than RAG-RSS interactions. To test this possibility, the affinity of these interactions was estimated by EMSA. Binding assays were set up in the same way as before, except that cRAG proteins were titrated ranging from 5 to 80 nM and no partner substrate was added. The fraction of bound substrate at each RAG concentration was measured by densitometry, and the K_d was estimated by plotting the log of RAG concentration against the fraction of substrate bound.

Previously, cRAG1 and cRAG2 have been found to bind a single RSS with a K_d of ~25 nM, as estimated by EMSA (Zhao et al., 2009), and ~13.5 nM in Ca^{2+} as measured by fluorescence anisotropy (Wang et al., 2012). The RAG-ESC binding affinity has not previously been assessed. Here, in Ca^{2+} and in the absence of a partner substrate, the cRAG proteins bind with a K_d of 12.8 nM for a 12RSS, 9.5 nM for a 23RSS, and 6.1 nM for an ESC (Fig. 4.6). Therefore, any differences in cleavage efficiency and symmetry are unlikely to be the result of differences in binding affinity.
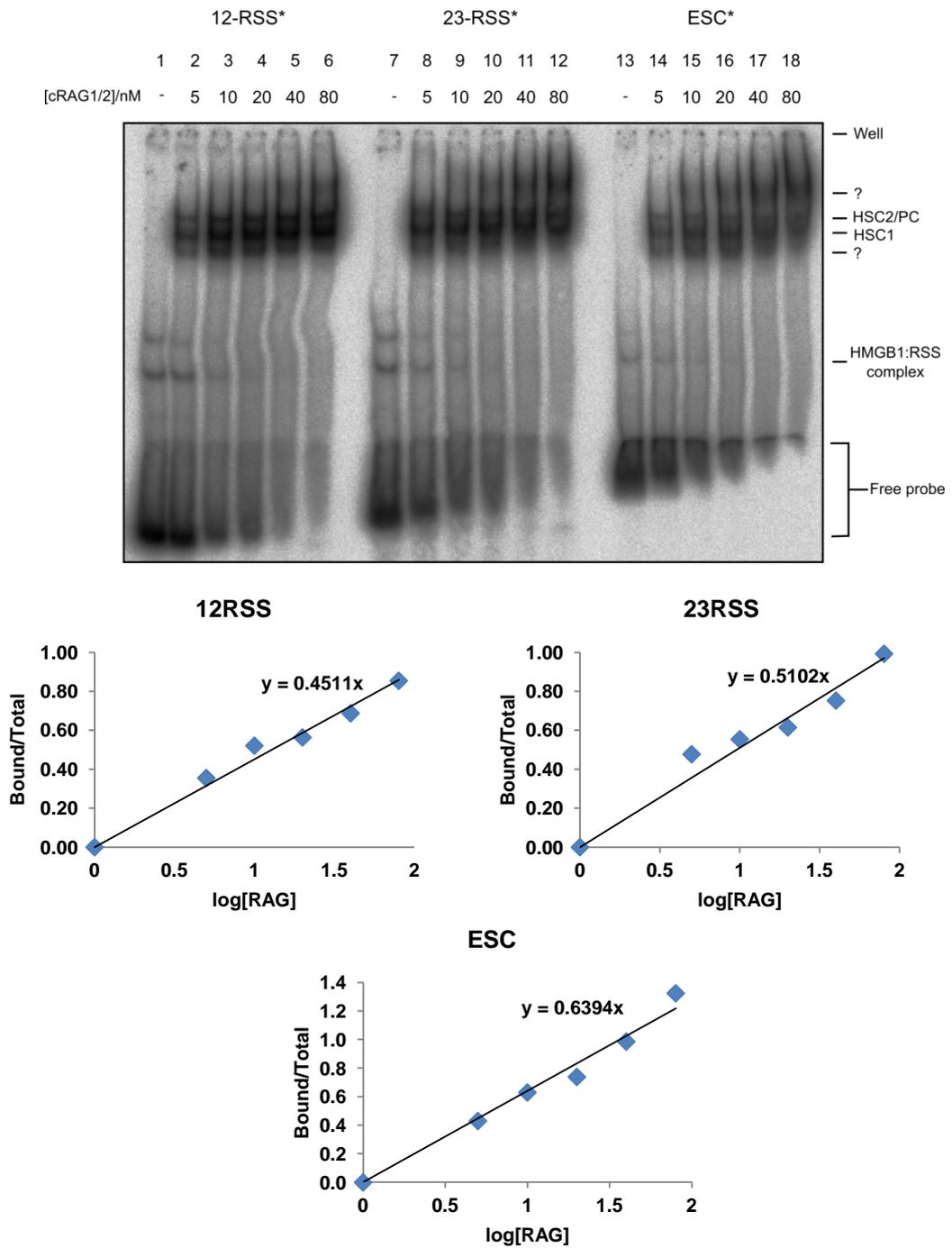
Figure 4.6 – Core RAG proteins bind to a 12RSS, 23RSS and ESC with a similar affinity. The affinity (Kd) of RAG proteins for the three different substrates was compared to see if binding affinity was a factor in asymmetric cleavage. 2 nM radiolabelled 12RSS, 23RSS or ESC oligonucleotide was incubated with core RAG proteins ranging in concentration from 5-80 nM, 100 nM HMGB1 in the presence of Ca^{2+}, on ice for 30 min and separated on a 4% polyacrylamide gel at 4°C. Gels were dried and scanned using a PhosphorImager. The complex seen below HSC1 could represent a RAG1-RSS complex, if the particular RAG prep contained more RAG1 than RAG2. To estimate the dissociation constant (Kd) with different substrates, the amount of bound and unbound substrate was measured by densitometry, and the ratio of bound substrate:total substrate was plotted against the log RAG concentration. The log[RAG] at which half the substrate is bound was calculated using the equation of the line, and then converted to the concentration in nM.

4.4 Mutating one side of the ESC restores symmetric substrate cleavage

Results from the *in vitro* binding assays comparing RSS and ESC complexes support the hypothesis that RAG proteins bind to both RSSs of the ESC simultaneously. To further test this idea, I mutated one of the RSSs in an ESC to see if this simultaneous RAG binding is prevented, thus restoring symmetric stimulation and cleavage of ESCs. Using the same *in vitro* reaction conditions established in Chapter 3, mutant ESC oligonucleotides with a deleted heptamer or nonamer were incubated with cRAG proteins, HMGB1 and an unlabelled partner substrate. Using an ESC where the heptamer of the 12RSS is deleted (ESC 12d7), the level of cleavage is low in the absence of partner (Fig. 4.7A, lane 2). Assuming that a 12/23 pairing with the intact 23RSS of the ESC should still be possible, an increasing amount of unlabelled 12RSS partner was added. A 12RSS stimulates cleavage by a maximum of 7.5-fold with a 10-fold excess of 12RSS (Fig. 4.6A, lanes 3-7). In contrast, cleavage of a consensus ESC is reduced upon addition of a 12RSS (Fig. 3.10B). Moreover, the mutated ESC 12d7 stimulates cleavage of a 12RSS 4.5-fold, presumably via synapsis with the intact 23RSS (Fig. 4.7B). Therefore, consistent with the idea that binding to both RSSs of the ESC blocks cleavage, deletion of the ESC 12RSS heptamer appears to have restored symmetric RAG-ESC-RSS cleavage.

Similarly, when the heptamer of the 23RSS is deleted (ESC 23d7), ESC cleavage is stimulated two-fold by addition of a 23RSS (Fig. 4.7C), and the ESC 23d7 oligonucleotide stimulates 23RSS cleavage 3.3-fold (Fig. 4.7D). Therefore, deletion of the ESC 23RSS heptamer enables the stimulation of ESC cleavage by a 23RSS, which was not the case with a consensus ESC (Fig. 3.10D). However, the cleavage efficiency in an ESC 23d7-23RSS pairing is still somewhat asymmetric and is less efficient than with the ESC12d7-12RSS pair.

Next, the effect of deleting the 23RSS nonamer in an ESC (ESC 23d9) was examined. As with the ESC heptamer mutants, ESC 23d9 cleavage was stimulated by a partner RSS (2.8-fold, Fig. 4.8A). Control reactions performed at the same time with a consensus ESC and 23RSS confirm that cleavage of a consensus ESC is reduced rather than stimulated (Fig. 4.8A, lanes 8 and 9). Furthermore, the ESC 23d9 mutant stimulates 23RSS cleavage 2.3-fold (Fig. 4.8B), and the overall level of cleavage efficiency at both substrates is very similar, indicating that mutation of the ESC 23RSS nonamer also restores symmetric cleavage to some extent.
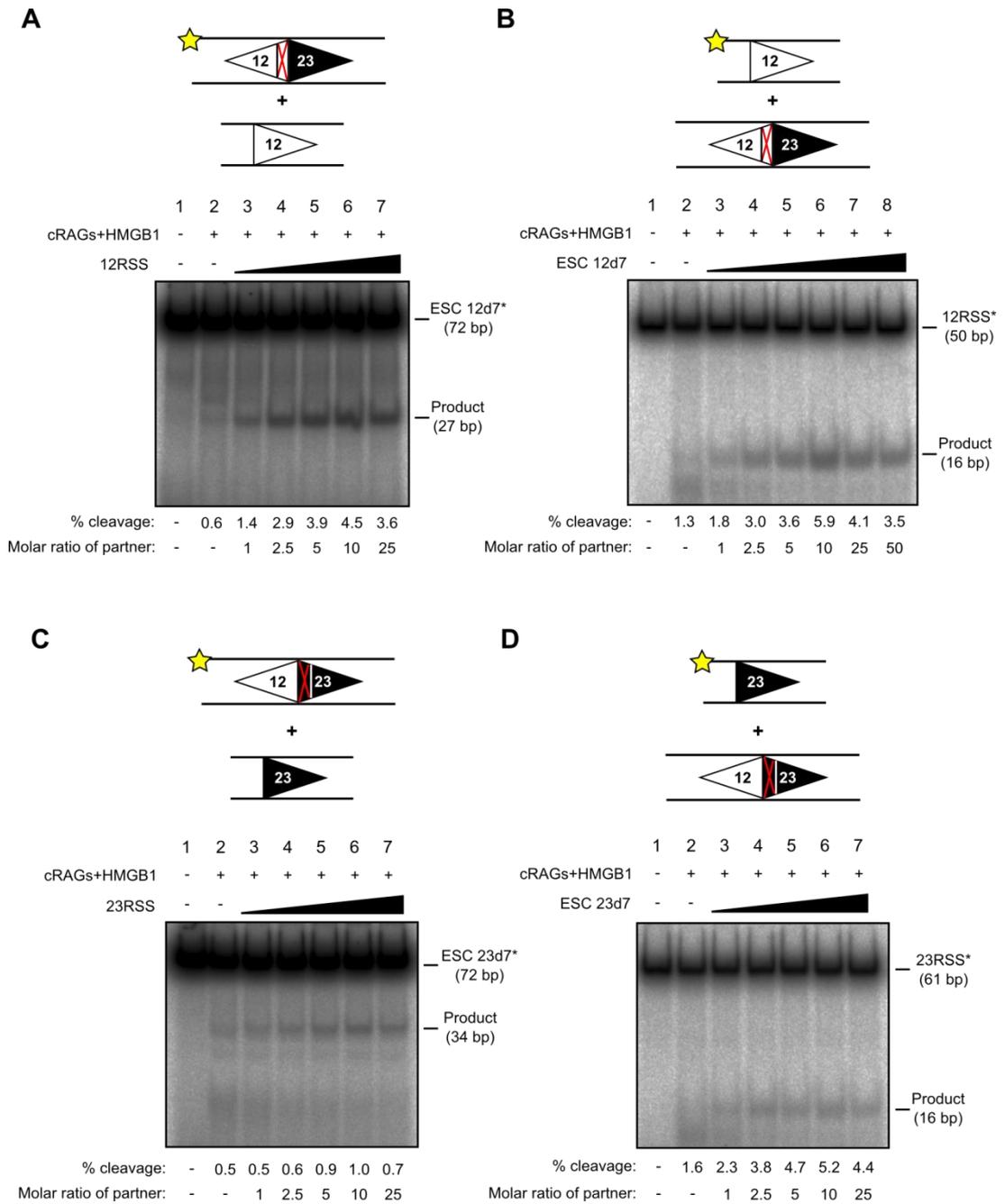
Figure 4.7 – Deleting one of the ESC heptamers restores symmetric cleavage. To determine whether symmetric cleavage could be restored by disrupting RAG binding to one side of the ESC, the heptamer of one RSS of the ESC was deleted. The ESC mutants were paired with an unlabelled substrate that would allow a 12/23 pairing between the RSS substrate and the intact RSS of the ESC and used in standard cleavage assays. **(A)** Labelled ESC 12d7 mutant and 12RSS **(B)** Labelled 12RSS and ESC 12d7 mutant **(C)** Labelled ESC 23d7 mutant and 23RSS **(D)** Labelled 23RSS and ESC 23d7 mutant.
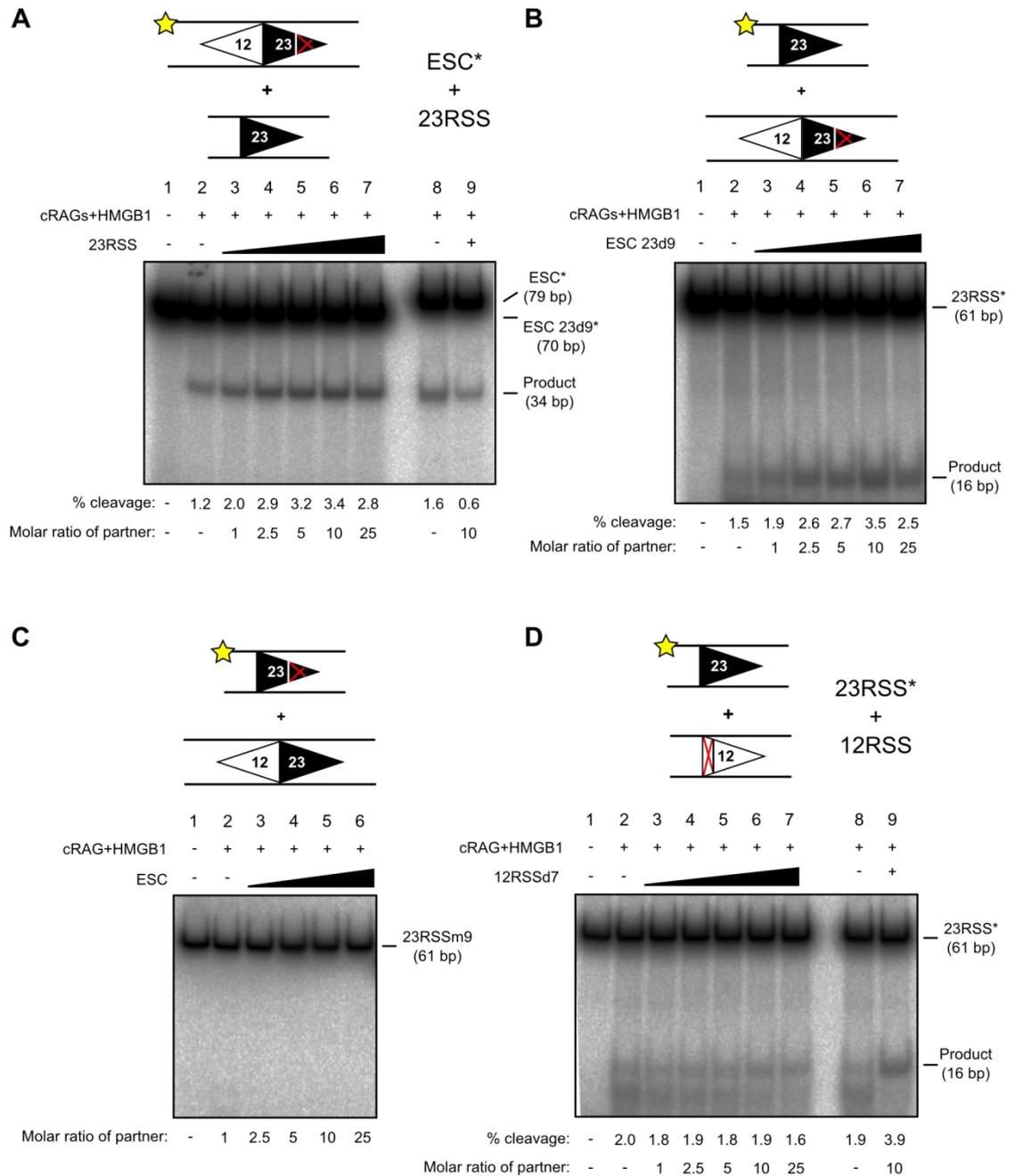
Figure 4.8 – Deleting the nonamer of the 23RSS in an ESC restores symmetric cleavage. (A) To see if symmetric cleavage could be restored by disrupting RAG binding to one side of the ESC, the nonamer of the ESC 23RSS was deleted. The ESC 23d9 mutant was paired with an unlabelled 23RSS to allow a 12/23 pairing between the intact 12RSS of the ESC. Wild-type ESC +/- 23RSS was included for comparison (lanes 8 and9) **(B)** As for (A), except the 23RSS is labelled. **(C)** A 23RSS with a mutated nonamer was incubated with a wild-type ESC to verify that mutating the nonamer of an RSS renders it unusable by RAG proteins. **(D)** A 12RSS with a mutated heptamer was incubated with a labelled 23RSS to verify that a lone nonamer cannot stimulate cutting at an RSS. Control reactions containing 23RSS +/- consensus 12RSS were carried out alongside (lanes 8 and 9).

In these experiments it was assumed that mutating either the conserved heptamer or nonamer elements of an ESC would render that RSS unusable as a RAG substrate. This assumption was tested using a 23RSS oligonucleotide where the nonamer sequence was mutated (not deleted, in this case), essentially leaving a heptamer sequence. No cleavage of this substrate is detected, and addition of a consensus ESC does not stimulate cleavage (Fig. 4.8C), indicating that the RAG proteins cannot cleave an RSS heptamer without the nonamer sequence. Similarly, a nonamer sequence alone is unable to stimulate cleavage at a 23RSS (Fig. 4.8D, lanes 1-7), whereas a consensus 12RSS stimulates cleavage (Fig. 4.8D, lanes 8 and 9). Taken together, these results indicate that mutating one RSS of an ESC renders that "side" of the ESC unusable and therefore supports the idea that RAG binding to both sides of the ESC somehow prevents symmetric cleavage and stimulation.

4.5   Analysis of RAG-ESC mutant complexes

To further test the hypothesis that the RAG proteins bind to both sides of the ESC, RAG binding to mutant ESC oligonucleotides was analysed by EMSA. If RAG binding to both sides of the ESCs simultaneously is represented by the higher molecular weight complex seen in Fig. 4.1, then mutating one of the RSSs should reduce the abundance of this complex or entirely prevent its formation.

Indeed, deleting the nonamer of the 23RSS in the ESC reduces the abundance of the higher molecular weight complex seen in Fig. 4.1 relative to HSC1 and HSC2/PC, consistent with the idea that the uppermost complex represents RAG proteins bound to both RSSs simultaneously (Fig. 4.9). Unlike Fig. 4.1, titration of the unlabelled partner RSS appears to result in eviction of HMGB1 from the HSC1 complex. This is evidenced by the increased abundance of an HMGB1:ESC 23d9 complex at higher concentrations of unlabelled partner (Fig. 4.9, lanes 8 and 9), and also by the increased migration of the smallest complex, which is roughly equivalent to the size difference between HSC1 and SC1 (Fig. 4.1A, compare lanes 3 and 4, and Fig. 4.4, compare lanes 3 and 4). It is unclear why addition of an unlabelled 23RSS partner would have this effect, but it is unique to the ESC 23d9 substrate. Since HMGB1 binds within the 23RSS spacer, deletion of the 23RSS nonamer in the ESC might reduce the affinity of the HMGB1:ESC interaction. Addition of a consensus 23RSS is then enough to evict HMGB1 from the complex.

