V(D)J Recombination: Mechanisms of Initiation

David G. Schatz1,* and Patrick C. Swanson2

1Department of Immunobiology and Howard Hughes Medical Institute, Yale University School of Medicine, New Haven, Connecticut 06520-8011; email: david.schatz@yale.edu
2Department of Medical Microbiology and Immunology, Creighton University Medical Center, Omaha, Nebraska 68178; email: patrickswanson@creighton.edu

Keywords
site-specific recombination, recombination signal sequence, RAG1, RAG2, HMGB1, protein-DNA complex

Abstract
V(D)J recombination assembles immunoglobulin and T cell receptor genes during lymphocyte development through a series of carefully orchestrated DNA breakage and rejoining events. DNA cleavage requires a series of protein-DNA complexes containing the RAG1 and RAG2 proteins and recombination signals that flank the recombining gene segments. In this review, we discuss recent advances in our understanding of the function and domain organization of the RAG proteins, the composition and structure of RAG-DNA complexes, and the pathways that lead to the formation of these complexes. We also consider the functional significance of RAG-mediated histone recognition and ubiquitin ligase activities, and the role played by RAG in ensuring proper repair of DNA breaks made during V(D)J recombination. Finally, we propose a model for the formation of RAG-DNA complexes that involves anchoring of RAG1 at the recombination signal nonamer and RAG2-dependent surveillance of adjoining DNA for suitable spacer and heptamer sequences.
INTRODUCTION
The discovery that immunoglobulin genes undergo somatic DNA rearrangement was a watershed moment in the field of immunology, stimulating research that continues to this day to identify and characterize the mechanisms that mediate and regulate this process. As the organization of the antigen receptor loci was revealed (see example in Figure 1b), it was recognized that functional V(D)J RSS and Tcrb locus organization. (a) The sequence, location, and numbering scheme of the heptamer and nonamer in a consensus 12RSS are shown overlaid on a triangle that represents the RSS orientation in subsequent figures. Most endogenous RSSs differ in sequence at one or more positions from the consensus, although positions 1–3 of the heptamer are almost perfectly conserved. The 23RSS contains 23 bp between the heptamer and nonamer. Spacer sequence is less well conserved than the heptamer or nonamer. Also indicated is the first nucleotide of the coding flank on the bottom strand (C1b), which is thought to be subject to base flipping prior to hairpin formation. (b) Organization of the Tcrb locus. The 12 and 23RSSs are identified by purple and yellow triangles, respectively, V, D, and J gene segments as green, blue, and gray rectangles, respectively, and constant regions as red rectangles. An AP1 site embedded in the 3′-Dβ1 23RSSs is indicated by a small red circle. The RSS composition and orientation for a representative gene segment within a given cluster are shown (e.g., all Jβ1 segments have a 5′ 12RSS); remaining RSSs are omitted for simplicity. Dβ-to-Jβ recombination (I) precedes Vβ-to-Dβ recombination (II). Despite being formally permissible by the 12/23 rule, direct Vβ-to-Jβ recombination is restricted by the “beyond 12/23 rule” (B12/23) (c) Sequence of the 3′-Dβ1 23RSS. The location of an embedded AP1 site is indicated by an overline; RAG-RSS heptamer-spacer contacts revealed by interference footprinting (130), which are conserved in the 3′-Dβ1 23-RSS and overlap with the putative API binding site, are identified above or below the nucleotide sequence.
rarrangements only occur between two gene segments with flanking recombination signal sequences (RSSs) that differ in the spacer length (either 12 or 23 bp) between the heptamer and nonamer elements (Figure 1a). This led to speculation that V(D)J recombination must involve the formation of one or more protein-DNA complexes that bridge the two RSSs to initiate the rearrangement process. The identification of the recombination activating genes, RAG1 and RAG2, and the eventual confirmation of their direct role in initiating V(D)J recombination by introducing site-specific DNA double-strand breaks (DSBs) spurred much additional work to dissect the molecular mechanisms of the reaction. This review discusses recent advances in our mechanistic understanding of how V(D)J recombination is initiated and how the RAG (recombination activating gene) proteins direct the repair of the DNA breaks they create. We do not provide detailed consideration of DNA repair mechanisms operative in V(D)J recombination, the evolutionary origins of the RAG proteins, or the mechanisms regulating the accessibility and positioning of antigen receptor loci during V(D)J recombination, but instead refer readers to other reviews on these topics (12, 46, 65, 86, 114). Likewise, to enable us to focus on more recent developments within each topic area, we will briefly overview earlier work in each of the major sections, referring readers to previous reviews for more in-depth coverage (31, 47, 51, 131).

OVERVIEW OF THE BIOCHEMICAL STEPS OF V(D)J RECOMBINATION

Based on extensive biochemical studies of the RAG proteins, characterization of V(D)J recombination intermediates, and identification of factors involved in the repair of these intermediates, one can conceptually divide V(D)J recombination into two distinct phases: a cleavage phase and a joining phase (47, 51) (Figure 2). During the cleavage phase, a 12RSS and a 23RSS are brought into close proximity by the RAG proteins to form a stable multi-subunit synaptic complex. Accumulating evidence suggests synaptic complex formation occurs through a stepwise capture model of assembly (28, 71, 98) that involves initial RAG binding and perhaps nicking of an RSS within a single RSS complex (SC), followed by subsequent capture of an appropriate partner RSS to form a synaptic or paired complex (PC) (Figure 2). Next, the RAG proteins nick the partner RSS and catalyze the coupled cleavage of both RSSs by direct transesterification using the 3′-OH exposed by nicking. This reaction yields a pair of blunt, 5′-phosphorylated signal ends and a pair of coding ends that terminate in covalently sealed DNA hairpin structures (92). After cleavage, the RAG proteins transiently retain the DNA ends in a cleaved signal complex (CSC) (61). Coding ends are thought to dissociate first, while the signal ends remain tightly bound to the RAG proteins in a signal end complex (SEC). Within these postcleavage complexes, the broken DNA transitions to the joining phase, during which the DNA ends are reorganized, processed, and repaired. Genetic and biochemical evidence suggests the catalytic subunit of DNA-dependent protein kinase (DNA-PKcs) and Artemis form a structure-specific endonuclease that resolves the DNA hairpin structures (89). Signal and coding ends are then processed and joined through the classical nonhomologous end joining (NHEJ) DNA repair pathway, involving at least five proteins, including Ku70, Ku80, XRCC4, DNA Ligase IV, and the Cernunnos/XLF protein (86). Members of the Pol X family of polymerases (TdT, polζ, and polη) contribute to the diversification and repair of coding joints through template-independent nucleotide addition and fill-in DNA synthesis (145). This process generally produces two types of joined products: a coding joint formed from joining the two gene coding segments together (often with small junctional nucleotide additions or deletions); and a signal joint formed by signal end ligation (often precisely) (Figure 2).
RAG1 AND RAG2: DOMAIN ORGANIZATION AND PURIFICATION

Protein Structure Overview

The RAG proteins are highly conserved among jawed vertebrate organisms. Because much of our understanding of RAG protein biochemistry has been gained using the murine RAG proteins, this review uses murine RAG1 and RAG2 amino acid numbering, unless otherwise noted. Murine RAG1 and RAG2 contain 1,040 and 527 amino acids, respectively (Figure 3). Substantial portions of both proteins can be removed without severely compromising V(D)J recombination activity on episomal or integrated substrates in cell culture (131 and references therein). The smallest functional truncation mutants, referred to as core RAG1 (384–1008) and RAG2 (1–387) (Figure 3), are more easily purified than their full-length counterparts, and hence have been (and continue to be) used in most studies of RAG protein biochemistry. However, the noncore regions of RAG1 and RAG2 clearly play critical roles in promoting physiological V(D)J recombination, because mice expressing only core forms of RAG1 or RAG2 exhibit defects in the efficiency, fidelity, and ordering of V(D)J rearrangement at endogenous loci (5, 29, 38, 85, 136). Successful purification of full-length RAG proteins has been reported more recently, providing important insights into the function of the noncore regions (43, 135, 140). Through extensive mutagenesis of both RAG proteins, a fairly...
A detailed picture of RAG1 and RAG2 structure-function relationships has emerged (Figure 3; see sidebar, RAG Mutants and Disease).