Unlike deletion of the nonamer, mutating the heptamer of either RSS in the ESC does not reduce the abundance of the higher molecular weight complex (Fig. 4.10). This is

perhaps not entirely unexpected, since the most crucial RAG-RSS interactions occur between the nonamer and the RAG1 nonamer binding domain (Yin et al., 2009), and RAG2 is thought to promote RAG1 binding of the heptamer (Swanson and Desiderio, 1999; Fugmann and Schatz, 2001; Qiu et al., 2001). Therefore, deleting the nonamer is more likely to disrupt RAG binding than deleting the heptamer. However, if RAG proteins still bind both sides of the ESC efficiently, why is asymmetric cleavage restored upon deletion of the heptamer (Fig. 4.7)? One possible explanation is that RAG1 binds the mutated ESC via the intact nonamer, but the lack of a heptamer means that RAG contacts do not extend toward the other RSS, which would otherwise prevent efficient cleavage and lead to asymmetry. Overall, these data support the model that RAG binding simultaneously to both sides of the ESC leads to asymmetric cleavage of ESCs.
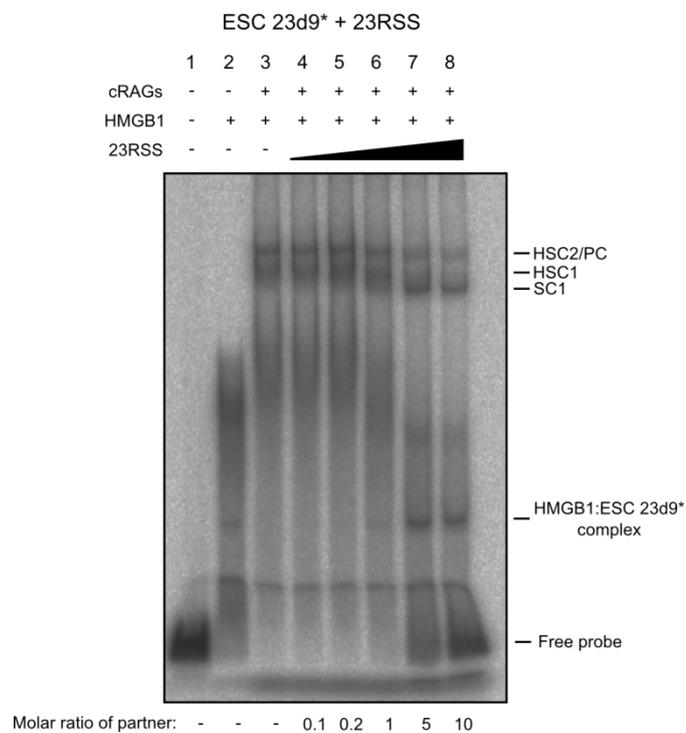
Figure 4.9 – Mutation of the 23RSS nonamer in an ESC reduces RAG binding at one side of the ESC. To see if deleting the nonamer of the 23RSS in an ESC prevents RAG to both sides of the ESC, an ESC 23d9 mutant oligonucleotide was incubated with cRAG proteins, HMGB1, and an unlabelled 23RSS and run on a 4% native polyacrylamide gel at 4°C. The higher molecular weight complex seen in Fig. 4.1 (lanes 16-20) is no longer visible, indicating that deletion of the 23RSS nonamer has prevented RAG binding to the 23RSS side of the ESC. HSC1/2 = HMGB1 single complex 1/2, PC = paired complex, SC1 = single complex 1

A

B
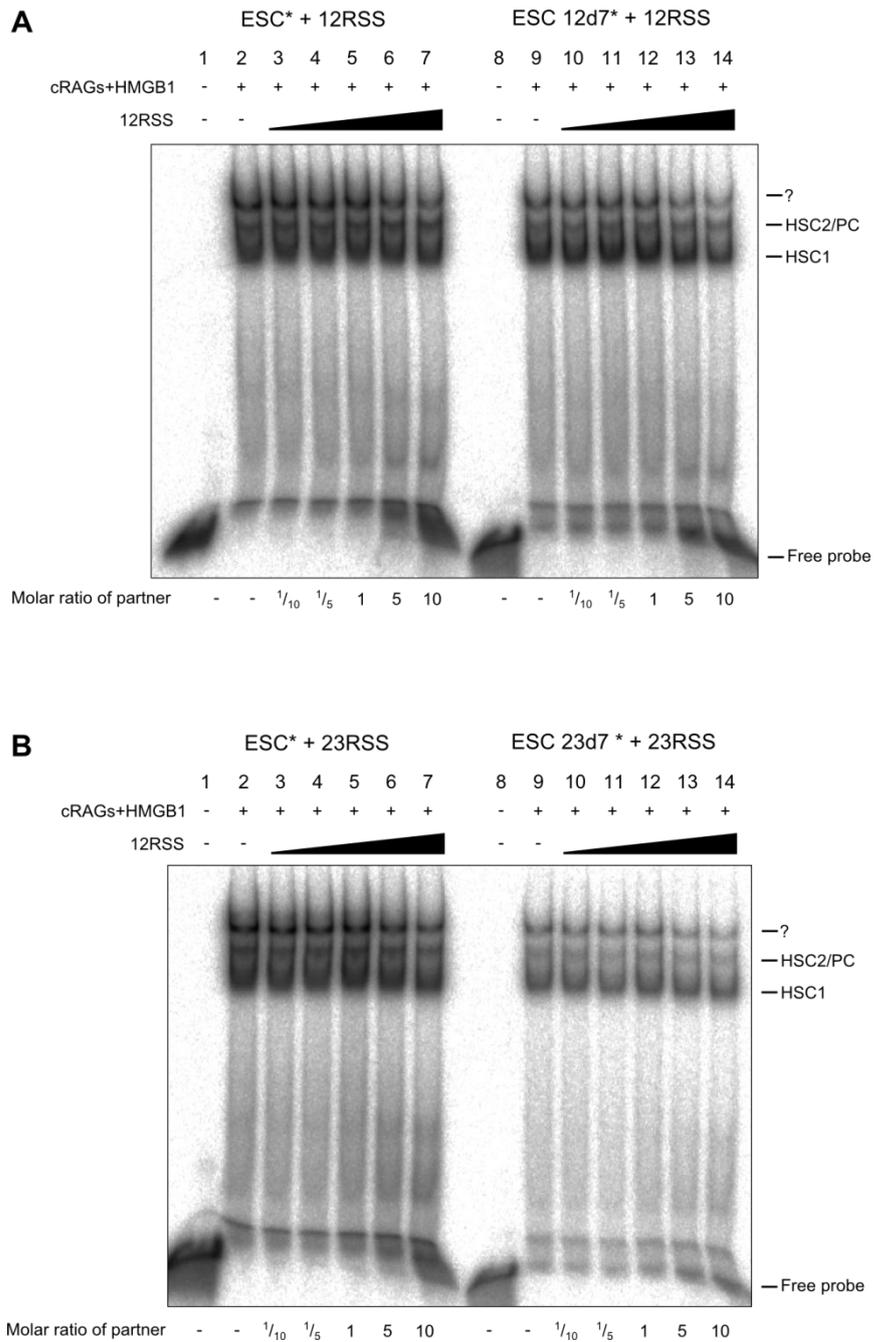
Figure 4.10 – Mutation of either heptamer in an ESC does not reduce RAG binding to both sides of the ESC. (A) To see if mutating one of the heptamers in an ESC reduces the abundance of the higher molecular weight complex, an ESC oligonucleotide where the heptamer of the 12RSS is deleted was incubated with a consensus 12RSS oligonucleotide, cRAG proteins and HMGB1, and run on a 4% polyacrylamide gel and compared with consensus ESC complex formation. **(B)** As for A, except either labelled consensus ESC or labelled ESC 23d7 mutant were used, paired with a 23RSS. HSC1/2 = HMGB1 single complex 1, PC = paired complex.

4.6   Analysis of postcleavage complex stability

The dangers posed by the cut-and-run reaction critically rely on the fate of the RAG-DNA complexes following cleavage. If a broken RSS is to trigger genome instability, it must first be released by the RAG proteins. To investigate the fate of the broken DNA ends, the stability of the RAG-RSS postcleavage complexes was analysed by time course assays where half of the reaction was separated on a 4% polyacrylamide gel to analyse RAG-substrate complexes, and the other half was separated on a 12% native polyacrylamide gel to analyse cleavage products. It should be noted that the labelled end of the 12 and 23RSS oligonucleotides is the coding end, and the signal end is unlabelled, so the data presented here show the stability of RAG-coding end interactions under conditions that facilitate cleavage (i.e. in the presence of Mg^{2+}). In the case of the ESC, the 12RSS end is labelled. Therefore, in a pairing with a 23RSS, the 23RSS half of the ESC is the "coding" end and the 12RSS is the signal end, and vice versa.

In the absence of a partner RSS, HSC1 and HSC2/PC (Fig. 4.2) are seen with a labelled 12RSS (Fig. 4.11A, lanes 1-6). Consistent with the data presented in Chapter 3, cleavage of a 12RSS is stimulated by addition of a 23RSS or ESC partner (Fig. 4.11A, bottom panel). The increased level of cleavage correlates with a slightly shorter half life for both HSC1 and HSC2 compared with 12RSS alone (Fig. 4.11B). The shorter half life of both HSC1 and HSC2 indicates that addition of a partner promotes the conversion of both complexes to a paired complex, which then releases the coding end of the 12RSS following cleavage. The fate of the signal end is unknown, since this end is unlabelled. To confirm that cleavage is essential for the observed decrease in complex abundance, the same experiment was carried out, but replacing Mg^{2+} with Ca^{2+} (Fig. 4.12). As in Fig. 4.11A, the abundances of HSC1 and HSC2 decrease with increasing cleavage (Fig. 4.12). When Ca^{2+} is used in place of Mg^{2+}, neither HSC1 nor HSC2 dissociates after 180 min (Fig. 4.12), thus confirming that cleavage of the RSS is required for the decrease in complex abundance. Furthermore, this is consistent with the idea that the RAG proteins release the 12RSS coding end following cleavage.

Similarly, addition of a 12RSS to a 23RSS stimulates cleavage by 3.6-fold (Fig. 4.13A). This increase in cleavage correlates with a shorter half life for both complexes (Fig. 4.13B), indicating that the 23RSS coding end is released by RAG proteins after cleavage. Addition of an ESC partner also stimulates 23RSS cleavage, but only by ~1.6-fold, and as such there is a smaller decrease in the abundance of the HSC1 and HSC2 complexes.

With a labelled ESC, the higher molecular weight complex previously described (Fig. 4.1) is no longer observed (Fig. 4.14A, top panel). A possible reason for this discrepancy is the use of divalent metal cation. Here, Mg^{2+} is used to facilitate cleavage, so that complex stability could be analysed. In Fig. 4.1, Ca^{2+} was used to preclude cleavage, therefore only precleavage synaptic complexes are observed. It is possible that in conditions that facilitate cleavage, the higher molecular weight complex is less stable and/or it quickly dissociates into one of the smaller molecular weight complexes. Addition of a 12RSS to an ESC results in a slight stimulation in cleavage over ESC alone, whereas a 23RSS does not (Fig. 4.14A, bottom panel). In both cases, the half life of HSC2 is longer than the half life of HSC1 (Fig. 4.14B). One possible interpretation of this observation is that addition of a partner promotes the conversion of HSC1 to HSC2 and PC, hence the shorter half life of HSC1. However, ESC cleavage is not greatly stimulated by the partner, leading to an accumulation of paired complexes, hence the longer half life of HSC2/PC with an ESC compared to a 12 or 23RSS. Taken together, these data indicate that an ESC-stimulated cleavage of RSS results in the release of the coding end, but the RAG proteins remain more stably bound to the ESC.
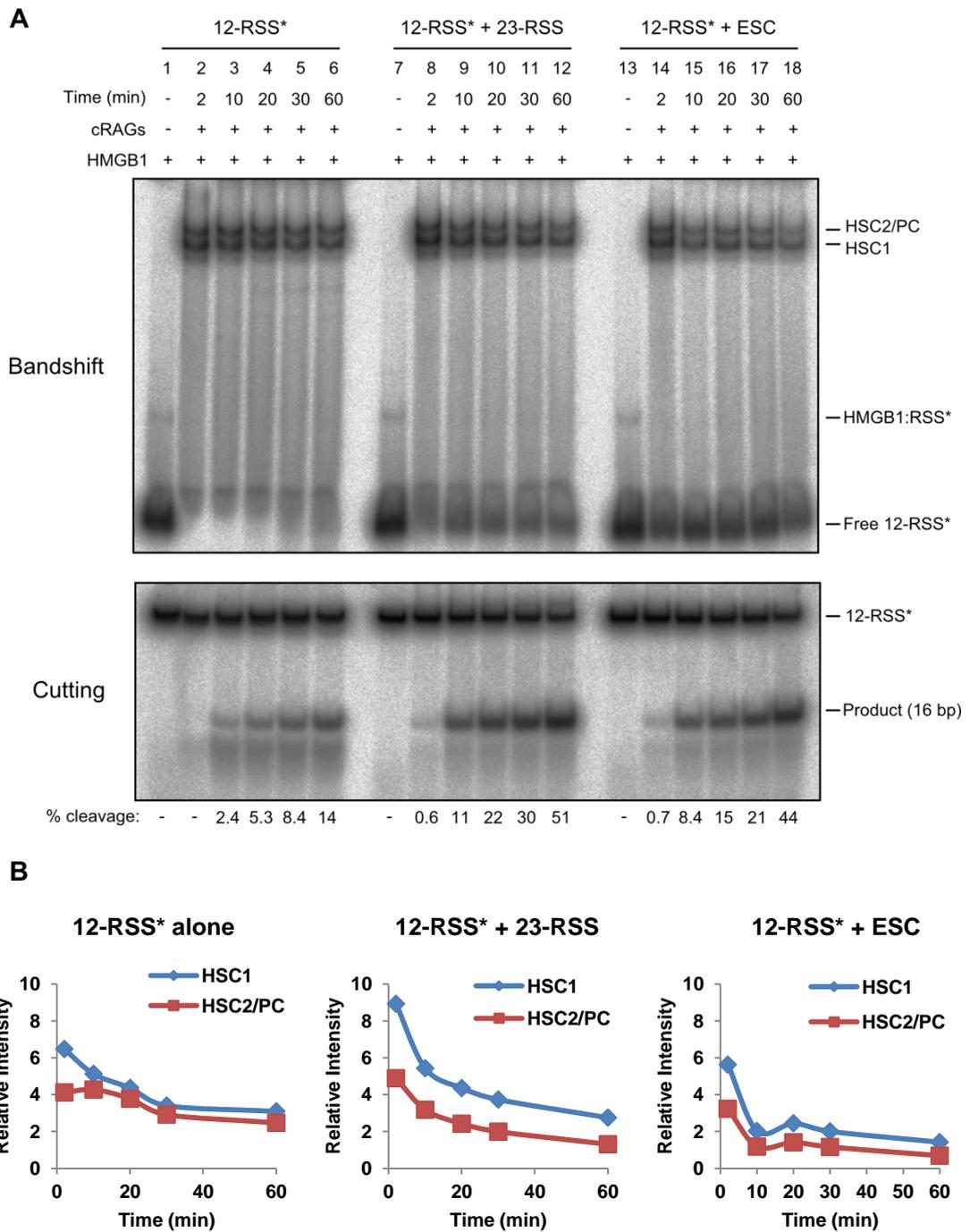
Figure 4.11 – RAG proteins release the 12RSS coding end following cleavage (A) Radiolabelled 12RSS oligonucleotide was incubated with cRAG proteins and HMGB1 in the presence of Mg^{2+} to facilitate RAG cleavage, either alone, with a 23RSS, or with an ESC. Reactions were incubated for the indicated time, quenched on ice then divided in two. One half was run on a 4% polyacrylamide gel to analyse complex stability, and the other half was run on a 12% polyacrylamide gel to analyse cleavage. **(B)** The abundance of each complex was measured by densitometry and plotted against time to estimate complex half life.
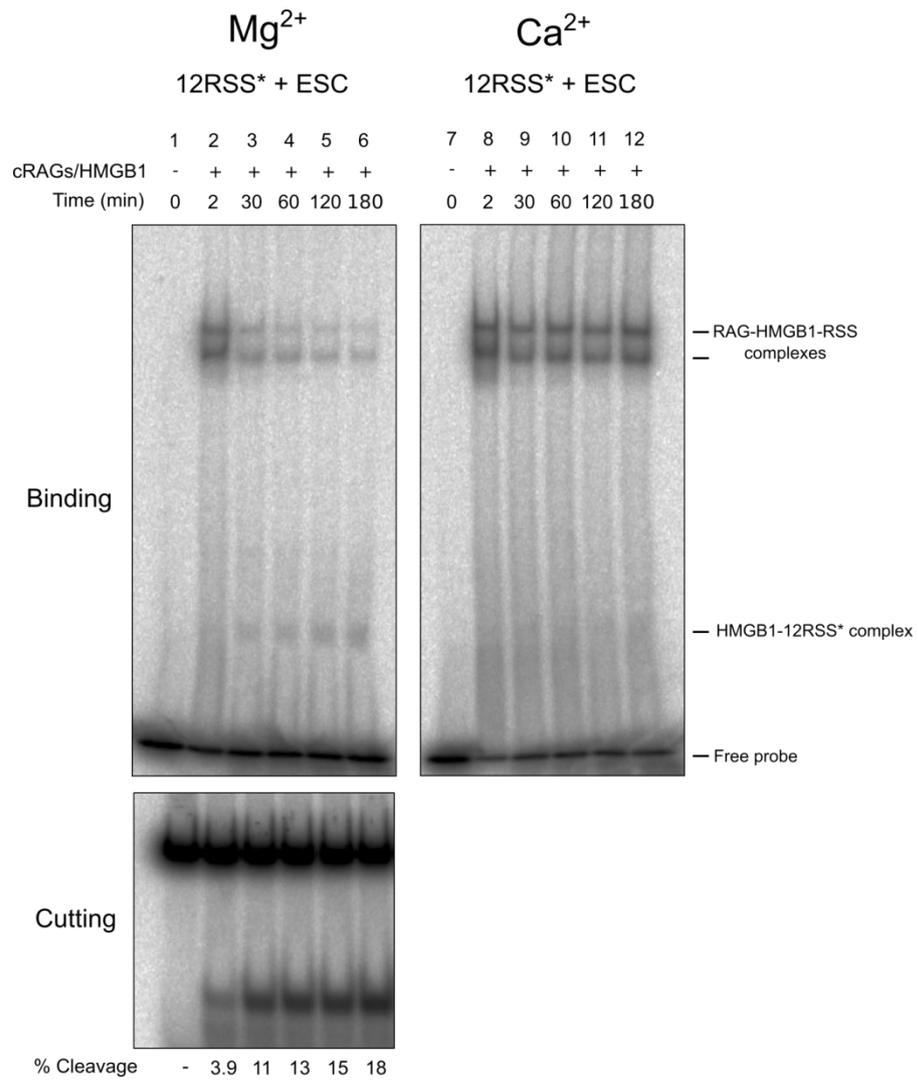
Figure 4.12 – RAG proteins release the 12RSS coding end following cleavage in Mg^{2+}, but remain bound in Ca^{2+}. A bandshift timecourse of 12RSS-ESC complexes was repeated to confirm that RAG proteins release the 12RSS coding end following cleavage. Identical reactions were carried out alongside, except that Ca^{2+} was used in place of Mg^{2+}. In Mg^{2+}, the RAG proteins release the 12RSS coding end concomitant with an increase in cleavage, but in Ca^{2+}, which precludes cleavage, the RAG proteins stay bound for over 180 min.
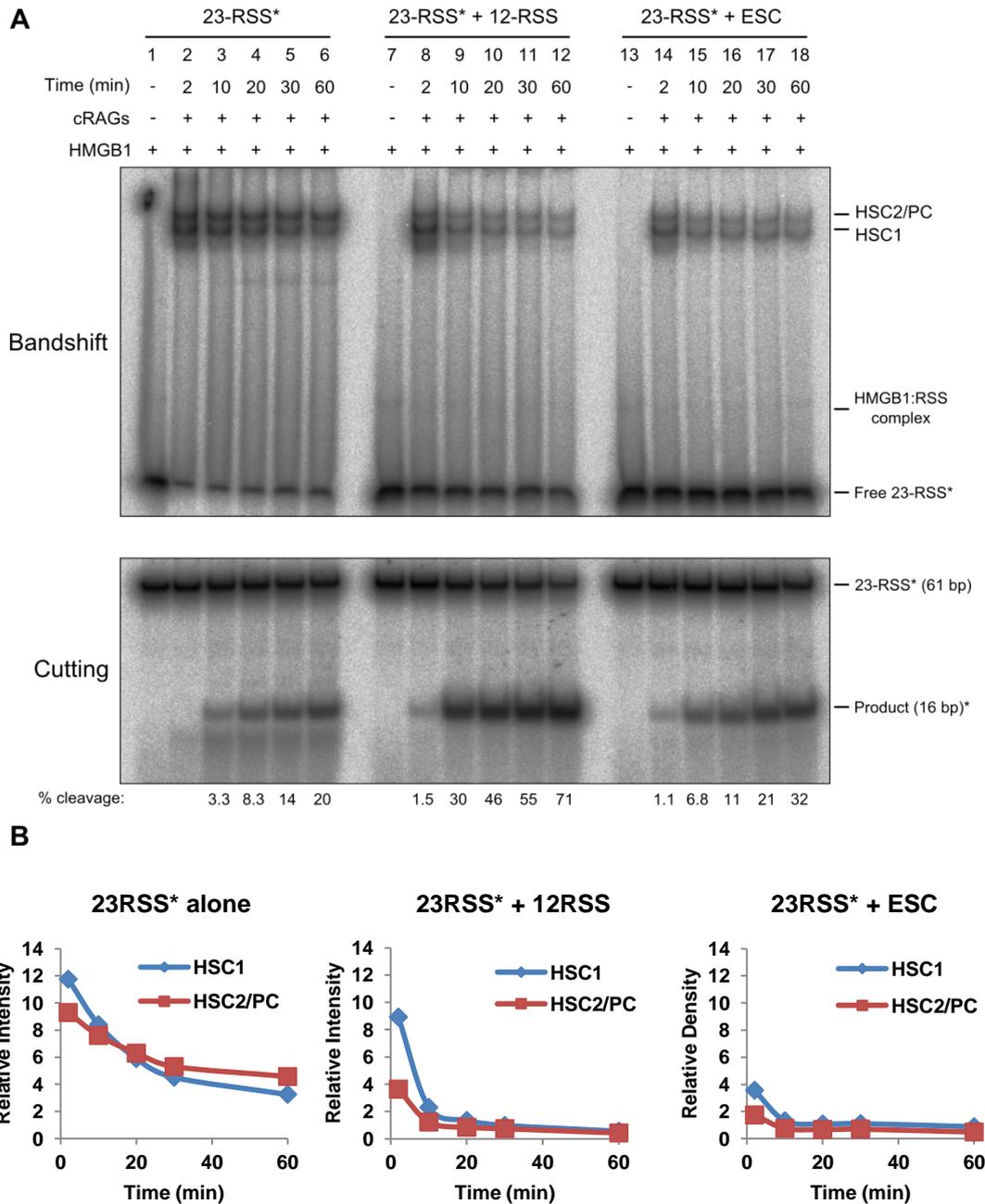
Figure 4.13 – RAG proteins release the 23RSS coding end following cleavage (A) A radiolabelled 23RSS oligonucleotide was incubated with cRAG proteins and HMGB1 in the presence of Mg^{2+} to facilitate RAG cleavage, either alone, with a 23RSS, or with an ESC. Reactions were incubated for the indicated time, quenched on ice then divided in two. One half was run on a 4% polyacrylamide gel to analyse complex stability, and the other half was run on a 12% polyacrylamide gel to analyse cleavage. **(B)** The abundance of each complex was measured by densitometry and plotted against time to estimate complex half life.
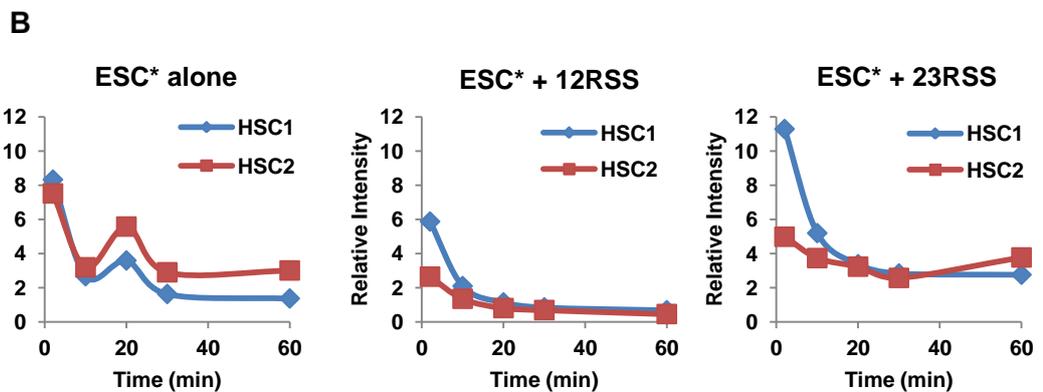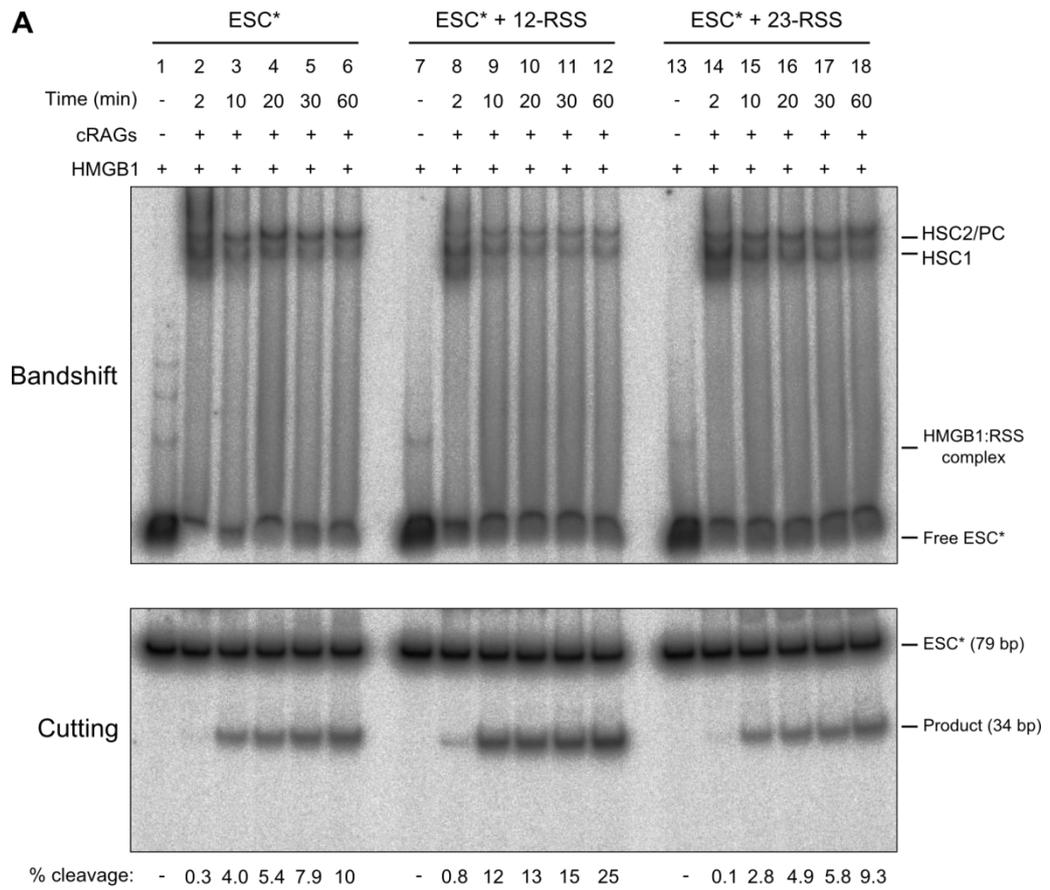
Figure 4.14 – RAG proteins stay bound to ESCs after cleavage (A) Radiolabelled ESC oligonucleotide was incubated with cRAG proteins and HMGB1 in the presence of Mg^{2+}, to facilitate RAG cleavage, either alone, with 23RSS, or with ESC. Reactions were incubated for the indicated time, quenched on ice then divided in two. One half was run on a 4% polyacrylamide gel to analyse complex stability, and the other half was run on a 12% polyacrylamide gel. **(B)** The abundance of each complex was measured by densitometry and plotted against time to estimate complex half life.