Full-length RAG1 contains distinct regions with different functional roles. The N-terminal noncore portion of RAG1 contains elements that regulate cellular protein levels (55, 93, 109), mediate interactions with several protein factors (25, 52, 90, 109), coordinate zinc ions, and enhance cellular V(D)J recombination activity (73, 93, 112). This region also includes a structurally characterized zinc-binding dimerization domain (ZDD) spanning residues 265–380 (110). The core portion of RAG1 can be subdivided into three distinct regions (31): (a) a structurally characterized nonamer binding domain (NBD) (residues 389–442) that forms a dimer and interacts primarily with the RSS nonamer (36, 126); (b) a central domain (residues 528–760) that exhibits nicking activity on ssDNA, which is enhanced by RAG2 (8), mediates heptamer contact (preferring single-stranded and nicked substrates) (103), and interacts with RAG2 partly though a C2H2 zinc finger domain (ZnF-B) (4); and (c) a C-terminal domain (CTD) (residues 761–979) that binds double-stranded DNA nonspecifically and cooperatively (8), mediates contact with the coding sequence flanking the RSS (97), and contains two zinc binding sites (57). In mixing experiments, the CTD is reported to suppress central domain ssDNA nicking activity, but not when substrates contain an ssDNA-dsDNA transition (153). However, this finding could be alternatively explained by competitive inhibition of DNA binding by the CTD on ssDNA substrates, whereas the lack of CTD inhibition using ssDNA-dsDNA substrates may reflect the stronger dsDNA binding preference of the CTD relative to the central domain. The far C terminus of RAG1 (residues 1009–1040) is reported to collaborate with the RAG2 C terminus to modulate RAG binding and cleavage activity (discussed below) (55).

Within RAG2, three key regions have been defined: (a) an N-terminal region comprising core RAG2 (1–387), containing six Kelch-like motifs (residues 1–350) proposed to adopt a six-bladed beta-propeller-like structure (21), the sixth of which reportedly mediates contacts with RAG1 (4); (b) a flexible acidic hinge region (residues 361–408); and (c) a C-terminal noncanonical plant homeodomain (PHD) (residues 414–487) shown to bind phosphoinositides (42) and histone H3 trimethylated at lysine 4 (hereafter termed H3K4me3) (87, 91). The far C terminus contains a conserved threonine residue (T490) that mediates phosphorylation-dependent and cell-regulated degradation of RAG2 via Skp2-SCF-mediated ubiquitination (68).

RAG MUTANTS AND DISEASE

The RAG proteins have been subjected to extensive mutagenesis and RAG mutants have been implicated in multiple disease entities. Early mutagenesis studies defined the catalytic core of the RAG1 and RAG2 proteins (51), and recent efforts have reduced the minimal region of RAG2 required to support levels of core RAG2 activity to residues 1–360 (24) (K. Dhar and P.C. Swanson, unpublished data). In addition, studies have defined the amino acid requirements for RAG enzymatic activity including catalytic (49, 76, 82), step-arrest (75, 105), joining-defective (62, 113), and gain-of-function mutants (78, 102), DNA binding activity (35, 48, 126, 132), RAG1-RAG2 interactions (4, 77, 94), and RAG interactions with other factors (3, 91, 109). Importantly, naturally occurring human RAG mutations have been identified that cause a complete or partial loss of physiological V(D)J recombination activity (see Related Resources and 103, 107) and are associated with a spectrum of severe immune deficiency disorders ranging from classical T-B-SCID to slightly milder variants, such as complete or incomplete Omenn syndrome, and RAG deficiency with γδ T cell expansion, granuloma formation, or maternofetal engraftment (103, 107). Representative RAG mutations are provided in Figure 3.
biochemical characterization without being appended to a fusion partner to enhance solubility (131). Bacterial, insect, and mammalian cell expression systems all yield RAG1 that is active in RSS binding and cleavage assays, whereas RAG2 is only recovered in highly active form from mammalian cells (131). However, coexpressed RAG1 and RAG2 recovered from insect or mammalian cells are active; the yield and activity of coexpressed RAG1 and RAG2 are reportedly higher than the individually expressed proteins (129). Early procedures for purifying core RAG1 or RAG2 relied on buffers containing high levels of NaCl (>0.5 M) (92). These conditions also permitted isolation of active full-length RAG1 coexposed with core RAG2 and vice-versa; however, RAG preparations containing full-length RAG1 or RAG2 did not bind substrate DNA uniformly (135). Milder purification conditions were subsequently identified that enable recovery of full-length RAG proteins with improved DNA binding characteristics and maintenance of protein-protein interactions (109).

Core RAG1 alone forms a stable dimer in solution under most in vitro conditions tested.
(10, 23, 133); higher-order oligomeric forms (tetramer and octamer) may be favored when under conditions of low ionic strength (0.2 M NaCl) and temperature (10°C) (32). RAG2 is reportedly purified as a mixture of monomeric and higher-order oligomeric forms, with the monomeric form predominating (10). Unlike core RAG1, core RAG2 by itself exhibits little, if any DNA binding activity (134). Core RAG1 and RAG2 associate with one another in solution without DNA, forming a 1:1 mixed tetramer complex (134). The tetramer configuration is further stabilized when the RAG2 C terminus is present (55). Based on this evidence, RAG1 and RAG2 most likely assemble a heterotetramer complex in vivo, but whether physiological synapsis and cleavage of RSS pairs in vivo are mediated by one or more (perhaps two) of these complexes remains unclear at present (discussed below).

**RAG-RSS SINGLE SITE AND SYNAPTIC COMPLEX ASSEMBLY AND ACTIVITY**

In this section, features of RAG-RSS single site and synaptic complexes are discussed, deferring for later sections a discussion of which RSS is targeted first by the RAG proteins and what mechanisms enforce the 12/23 rule and “beyond 12/23 rule” (B12/23) of V(D)J recombination. Much of what is known about RAG-RSS complex assembly and activity has been gained from studies using core RAG proteins and short oligonucleotide RSS substrates. These studies are not discussed in depth here, as most of this material has been reviewed previously (47, 51, 131, 134).

**SINGLE RAG-RSS COMPLEXES**

**RAG1-RSS Complexes**

Core RAG1 alone binds an isolated RSS as a dimer (134) with an apparent Kd ranging from 29–41 nM, as measured by fluorescence anisotropy (23, 153), to 92–114 nM, as measured by electrophoretic mobility shift assay (EMSA) (23, 111, 153). RAG1 subunits do not spontaneously reassort during the process of DNA binding (133). Minor higher-order oligomeric complexes of uncertain biological significance are also variably detected by EMSA but have not been extensively analyzed. RAG1 binding to the RSS induces conformational changes in RAG1 (23) and promotes DNA bending (3). DNA binding and footprinting studies of dimeric core RAG1-RSS complexes suggest that RSS binding is primarily mediated by the RAG1 NBD; the DNA interactions are...
Hairpin formation: second step of DNA cleavage by RAG in which the 3′-hydroxyl of the nicked strand attacks the other strand.

mostly centered on the nonamer and flanking spacer sequence and involve extensive contacts with the phosphodiester backbone (131). These findings are further substantiated and extended by recent structural analysis of the RAG1 NBD (residues 389–464) bound to a DNA duplex containing the 3′ end of the spacer (4 base pairs) and a consensus nonamer (146). In this complex, the NBD forms a symmetrical homo-dimer that is bound to two DNA molecules, with each of the two subunits contacting both DNA molecules. The AT-hook-like GGRPR motif at the N terminus of the NBD binds in the minor groove of the nonamer at positions 5–7 (see Figure 1a), with R391 mediating base-specific contacts (T5 and T6; bottom strand), and R393 contacting the phosphodiester backbone at T5. The polypeptide extends over the phosphodiester backbone between T4 and T5, with the remaining polypeptide arranged into three helices. The first helix is kinked, with the N-terminal portion (residues 400–407) occupying the major groove over a highly conserved guanine at nonamer position 2 (bottom strand). This residue is contacted by R402, which is consistent with the sensitivity of G2 to methylation interference (131). The phosphate backbone is contacted by R407, which also plays a structural role in stabilizing the loop that connects the GGRPR motif to helix H1. The first half of the C-terminal portion of helix 1 (residues 408–422) contains basic residues that either directly contact (K405 and K412) or approach (R401 and R409) the phosphate backbone in the 3′ end of the spacer (S8–S12 on bottom strand; see Figure 1a). Helix 2 (residues 426–441) and helix 3 (residues 444–454) form a four-helix bundle with their counterparts on the other NBD subunit to create an extensive dimer interface ( burying 1900 Å² of each monomer). Interestingly, helix 2 crosses over to approach the DNA bound by the other subunit at its C terminus, with residues Asn443 and His445 positioned near the backbone phosphates at S8 and S9 (bottom strand). Taken together, the extensive phosphate contacts involved in stabilizing RAG1-RSS complexes likely explain why core RAG1 binding exhibits only moderate discrimination between specific and nonspecific DNA (~tenfold) (23) and some sensitivity to ionic strength in DNA binding assays (153).