C) Discussion

In this chapter, I investigated the molecular basis of asymmetric RSS/ESC cleavage. A plausible hypothesis is that RAG proteins bind to both sides of the ESC simultaneously, since it contains two RSSs, and that this simultaneous binding prevents the RAG proteins from efficiently cleaving the ESC. EMSA data show that core RAG proteins form an additional higher molecular weight complex in the presence of an ESC, consistent with this hypothesis. Additionally, estimates of RAG-RSS and RAG-ESC binding affinity by EMSA showed that all three substrates (12/23/ESC) are bound with roughly the same affinity, ruling out the possibility that asymmetric cleavage is caused by differences in binding affinity. Mutating one of the RSSs in an ESC restored symmetric cleavage with RSS (albeit with varying levels of efficacy), lending further support to the model of simultaneous RAG binding. Finally, the stability of RAG postcleavage complexes was examined by a time course EMSA and cleavage assay. The results of these assays indicate that the RAG proteins do not remain bound to the cleaved coding end of an RSS after stimulation by either an RSS or an ESC, since the half life of the RAG-RSS complexes is shorter when there is a higher level of cleavage. Furthermore, the RAG proteins appear to remain bound to the uncleaved ESC.

Taken together, the EMSA, DNase I footprinting and ESC mutant cleavage data support the simultaneous binding model where the RAG proteins bind to both sides of the ESC at the same time, thus preventing symmetric cleavage. However, an inconsistency with this model is that the higher-order complex is not seen at an RSS substrate in the presence of an ESC. If a RAG1-RAG2 heterotetramer binds to both an RSS and an ESC in a paired complex, and also recruits additional RAG molecules via the second RSS of the ESC, the larger complex should be seen at the RSS also. As mentioned earlier, it is possible that formation of the PC is very transient, and as such they are not observed in an EMSA, or that the PC is unstable during electrophoresis. Investigating the catalytic activity of each complex using an in-gel cleavage assay (Swanson, 2002a) would be invaluable in shedding light on its exact nature.

The stability of RAG postcleavage ESC-RSS and RSS-RSS complexes was analysed using an EMSA timecourse. These assays were performed in the absence of any end-joining machinery and therefore they are not an accurate representation of the fate of cleaved substrates *in vivo*. Nevertheless, they provide insight into the intrinsic stability of the different postcleavage complexes when one substrate is an ESC. Performing these *in vitro* reactions in repair-proficient conditions would require either a) the use of cell extract, which would inevitably introduce many other unwanted factors, or b)

purified proteins from the NHEJ pathway. The entire complement of proteins from the NHEJ pathway has indeed been successfully purified and used to recapitulate end-joing *in vitro* (Lu et al., 2008), but since this involves the purification of 13 different proteins, it is beyond the scope of this project.

Once the RSSs are cleaved by the RAG proteins, the coding ends dissociate *in vitro*, but the signal ends remain tightly bound in a signal end complex (SEC; (Agrawal and Schatz, 1997; Wang et al., 2012). The SEC is resistant to nucleases and end joining proteins (Jones and Gellert, 2001), indicating that the signal ends cannot be joined to form an ESC until the cell cycle-dependent degradation of RAG2. Here, the dissociation of the coding end was investigated. The finding that the coding end dissociates following cleavage is consistent with previous studies. However, the fate of the signal ends was not investigated, but this could be done simply by labelling the opposite end of the oligonucleotide. In normal V(D)J recombination, the two coding ends would be in close proximity and are swiftly repaired by the NHEJ pathway (Lieber, 2010). However, in cut-and-run, since the ESC is not cleaved there is only one hairpinned coding end and one signal end. In the absence of a second coding end to be joined with, the coding end could become available for repair by more error-prone pathways such as alternative NHEJ, which would give greater scope for aberrantly joining the broken coding end to another chromosome.

One study indicated that the non-core regions of RAG1 and RAG2 promote the retention of hairpins following cleavage *in vitro* (Kumar and Swanson, 2009). It is possible that the coding end could be retained for longer following asymmetric cleavage by full-length RAG proteins *in vitro*, therefore it would be useful to investigate postcleavage complex stability with full-length RAG proteins. In any case, the most dangerous outcome following asymmetric cleavage is that the broken RSS is released by the RAG proteins (i.e. the RAG proteins cut-and-run), since potentially leukaemogenic chromosome translocations could result if it is repaired incorrectly.

Chapter 5

Testing the cut-and-run hypothesis *in vivo*

A)   Introduction

In the first two chapters, I presented evidence that cleavage of RSSs and ESCs by purified cRAG proteins *in vitro* is asymmetric, and that the molecular basis of asymmetric cutting is most likely that RAG proteins bind to both RSSs of an ESC at the same time, thus preventing efficient cleavage of the ESC. The aim of this chapter is to investigate whether or not asymmetric cleavage occurs *in vivo*, and therefore whether cut-and-run is likely to occur in developing B and T cells.

The transfection of nonlymphoid cells with extrachromosomal recombination substrates to investigate recombination activity is well established (Lieber et al., 1988; Sadofsky et al., 1993; Han et al., 1999; Vanura et al., 2007). In fact, their use predates *in vitro* experiments with short oligonucleotides (Hesse et al., 1987; McBlane et al., 1995). A number of these studies have demonstrated that intermolecular V(D)J recombination (i.e. recombination between RSSs located on different DNA molecules) is impaired relative to recombination between RSSs located on the same DNA molecule (Hesse et al., 1987; Tevelev and Schatz, 2000). Further studies showed joining of signal ends *in trans* occurs efficiently, but joining of coding ends is very rare, indicating that intermolecular recombination is inhibited at the level of coding joint formation (Han et al., 1999). However, binding and cleavage of RSSs *in trans* is as efficient as RSSs located *in cis*, and stimulation of cleavage adheres to the 12/23 rule (Steen et al., 1996; Han et al., 1999). Synapsis of an ESC and a genomic RSS, either cryptic or legitimate, is an inherently intermolecular interaction. In the proposed cut-and-run reaction, synapsis between an ESC and an RSS results in cleavage of the RSS but not the ESC, and does not involve a joining step. Therefore, the plasmid-based assay would be suitable for examining whether cleavage an RSS-ESC pair is asymmetric *in vivo*.

To test if RSS-ESC cleavage is indeed asymmetric *in vivo*, nonlymphoid cells were transfected with RAG expression vectors and different combinations of RAG substrates and quantitative PCR (qPCR) was used to assess the level of cleavage. If cut-and-run occurs *in vivo*, then the cleavage of these transfected substrates should be asymmetric. Indeed, I find that cleavage of co-transfected 12 and 23RSSs is symmetric, but importantly, when an RSS is co-transfected with an ESC, the RSS is cleaved but there is essentially no cleavage of the ESC.

Whilst ESCs cannot replicate, they appear to be remarkably stable structures *in vivo* that persist until they are diluted out by subsequent cell divisions. Indeed, the stability of ESCs has led to their development as a marker for estimating recent thymic activity (i.e. T cell production; Hazenberg et al., 2001) and also as a forensic tool to estimate the age of an individual (Zubakov et al., 2010). Furthermore, several studies have shown that the RAG proteins remain tightly bound to cleaved signal ends (Agrawal and Schatz, 1997; Jones and Gellert, 2001), and that signal ends are not joined to form ESCs until RAG expression is downregulated (Livak and Schatz, 1996; Jiang et al., 2004). This could either be because the joining machinery cannot access the tightly-bound signal ends, or because the RAG proteins re-cleave the ESCs, potentially via the nick-nick mechanism (Neiditch et al., 2002). Since a prerequisite of cut-and-run is an intact ESC, I also tested the *in vivo* state of a *bona fide* ESC from the Igλ locus. Using qPCR assays on DNA extracted from mouse pro-B cells, I find that this ESC is largely intact in these cells, which is therefore consistent with the cut-and-run hypothesis.

B)   Results

5.1   Cleavage of extrachromosomal substrates

To test if cleavage of ESC and RSSs is asymmetric *in vivo*, NIH/3T3 mouse fibroblast cells were transfected with four plasmids: an untagged full-length RAG1 expression vector (pJH548; Sadofsky et al., 1992), either a full-length or core RAG2 expression vector (pEFXC-RAG2 and pEFXC-cRAG2, respectively, both untagged), and two substrate plasmids, carrying either a 12RSS (pJβ2.7+), 23RSS (p23+) or an ESC (pJδ1+; Fig. 5.1A). The 12RSS and ESC plasmids were already available in the Boyes laboratory, but a 23RSS-containing plasmid had to be constructed *de novo*. This was done by cloning a consensus 23RSS sequence (identical to the sequence of the oligonucleotide substrate used in Chapters 3 and 4) into pJβ2.7+, replacing the 12RSS (see Section 2.43). Although the signal sequences are located on plasmids and not in the context of a chromosome, transiently transfected plasmids are chromatinised to some degree (Cereghini and Yaniv, 1984; Reeves et al., 1985). Therefore, all substrate plasmids contain the strong human cytomegalovirus (CMV) promoter sequence upstream of the RSS to enable transcription through the signal sequence , since transcription is known to enhance recombination by improving accessibility via several

A

B
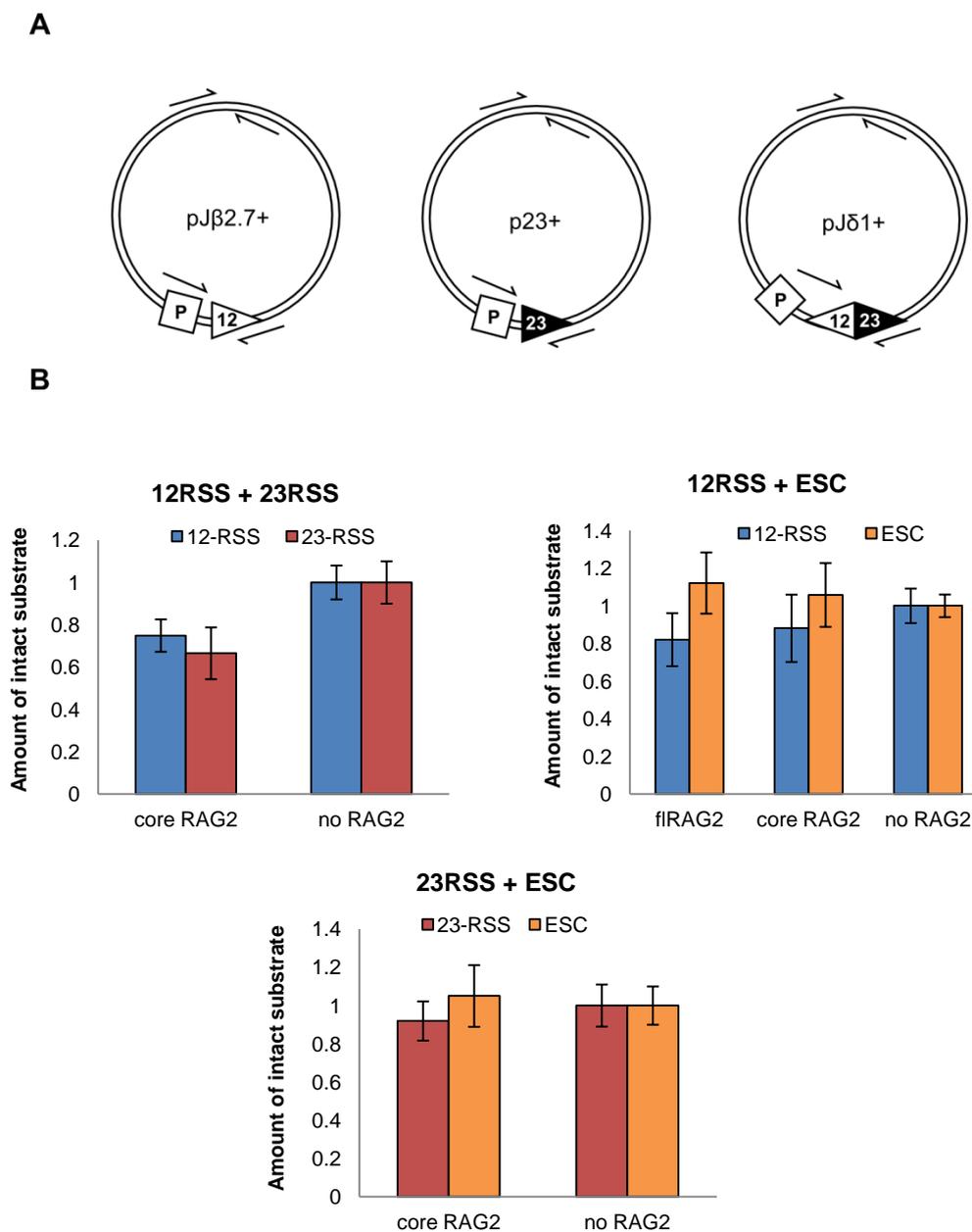
Figure 5.1 – Cleavage of transfected RSS/ESC substrates is asymmetric in vivo. To assess whether ESC/RSS cleavage is asymmetric *in vivo*, NIH/3T3 cells were transfected with a RAG1 expression vector, a RAG2 expression vector (either core or full-length, as indicated), and two substrate plasmids. Plasmids were extracted after 48 h and the level of cleavage was measured by qPCR to detect the amount of intact substrate. **(A)** Schematic of substrate plasmids used, indicating the primer pairs used to detect cleavage (see Section 2.44 for primer details). A primer pair that amplifies across the RSS was used to measure the level of intact substrate, and a primer that amplifies a unique region in the backbone of each plasmid was used to detect the overall level of substrate for normalisation. **(B)** Cleavage of cotransfected 12 and 23RSS substrates is symmetric, whereas cotransfection of 12 or 23RSS substrates with an ESC substrate leads to cleavage of the RSS, but not the ESC, indicating that cleavage of extrachromosomal ESC/RSS substrates is asymmetric *in vivo*. Data were normalised to unique regions within each plasmid, and the values given are relative to no RAG2 controls. Error bars show the SEM of separate transfections. 12+23, n=2; 12+ESC, n=3; 23+ESC, n=2. The data are an amalgamation of work done by Adam Kupinski and myself.

mechanisms, including the eviction of H2A/H2B dimers from nucleosomes (Bevington and Boyes, 2013). Transfected cells were incubated for 48 h, and plasmid DNA was recovered via Hirt extraction (Hirt, 1967) and used in a qPCR assay. The level of cleavage was measured using a primer pair that amplifies across the RAG cleavage site so that only intact substrate is amplified. The data were then normalised by qPCR amplification of a sequence unique to each substrate plasmid (Fig. 5.1A).

When cells are transfected with RAG1, RAG2, and 12RSS and 23RSS plasmids, the amount of intact 12RSS detected decreases by 26%, and the amount of intact 23RSS detected decreases by 34%, relative to samples where RAG2 is not transfected (Fig. 5.1B), consistent with previously published data (Steen et al., 1996; Han et al., 1999). When transfected with a 12RSS and an ESC, 12-18% of the 12RSS is cleaved but no cleavage of ESC plasmid is detected (Fig. 5.1B). Similarly, when transfected with 23RSS and ESC substrates, the 23RSS is cleaved by 10%, but the ESC is not cleaved relative to samples lacking RAG2. Therefore, these data indicate that cleavage of two RSSs is symmetric, but RSS-ESC cleavage is asymmetric *in vivo*, consistent with the cut-and-run hypothesis.

Since purified full-length RAG proteins did not show asymmetric cleavage of oligonucleotide substrates *in vitro* (Fig. 3.16), the cleavage activity of core and full-length RAG proteins *in vivo* was compared. Importantly, cleavage of co-transfected 12RSS and ESC substrates is asymmetric with both core and full-length RAG proteins *in vivo* (Fig. 5.1B), consistent with full-length RAG proteins catalysing the cut-and-run reaction. Results from these qPCR assays indicate that there is absolutely no cleavage of the ESC.

The qPCR assay is unable to differentiate between nicked DNA and hairpinned double-strand break products, since neither will be amplified. Therefore, ligation-mediated PCR (LM-PCR) was used to verify that DSBs, and not just nicks, are formed at the 12 and 23RSS substrates. Here, the linker will only ligate to a DSB, not a single-strand nick. As can be seen in Fig. 5.2, an LM-PCR product is formed at the cut 12RSS in the presence of both a 23RSS and ESC partner. Furthermore, the product is the same size as for control DNA digested with EcoRV. Notably, the uncut ESC does not generate an LM-PCR product, but the positive control sample does (ApaLI-cut, and filled in with Klenow polymerase), demonstrating a cleaved ESC, if formed, could be detected by LM-PCR. Overall, these data indicate that DSBs are formed at the plasmid substrates, and it is likely that this is why a decrease in PCR product is seen in Fig. 5.1.
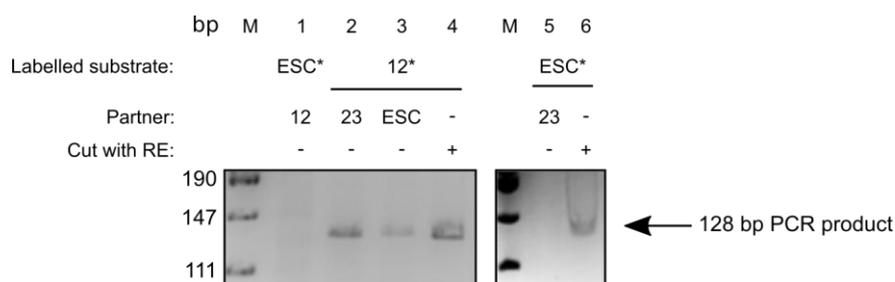
Figure 5.2 – LM-PCR of extrachromosomal substrate cleavage products. The plasmid samples from Figure 5.1 were used in an LM-PCR assay to verify that the plasmids were fully cleaved to form DSBs, and not nicked, since the qPCR assay cannot differentiate between DSBs and nicked DNA. Radioactive LM-PCR products were run on a 1% agarose gel, dried and visualised using a phosphorimager. The labelled substrate is indicated above the gel with an asterisk. R indicates positive control samples that were cut with a restriction enzyme. A product of 128 bp is detected only with the 12RSS samples, and not with the ESC sample. Positive control samples cut with EcoRV also yield products of the expected size. See section 2.47 for LM-PCR primer details.

5.2 An endogenous ESC is intact in mouse pro-B cells

The Igλ locus is the simplest of the seven Ig loci, containing just six V and J gene segments, and only one known recombination enhancer, making it an ideal model for investigating V(D)J recombination (Haque et al., 2013). Recombination of Igλ is activated at the pro-B/pre-B cell transition by a single transcription factor, IRF4, which activates recombination by binding to the Igλ enhancer and triggering non-coding transcription (Eisenbeis et al., 1993; Muljo and Schlissel, 2003). The Boyes laboratory have produced a transgenic mouse line where IRF4 is under the control of the pro-B-cell-specific λ5/Vpre-B promoter and locus control region (LCR). In the pro-B cells of these transgenic mice, IRF4 is expressed at the same level as in pre-B cells. Consequently, this single change increases the level of recombination in pro-B-cells to pre-B-cell levels, thus demonstrating that IRF4 is the only transcription factor required to fully activate recombination of the Igλ locus (Bevington and Boyes, 2013).

This transgenic line (PIP3) provides a convenient model for investigating the *in vivo* state of a *bona fide* ESC. Approximately 60% of all rearrangements at the Igλ locus occur between just two gene segments, namely Vλ1 and JCλ1 (Fig. 5.3; Boudinot et al., 1995). Therefore, the most common ESC in pro-B cells is likely to be the Vλ1-JCλ1 ESC and this provides a good substrate for examining ESC cleavage *in vivo*. Furthermore, the amount of Vλ1-JCλ1 ESC should correlate with the level of Vλ1-JCλ1 recombination. This can be tested by comparing nontransgenic samples with transgenic pro-B cell samples, which should display a ~4-fold increase in recombination over nontransgenic samples (Bevington and Boyes, 2013).

Conveniently, the sequence of the adjoined RSSs in an ESC creates an ApaLI site. Therefore, ApaLI-digested DNA can be used as a negative control.

Samples from nontransgenic (NTG) and transgenic (PIP3) mouse pro B cells were prepared by James Scott for unrelated experiments, but he kindly provided me with an aliquot of each sample for these analyses. Pro-B cells from the bone marrow of mouse femurs were purified by fluorescence-activated cell sorting (FACS) using two pro-B-cell-specific markers, FITC-CD19 and PE-CD43. DNA purified from double-positive cells was analysed by qPCR using primers specific to recombined Vλ1-JCλ1 and the Vλ1-JCλ1 ESC (Fig. 5.3A). The data were then normalised to the level of GAPDH in each sample.

In the PIP3 transgenic pro-B cells, the level of recombination is 5-fold greater than in NTG pro-B cells (Fig. 5.3B), similar to previously published data (Bevington and Boyes, 2013). The level of recombination decreases with increasing amounts of interleukin 7 (IL-7) in the growth medium. IL-7 is essential for promoting the proliferation and survival of pro-B cells, and to inhibit pro-B to pre-B differentiation (Namen et al., 1988; Lu et al., 1999; Milne et al., 2004). However, IL-7 signalling also inhibits recombination by downregulating RAG1 and RAG2 expression. This is achieved via activation of the phosphoinositide 3 kinase-AKT pathway, which phosphorylates the forkhead box protein O (FOXO) proteins that are directly responsible for activating RAG1 and RAG2 expression (Amin and Schlissel, 2008; Herzog et al., 2008; Clark et al., 2014). Therefore, a growth medium concentration of 0.5 µg/ml IL-7 represents the best trade-off between the survival and proliferation of pro-B cells and the inhibition of recombination, as concentrations greater than this inhibit recombination (Fig. 5.3B).

Importantly, there is a good correlation between the levels of ESC and recombination in each undigested sample and at each concentration of IL-7, indicating that the ESC is intact in these cells (Fig. 5.3B). In samples digested with ApaLI, the level of recombination is the same as in undigested samples, since ApaLI does not cut in the recombination amplicon. However, the level of ESC is greatly reduced in each sample, verifying that the ESC is being amplified specifically in this assay. Together, these data suggest that an endogenous ESC is largely intact.

A

B
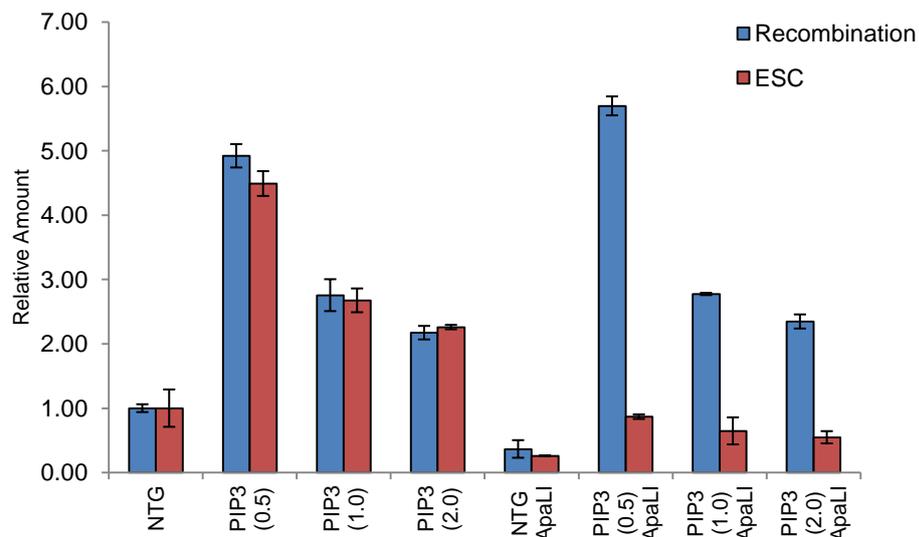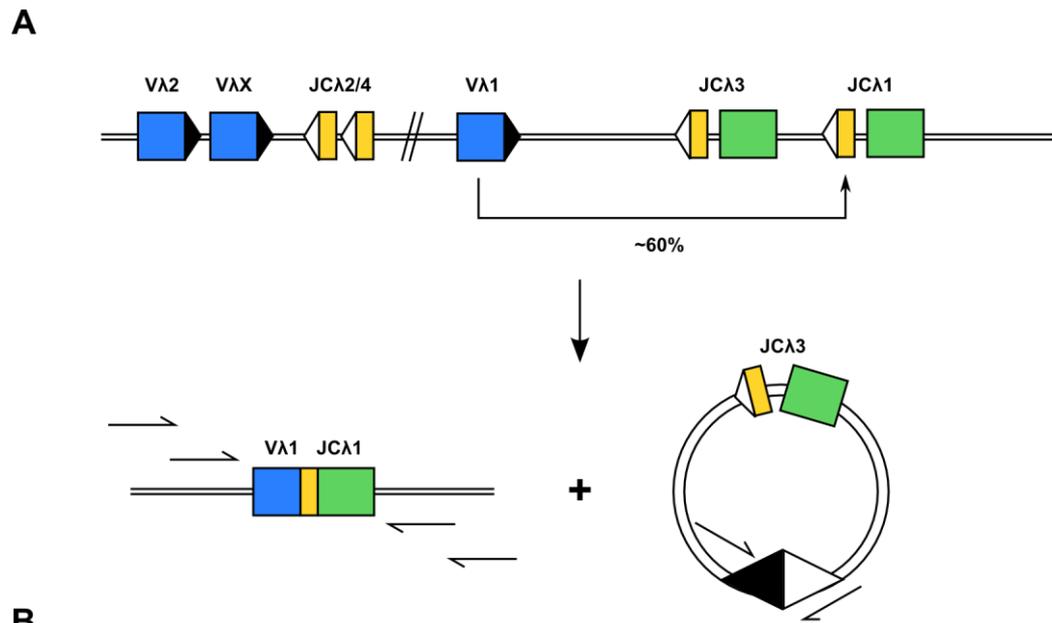
Figure 5.3 – An endogenous ESC from the Igλ locus is intact in vivo (A) A simplified schematic of the murine Igλ locus. The lambda light chain is the simplest immunoglobulin locus, and ~60% of rearrangements occur between gene segments Vλ1 and JCλ1 (Boudinot et al., 1995). The primers used to detect the level of Vλ1-JCλ1 recombination and ESC are indicated by the half-head arrows (see Section 2.48 for specific primers used). Specific detection of recombination required an initial PCR step (outer primers), the products of which were then used in nested qPCR (inner primers). **(B)** To test whether the Vλ1-JCλ1 ESC is intact *in vivo*, DNA was extracted from nontransgenic (NTG) and transgenic (PIP3) mouse pro-B cells that overexpress IRF4 to induce early recombination. The number in brackets for each PIP3 sample indicates the concentration of IL-7 in the growth medium, in μg/ml. Samples were analysed by qPCR using the primer pairs shown in (A), and normalised to GAPDH. The values given are relative to NTG, which is set at 1. The level of ESC correlates with the level of recombination. Digestion with ApaLI, which cleaves at the ESC heptamer-heptamer junction, greatly reduces the level of ESC. Together, these data imply that the Vλ1-JCλ1 ESC is fully intact *in vivo*. Error bars show the SEM of 2 independent experiments.

C)   Discussion

Cleavage of extrachromosomal substrates is asymmetric *in vivo*

In this chapter, I investigated the likelihood that cut-and-run occurs *in vivo* in two ways. Firstly, using transfected extrachromosomal substrates and RAG expression vectors in non-lymphoid cells, I found that 12/23 cleavage is symmetric *in vivo*, whereas RSS/ESC cleavage is asymmetric. These findings are supported by the fact that no cleaved ESCs could be detected using LM-PCR. Therefore, these results indicate the asymmetric cleavage observed in Chapter 3 is not an *in vitro* artefact that is peculiar to purified RAG proteins or short DNA oligonucleotide substrates. The transfected RAG constructs do not contain any affinity tags yet still display the same asymmetric cleavage activity as the MBP-tagged RAG proteins used *in vitro*, indicating that the large MBP tag is unlikely to have influenced this aspect of RAG activity in the data presented in Chapters 3 and 4. Furthermore, the use of plasmid substrates is a more accurate reflection of how cut and-run would occur *in vivo*, on longer, and in the case of the ESC, circular DNA molecules.

It is possible that cleavage of the extrachromosomal substrates took place within a single complex, rather than a paired complex. To confirm that cleavage of the transfected substrates adheres to the 12/23 rule, a lone 12 or 23RSS should be transfected, and the level of cleavage compared with the level of cleavage in the presence of a suitable partner. In addition, transfection experiments using further ESC mutants should be performed analogous to the *in vitro* experiments described in Chapter 4, to see whether asymmetric cleavage *in vivo* is also a result of RAG binding to both sides of the ESC, as it is *in vitro*.

The possibility that the ESC reintegrates with the RSS plasmid, or indeed into a genomic location, was not addressed. It is possible that reintegration occurs to some extent, since it is a similar experimental system that was originally used to detect reintegration (Vanura et al., 2007). However, the fact that the amount of intact ESC did not differ from samples containing no RAG2 (Fig. 5.1) indicates that if any reintegration does occur, it occurs only at a very low level. Nevertheless, the level of reintegration should be assessed in order to make a direct comparison between the potential frequency of cut-and-run and reintegration *in vivo*.

An endogenous ESC detected in murine pro-B cells is intact

Testing whether cut-and-run occurs in B cells is not straightforward, and the possible ways in which this could be done are addressed in Chapter 6. To begin to address this question, I investigated whether an endogenous ESC is intact in B cells, for two reasons: a) an intact ESC is required for cut and run, and b) previous studies indicate that signal ends are abundant in developing B and T cells, suggesting that ESCs are not intact (Livak and Schatz, 1996; Jiang et al., 2004). Furthermore, the RAG proteins remain stably bound to signal ends, which persist until the RAG proteins are downregulated (Ramsden and Gellert, 1995; Livak and Schatz, 1996; Jiang et al., 2004).

I investigated the *in vivo* state of a *bona fide* ESC using transgenic and nontransgenic mouse pro-B cells. Importantly, in this system, pro-B cells are cultured for 5 days prior to recombination taking place. This slowly increases between day 5 and day 7, when the cells were isolated for analysis (Bevington and Boyes, 2013). Therefore, the ESC is isolated relatively soon after recombination. qPCR analyses indicate that the level of a specific ESC correlates well with the corresponding coding joint, suggesting that, in contrast to previous studies, the ESC is intact in these cells. However, a limitation of the approach taken here is that the level of signal ends was not assayed. If the ESC is cleaved/unjoined by the same amount in each sample, this would give the same ratio of recombination:ESC in each case. One way of addressing this issue would be to use LM-PCR to detect cleaved/unjoined ESC in these pro-B samples, in the same way LM-PCR was used to detect broken ESCs in the transfection experiments (Fig. 5.2).

Several studies indicate that signal ends can only form ESCs once RAG expression is downregulated (Livak and Schatz, 1997; Jiang et al., 2004). This suggests that in the pro-B cells used in the present experiments, either a) ESC formation can occur to an extent in the presence of RAG proteins, or b) the RAG proteins are downregulated following light chain recombination in these cells. There are two phases of B cell development at which RAG protein expression is downregulated. The first is following successful *IgH* recombination and expression of the pre-B cell receptor (Grawunder et al., 1995). The second is when cell surface IgM is expressed following successful *IgL* recombination, which signals for the downregulation of RAG expression (Wilson et al., 1994). In addition, RAG2 is degraded via the ubiquitin-proteasome pathway before entry into S phase (Jiang et al., 2005). Whether cell-surface IgM is expressed in the transgenic cells used in this study has not been determined. It is possible that once the

Igλ locus successfully rearranges in these cells, cell-surface IgM is expressed, leading to the downregulation of RAG expression and the joining of signal ends to form ESCs.

If ESCs cannot form until the RAG expression is downregulated, what scope is there for RAG proteins to catalyse cut-and-run *in vivo*? There are two conceivable scenarios where the RAG proteins and ESCs could co-exist in B and T cells. The first is if the first rearrangement of an allele is unproductive. In most cells the first rearrangement of an Ig locus is unproductive, and the cell rearranges the second allele (Rajewsky, 1996). Therefore, in most lymphocytes, the RAG proteins will coexist with an ESC from the first unproductive rearrangement. However, if signal ends cannot be joined to form ESCs whilst RAG proteins are still expressed, then this scenario might occur only rarely.

During B cell development the *IgH* locus is rearranged before the *IgL* loci, and in T cells the TCRβ locus rearranges before TCRα (Bassing et al., 2002). Therefore, the second scenario where RAG proteins and ESCs (in this case from the IgH/TCRβ locus) could coexist is immediately after rearranging the IgH/TCRβ chain, and whilst the IgL/TCRα chain is rearranging in the pro-B to pre-B or CD4⁻/CD8⁻ to CD4⁺/CD8⁺ transition. In this case, the RAG proteins are downregulated after the IgH/TCRβ rearrangement, and are then upregulated for IgL/TCRα rearrangement, which would give the signal ends an opportunity to become ESCs. In any case, the data presented in this chapter indicate that asymmetric cleavage is not merely an *in vitro* artefact, and that a *bona fide* ESC is intact in mouse pro-B cells, a prerequisite for cut-and-run. Whether or not the production of intact ESCs coincides with RAG expression remains to be seen.

Chapter 6

Discussion

6.1   Cut-and-run

An integral part of V(D)J recombination is the excision of intervening DNA between recombining gene segments to form an ESC by-product. Sealing the two signal ends in an ESC was previously thought to be a way of sequestering the potentially reactive 3'-OH group from undergoing transposition reactions (Arnal and Roth, 2007). However, it was later discovered that transposition is almost completely inhibited by the RAG2 C-terminus, and happens rarely *in vivo* (Elkin et al., 2003; Messier et al., 2003; Messier et al., 2006). Two reports published around the same time then indicated that ESCs are far from inert, and can be rebound by the RAG proteins by virtue of the two RSSs, and reintegrated back into the genome at RSSs or cRSSs (Curry et al., 2007; Vanura et al., 2007). Reintegration of ESCs is potentially a dangerous reaction, specifically if reintegration of an ESC occurs adjacent to an oncogene (Fig. 1.11B). Evidence from transgenic mice indicated that the RAG2 C-terminus inhibits reintegration 7-fold (Curry et al., 2007). Therefore, the initial aim of this project was to investigate the mechanism of ESC reintegration and the mechanism by which the RAG2 C-terminus inhibits reintegration, with a view to developing inhibitors of RAG-ESC binding to prevent reintegration entirely. However, during verification and optimisation of the reaction conditions, I discovered that cleavage of an ESC-RSS pair is asymmetric (Chapter 3). If cleavage is truly asymmetric *in vivo*, then in the majority of ESC-RSS pairings the ESC will remain uncleaved. Since the RSS appears to be released following cleavage, cleavage at the RSS could reoccur until the ESC becomes cleaved, or until the RAG proteins are downregulated, thus generating multiple genomic DSBs in developing lymphocytes. In turn, these could potentially be substrates for chromosome translocations, leading to leukaemia.

In vivo, recombination only occurs between gene segments that are flanked by different length RSSs, a requirement known as the 12/23 rule (Tonegawa, 1983). Under conditions that recapitulate the 12/23 rule *in vitro* (specifically, short incubation times and the presence of Mg^{2+} and DMSO), I found that cleavage of an RSS/ESC complex is asymmetric (Fig. 3.11). Asymmetry is not a factor of RAG concentration, since cleavage is asymmetric at a range of RAG concentrations (Fig. 3.12). Analysis of cleavage products by denaturing gel electrophoresis revealed that cleavage of ESCs is blocked at the nicking step, rather than the binding or hairpinning step (Fig. 3.18).

Furthermore, the asymmetry is more pronounced with a 23RSS than with a 12RSS. A low level of hairpinning of ESCs is seen in the presence of a 12RSS, but cleavage of ESCs in the presence of a 23RSS also appears to be blocked at the hairpinning step.

Since an ESC contains two RSSs in a head-to-head configuration, a plausible hypothesis is that RAG proteins bind to both RSSs of the ESC simultaneously, thus blocking efficient binding of the heptamer-heptamer border and preventing efficient cleavage. This prediction was borne out by EMSA experiments which showed that the RAG proteins form a larger complex with ESCs than with RSSs (Fig. 4.1), and by cleavage experiments with mutant ESCs, which showed that deleting either the conserved heptamer or nonamer on one side of the ESC restores symmetric cleavage by preventing simultaneous binding (Section 4.4).

Since the danger posed by cut-and-run relies on the fate of the broken DNA ends following cleavage, the stability of postcleavage complexes was analysed by an EMSA timecourse (Section 4.6). These data imply that the coding end of an RSS is released by the RAG proteins following cleavage, whereas the ESC is retained. Experiments with extrachromosomal substrates showed that asymmetric cleavage of ESCs and RSSs also occurs *in vivo*, and analysis of an endogenous ESC in pro-B cells indicates that the ESC is largely intact, a prerequisite for the cut-and-run phenomenon. Together, these experiments provide evidence in favour of the cut-and-run hypothesis.

6.2   Asymmetric cleavage

Cleavage of an RSS by the RAG proteins occurs in two steps. First, a nick is introduced precisely at the heptamer-coding sequence boundary (McBlane et al., 1995). Nicking of RSSs can occur independently of synapsis with a suitable partner, both *in vitro* and *in vivo* (Eastman and Schatz, 1997; Yu and Lieber, 2000). Synapsis with a suitable partner RSS is essential for hairpin formation, which proceeds via transesterification of the opposite strand. Analysis of the RAG cleavage products on a denaturing gel shows that a high level of nicking occurs at both 12 and 23RSSs in the absence of a partner substrate, and that addition of a suitable partner stimulates hairpin formation, consistent with previous studies (Eastman and Schatz, 1997; Yu and Lieber, 2000).

The level of nicking at a lone ESC is much lower than at a lone RSS, indicating that ESC cleavage is blocked at the nicking step (Fig. 3.18). Several cases have been reported where RAG cleavage is blocked at the nicking step. First, RAG cleavage is

blocked at the nicking step when the aromatic residues thought to stabilise a flipped out base are mutated (Grundy et al., 2007). Furthermore, DNA is only known to inhibit RAG cleavage at the nicking step when certain base pairs of the coding flank sequence are replaced with abasic nucleotides (Grundy et al., 2007). Cleavage efficiency can be greatly influenced by the flanking coding sequence, but neither RSS in the ESC constitutes a bad flank for the other (Ramsden et al., 1996). Given that sequencing of the RAG expression vectors showed no mutations, and DNA substrates with abasic residues were not used here, the source of blocking at the nicking step must be unrelated to the RAG proteins themselves or the specific sequence of the ESC.

A plausible hypothesis is that the RAG proteins bind to both RSSs of the ESC simultaneously and compete (unsuccessfully) for binding at the heptamer boundary. The low level of nicking seen at the ESC means that binding is not blocked entirely, suggesting that there is a degree of "breathing" by the complex (Fig. 3.18). Synapsis with a 12RSS partially relieves the block since there is an increase in hairpin formation (Fig. 3.18). Synapsis with a 23RSS results in a slight increase in the level of nicking, but does not allow for hairpin formation (Fig. 3.18). This lack of hairpinning indicates that there is an additional block at the hairpinning step, or that the block at the nicking step extends to the hairpin step with a 23RSS. The reason for the block in ESC hairpinning when synapsed with a 23RSS is unclear, especially since the molecular basis of the 12/23 rule is unknown. In some cases, 23RSS cleavage is stimulated less well by an ESC than by a 12RSS (i.e. asymmetric cleavage is greater with a 23RSS). This is consistent with a study which found that reintegration occurs *in vivo* at Vκ 12RSSs but not Jκ 23RSSs (Curry et al., 2007). One possibility is that the 23/ESC complex is less stable, but this is not supported by the bandshift timecourse assays, since a 23-ESC complex is as stable as a 23-12 complex (Fig. 4.13 and 4.14). Another possible explanation is that in an ESC-23RSS pairing, RAG proteins do not undergo the necessary conformational change to catalyse efficient hairpin formation by aligning a flipped out base with aromatic residues of the RAG1 active site.