RAG1-RAG2 Single RSS Complexes

Purified RAG1 mixed or coexpressed with RAG2 forms one or two discrete RAG-RSS complexes, generally called SC1 and SC2, which are detectable by EMSA and exhibit similar catalytic activity in vitro (134). In reactions containing Mg²⁺, this activity is limited mostly to RSS nicking. However, in Mn²⁺, the RAG proteins robustly catalyze both cleavage steps (nicking and hairpin formation) in the absence of synapsis. In the SC complexes, RAG1 likely retains the dimeric configuration observed in RAG1-RSS complexes, although there is not complete agreement on this issue (134). By contrast, RAG2 stoichiometry differs between SC1 and SC2, with the former and latter containing one or two RAG2 subunits, respectively. The abundance and distribution of the two complexes depend in part on fusion partner composition (GST-RAG2 preferentially forms SC2, perhaps due to the intrinsic dimerization potential of GST) (32), RAG expression and purification strategies (coexpression and purification under mild conditions favors SC2) (109, 129), and the absence or presence of the RAG2 C terminus (SC2 is favored when the RAG2 C terminus is present) (109).

Compared to RAG1 alone, RAG1 and RAG2 together (hereafter called RAG) bind a single RSS with greater affinity ( apparent KD of 25 nM by EMSA) (153) and specificity (60, 96, 132). Enhanced binding can be attributed to additional base-specific and phosphate contacts that overlap with and extend from those observed in RAG1-RSS complexes to include the 5′-end of the spacer, the heptamer, and the heptamer-coding junction, and are biased toward one side of the DNA helix (131). RAG-RSS complex formation also promotes DNA conformational changes in the spacer and heptamer regions that are evidenced by...
hypersensitivity toward chemical and enzymatic probes (131). Most of these contacts are probably mediated by the central and C-terminal domains of RAG1, which contain three residues critical for RAG cleavage activity, but are otherwise dispensable for DNA binding (D600, D708, and E962, see Figure 3) (49, 76, 82). Evidence suggests these active site residues, and by extension the protein-DNA contacts mediated by the central and C-terminal domains that contain them, are contributed \textit{in trans} to those mediated by the GGRPR motif of the nonamer-bound NBD in the RAG1 dimer (128). This model is substantiated by structural studies of a RAG1 NBD-DNA complex discussed above and places the likely \textit{cis-trans} crossover point in helix 2 of the NBD. Two lines of evidence suggest that direct DNA contact by RAG2 in RAG-RSS complexes is a strong possibility, despite its lack of intrinsic DNA binding activity. First, photo-crosslinking studies show that RAG2 localizes proximal to the site of DNA cleavage (41, 133). Second, RAG2 mutations in basic residues have been identified that specifically impair RAG-RSS complex formation without disrupting RAG1-RAG2 interactions (see Figure 3) (48). Nevertheless, any RAG2 interactions with DNA are likely quite restricted, and on balance RAG2 probably serves more as a cofactor in promoting RAG1-RSS interactions than extensively contributing to RSS recognition.

**HMGB1 and HMGB2**

Although the RAG proteins alone are sufficient to mediate RSS binding and cleavage, the high mobility group box proteins HMGB1 and HMGB2 were found to stimulate RAG binding and cleavage activity on DNA substrates containing a 23RSS, but not a 12RSS, and to promote coupled cleavage of 12RSS and 23RSS paired \textit{in cis} (141). HMGB1 and HMGB2 are structurally very similar and functionally redundant in RAG biochemical assays; for simplicity, we will only refer to HMGB1 unless otherwise specified.