The blocking hypothesis could be further tested by introducing a spacer of several bp between the two heptamers of the ESC. The prediction here would be that the higher molecular weight complex would still be abundant with this ESC, but that symmetric cleavage would be restored. A possible scenario that could explain the lack of ESC hairpin formation when paired with a 23RSS (Fig. 3.18) is that the second "bystander" RSS in the ESC somehow influences hairpinning. It might be that when the 23RSS of an ESC is paired with a 12RSS, the 12RSS bystander in the ESC allows hairpinning, but when an ESC is paired with a 23RSS, the 23RSS bystander somehow blocks

hairpinning. This could be tested by pairing a 12-12 ESC with a 23RSS and a 23-23 ESC with a 12RSS partner.

6.3   RAG binding to both RSSs of an ESC prevents efficient cleavage

EMSAs showed that an extra complex of higher molecular weight is more abundant with a labelled ESC than a labelled RSS (Fig. 4.1), consistent with the hypothesis that the RAG proteins bind to both RSSs of the ESC simultaneously. If the source of asymmetry is double occupancy, then the ESC-RSS paired complex (PC) has to contain an RSS partner and the RAG proteins bound to both RSSs of the ESC. Whilst the presence of a higher molecular weight complex is consistent with double occupancy/simultaneous binding, the protein/DNA content and cleavage activity of this complex is unknown. This could be addressed in two ways. First, to deduce the protein stoichiometry of the upper complex, differential tagging of the RAG1 and RAG2 proteins combined with supershift EMSA could be performed. Secondly, the cleavage activity of the additional ESC complex could be analysed using an in-gel cleavage assay (Swanson, 2002a). In this assay, complexes are formed in Ca^{2+} (to prevent substrate cleavage) and separated on a gel. Discrete complexes are excised from the gel and incubated in Mg^{2+}-containing buffer to instigate substrate cleavage and the products are separated on a second gel. This would show whether or not the upper complex is competent for asymmetric cleavage.

If the doubly-occupied ESC is involved in asymmetric cleavage, then it would be expected to form the same higher molecular weight complex when paired with a labelled 12 or 23RSS (i.e. the ESC would supershift a labelled RSS), but this was not observed (Fig. 4.1). However, the formation of a PC could not be detected with a standard 12/23RSS pair. Detection of the PC is confounded by the fact that HSC2 and PC have the same protein content and therefore co-migrate. I attempted to demonstrate PC formation by streptavidin supershift in Chapter 4 (Fig. 4.4) but was unsuccessful, possibly because the PC is only transient, or it is unstable under the electrophoresis conditions used. If either or both of these are true, the PC could be captured by cross-linking then used in a supershift EMSA. In this case, pairing of the doubly-occupied ESC with a labelled 12/23 RSS would be expected to give a higher molecular weight complex.

6.4   The fate of broken ends following asymmetric cleavage

Time course analysis of complex stability *in vitro* indicates that the RAG proteins release the coding end of an RSS after cleavage (Section 4.6). In these experiments, only the coding end was labelled, therefore they should be repeated with oligonucleotides where the signal end is labelled to deduce its fate. The stability of postcleavage complexes could also be analysed using single molecule total internal reflection fluorescence microscopy (smTIRFm). Use of smTIRFm would enable analysis of individual complexes, as opposed to EMSAs, where a heterogeneous population of molecules is simultaneously examined. In the smTIRF approach, RAG proteins would be immobilised on a cover slip and bound to an RSS-containing oligonucleotide labelled with a FRET pair, such that cleavage of the oligonucleotide would result in loss of the FRET signal (Wang et al., 2012). In the presence of calcium, an ESC labelled with a third fluorophore would be added to assemble RSS-ESC complexes, and complexes containing all three fluorophores would be analysed. Then, magnesium would be added to trigger cleavage and loss of the FRET signal would be used to follow dissociation of the cleaved RSS. The dissociation of the coding end, signal end and ESC would then be monitored by virtue of their different fluorophores.

6.5   Full length RAG proteins

A considerable amount of time and effort was expended in expressing and purifying flRAG proteins. Unfortunately, once successfully expressed and purified from HEK-293T cells, they were not concentrated enough for binding studies, and data from cleavage assays indicated that ESC-RSS cleavage was not asymmetric (observed with three individual preparations). One possible reason why the *in vitro* cleavage activity of flRAG proteins is inconsistent with that of the cRAG proteins is that the flRAG proteins were not concentrated enough to occupy both RSSs of the ESC. The fact that flRAG proteins were not concentrated enough to saturate the probe in an EMSA is consistent with this. However, cleavage of transfected substrates in nonlymphoid mammalian cells shows that both full-length and core RAG proteins display asymmetric cleavage of ESC and RSS substrates *in vivo*.

A comparison of core and full-length RAG proteins is crucial because the RAG2 C-terminus has been shown to inhibit reintegration 7-fold (Curry et al., 2007). Since reintegration and cut-and-run both involve binding an ESC, it is also possible that the RAG2 C-terminus also inhibits cut-and-run. The mechanism by which the RAG2 C-terminus inhibits reintegration is unknown. The PHD finger of the RAG2 C-terminus

directs the RAG recombinase to actively transcribed regions by binding to H3K4me3 (Matthews et al., 2007; Ramon-Maiques et al., 2007), and can bind to H3K4me3 at any location in the genome in the absence of RAG1. However, RAG1 and RAG2 binding only coincides at RSSs that contain H3K4me3 (Ji et al., 2010). Therefore, it is possible that the RAG2 C-terminus inhibits reintegration by directing the RAG proteins away from ESCs, which might lose their H3K4me3 after excision from the genome. However, the ESC plasmid substrate used in the *in vivo* cleavage assay contained the CMV promoter upstream of the signal joint, to enable transcription (Fig. 5.1). Therefore, the ESC likely contained H3K4me3, but was cleaved asymmetrically by flRAG proteins regardless. Preliminary data from the Boyes laboratory suggest that the RAG2 C-terminus inhibits the end-joining step of reintegration. However, since cut-and run does not involve joining ESC ends with genomic ends (and it releases the RSS ends), it is unlikely that the C-terminus would inhibit cut-and-run in the same way.

6.6   Does cut-and run occur in B and T cells?

Definitive proof that cut-and-run occurs *in vivo* is inherently difficult to obtain. Unlike chromosome translocations and reintegration, which result in detectable DNA rearrangements and insertions, cut-and-run would leave no trace other than a broken RSS. In addition, there would be nothing specific about the broken RSS to indicate that it was caused by synapsis with an ESC, since the mechanism by which the RSS is cleaved is the same as standard V(D)J recombination. Furthermore, if the cleaved RSS was subsequently used in a chromosome translocation, there would be no indication that the breakpoint was specifically caused by an ESC. Nevertheless, there are several ways in which the likelihood of cut-and-run could be assessed.

Extrachromosomal substrate assays indicate that cleavage of RSSs and ESCs is asymmetric *in vivo*, and therefore is not an *in vitro* artefact (Fig. 5.1). In these experiments RAG levels are likely to be very high since RAG expression is under the control of the strong EF-1α promoter. To verify that asymmetric cutting occurs at physiological levels of RAG proteins, the same plasmid substrates used in Chapter 5 would be used to transfect a cell line that endogenously expresses the RAG proteins. One such cell line is the human *ETV6-RUNX1* positive leukaemic cell line, REH (Uphoff et al., 1997; Wagner et al., 2004). Since the level of cleavage will be lower in these cells than in transfected NIH/3T3 cells, one of the substrate plasmids would contain a green fluorescent protein transfection marker. This would allow for FACS purification of successfully transfected cells, which in turn should improve the detection of substrate cleavage. Analysis of cleavage would be done using a qPCR assay in the

same way as in Chapter 5.

In order for ESCs to pose a threat, they must be present in a lymphocyte at the same time as the RAG proteins. Previous studies indicate that signal ends are tightly bound by RAG proteins, and are unlikely to form ESCs until the RAG proteins are downregulated (Livak and Schatz, 1996; Zhu et al., 1996; Agrawal and Schatz, 1997; Jiang et al., 2004). Analysis of a single endogenous ESC in mouse pro-B cells indicates that at least some degree of ESC formation occurs (Fig. 5.3), but whether formation of the ESC coincides with RAG expression was not investigated. As discussed in Chapter 5, there are two instances where RAG expression could coincide with ESC formation. The first is following unproductive rearrangement of the first Ig/TCR allele, and the second is following rearrangement of the IgH/TCRβ locus, whilst the IgL/TCRα locus is rearranging. RAG binding of the Vλ1JCλ1 ESC, which is formed in ~60% of cells that undergo *IgL* rearrangement, could be tested by ChIP with an anti-RAG1 antibody followed by qPCR analysis of the recovered DNA with primers specific to the head-to-head RSSs. If RAG expression and ESC formation coincide, and ESCs are bound by RAG proteins, intact ESCs should be pulled down with RAG1. If ESCs and RAG expression do not coincide, only signal ends will be pulled down.

A transgenic mouse line recently developed in the Boyes laboratory could be used to further test this. These mice express IRF4 fused to the estrogen receptor (ER) to form IRF4-ER. This fusion protein is under the control of the λ5/VpreB promoter and LCR, (as for the mice described in Chapter 5), therefore it is only expressed in pro-B cells (Sabbattini et al., 2001). Notably, increased levels of IRF4 in pro-B cells is all that is required to full activate *IgL* recombination (Bevington and Boyes, 2013). Activity of IRF4-ER is induced by the addition of tamoxifen to the growth medium, which then triggers early Igλ recombination in pro-B cells. The ability to control the onset of Igλ recombination allows for temporal analyses. *IgL* recombination could be induced in B cells from these mice, and then the Vλ1-Jλ1 signal ends could be tracked by LM-PCR at intervals to deduce when joining of signal ends occurs. This could be combined with qPCR/western blot analysis of RAG expression following induction to see if there is any overlap between RAG expression and ESC formation.

6.7   Do ESCs form synaptic complexes with genomic RSSs?

In addition to the asymmetric cleavage of ESCs, and the coexpression of RAG proteins and ESCs, the ESC has to form a complex with a genomic RSS for cut-and-run to occur. Evidence from a study investigating reintegration of an ESC at genomic loci indicates that this is likely to be the case (Curry et al., 2007). To gain a comprehensive map of the genomic sites that are bound by ESCs, a combination of ChIP-Seq and a transcription activator-like effector (TALE) designed to recognise an ESC could be used. TALEs can be designed to bind any sequence of interest (Scott et al., 2014). Therefore, an expression vector for HA-tagged TALE that binds next to the head-to-head RSSs in an ESC could be cotransfected into mammalian cells with the ESC. TALE-ChIP-Seq could then be used to deduce the genomic binding locations of the ESC. If ESCs do indeed undergo cut-and-run reactions, the binding sites would be expected to map to RSSs and cRSSs.

In a second approach to determine whether ESCs bind to genomic RSSs and trigger DBSs, the incidence of γH2AX foci could be examined. Within one minute of a chromosomal DSB forming, the histone variant H2AX is phosphorylated at Ser139, creating γH2AX (Rogakou et al., 1998). γH2AX foci spread for several Mb around the site of the DSB (Rogakou et al., 1999) and are crucial in signalling the formation of DSBs. RAG-induced DSBs at genomic RSSs are no different, leading to rapid formation of γH2AX foci (Chen et al., 2000). Therefore, it would be possible to see whether ESCs trigger DSBs in the genome by transfecting cells with an ESC plasmid, and measuring the incidence of γH2AX foci relative to untransfected cells. There are multiple methods for detecting γH2AX, but immunofluorescence is the best established technique (Nakamura et al., 2006).

The incidence of γH2AX foci caused by ESCs could be tested in three different ways. First, nonlymphoid mammalian cells such as COS-7 cells could be transfected with an ESC and RAG expression vectors, and the resulting γH2AX foci compared with untransfected cells. This would give the highest chance of detecting ESC-induced DSBs since the RAG proteins would be highly expressed, but the ESC would only be able to synapse with cryptic RSSs since the Ig loci are inactive in these cells. The second way of testing ESC-induced DSBs would be to transfect the lymphoid REH cell line with an ESC, but no RAG expression vectors. This would allow for break analysis in a more relevant cell line where the Ig loci are accessible, and where RAG proteins are expressed at physiological level.

The third approach makes use of the IRF4-ER-expressing mice described above. To gain insight into the physiological level of cut-and-run resulting from V(D)J recombination, γH2AX foci could be measured following induction of Igλ recombination and compared with uninduced cells at intervals following induction of recombination. Since most Igλ recombination occurs between Vλ1 and JCλ1, increases in γH2AX due to V(D)J recombination-induced breaks will be small. Therefore, a large increase in γH2AX foci might indicate ESC-induced breaks. The site of γH2AX foci could then be mapped by ChIP-seq and compared with site of ESC binding (above) to determine whether any of the DSBs were likely to be ESC-induced. Together, data from experiments that test if a) RAG and ESC expression coincides, b) if the ESC is bound by RAG proteins at genomic RSSs, and c) if the presence of the ESC increases the incidence of γH2AX foci at RSSs, would enable a comprehensive assessment of whether cut-and-run truly occurs *in vivo*.

6.8   Implications of cut-and-run

a) Cut-and-run could be a major way of causing orphan DNA breaks during V(D)J recombination.

Chromosome translocations caused by aberrant recombinase activity in the production of antibodies and T cell receptors are a hallmark of many leukaemias and lymphomas (Kuppers, 2005). In this thesis, I have identified a new way in which aberrant RAG activity could produce orphan DSBs that have the potential to participate in carcinogenic chromosome translocations. It is possible that many of the DSBs resulting from cut-and-run are efficiently repaired via NHEJ. Nevertheless, the high frequency of chromosome translocations in many lymphoid cancers proves that many DSBs, regardless of their origin, are capable of escaping the cell's repair machinery. Therefore, cut-and-run could be a potent source of translocation partners at the Ig/TCR loci.

One of the most common forms of translocation caused by V(D)J recombination is end donation (Roth, 2003; Kuppers, 2005). In end donation, a broken RSS formed by RAG cleavage is joined with a DSB formed by an unrelated process, such as ionizing radiation. It is unclear how end donation occurs, since RSS cleavage should take place within the context of a paired complex. Whilst it is true that the coding ends are bound less tightly than signal ends *in vitro*, this weaker association might be necessary for efficient resolution of coding ends via the NHEJ pathway. This is supported by certain RAG mutants that are competent for cleavage but deficient for joining, possibly by binding too tightly to the coding ends (Roth, 2003). Since cut-and-run occurs at a single

RSS during each cutting cycle, it is possible that cut-and-run is a potent source of "single" broken RSSs that can undergo an end donation-type translocation.

The mechanism by which broken RSSs are transferred to the NHEJ machinery is poorly understood. Mutational analyses of the RAG proteins revealed several joining-deficient mutants, indicating that the RAG proteins play a critical scaffolding role to facilitate end-joining (Qiu et al., 2001; Yarnell Schultz et al., 2001; Huye et al., 2002). Moreover, the acidic hinge of the RAG2 C-terminus helps to shepherd the DSBs toward the classical NHEJ pathway, and away from the error-prone alternative NHEJ pathway (Coussens et al., 2013; Gigi et al., 2014). It is unclear how asymmetric cleavage might affect the end-joining process. Since only one coding end and one signal end is produced following asymmetric cleavage, instead of two of each, this might negatively influence the end-joining. If the RAG proteins released the coding end but remain bound to the signal end as previous studies suggest, only the coding end will be immediately available for joining. This coding end might be bound by the nonhomologous end joining machinery, and in the absence of a suitable partner it is perhaps more likely to undergo translocation. Alternatively, the coding end might be transferred to the alternative NHEJ pathway, or perhaps not transferred to any end-joining machinery. More thorough analyses of the postcleavage complex following asymmetric cleavage, as described above, are needed to address these issues.

A specific example where cut-and-run could play a significant role is in leukaemia patients bearing the *ETV6/RUNX1* (*TEL/AML1*) translocation, the most common translocation found in childhood acute lymphocytic leukaemia (ALL; Pui et al., 2004). This translocation often arises *in utero*, but it cannot trigger malignancy in the absence of other aberrations. Instead, the *ETV6/RUNX1* translocation partially stalls B cell progression at a stage where the RAG proteins are still expressed (Tsuzuki et al., 2004), thus providing an extended window of opportunity for secondary recombination reactions. The majority of these aberrant recombination reactions occur at cryptic RSSs adjacent to transcriptionally active B cell development genes, thereby stalling B-cell progression even further, which can eventually result in full malignant transformation (Papaemmanuil et al., 2014). In patients where full malignant transformation has occurred, 27% of the chromosome aberrations contain an RSS or cRSS at only one side of translocation (Papaemmanuil et al., 2014), consistent with cut-and-run providing broken RSSs for end donation. If cut-and-run is the cause of these breaks, then their location would correspond with ESC binding in the experiments described above. Moreover, slowly proliferating pro-B cells deriving from the original transformed cell are continuously produced, meaning that a cell population

with the potential to become transformed is constantly present. Therefore, inhibitors that prevent the recombination by-product from triggering genome instability could be very useful in preventing progression from the pre-leukaemic state in these patients.

b) Cut-and-run may have a protective role in preventing reintegration

An alternative interpretation of this work is that rather than disrupting genomic stability and contributing to lymphoid transformation, cut-and-run actually plays a protective role in preventing reintegration. As described in the introduction, reintegration has the potential to upregulate oncogene expression if reintegration occurs at an adjacent cryptic RSS. Reintegration of an ESC is likely to be a largely irreversible process, since the introduction of non-templated base pairs at the hybrid joint is likely to render that RSS useless (Fig. 1.8).

Whilst cut-and-run creates DSBs that could participate in chromosome translocations, these DSBs have the potential to be repaired correctly by the cell. From this perspective, cut-and-run may represent the lesser of two evils. Since cut-and-run is more likely to occur than reintegration, by virtue of asymmetric cleavage, cut-and-run might actually work together with the inhibitory effect of the RAG2 C-terminus to prevent the potentially more damaging reintegration reaction.

6.9   Conclusions

Aberrant RAG activity is a hallmark of lymphoid cancers. Whilst investigating the mechanism of reintegration of the ESC by-product, I found that cleavage of ESC is actually remarkably inefficient compared with RSS cleavage. Since reintegration requires opening of the ESC, a more likely outcome from ESC-RSS synapsis is that the RSS is cleaved, but the ESC remains intact. This leaves the ESC free to stimulate multiple DSBs within the genome of developing B and T lymphocytes, thus posing a major threat to genomic stability in a reaction called cut-and-run.

Asymmetric cleavage of an ESC-RSS pair is most likely the result of RAG binding to both sides of the ESC simultaneously, and initial experiments indicate that cut-and-run could occur *in vivo*. This has significant implications for the generation of lymphoid cancers. In particular, it could be a potent source of broken RSSs that undergo chromosome translocations, and might play a large role in the progression of leukaemias that bear the ETV6/RUNX1 translocation.

References

Abarrategui, I. & Krangel, M.S. (2006). Regulation of T cell receptor-alpha gene recombination by transcription. *Nat Immunol* **7**, 1109-1115.

Abarrategui, I. & Krangel, M.S. (2007). Noncoding transcription controls downstream promoters to regulate T-cell receptor alpha recombination. *Embo J* **26**, 4380-4390.

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

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

Agrawal, A. & 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**, 43-53.

Aidinis, V., Dias, D.C., Gomez, C.A., Bhattacharyya, D., Spanopoulou, E., & Santagata, S. (2000). Definition of minimal domains of interaction within the recombination-activating genes 1 and 2 recombinase complex. *J Immunol* **164**, 5826-5832.

Akamatsu, Y., Monroe, R., Dudley, D.D., Elkin, S.K., Gartner, F., Talukder, S.R., Takahama, Y., Alt, F.W., Bassing, C.H., & Oettinger, M.A. (2003). Deletion of the RAG2 C terminus leads to impaired lymphoid development in mice. *Proc Natl Acad Sci USA* **100**, 1209-1214.

Akamatsu, Y. & Oettinger, M.A. (1998). Distinct roles of RAG1 and RAG2 in binding the V(D)J recombination signal sequences. *Mol Cell Biol* **18**, 4670-4678.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Robert, K., & Walter, P., Molecular Biology of the Cell, Available at http://www.ncbi.nlm.nih.gov/books/NBK21070/, (2002).

Alt, F.W., Yancopoulos, G.D., Blackwell, T.K., Wood, C., Thomas, E., Boss, M., Coffman, R., Rosenberg, N., Tonegawa, S., & Baltimore, D. (1984). Ordered rearrangement of immunoglobulin heavy chain variable region segments. *Embo J* **3**, 1209-1219.

Amin, R.H. & Schlissel, M.S. (2008). Foxo1 directly regulates the transcription of recombination-activating genes during B cell development. *Nat Immunol* **9**, 613-622.

Arbuckle, J.L., Fauss, L.A., Simpson, R., Ptaszek, L.M., & Rodgers, K.K. (2001). Identification of two topologically independent domains in RAG1 and their role in macromolecular interactions relevant to V(D)J recombination. *J Biol Chem* **276**, 37093-37101.

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

Arnal, S.M. & Roth, D.B. (2007). Excised V(D)J recombination byproducts threaten genomic integrity. *Trends Immunol* **28**, 289-292.

Bailin, T., Mo, X., & Sadofsky, M.J. (1999). A RAG1 and RAG2 tetramer complex is active in cleavage in V(D)J recombination. *Mol Cell Biol* **19**, 4664-4671.

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

Bassing, C.H., Swat, W., & Alt, F.W. (2002). The mechanism and regulation of chromosomal V(D)J recombination. *Cell* **109 Suppl**, S45-55.

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

Bellon, S.F., Rodgers, K.K., Schatz, D.G., Coleman, J.E., & Steitz, T.A. (1997). Crystal structure of the RAG1 dimerization domain reveals multiple zinc-binding motifs including a novel zinc binuclear cluster. *Nat Struct Biol* **4**, 586-591.

Belotserkovskaya, R., Oh, S., Bondarenko, V.A., Orphanides, G., Studitsky, V.M., & Reinberg, D. (2003). FACT facilitates transcription-dependent nucleosome alteration. *Science* **301**, 1090-1093.

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

Bergeron, S., Madathiparambil, T., & 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. *J Biol Chem* **280**, 31314-31324.

Bernstein, B.E., Kamal, M., Lindblad-Toh, K., Bekiranov, S., Bailey, D.K., Huebert, D.J., McMahon, S., Karlsson, E.K., Kulbokas, E.J., 3rd, Gingeras, T.R. et al. (2005). Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* **120**, 169-181.

Bevington, S. & Boyes, J. (2013). Transcription-coupled eviction of histones H2A/H2B governs V(D)J recombination. *EMBO J.*

Bischerour, J., Lu, C., Roth, D.B., & 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. *Mol Cell Biol* **29**, 5889-5899.

Blackwell, T.K., Moore, M.W., Yancopoulos, G.D., Suh, H., Lutzker, S., Selsing, E., & Alt, F.W. (1986). Recombination between immunoglobulin variable region gene segments is enhanced by transcription. *Nature* **324**, 585-589.

Boehm, T., Baer, R., Lavenir, I., Forster, A., Waters, J.J., Nacheva, E., & Rabbitts, T.H. (1988). The mechanism of chromosomal translocation t(11;14) involving the T-cell receptor C delta locus on human chromosome 14q11 and a transcribed region of chromosome 11p15. *EMBO J* **7**, 385-394.

Born, W.K., Reardon, C.L., & O'Brien, R.L. (2006). The function of gammadelta T cells in innate immunity. *Curr Opin Immunol* **18**, 31-38.

Boudinot, P., Rueff-Juy, D., Drapier, A.M., Cazenave, P.A., & Sanchez, P. (1995). Various V-J rearrangement efficiencies shape the mouse lambda B cell repertoire. *Eur J Immunol* **25**, 2499-2505.

Brandle, D., Muller, C., Rulicke, T., Hengartner, H., & Pircher, H. (1992). Engagement of the T-cell receptor during positive selection in the thymus down-regulates RAG-1 expression. *Proc Natl Acad Sci USA* **89**, 9529-9533.

Brandt, V.L. & Roth, D.B. (2009). Recent insights into the formation of RAG-induced chromosomal translocations. *Adv Exp Med Biol* **650**, 32-45.

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. et al. (2006). ATM stabilizes DNA double-strand-break complexes during V(D)J recombination. *Nature* **442**, 466-470.

Bujacz, G., Alexandratos, J., Wlodawer, A., Merkel, G., Andrake, M., Katz, R.A., & Skalka, A.M. (1997). Binding of different divalent cations to the active site of avian sarcoma virus integrase and their effects on enzymatic activity. *J Biol Chem* **272**, 18161-18168.

Calin, G.A. & Croce, C.M. (2007). Chromosomal rearrangements and microRNAs: a new cancer link with clinical implications. *J Clin Invest* **117**, 2059-2066.

Calin, G.A., Sevignani, C., Dumitru, C.D., Hyslop, T., Noch, E., Yendamuri, S., Shimizu, M., Rattan, S., Bullrich, F., Negrini, M. et al. (2004). Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *Proc Natl Acad Sci USA* **101**, 2999-3004.

Callebaut, I. & 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. *Cell Mol Life Sci* **54**, 880-891.

Calogero, S., Grassi, F., Aguzzi, A., Voigtlander, T., Ferrier, P., Ferrari, S., & Bianchi, M.E. (1999). The lack of chromosomal protein Hmg1 does not disrupt cell growth but causes lethal hypoglycaemia in newborn mice. *Nat Genet* **22**, 276-280.

Casali, P., Pal, Z., Xu, Z., & Zan, H. (2006). DNA repair in antibody somatic hypermutation. *Trends Immunol* **27**, 313-321.

Cereghini, S. & Yaniv, M. (1984). Assembly of transfected DNA into chromatin: structural changes in the origin-promoter-enhancer region upon replication. *EMBO J* **3**, 1243-1253.

Chaumeil, J., Micsinai, M., Ntziachristos, P., Roth, D.B., Aifantis, I., Kluger, Y., Deriano, L., & Skok, J.A. (2013). The RAG2 C-terminus and ATM protect genome integrity by controlling antigen receptor gene cleavage. *Nat Commun* **4**, 2231.

Chen, H.T., Bhandoola, A., Difilippantonio, M.J., Zhu, J., Brown, M.J., Tai, X., Rogakou, E.P., Brotz, T.M., Bonner, W.M., Ried, T. et al. (2000). Response to RAG-mediated VDJ cleavage by NBS1 and gamma-H2AX. *Science* **290**, 1962-1965.

Cheroutre, H. & Husain, M.M. (2013). CD4 CTL: living up to the challenge. *Semin Immunol* **25**, 273-281.

Chowdhury, D. & Sen, R. (2004). Regulation of immunoglobulin heavy-chain gene rearrangements. *Immunol Rev* **200**, 182-196.

Ciubotaru, M., Ptaszek, L.M., Baker, G.A., Baker, S.N., Bright, F.V., & Schatz, D.G. (2003). RAG1-DNA binding in V(D)J recombination. Specificity and DNA-induced conformational changes revealed by fluorescence and CD spectroscopy. *J Biol Chem* **278**, 5584-5596.

Ciubotaru, M., Trexler, A.J., Spiridon, L.N., Surleac, M.D., Rhoades, E., Petrescu, A.J., & Schatz, D.G. (2013). RAG and HMGB1 create a large bend in the 23RSS in the V(D)J recombination synaptic complexes. *Nucleic Acids Res* **41**, 2437-2454.

Clark, M.R., Mandal, M., Ochiai, K., & Singh, H. (2014). Orchestrating B cell lymphopoiesis through interplay of IL-7 receptor and pre-B cell receptor signalling. *Nat Rev Immunol* **14**, 69-80.

Corneo, B., Wendland, R.L., Deriano, L., Cui, X., Klein, I.A., Wong, S.Y., Arnal, S., Holub, A.J., Weller, G.R., Pancake, B.A. et al. (2007). Rag mutations reveal robust alternative end joining. *Nature* **449**, 483-486.

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

Cui, X. & Meek, K. (2007). Linking double-stranded DNA breaks to the recombination activating gene complex directs repair to the nonhomologous end-joining pathway. *Proc Natl Acad Sci USA* **104**, 17046-17051.

Cuomo, C.A., Mundy, C.L., & Oettinger, M.A. (1996). DNA sequence and structure requirements for cleavage of V(D)J recombination signal sequences. *Mol Cell Biol* **16**, 5683-5690.

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

Curry, J.D., Geier, J.K., & Schlissel, M.S. (2005). Single-strand recombination signal sequence nicks in vivo: evidence for a capture model of synapsis. *Nat Immunol* **6**, 1272-1279.

Curry, J.D., Schulz, D., Guidos, C.J., Danska, J.S., Nutter, L., Nussenzweig, A., & Schlissel, M.S. (2007). Chromosomal reinsertion of broken RSS ends during T cell development. *J Exp Med* **204**, 2293-2303.

Dalla-Favera, R., Bregni, M., Erikson, J., Patterson, D., Gallo, R.C., & Croce, C.M. (1982). Human c-myc onc gene is located on the region of chromosome 8 that is translocated in Burkitt lymphoma cells. *Proc Natl Acad Sci USA* **79**, 7824-7827.

Davies, D.R., Goryshin, I.Y., Reznikoff, W.S., & Rayment, I. (2000). Three-dimensional structure of the Tn5 synaptic complex transposition intermediate. *Science* **289**, 77-85.

De Silva, N.S., Simonetti, G., Heise, N., & Klein, U. (2012). The diverse roles of IRF4 in late germinal center B-cell differentiation. *Immunol Rev* **247**, 73-92.

Degner, S.C., Verma-Gaur, J., Wong, T.P., Bossen, C., Iverson, G.M., Torkamani, A., Vettermann, C., Lin, Y.C., Ju, Z., Schulz, D. et al. (2011). CCCTC-binding factor (CTCF) and cohesin influence the genomic architecture of the Igh locus and antisense transcription in pro-B cells. *Proc Natl Acad Sci USA* **108**, 9566-9571.

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

Deriano, L. & Roth, D.B. (2013). Modernizing the nonhomologous end-joining repertoire: alternative and classical NHEJ share the stage. *Annu Rev Genet* **47**, 433-455.

Di Noia, J.M. & Neuberger, M.S. (2007). Molecular mechanisms of antibody somatic hypermutation. *Annu Rev Biochem* **76**, 1-22.

Difilippantonio, M.J., McMahan, C.J., Eastman, Q.M., Spanopoulou, E., & Schatz, D.G. (1996). RAG1 mediates signal sequence recognition and recruitment of RAG2 in V(D)J recombination. *Cell* **87**, 253-262.

Drejer-Teel, A.H., Fugmann, S.D., & Schatz, D.G. (2007). The beyond 12/23 restriction is imposed at the nicking and pairing steps of DNA cleavage during V(D)J recombination. *Mol Cell Biol* **27**, 6288-6299.

Duber, S., Engel, H., Rolink, A., Kretschmer, K., & Weiss, S. (2003). Germline transcripts of immunoglobulin light chain variable regions are structurally diverse and differentially expressed. *Mol Immunol* **40**, 509-516.

Duquette, M.L., Pham, P., Goodman, M.F., & Maizels, N. (2005). AID binds to transcription-induced structures in c-MYC that map to regions associated with translocation and hypermutation. *Oncogene* **24**, 5791-5798.

Earl, P.L., Moss, B., Wyatt, L.S., & Carroll, M.W. (2001). Generation of recombinant vaccinia viruses. *Curr Protoc Protein Sci* **Chapter 5**, Unit5 13.

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

Eisenbeis, C.F., Singh, H., & Storb, U. (1993). PU.1 is a component of a multiprotein complex which binds an essential site in the murine immunoglobulin lambda 2-4 enhancer. *Mol Cell Biol* **13**, 6452-6461.

Elkin, S.K., Ivanov, D., Ewalt, M., Ferguson, C.G., Hyberts, S.G., Sun, Z.Y., Prestwich, G.D., Yuan, J., Wagner, G., Oettinger, M.A. et al. (2005). A PHD finger motif in the C terminus of RAG2 modulates recombination activity. *J Biol Chem* **280**, 28701-28710.

Elkin, S.K., Matthews, A.G., & Oettinger, M.A. (2003). The C-terminal portion of RAG2 protects against transposition in vitro. *EMBO J* **22**, 1931-1938.

Ellington, A. & Pollard, J.D., Jr. (2001). Purification of oligonucleotides using denaturing polyacrylamide gel electrophoresis. *Curr Protoc Mol Biol* **Chapter 2**, Unit2 12.

Elliott, B., Richardson, C., & Jasin, M. (2005). Chromosomal translocation mechanisms at intronic alu elements in mammalian cells. *Mol Cell* **17**, 885-894.

Elroy-Stein, O. & Moss, B. (2001). Gene expression using the vaccinia virus/T7 RNA polymerase hybrid system. *Curr Protoc Mol Biol* **Chapter 16**, Unit16 19.

Engel, H., Rolink, A., & Weiss, S. (1999). B cells are programmed to activate kappa and lambda for rearrangement at consecutive developmental stages. *Eur J Immunol* **29**, 2167-2176.

Espinoza, C.R. & Feeney, A.J. (2005). The extent of histone acetylation correlates with the differential rearrangement frequency of individual VH genes in pro-B cells. *J Immunol* **175**, 6668-6675.

Featherstone, K., Wood, A.L., Bowen, A.J., & Corcoran, A.E. (2010). The mouse immunoglobulin heavy chain V-D intergenic sequence contains insulators that may regulate ordered V(D)J recombination. *J Biol Chem* **285**, 9327-9338.

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. et al. (2002). Gene expression signatures define novel oncogenic pathways in T cell acute lymphoblastic leukemia. *Cancer Cell* **1**, 75-87.

Fitzsimmons, S.P., Bernstein, R.M., Max, E.E., Skok, J.A., & Shapiro, M.A. (2007). Dynamic changes in accessibility, nuclear positioning, recombination, and transcription at the Ig kappa locus. *J Immunol* **179**, 5264-5273.

Franchini, D.M., Benoukraf, T., Jaeger, S., Ferrier, P., & Payet-Bornet, D. (2009). Initiation of V(D)J recombination by Dbeta-associated recombination signal sequences: a critical control point in TCRbeta gene assembly. *PLoS ONE* **4**, e4575.

Fugmann, S.D., Lee, A.I., Shockett, P.E., Villey, I.J., & Schatz, D.G. (2000a). The RAG proteins and V(D)J recombination: complexes, ends, and transposition. *Annu Rev Immunol* **18**, 495-527.

Fugmann, S.D. & Schatz, D.G. (2001). Identification of basic residues in RAG2 critical for DNA binding by the RAG1-RAG2 complex. *Mol Cell* **8**, 899-910.

Fugmann, S.D., Villey, I.J., Ptaszek, L.M., & Schatz, D.G. (2000b). Identification of two catalytic residues in RAG1 that define a single active site within the RAG1/RAG2 protein complex. *Mol Cell* **5**, 97-107.

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

Gay, D., Saunders, T., Camper, S., & Weigert, M. (1993). Receptor editing: an approach by autoreactive B cells to escape tolerance. *J Exp Med* **177**, 999-1008.

Geier, J.K. & Schlissel, M.S. (2006). Pre-BCR signals and the control of Ig gene rearrangements. *Semin Immunol* **18**, 31-39.

Gellert, M. (2002). V(D)J recombination: RAG proteins, repair factors, and regulation. *Annu Rev Biochem* **71**, 101-132.

Gigi, V., Lewis, S., Shestova, O., Mijuskovic, M., Deriano, L., Meng, W., Luning Prak, E.T., & Roth, D.B. (2014). RAG2 mutants alter DSB repair pathway choice in vivo and illuminate the nature of 'alternative NHEJ'. *Nucleic Acids Res* **42**, 6352-6364.

Golding, A., Chandler, S., Ballestar, E., Wolffe, A.P., & Schlissel, M.S. (1999). Nucleosome structure completely inhibits in vitro cleavage by the V(D)J recombinase. *Embo J* **18**, 3712-3723.

Goldmit, M., Ji, Y., Skok, J., Roldan, E., Jung, S., Cedar, H., & Bergman, Y. (2005). Epigenetic ontogeny of the Igk locus during B cell development. *Nat Immunol* **6**, 198-203.

Gordon, S.J., Saleque, S., & Birshtein, B.K. (2003). Yin Yang 1 is a lipopolysaccharide-inducible activator of the murine 3' Igh enhancer, hs3. *J Immunol* **170**, 5549-5557.

Gorman, J.R. & Alt, F.W. (1998). Regulation of immunoglobulin light chain isotype expression. *Adv Immunol* **69**, 113-181.

Gorman, J.R., van der Stoep, N., Monroe, R., Cogne, M., Davidson, L., & Alt, F.W. (1996). The Ig(kappa) enhancer influences the ratio of Ig(kappa) versus Ig(lambda) B lymphocytes. *Immunity* **5**, 241-252.

Gottlieb, T.M. & Jackson, S.P. (1993). The DNA-dependent protein kinase: requirement for DNA ends and association with Ku antigen. *Cell* **72**, 131-142.

Grawunder, U., Leu, T.M., Schatz, D.G., Werner, A., Rolink, A.G., Melchers, F., & Winkler, T.H. (1995). Down-regulation of RAG1 and RAG2 gene expression in preB cells after functional immunoglobulin heavy chain rearrangement. *Immunity* **3**, 601-608.

Grazini, U., Zanardi, F., Citterio, E., Casola, S., Goding, C.R., & McBlane, F. (2010). The RING domain of RAG1 ubiquitylates histone H3: a novel activity in chromatin-mediated regulation of V(D)J joining. *Mol Cell* **37**, 282-293.

Grundy, G.J., Hesse, J.E., & Gellert, M. (2007). Requirements for DNA hairpin formation by RAG1/2. *Proc Natl Acad Sci USA* **104**, 3078-3083.

Grundy, G.J., Ramon-Maiques, S., Dimitriadis, E.K., Kotova, S., Biertumpfel, C., Heymann, J.B., Steven, A.C., Gellert, M., & Yang, W. (2009). Initial Stages of V(D)J Recombination: The Organization of RAG1/2 and RSS DNA in the Postcleavage Complex. *Mol Cell* **35**, 217-227.

Grundy, G.J., Yang, W., & Gellert, M. (2010). Autoinhibition of DNA cleavage mediated by RAG1 and RAG2 is overcome by an epigenetic signal in V(D)J recombination. *Proc Natl Acad Sci USA* **107**, 22487-22492.

Gu, Y., Seidl, K.J., Rathbun, G.A., Zhu, C., Manis, J.P., van der Stoep, N., Davidson, L., Cheng, H.L., Sekiguchi, J.M., Frank, K. et al. (1997). Growth retardation and leaky SCID phenotype of Ku70-deficient mice. *Immunity* **7**, 653-665.

Guo, C., Gerasimova, T., Hao, H., Ivanova, I., Chakraborty, T., Selimyan, R., Oltz, E.M., & Sen, R. (2011a). Two forms of loops generate the chromatin conformation of the immunoglobulin heavy-chain gene locus. *Cell* **147**, 332-343.

Guo, C., Yoon, H.S., Franklin, A., Jain, S., Ebert, A., Cheng, H.L., Hansen, E., Despo, O., Bossen, C., Vettermann, C. et al. (2011b). CTCF-binding elements mediate control of V(D)J recombination. *Nature* **477**, 424-430.

Han, J.O., Steen, S.B., & Roth, D.B. (1999). Intermolecular V(D)J recombination is prohibited specifically at the joining step. *Mol Cell* **3**, 331-338.

Haque, S.F., Bevington, S.L., & Boyes, J. (2013). The Eλ Enhancer is Essential for V(D)J Recombination of the Murine Immunoglobulin Lambda Light Chain Locus. *Biochem Biophys Res Commun.*

Hazenberg, M.D., Verschuren, M.C., Hamann, D., Miedema, F., & van Dongen, J.J. (2001). T cell receptor excision circles as markers for recent thymic emigrants: basic aspects, technical approach, and guidelines for interpretation. *J Mol Med (Berl)* **79**, 631-640.

Herzog, S., Hug, E., Meixlsperger, S., Paik, J.H., DePinho, R.A., Reth, M., & Jumaa, H. (2008). SLP-65 regulates immunoglobulin light chain gene recombination through the PI(3)K-PKB-Foxo pathway. *Nat Immunol* **9**, 623-631.

Hesse, J.E., Lieber, M.R., Gellert, M., & Mizuuchi, K. (1987). Extrachromosomal DNA substrates in pre-B cells undergo inversion or deletion at immunoglobulin V-(D)-J joining signals. *Cell* **49**, 775-783.

Hesslein, D.G., Pflugh, D.L., Chowdhury, D., Bothwell, A.L., Sen, R., & Schatz, D.G. (2003). Pax5 is required for recombination of transcribed, acetylated, 5' IgH V gene segments. *Genes Dev* **17**, 37-42.

Hiom, K. & Gellert, M. (1998). Assembly of a 12/23 paired signal complex: a critical control point in V(D)J recombination. *Mol Cell* **1**, 1011-1019.

Hiom, K., Melek, M., & Gellert, M. (1998). DNA transposition by the RAG1 and RAG2 proteins: a possible source of oncogenic translocations. *Cell* **94**, 463-470.

Hirt, B. (1967). Selective extraction of polyoma DNA from infected mouse cell cultures. *J Mol Biol* **26**, 365-369.

Holtmeier, W. & Kabelitz, D. (2005). gammadelta T cells link innate and adaptive immune responses. *Chem Immunol Allergy* **86**, 151-183.

Huang, J. & Muegge, K. (2001). Control of chromatin accessibility for V(D)J recombination by interleukin-7. *J Leukoc Biol* **69**, 907-911.

Huye, L.E., Purugganan, M.M., Jiang, M.M., & 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. *Mol Cell Biol* **22**, 3460-3473.

Ichihara, Y., Matsuoka, H., & Kurosawa, Y. (1988). Organization of human immunoglobulin heavy chain diversity gene loci. *EMBO J* **7**, 4141-4150.

Inlay, M., Alt, F.W., Baltimore, D., & Xu, Y. (2002). Essential roles of the kappa light chain intronic enhancer and 3' enhancer in kappa rearrangement and demethylation. *Nat Immunol* **3**, 463-468.

Jackson, A.M. & Krangel, M.S. (2006). Turning T-cell receptor beta recombination on and off: more questions than answers. *Immunol Rev* **209**, 129-141.

Jager, U., Bocskor, S., Le, T., Mitterbauer, G., Bolz, I., Chott, A., Kneba, M., Mannhalter, C., & Nadel, B. (2000). Follicular lymphomas' BCL-2/IgH junctions contain templated nucleotide insertions: novel insights into the mechanism of t(14;18) translocation. *Blood* **95**, 3520-3529.

Ji, Y., Resch, W., Corbett, E., Yamane, A., Casellas, R., & Schatz, D.G. (2010). The in vivo pattern of binding of RAG1 and RAG2 to antigen receptor loci. *Cell* **141**, 419-431.

Jiang, H., Chang, F.C., Ross, A.E., Lee, J., Nakayama, K., & Desiderio, S. (2005). Ubiquitylation of RAG-2 by Skp2-SCF links destruction of the V(D)J recombinase to the cell cycle. *Mol Cell* **18**, 699-709.

Jiang, H., Ross, A.E., & Desiderio, S. (2004). Cell cycle-dependent accumulation in vivo of transposition-competent complexes between recombination signal ends and full-length RAG proteins. *J Biol Chem* **279**, 8478-8486.

Johansen, F.E., Braathen, R., & Brandtzaeg, P. (2000). Role of J chain in secretory immunoglobulin formation. *Scand J Immunol* **52**, 240-248.

Johnson, K., Angelin-Duclos, C., Park, S., & Calame, K.L. (2003). Changes in histone acetylation are associated with differences in accessibility of V(H) gene segments to V-DJ recombination during B-cell ontogeny and development. *Mol Cell Biol* **23**, 2438-2450.

Johnson, K., Hashimshony, T., Sawai, C.M., Pongubala, J.M., Skok, J.A., Aifantis, I., & Singh, H. (2008). Regulation of immunoglobulin light-chain recombination by the transcription factor IRF-4 and the attenuation of interleukin-7 signaling. *Immunity* **28**, 335-345.

Jones, J.M., Bhattacharyya, A., Simkus, C., Vallieres, B., Veenstra, T.D., & Zhou, M. (2011). The RAG1 V(D)J recombinase/ubiquitin ligase promotes ubiquitylation of acetylated, phosphorylated histone 3.3. *Immunol Lett* **136**, 156-162.