HMGB1 contains tandem HMGB domains, called A and B, followed by a basic linker and an acidic tail (137). Each \~80 amino acid HMGB domain contains three \( \alpha \)-helixes that form an L-shaped structure; box A favors binding to distorted DNA structures, whereas box B can bind and bend linear DNA. Although HMGB1 exhibits both RAG-independent RSS binding (130) and RSS-independent association with the RAG1 NBD (3), these interactions are not independently robust, but are likely enhanced by ternary complex formation (A.J. Little and D.G. Schatz, unpublished results). DNA footprinting and photo-crosslinking studies, together with more recent structure-function analyses of HMGB1 (see 131, 134, and references therein), suggest a model in which the HMGB domains play separable roles to stimulate RAG-mediated 23-RSS binding and cleavage: One HMGB domain may be positioned proximal to the nonamer to promote RAG-mediated bending of the spacer, and the other may be positioned in or near the heptamer to bind distorted DNA structures induced by RAG binding and/or RSS nicking. Based on their known DNA binding and bending properties (137), boxes A and B are speculated to occupy sites in or near the heptamer and nonamer, respectively. The tandem arrangement of the HMGB domains may provide bridging needed between the two sites in the SC to stimulate binding, but once synopsis with a partner RSS has been achieved, a conformational change may occur to stabilize the complex and obviate the function of one of the two HMGB domains (79). The HMGB1 stoichiometry in precleavage RAG-RSS complexes remains to be formally defined, but recent biochemical characterization of the SEC suggests the ratio of RAG1:RAG2:HMGB1 is approximately 2:2:1 in this complex (54).

**SYNAPTIC RAG-RSS COMPLEXES**

**RAG1/RAG2 Precleavage Paired RSS Complexes Assembled on Short DNA Fragments**

As discussed above, single RSS-RAG complexes assembled in \( Mg^{2+} \) support mainly RSS
nicking. By adding appropriate partner RSS in the presence of HMGB1, however, the RAG proteins catalyze hairpin formation at much higher levels in solution, and under these conditions form a PC detectable by EMSA with slightly slower electrophoretic mobility and higher intrinsic cleavage activity (~fivefold) than its counterpart lacking partner RSS (131). Time-course studies show that under conditions favoring synopsis with consensus oligonucleotide RSS substrates, nicking occurs rapidly and largely independently of synopsis (149) with similar unireactant kinetics and catalytic rate constants for both 12- and 23-RSS substrates (0.5–0.6 min\(^{-1}\)) (149), whereas hairpin formation occurs at an ~150-fold slower rate (~0.004 min\(^{-1}\)) (147). On long DNA fragments, however, the nicking rate may be stimulated slightly (~twofold) by synopsis (40, 80). Using EMSA-based approaches, subsequent studies found that SC2 is likely converted to the PC through the capture of an appropriate partner RSS without altering the complement of RAG proteins in the complex (98). Evidence favors a RAG1:RAG2 heterotetramer configuration in the PC (54, 129), with RAG2, but not RAG1, capable of reassorting during assembly (98, 129). One important caveat to consider is that the PCs analyzed in these studies were assembled on short DNA fragments. As discussed below, DNA fragment length may influence RAG protein stoichiometry and RSS recognition in the precleavage synaptic complex. Nevertheless, characterization of the PC in these experimental systems has provided valuable insights into the mechanisms that regulate PC formation and activity, and the major findings from these studies are reviewed here.

DNA footprinting studies indicate the RAG proteins afford greater protection from DNAse I cleavage to the heptamer and coding sequence when they are in the PC than when in the SC (100). In the PC, the RAG proteins protect about 12 bp of the coding flank on the bottom strand from DNAse I cleavage (99), but confer DNAse I hypersensitivity on the top strand about 12 bp upstream of the heptamer (100). Interestingly, when both RSSs are nicked, bottom strand protection from DNAse I is extended another ~4 nt into the coding region (99), and protection of the first and third G residues on the bottom strand in the heptamer from methylation by dimethyl sulfate is observed (102). This observation suggests that synopsis is accompanied by additional protein and/or DNA conformational changes that promote more intimate RAG interactions with the cleavage site, which are further enhanced upon RSS nicking. Because the central domain of RAG1 exhibits sequence-specific recognition of the heptamer and ssDNA binding activity (103), the most plausible scenario is that RAG binding imparts local melting of the DNA strands in and near the heptamer that is mediated and stabilized by the RAG1 central domain, possibly with assistance by HMGB1. Although DNA footprinting data suggest melting may be partly induced before substrate nicking, biochemical studies showing that base-pair mismatches or abasic sites incorporated at the coding flank stimulate hairpin formation more than nicking (27, 53, 107) suggest that more pronounced unpairing likely occurs after the RSS is nicked, presumably to orient the DNA strands in a configuration to catalyze hairpin formation.