Jones, J.M. & Gellert, M. (2001). Intermediates in V(D)J recombination: a stable RAG1/2 complex sequesters cleaved RSS ends. *Proc Natl Acad Sci USA* **98**, 12926-12931.

Jones, J.M. & Gellert, M. (2002). Ordered assembly of the V(D)J synaptic complex ensures accurate recombination. *EMBO J* **21**, 4162-4171.

Jones, J.M. & Gellert, M. (2003). Autoubiquitylation of the V(D)J recombinase protein RAG1. *Proc Natl Acad Sci USA* **100**, 15446-15451.

Jones, J.M. & Simkus, C. (2009). The roles of the RAG1 and RAG2 "non-core" regions in V(D)J recombination and lymphocyte development. *Arch Immunol Ther Exp (Warsz)* **57**, 105-116.

Kaplan, M.H. (2013). Th9 cells: differentiation and disease. *Immunol Rev* **252**, 104-115.

Kassmeier, M.D., Mondal, K., Palmer, V.L., Raval, P., Kumar, S., Perry, G.A., Anderson, D.K., Ciborowski, P., Jackson, S., Xiong, Y. et al. (2012). VprBP binds full-length RAG1 and is required for B-cell development and V(D)J recombination fidelity. *EMBO J* **31**, 945-958.

Kelsoe, G. (1996). The germinal center: a crucible for lymphocyte selection. *Semin Immunol* **8**, 179-184.

Kibbe, W.A. (2007). OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Res* **35**, W43-46.

Kim, D.R., Dai, Y., Mundy, C.L., Yang, W., & Oettinger, M.A. (1999). Mutations of acidic residues in RAG1 define the active site of the V(D)J recombinase. *Genes Dev* **13**, 3070-3080.

Kirch, S.A., Sudarsanam, P., & Oettinger, M.A. (1996). Regions of RAG1 protein critical for V(D)J recombination. *Eur J Immunol* **26**, 886-891.

Kirsch, I.R., Morton, C.C., Nakahara, K., & Leder, P. (1982). Human immunoglobulin heavy chain genes map to a region of translocations in malignant B lymphocytes. *Science* **216**, 301-303.

Korn, T., Bettelli, E., Oukka, M., & Kuchroo, V.K. (2009). IL-17 and Th17 Cells. *Annu Rev Immunol* **27**, 485-517.

Kosak, S.T., Skok, J.A., Medina, K.L., Riblet, R., Le Beau, M.M., Fisher, A.G., & Singh, H. (2002). Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development. *Science* **296**, 158-162.

Kovalchuk, A.L., Kim, J.S., Park, S.S., Coleman, A.E., Ward, J.M., Morse, H.C., 3rd, Kishimoto, T., Potter, M., & Janz, S. (2002). IL-6 transgenic mouse model for extraosseous plasmacytoma. *Proc Natl Acad Sci USA* **99**, 1509-1514.

Kramerov, D.A. & Vassetzky, N.S. (2005). Short retroposons in eukaryotic genomes. *Int Rev Cytol* **247**, 165-221.

Kriatchko, A.N., Anderson, D.K., & Swanson, P.C. (2006). Identification and characterization of a gain-of-function RAG-1 mutant. *Mol Cell Biol* **26**, 4712-4728.

Kridel, R., Sehn, L.H., & Gascoyne, R.D. (2012). Pathogenesis of follicular lymphoma. *J Clin Invest* **122**, 3424-3431.

Kumar, S. & Swanson, P.C. (2009). Full-length RAG1 promotes contact with coding and intersignal sequences in RAG protein complexes bound to recombination signals paired in cis. *Nucleic Acids Res* **37**, 2211-2226.

Kuppers, R. (2005). Mechanisms of B-cell lymphoma pathogenesis. *Nat Rev Cancer* **5**, 251-262.

Kuppers, R. & Dalla-Favera, R. (2001). Mechanisms of chromosomal translocations in B cell lymphomas. *Oncogene* **20**, 5580-5594.

Kurahashi, H., Inagaki, H., Ohye, T., Kogo, H., Kato, T., & Emanuel, B.S. (2006). Palindrome-mediated chromosomal translocations in humans. *DNA Repair (Amst)* **5**, 1136-1145.

Kwon, J., Imbalzano, A.N., Matthews, A., & Oettinger, M.A. (1998). Accessibility of nucleosomal DNA to V(D)J cleavage is modulated by RSS positioning and HMG1. *Mol Cell* **2**, 829-839.

Landree, M.A., Wibbenmeyer, J.A., & 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 Dev* **13**, 3059-3069.

Lee, G.S., Neiditch, M.B., Salus, S.S., & 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**, 171-184.

Lennon, G.G. & Perry, R.P. (1985). C mu-containing transcripts initiate heterogeneously within the IgH enhancer region and contain a novel 5'-nontranslatable exon. *Nature* **318**, 475-478.

Lewis, S.M. (1994). The mechanism of V(D)J joining: lessons from molecular, immunological, and comparative analyses. *Adv Immunol* **56**, 27-150.

Lewis, S.M., Agard, E., Suh, S., & Czyzyk, L. (1997). Cryptic signals and the fidelity of V(D)J joining. *Mol Cell Biol* **17**, 3125-3136.

Li, Z., Dordai, D.I., Lee, J., & 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**, 575-589.

Li, Z., Woo, C.J., Iglesias-Ussel, M.D., Ronai, D., & Scharff, M.D. (2004). The generation of antibody diversity through somatic hypermutation and class switch recombination. *Genes Dev* **18**, 1-11.

Lieber, M.R. (2010). The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu Rev Biochem* **79**, 181-211.

Lieber, M.R., Hesse, J.E., Mizuuchi, K., & Gellert, M. (1988). Studies of V(D)J recombination with extrachromosomal substrates. *Curr Top Microbiol Immunol* **137**, 94-99.

Little, A.J., Corbett, E., Ortega, F., & Schatz, D.G. (2013). Cooperative recruitment of HMGB1 during V(D)J recombination through interactions with RAG1 and DNA. *Nucleic Acids Res* **41**, 3289-3301.

Liu, H., Schmidt-Supprian, M., Shi, Y., Hobeika, E., Barteneva, N., Jumaa, H., Pelanda, R., Reth, M., Skok, J., Rajewsky, K. et al. (2007a). Yin Yang 1 is a critical regulator of B-cell development. *Genes Dev* **21**, 1179-1189.

Liu, Y., Subrahmanyam, R., Chakraborty, T., Sen, R., & Desiderio, S. (2007b). A plant homeodomain in RAG-2 that binds Hypermethylated lysine 4 of histone H3 is necessary for efficient antigen-receptor-gene rearrangement. *Immunity* **27**, 561-571.

Livak, F. & Schatz, D.G. (1996). T-cell receptor alpha locus V(D)J recombination by-products are abundant in thymocytes and mature T cells. *Mol Cell Biol* **16**, 609-618.

Livak, F. & Schatz, D.G. (1997). Identification of V(D)J recombination coding end intermediates in normal thymocytes. *J Mol Biol* **267**, 1-9.

Lu, C.P., Sandoval, H., Brandt, V.L., Rice, P.A., & Roth, D.B. (2006). Amino acid residues in Rag1 crucial for DNA hairpin formation. *Nat Struct Mol Biol* **13**, 1010-1015.

Lu, H., Shimazaki, N., Raval, P., Gu, J., Watanabe, G., Schwarz, K., Swanson, P.C., & Lieber, M.R. (2008). A biochemically defined system for coding joint formation in V(D)J recombination. *Mol Cell* **31**, 485-497.

Lu, L., Chaudhury, P., & Osmond, D.G. (1999). Regulation of cell survival during B lymphopoiesis: apoptosis and Bcl-2/Bax content of precursor B cells in bone marrow of mice with altered expression of IL-7 and recombinase-activating gene-2. *J Immunol* **162**, 1931-1940.

Ma, C.S. & Deenick, E.K. (2014). Human T follicular helper (Tfh) cells and disease. *Immunol Cell Biol* **92**, 64-71.

Ma, Y., Pannicke, U., Schwarz, K., & Lieber, M.R. (2002). Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination. *Cell* **108**, 781-794.

MacLennan, I.C. (2005). Germinal centers still hold secrets. *Immunity* **22**, 656-657.

Maes, J., O'Neill, L.P., Cavelier, P., Turner, B.M., Rougeon, F., & Goodhardt, M. (2001). Chromatin remodeling at the Ig loci prior to V(D)J recombination. *J Immunol* **167**, 866-874.

Malarkey, C.S. & Churchill, M.E. (2012). The high mobility group box: the ultimate utility player of a cell. *Trends Biochem Sci* **37**, 553-562.

Marculescu, R., Le, T., Simon, P., Jaeger, U., & Nadel, B. (2002). V(D)J-mediated translocations in lymphoid neoplasms: a functional assessment of genomic instability by cryptic sites. *J Exp Med* **195**, 85-98.

Mathieu, N., Hempel, W.M., Spicuglia, S., Verthuy, C., & Ferrier, P. (2000). Chromatin remodeling by the T cell receptor (TCR)-beta gene enhancer during early T cell development: Implications for the control of TCR-beta locus recombination. *J Exp Med* **192**, 625-636.

Matthews, A.G., Kuo, A.J., Ramon-Maiques, S., Han, S., Champagne, K.S., Ivanov, D., Gallardo, M., Carney, D., Cheung, P., Ciccone, D.N. et al. (2007). RAG2 PHD finger couples histone H3 lysine 4 trimethylation with V(D)J recombination. *Nature* **450**, 1106-1110.

McBlane, F. & Boyes, J. (2000). Stimulation of V(D)J recombination by histone acetylation. *Curr Biol* **10**, 483-486.

McBlane, J.F., van Gent, D.C., Ramsden, D.A., Romeo, C., Cuomo, C.A., Gellert, M., & 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**, 387-395.

McMurry, M.T. & Krangel, M.S. (2000). A role for histone acetylation in the developmental regulation of VDJ recombination [see comments]. *Science* **287**, 495-498.

Melek, M., Jones, J.M., O'Dea, M.H., Pais, G., Burke, T.R., Jr., Pommier, Y., Neamati, N., & Gellert, M. (2002). Effect of HIV integrase inhibitors on the RAG1/2 recombinase. *Proc Natl Acad Sci USA* **99**, 134-137.

Messier, T.L., O'Neill, J.P., & Finette, B.A. (2006). V(D)J recombinase mediated inter-chromosomal HPRT alterations at cryptic recombination signal sequences in peripheral human T cells. *Hum Mutat* **27**, 829.

Messier, T.L., O'Neill, J.P., Hou, S.M., Nicklas, J.A., & Finette, B.A. (2003). In vivo transposition mediated by V(D)J recombinase in human T lymphocytes. *EMBO J* **22**, 1381-1388.

Mestecky, J., Zikan, J., & Butler, W.T. (1971). Immunoglobulin M and secretory immunoglobulin A: presence of a common polypeptide chain different from light chains. *Science* **171**, 1163-1165.

Milne, C.D., Fleming, H.E., & Paige, C.J. (2004). IL-7 does not prevent pro-B/pre-B cell maturation to the immature/sIgM(+) stage. *Eur J Immunol* **34**, 2647-2655.

Mizushima, S. & Nagata, S. (1990). pEF-BOS, a powerful mammalian expression vector. *Nucleic Acids Res* **18**, 5322.

Mo, X., Bailin, T., & Sadofsky, M.J. (1999). RAG1 and RAG2 cooperate in specific binding to the recombination signal sequence in vitro. *J Biol Chem* **274**, 7025-7031.

Morshead, K.B., Ciccone, D.N., Taverna, S.D., Allis, C.D., & 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. *Proc Natl Acad Sci USA* **100**, 11577-11582.

Moss, B., Elroy-Stein, O., Mizukami, T., Alexander, W.A., & Fuerst, T.R. (1990). Product review. New mammalian expression vectors. *Nature* **348**, 91-92.

Mucida, D., Husain, M.M., Muroi, S., van Wijk, F., Shinnakasu, R., Naoe, Y., Reis, B.S., Huang, Y., Lambolez, F., Docherty, M. et al. (2013). Transcriptional reprogramming of mature CD4(+) helper T cells generates distinct MHC class II-restricted cytotoxic T lymphocytes. *Nat Immunol* **14**, 281-289.

Muljo, S.A. & Schlissel, M.S. (2003). A small molecule Abl kinase inhibitor induces differentiation of Abelson virus-transformed pre-B cell lines. *Nat Immunol* **4**, 31-37.

Mundy, C.L., Patenge, N., Matthews, A.G., & Oettinger, M.A. (2002). Assembly of the RAG1/RAG2 synaptic complex. *Mol Cell Biol* **22**, 69-77.

Murphy, K.M., Travers, P., & Walport, M., in *NCBI bookshelf* (Garland Pub., London, 2007).

Nagawa, F., Hirose, S., Nishizumi, H., Nishihara, T., & Sakano, H. (2004). Joining mutants of RAG1 and RAG2 that demonstrate impaired interactions with the coding-end DNA. *J Biol Chem* **279**, 38360-38368.

Nagawa, F., Ishiguro, K., Tsuboi, A., Yoshida, T., Ishikawa, A., Takemori, T., Otsuka, A.J., & Sakano, H. (1998a). Footprint analysis of the RAG protein recombination signal sequence complex for V(D)J type recombination. *Mol Cell Biol* **18**, 655-663.

Nagawa, F., Ishiguro, K., Tsuboi, A., Yoshida, T., Ishikawa, A., Takemori, T., Otsuka, A.J., & Sakano, H. (1998b). Footprint analysis of the RAG protein recombination signal sequence complex for V(D)J type recombination. *Mol Cell Biol* **18**, 655-663.

Nagawa, F., Kodama, M., Nishihara, T., Ishiguro, K., & Sakano, H. (2002). Footprint analysis of recombination signal sequences in the 12/23 synaptic complex of V(D)J recombination. *Mol Cell Biol* **22**, 7217-7225.

Nakamura, A., Sedelnikova, O.A., Redon, C., Pilch, D.R., Sinogeeva, N.I., Shroff, R., Lichten, M., & Bonner, W.M. (2006). Techniques for gamma-H2AX detection. *Methods Enzymol* **409**, 236-250.

Nambiar, M. & Raghavan, S.C. (2012). Mechanism of fragility at BCL2 gene minor breakpoint cluster region during t(14;18) chromosomal translocation. *J Biol Chem* **287**, 8688-8701.

Namen, A.E., Lupton, S., Hjerrild, K., Wignall, J., Mochizuki, D.Y., Schmierer, A., Mosley, B., March, C.J., Urdal, D., & Gillis, S. (1988). Stimulation of B-cell progenitors by cloned murine interleukin-7. *Nature* **333**, 571-573.

Neiditch, M.B., Lee, G.S., Huye, L.E., Brandt, V.L., & Roth, D.B. (2002). The V(D)J recombinase efficiently cleaves and transposes signal joints. *Mol Cell* **9**, 871-878.

Nelsen, B., Tian, G., Erman, B., Gregoire, J., Maki, R., Graves, B., & Sen, R. (1993). Regulation of lymphoid-specific immunoglobulin mu heavy chain gene enhancer by ETS-domain proteins. *Science* **261**, 82-86.

Ng, H.H., Robert, F., Young, R.A., & Struhl, K. (2003). Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. *Mol Cell* **11**, 709-719.

Niedziela-Majka, A., Rymarczyk, G., Kochman, M., & Ozyhar, A. (1998). GST-Induced dimerization of DNA-binding domains alters characteristics of their interaction with DNA. *Protein Expr Purif* **14**, 208-220.

Nightingale, K.P., Baumann, M., Eberharter, A., Mamais, A., Becker, P.B., & Boyes, J. (2007). Acetylation increases access of remodelling complexes to their nucleosome targets to enhance initiation of V(D)J recombination. *Nucleic Acids Res* **35**, 6311-6321.

Nishihara, T., Nagawa, F., Imai, T., & Sakano, H. (2008). RAG-heptamer interaction in the synaptic complex is a crucial biochemical checkpoint for the 12/23 recombination rule. *J Biol Chem* **283**, 4877-4885.

Nussenzweig, A. & Nussenzweig, M.C. (2010). Origin of chromosomal translocations in lymphoid cancer. *Cell* **141**, 27-38.

Nutt, S.L., Heavey, B., Rolink, A.G., & Busslinger, M. (1999). Commitment to the B-lymphoid lineage depends on the transcription factor Pax5. *Nature* **401**, 556-562.

Oettinger, M.A., Schatz, D.G., Gorka, C., & Baltimore, D. (1990). RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. *Science* **248**, 1517-1523.

Orphanides, G., Wu, W.H., Lane, W.S., Hampsey, M., & Reinberg, D. (1999). The chromatin-specific transcription elongation factor FACT comprises human SPT16 and SSRP1 proteins. *Nature* **400**, 284-288.

Papaemmanuil, E., Rapado, I., Li, Y., Potter, N.E., Wedge, D.C., Tubio, J., Alexandrov, L.B., Van Loo, P., Cooke, S.L., Marshall, J. et al. (2014). RAG-mediated recombination is the predominant driver of oncogenic rearrangement in ETV6-RUNX1 acute lymphoblastic leukemia. *Nat Genet* **46**, 116-125.

Pasqualucci, L., Migliazza, A., Fracchiolla, N., William, C., Neri, A., Baldini, L., Chaganti, R.S., Klein, U., Kuppers, R., Rajewsky, K. et al. (1998). BCL-6 mutations in normal germinal center B cells: evidence of somatic hypermutation acting outside Ig loci. *Proc Natl Acad Sci USA* **95**, 11816-11821.

Pasqualucci, L., Neumeister, P., Goossens, T., Nanjangud, G., Chaganti, R.S., Kuppers, R., & Dalla-Favera, R. (2001). Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. *Nature* **412**, 341-346.

Peak, M.M., Arbuckle, J.L., & Rodgers, K.K. (2003). The central domain of core RAG1 preferentially recognizes single-stranded recombination signal sequence heptamer. *J Biol Chem* **278**, 18235-18240.

Peled, J.U., Kuang, F.L., Iglesias-Ussel, M.D., Roa, S., Kalis, S.L., Goodman, M.F., & Scharff, M.D. (2008a). The biochemistry of somatic hypermutation. *Annu Rev Immunol* **26**, 481-511.

Peled, J.U., Kuang, F.L., Iglesias-Ussel, M.D., Roa, S., Kalis, S.L., Goodman, M.F., & Scharff, M.D. (2008b). The biochemistry of somatic hypermutation. *Annual Review of Immunology* **26**, 481-511.

Perkins, E.J., Kee, B.L., & Ramsden, D.A. (2004). Histone 3 lysine 4 methylation during the pre-B to immature B-cell transition. *Nucleic Acids Res* **32**, 1942-1947.

Pijlman, G.P., van den Born, E., Martens, D.E., & Vlak, J.M. (2001). Autographa californica baculoviruses with large genomic deletions are rapidly generated in infected insect cells. *Virology* **283**, 132-138.

Pijlman, G.P., van Schijndel, J.E., & Vlak, J.M. (2003). Spontaneous excision of BAC vector sequences from bacmid-derived baculovirus expression vectors upon passage in insect cells. *J Gen Virol* **84**, 2669-2678.

Pokholok, D.K., Harbison, C.T., Levine, S., Cole, M., Hannett, N.M., Lee, T.I., Bell, G.W., Walker, K., Rolfe, P.A., Herbolsheimer, E. et al. (2005). Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* **122**, 517-527.

Poltoratsky, V., Goodman, M.F., & Scharff, M.D. (2000). Error-prone candidates vie for somatic mutation. *J Exp Med* **192**, F27-30.

Potter, M. & Wiener, F. (1992). Plasmacytomagenesis in mice: model of neoplastic development dependent upon chromosomal translocations. *Carcinogenesis* **13**, 1681-1697.

Pui, C.H., Relling, M.V., & Downing, J.R. (2004). Acute lymphoblastic leukemia. *N Engl J Med* **350**, 1535-1548.

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. et al. (1999). Notch1 expression in early lymphopoiesis influences B versus T lineage determination. *Immunity* **11**, 299-308.

Qiu, J.X., Kale, S.B., Yarnell Schultz, H., & Roth, D.B. (2001). Separation-of-function mutants reveal critical roles for RAG2 in both the cleavage and joining steps of V(D)J recombination. *Mol Cell* **7**, 77-87.

Rabkin, C.S. & Janz, S. (2008). Mechanisms and consequences of chromosomal translocation. *Cancer Epidemiol Biomarkers Prev* **17**, 1849-1851.

Radtke, F., Wilson, A., Stark, G., Bauer, M., van Meerwijk, J., MacDonald, H.R., & Aguet, M. (1999). Deficient T cell fate specification in mice with an induced inactivation of Notch1. *Immunity* **10**, 547-558.

Raghavan, S.C., Houston, S., Hegde, B.G., Langen, R., Haworth, I.S., & Lieber, M.R. (2004a). Stability and strand asymmetry in the non-B DNA structure at the bcl-2 major breakpoint region. *J Biol Chem* **279**, 46213-46225.

Raghavan, S.C., Kirsch, I.R., & Lieber, M.R. (2001). Analysis of the V(D)J recombination efficiency at lymphoid chromosomal translocation breakpoints. *J Biol Chem* **276**, 29126-29133.

Raghavan, S.C., Swanson, P.C., Ma, Y., & 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. *Mol Cell Biol* **25**, 5904-5919.

Raghavan, S.C., Swanson, P.C., Wu, X., Hsieh, C.L., & Lieber, M.R. (2004b). A non-B-DNA structure at the Bcl-2 major breakpoint region is cleaved by the RAG complex. *Nature* **428**, 88-93.

Rajewsky, K. (1996). Clonal selection and learning in the antibody system. *Nature* **381**, 751-758.

Ramiro, A.R., Jankovic, M., Callen, E., Difilippantonio, S., Chen, H.T., McBride, K.M., Eisenreich, T.R., Chen, J., Dickins, R.A., Lowe, S.W. et al. (2006). Role of genomic instability and p53 in AID-induced c-myc-Igh translocations. *Nature* **440**, 105-109.

Ramon-Maiques, S., Kuo, A.J., Carney, D., Matthews, A.G., Oettinger, M.A., Gozani, O., & Yang, W. (2007). The plant homeodomain finger of RAG2 recognizes histone H3 methylated at both lysine-4 and arginine-2. *Proc Natl Acad Sci USA* **104**, 18993-18998.

Ramsden, D.A. & Gellert, M. (1995). Formation and resolution of double-strand break intermediates in V(D)J rearrangement. *Genes Dev* **9**, 2409-2420.

Ramsden, D.A., McBlane, J.F., van Gent, D.C., & Gellert, M. (1996). Distinct DNA sequence and structure requirements for the two steps of V(D)J recombination signal cleavage. *EMBO J* **15**, 3197-3206.

Ramsden, D.A., van Gent, D.C., & Gellert, M. (1997). Specificity in V(D)J recombination: new lessons from biochemistry and genetics. *Curr Opin Immunol* **9**, 114-120.

Reeves, R., Gorman, C.M., & Howard, B. (1985). Minichromosome assembly of non-integrated plasmid DNA transfected into mammalian cells. *Nucleic Acids Res* **13**, 3599-3615.

Rivera, R.R., Stuiver, M.H., Steenbergen, R., & Murre, C. (1993). Ets proteins: new factors that regulate immunoglobulin heavy-chain gene expression. *Mol Cell Biol* **13**, 7163-7169.

Robbiani, D.F., Bothmer, A., Callen, E., Reina-San-Martin, B., Dorsett, Y., Difilippantonio, S., Bolland, D.J., Chen, H.T., Corcoran, A.E., Nussenzweig, A. et al. (2008). AID is required for the chromosomal breaks in c-myc that lead to c-myc/IgH translocations. *Cell* **135**, 1028-1038.

Robbiani, D.F. & Nussenzweig, M.C. (2013). Chromosome translocation, B cell lymphoma, and activation-induced cytidine deaminase. *Annu Rev Pathol* **8**, 79-103.

Rodgers, K.K., Bu, Z., Fleming, K.G., Schatz, D.G., Engelman, D.M., & Coleman, J.E. (1996). A zinc-binding domain involved in the dimerization of RAG1. *J Mol Biol* **260**, 70-84.

Rodgers, K.K., Villey, I.J., Ptaszek, L., Corbett, E., Schatz, D.G., & Coleman, J.E. (1999). A dimer of the lymphoid protein RAG1 recognizes the recombination signal sequence and the complex stably incorporates the high mobility group protein HMG2. *Nucleic Acids Res* **27**, 2938-2946.

Rogakou, E.P., Boon, C., Redon, C., & Bonner, W.M. (1999). Megabase chromatin domains involved in DNA double-strand breaks in vivo. *J Cell Biol* **146**, 905-916.

Rogakou, E.P., Pilch, D.R., Orr, A.H., Ivanova, V.S., & Bonner, W.M. (1998). DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. *J Biol Chem* **273**, 5858-5868.

Roldan, E., Fuxa, M., Chong, W., Martinez, D., Novatchkova, M., Busslinger, M., & Skok, J.A. (2005). Locus 'decontraction' and centromeric recruitment contribute to allelic exclusion of the immunoglobulin heavy-chain gene. *Nat Immunol* **6**, 31-41.

Roth, D.B. (2003). Restraining the V(D)J recombinase. *Nat Rev Immunol* **3**, 656-666.

Roth, D.B. & Roth, S.Y. (2000). Unequal access: regulating V(D)J recombination through chromatin remodeling. *Cell* **103**, 699-702.

Sabbattini, P., Lundgren, M., Georgiou, A., Chow, C., Warnes, G., & Dillon, N. (2001). Binding of Ikaros to the lambda5 promoter silences transcription through a mechanism that does not require heterochromatin formation. *EMBO J* **20**, 2812-2822.

Sadofsky, M.J., Hesse, J.E., & Gellert, M. (1994). Definition of a core region of RAG-2 that is functional in V(D)J recombination. *Nucleic Acids Res* **22**, 1805-1809.

Sadofsky, M.J., Hesse, J.E., McBlane, J.F., & Gellert, M. (1993). Expression and V(D)J recombination activity of mutated RAG-1 proteins. *Nucleic Acids Res* **21**, 5644-5650.

Sakamoto, S., Wakae, K., Anzai, Y., Murai, K., Tamaki, N., Miyazaki, M., Miyazaki, K., Romanow, W.J., Ikawa, T., Kitamura, D. et al. (2012). E2A and CBP/p300 act in synergy to promote chromatin accessibility of the immunoglobulin kappa locus. *J Immunol* **188**, 5547-5560.

Santagata, S., Aidinis, V., & Spanopoulou, E. (1998). The effect of Me^{2+} cofactors at the initial stages of V(D)J recombination. *J Biol Chem* **273**, 16325-16331.

Sawchuk, D.J., Weis-Garcia, F., Malik, S., Besmer, E., Bustin, M., Nussenzweig, M.C., & Cortes, P. (1997). V(D)J recombination: modulation of RAG1 and RAG2 cleavage activity on 12/23 substrates by whole cell extract and DNA-bending proteins. *J Exp Med* **185**, 2025-2032.

Sayegh, C.E., Jhunjhunwala, S., Riblet, R., & Murre, C. (2005). Visualization of looping involving the immunoglobulin heavy-chain locus in developing B cells. *Genes Dev* **19**, 322-327.

Schatz, D.G. & Ji, Y. (2011). Recombination centres and the orchestration of V(D)J recombination. *Nat Rev Immunol* **11**, 251-263.

Schatz, D.G. & Swanson, P.C. (2011). V(D)J recombination: mechanisms of initiation. *Annu Rev Genet* **45**, 167-202.

Schlissel, M.S. (2004). Regulation of activation and recombination of the murine Igkappa locus. *Immunol Rev* **200**, 215-223.

Schlissel, M.S. & Baltimore, D. (1989). Activation of immunoglobulin kappa gene rearrangement correlates with induction of germline kappa gene transcription. *Cell* **58**, 1001-1007.

Sciammas, R. & Davis, M.M. (2004). Modular nature of Blimp-1 in the regulation of gene expression during B cell maturation. *J Immunol* **172**, 5427-5440.

Scott, J.N., Kupinski, A.P., & Boyes, J. (2014). Targeted genome regulation and modification using transcription activator-like effectors. *FEBS J.*

Shaffer, A.L., Emre, N.C., Lamy, L., Ngo, V.N., Wright, G., Xiao, W., Powell, J., Dave, S., Yu, X., Zhao, H. et al. (2008). IRF4 addiction in multiple myeloma. *Nature* **454**, 226-231.

Shaffer, A.L., Lin, K.I., Kuo, T.C., Yu, X., Hurt, E.M., Rosenwald, A., Giltnane, J.M., Yang, L., Zhao, H., Calame, K. et al. (2002). Blimp-1 orchestrates plasma cell differentiation by extinguishing the mature B cell gene expression program. *Immunity* **17**, 51-62.

Shaffer, A.L., Shapiro-Shelef, M., Iwakoshi, N.N., Lee, A.H., Qian, S.B., Zhao, H., Yu, X., Yang, L., Tan, B.K., Rosenwald, A. et al. (2004). XBP1, downstream of Blimp-1, expands the secretory apparatus and other organelles, and increases protein synthesis in plasma cell differentiation. *Immunity* **21**, 81-93.

Shapiro-Shelef, M., Lin, K.I., McHeyzer-Williams, L.J., Liao, J., McHeyzer-Williams, M.G., & Calame, K. (2003). Blimp-1 is required for the formation of immunoglobulin secreting plasma cells and pre-plasma memory B cells. *Immunity* **19**, 607-620.

Shimazaki, N., Askary, A., Swanson, P.C., & Lieber, M.R. (2012). Mechanistic basis for RAG discrimination between recombination sites and the off-target sites of human lymphomas. *Mol Cell Biol* **32**, 365-375.

Shimazaki, N., Tsai, A.G., & Lieber, M.R. (2009). H3K4me3 stimulates the V(D)J RAG complex for both nicking and hairpinning in trans in addition to tethering in cis: implications for translocations. *Mol Cell* **34**, 535-544.

Shlyakhtenko, L.S., Gilmore, J., Kriatchko, A.N., Kumar, S., Swanson, P.C., & Lyubchenko, Y.L. (2009). Molecular mechanism underlying RAG1/RAG2 synaptic complex formation. *J Biol Chem* **284**, 20956-20965.

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., & 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**, 1985-1989.

Sikes, M.L., Meade, A., Tripathi, R., Krangel, M.S., & Oltz, E.M. (2002). Regulation of V(D)J recombination: a dominant role for promoter positioning in gene segment accessibility. *Proc Natl Acad Sci U S A* **99**, 12309-12314.

Silver, D.P., Spanopoulou, E., Mulligan, R.C., & Baltimore, D. (1993). Dispensable sequence motifs in the RAG-1 and RAG-2 genes for plasmid V(D)J recombination. *Proc Natl Acad Sci USA* **90**, 6100-6104.

Simkus, C., Anand, P., Bhattacharyya, A., & Jones, J.M. (2007). Biochemical and folding defects in a RAG1 variant associated with Omenn syndrome. *J Immunol* **179**, 8332-8340.

Simkus, C., Makiya, M., & Jones, J.M. (2009). Karyopherin alpha 1 is a putative substrate of the RAG1 ubiquitin ligase. *Mol Immunol* **46**, 1319-1325.

Skok, J.A., Gisler, R., Novatchkova, M., Farmer, D., de Laat, W., & Busslinger, M. (2007). Reversible contraction by looping of the Tcra and Tcrb loci in rearranging thymocytes. *Nat Immunol* **8**, 378-387.

Sobacchi, C., Marrella, V., Rucci, F., Vezzoni, P., & Villa, A. (2006). RAG-dependent primary immunodeficiencies. *Hum Mutat* **27**, 1174-1184.

Spanopoulou, E., Cortes, P., Shih, C., Huang, C.M., Silver, D.P., Svec, P., & Baltimore, D. (1995). Localization, interaction, and RNA binding properties of the V(D)J recombination-activating proteins RAG1 and RAG2. *Immunity* **3**, 715-726.

Spanopoulou, E., Zaitseva, F., Wang, F.H., Santagata, S., Baltimore, D., & Panayotou, G. (1996). The homeodomain region of Rag-1 reveals the parallel mechanisms of bacterial and V(D)J recombination. *Cell* **87**, 263-276.

Stavnezer, J. (1996). Immunoglobulin class switching. *Curr Opin Immunol* **8**, 199-205.

Stavnezer, J., Guikema, J.E., & Schrader, C.E. (2008). Mechanism and regulation of class switch recombination. *Annu Rev Immunol* **26**, 261-292.

Steen, S.B., Gomelsky, L., & Roth, D.B. (1996). The 12/23 rule is enforced at the cleavage step of V(D)J recombination in vivo. *Genes Cells* **1**, 543-553.

Stros, M. (1998). DNA bending by the chromosomal protein HMG1 and its high mobility group box domains. Effect of flanking sequences. *J Biol Chem* **273**, 10355-10361.

Swanson, P.C. (2001). The DDE motif in RAG-1 is contributed in trans to a single active site that catalyzes the nicking and transesterification steps of V(D)J recombination. *Mol Cell Biol* **21**, 449-458.

Swanson, P.C. (2002a). Fine structure and activity of discrete RAG-HMG complexes on V(D)J recombination signals. *Mol Cell Biol* **22**, 1340-1351.

Swanson, P.C. (2002b). A RAG-1/RAG-2 tetramer supports 12/23-regulated synapsis, cleavage, and transposition of V(D)J recombination signals. *Mol Cell Biol* **22**, 7790-7801.

Swanson, P.C. (2004). The bounty of RAGs: recombination signal complexes and reaction outcomes. *Immunol Rev* **200**, 90-114.

Swanson, P.C. & Desiderio, S. (1998). V(D)J recombination signal recognition: distinct, overlapping DNA-protein contacts in complexes containing RAG1 with and without RAG2. *Immunity* **9**, 115-125.

Swanson, P.C. & Desiderio, S. (1999). RAG-2 promotes heptamer occupancy by RAG-1 in the assembly of a V(D)J initiation complex. *Mol Cell Biol* **19**, 3674-3683.

Swanson, P.C., Kumar, S., & Raval, P. (2009). Early steps of V(D)J rearrangement: insights from biochemical studies of RAG-RSS complexes. *Adv Exp Med Biol* **650**, 1-15.

Tabor, S. & Richardson, C.C. (1989). Effect of manganese ions on the incorporation of dideoxynucleotides by bacteriophage T7 DNA polymerase and Escherichia coli DNA polymerase I. *Proc Natl Acad Sci USA* **86**, 4076-4080.

Takeda, S., Zou, Y.R., Bluethmann, H., Kitamura, D., Muller, U., & Rajewsky, K. (1993). Deletion of the immunoglobulin kappa chain intron enhancer abolishes kappa chain gene rearrangement in cis but not lambda chain gene rearrangement in trans. *Embo J* **12**, 2329-2336.

Teitell, M.A. & Pandolfi, P.P. (2009). Molecular genetics of acute lymphoblastic leukemia. *Annu Rev Pathol* **4**, 175-198.

Teng, G. & Papavasiliou, F.N. (2007). Immunoglobulin somatic hypermutation. *Annu Rev Genet* **41**, 107-120.

Tevelev, A. & Schatz, D.G. (2000). Intermolecular V(D)J recombination. *J Biol Chem* **275**, 8341-8348.

Tiegs, S.L., Russell, D.M., & Nemazee, D. (1993). Receptor editing in self-reactive bone marrow B cells. *J Exp Med* **177**, 1009-1020.

Tonegawa, S. (1983). Somatic generation of antibody diversity. *Nature* **302**, 575-581.

Torrano, V., Procter, J., Cardus, P., Greaves, M., & Ford, A.M. (2011). ETV6-RUNX1 promotes survival of early B lineage progenitor cells via a dysregulated erythropoietin receptor. *Blood* **118**, 4910-4918.

Tripathi, R., Jackson, A., & Krangel, M.S. (2002). A change in the structure of Vbeta chromatin associated with TCR beta allelic exclusion. *J Immunol* **168**, 2316-2324.

Tsai, A.G. & Lieber, M.R. (2010). Mechanisms of chromosomal rearrangement in the human genome. *BMC Genomics* **11 Suppl 1**, S1.

Tsuzuki, S., Seto, M., Greaves, M., & Enver, T. (2004). Modeling first-hit functions of the t(12;21) TEL-AML1 translocation in mice. *Proc Natl Acad Sci USA* **101**, 8443-8448.

Tycko, B. & Sklar, J. (1990). Chromosomal translocations in lymphoid neoplasia: a reappraisal of the recombinase model. *Cancer Cells* **2**, 1-8.

Unniraman, S., Zhou, S., & Schatz, D.G. (2004). Identification of an AID-independent pathway for chromosomal translocations between the Igh switch region and Myc. *Nat Immunol* **5**, 1117-1123.

Uphoff, C.C., MacLeod, R.A., Denkmann, S.A., Golub, T.R., Borkhardt, A., Janssen, J.W., & Drexler, H.G. (1997). Occurrence of TEL-AML1 fusion resulting from (12;21) translocation in human early B-lineage leukemia cell lines. *Leukemia* **11**, 441-447.

van Gent, D.C., Hiom, K., Paull, T.T., & Gellert, M. (1997). Stimulation of V(D)J cleavage by high mobility group proteins. *EMBO J* **16**, 2665-2670.

van Gent, D.C., McBlane, J.F., Ramsden, D.A., Sadofsky, M.J., Hesse, J.E., & Gellert, M. (1996a). Initiation of V(D)J recombinations in a cell-free system by RAG1 and RAG2 proteins. *Curr Top Microbiol Immunol* **217**, 1-10.

van Gent, D.C., Mizuuchi, K., & Gellert, M. (1996b). Similarities between initiation of V(D)J recombination and retroviral integration. *Science* **271**, 1592-1594.

van Gent, D.C., Ramsden, D.A., & Gellert, M. (1996c). The RAG1 and RAG2 proteins establish the 12/23 rule in V(D)J recombination. *Cell* **85**, 107-113.

Vanasse, G.J., Concannon, P., & Willerford, D.M. (1999). Regulated genomic instability and neoplasia in the lymphoid lineage. *Blood* **94**, 3997-4010.

Vanura, K., Montpellier, B., Le, T., Spicuglia, S., Navarro, J.M., Cabaud, O., Roulland, S., Vachez, E., Prinz, I., Ferrier, P. et al. (2007). In vivo reinsertion of excised episomes by the V(D)J recombinase: a potential threat to genomic stability. *PLoS Biol* **5**, e43.

Vignali, D.A., Collison, L.W., & Workman, C.J. (2008). How regulatory T cells work. *Nat Rev Immunol* **8**, 523-532.

Wagner, H.J., Scott, R.S., Buchwald, D., & Sixbey, J.W. (2004). Peripheral blood lymphocytes express recombination-activating genes 1 and 2 during Epstein-Barr virus-induced infectious mononucleosis. *J Infect Dis* **190**, 979-984.

Wang, G., Dhar, K., Swanson, P.C., Levitus, M., & Chang, Y. (2012). Real-time monitoring of RAG-catalyzed DNA cleavage unveils dynamic changes in coding end association with the coding end complex. *Nucleic Acids Res* **40**, 6082-6096.

Wang, X., Xiao, G., Zhang, Y., Wen, X., Gao, X., Okada, S., & Liu, X. (2008). Regulation of Tcrb recombination ordering by c-Fos-dependent RAG deposition. *Nat Immunol* **9**, 794-801.

Weir, H.M., Kraulis, P.J., Hill, C.S., Raine, A.R., Laue, E.D., & Thomas, J.O. (1993). Structure of the HMG box motif in the B-domain of HMG1. *EMBO J* **12**, 1311-1319.

Welzel, N., Le, T., Marculescu, R., Mitterbauer, G., Chott, A., Pott, C., Kneba, M., Du, M.Q., Kusec, R., Drach, J. et al. (2001). Templated nucleotide addition and immunoglobulin JH-gene utilization in t(11;14) junctions: implications for the mechanism of translocation and the origin of mantle cell lymphoma. *Cancer Res* **61**, 1629-1636.

West, K.L., Singha, N.C., De Ioannes, P., Lacomis, L., Erdjument-Bromage, H., Tempst, P., & 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**, 203-212.

West, R.B. & Lieber, M.R. (1998). The RAG-HMG1 complex enforces the 12/23 rule of V(D)J recombination specifically at the double-hairpin formation step. *Mol Cell Biol* **18**, 6408-6415.

Wilson, A., Held, W., & MacDonald, H.R. (1994). Two waves of recombinase gene expression in developing thymocytes. *J Exp Med* **179**, 1355-1360.

Workman, J.L. (2006). Nucleosome displacement in transcription. *Genes Dev* **20**, 2009-2017.

Xu, C.R. & 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. *J Immunol* **182**, 1362-1369.

Xu, Y., Davidson, L., Alt, F.W., & Baltimore, D. (1996). Deletion of the Ig kappa light chain intronic enhancer/matrix attachment region impairs but does not abolish V kappa J kappa rearrangement. *Immunity* **4**, 377-385.

Yamane, H. & Paul, W.E. (2013). Early signaling events that underlie fate decisions of naive CD4(+) T cells toward distinct T-helper cell subsets. *Immunol Rev* **252**, 12-23.

Yancopoulos, G.D. & Alt, F.W. (1985). Developmentally controlled and tissue-specific expression of unrearranged VH gene segments. *Cell* **40**, 271-281.

Yarnell Schultz, H., Landree, M.A., Qiu, J.X., Kale, S.B., & Roth, D.B. (2001). Joining-deficient RAG1 mutants block V(D)J recombination in vivo and hairpin opening in vitro. *Mol Cell* **7**, 65-75.

Ye, S.K., Agata, Y., Lee, H.C., Kurooka, H., Kitamura, T., Shimizu, A., Honjo, T., & Ikuta, K. (2001). The IL-7 receptor controls the accessibility of the TCRgamma locus by Stat5 and histone acetylation. *Immunity* **15**, 813-823.

Yin, F.F., Bailey, S., Innis, C.A., Ciubotaru, M., Kamtekar, S., Steitz, T.A., & Schatz, D.G. (2009). Structure of the RAG1 nonamer binding domain with DNA reveals a dimer that mediates DNA synapsis. *Nat Struct Mol Biol* **16**, 499-508.

Yoshida, T., Tsuboi, A., Ishiguro, K., Nagawa, F., & Sakano, H. (2000). The DNA-bending protein, HMG1, is required for correct cleavage of 23 bp recombination signal sequences by recombination activating gene proteins in vitro. *Int Immunol* **12**, 721-729.

Yu, K. & Lieber, M.R. (1999). Mechanistic basis for coding end sequence effects in the initiation of V(D)J recombination. *Mol Cell Biol* **19**, 8094-8102.

Yu, K. & Lieber, M.R. (2000). The nicking step in V(D)J recombination is independent of synapsis: implications for the immune repertoire. *Mol Cell Biol* **20**, 7914-7921.

Zachau, H.G. (1993). The immunoglobulin kappa locus-or-what has been learned from looking closely at one-tenth of a percent of the human genome. *Gene* **135**, 167-173.

Zhang, L., Reynolds, T.L., Shan, X., & Desiderio, S. (2011). Coupling of V(D)J recombination to the cell cycle suppresses genomic instability and lymphoid tumorigenesis. *Immunity* **34**, 163-174.

Zhao, S., Gwyn, L.M., De, P., & Rodgers, K.K. (2009). A non-sequence-specific DNA binding mode of RAG1 is inhibited by RAG2. *J Mol Biol* **387**, 744-758.

Zhu, C., Bogue, M.A., & Roth, D.B. (1996). Thymocyte differentiation in gamma-irradiated severe-combined immunodeficient mice: characterization of intermediates and products of V(D)J recombination at the T cell receptor alpha locus. *Eur J Immunol* **26**, 2859-2865.

Zubakov, D., Liu, F., van Zelm, M.C., Vermeulen, J., Oostra, B.A., van Duijn, C.M., Driessen, G.J., van Dongen, J.J., Kayser, M., & Langerak, A.W. (2010). Estimating human age from T-cell DNA rearrangements. *Curr Biol* **20**, R970-971.