More recently, direct evidence of base unpairing has been obtained using permanganate oxidation to detect extrahelical thymidine residues. Two independent studies showed that the terminal nucleotide on the bottom strand of the coding flank (C1b) exhibits hypersensitivity to permanganate oxidation when the RAG proteins are incubated with a 12/23 pair of nicked oligonucleotide RSS substrates under conditions favoring synopsis; no hypersensitivity is detected if one or both substrates are intact (16, 102). This pattern of sensitivity is reminiscent of what is observed with the Tn5 and Tn10 transposases, which, like the RAG proteins, generate DNA hairpin intermediates during the recombination process and create a hypersensitive thymidine in the strand opposite the nicking site (14, 15). Structural studies of the Tn5 postcleavage synaptic complex revealed that the extrahelical thymidine is
Base flipping: the movement of a base from its normal intrahelical position in double-stranded DNA to a position outside of the helix.

The Signal End Complex

Once the RAG proteins complete the RSS cleavage steps, the coding ends dissociate in vitro, but the signal ends remain tightly bound to the RAG proteins in the SEC (2). This complex is more resistant than precleavage complexes to dissociation, and the RAG proteins protect the ends from access by nucleases or joining factors (70). These features are reflected by strong protection of the entire RSS to chemical and enzymatic probes by DNA footprinting (131). Together, these findings suggest the SEC remains intact until it is physically disassembled, perhaps involving the targeted degradation of RAG2 (35, 95). The SEC is reported to contain a RAG1:RAG2 heterotetramer (54, 129) and one HMGB1 molecule (54). Characterization of the SEC by electron microscopy (EM) and atomic force microscopy (AFM) by Grundy et al. (54) suggests the RAG protein core forms a pseudosymmetric anchor-shaped complex containing two fluke-like extensions spanning about 150–160 Å tip-to-tip and central shank ~80–90 Å wide and ~125 Å long. Selective immunolabeling experiments suggest RAG2 and the RAG1 NBD localize to the flukes and shank, respectively, of the anchor-shaped complex. SECs formed with DNA fragments containing long nonamer end extensions show that the DNA adopts a parallel configuration in ~80% of complexes, and exits the SEC from adjacent points at the end opposite of where RAG2 is localized (see Figure 4).

DNA Length-Dependent Effects on RAG Stoichiometry and DNA Organization in Synaptic Complexes

Although a large body of evidence suggests a RAG1/2 heterotetramer core complex is capable of binding and cleaving pairs of oligonucleotide RSS substrates in vitro, and forming an SEC with short and long DNA fragments, there remains some question about how closely models drawn from these studies reflect RAG-mediated synapsis of longer, more physiologically relevant DNA...
substrates. The Roth laboratory provided the first direct evidence showing the RAG proteins differentially bind and cleave short and long DNA fragments. Specifically, Huye & Roth (63) found that the RAG proteins exhibit a substantial bias against cleaving a 12/23 pair of RSSs \textit{in trans} (on separate DNA molecules) when the RSSs are embedded in long DNA fragments but not when they are incorporated in short oligonucleotide substrates.
Additional evidence that DNA fragment length influences how the RAG proteins mediate synapsis comes from recent methylation interference and direct 1,10 copper-phenanthroline (Cu-OP) footprinting studies of RAG-RSS complexes assembled on long DNA fragments containing RSSs paired in cis (80). This study found evidence that the RAG proteins (a) mediate contact with the top strand of the coding flank ∼14 nt from the cleavage site [close to a DNAse I hypersensitive site detected in the PC by Nagawa et al. (100)] and induce a conformational change evidenced by a shift in CuOP sensitivity that is propagated distally through another turn of the DNA helix, and which is partially alleviated by substrate nicking; (b) induce conformational changes in the intersignal sequence that are coupled to substrate nicking and promoted by full-length RAG1; and (c) exhibit a DNA length-dependency in cleavage (hairpin) efficiency that involves putative coding end contacts with the C terminus of RAG1. The last finding is evidenced by the observation that a K980A RAG1 mutant exhibits selectively impaired hairpin formation, but not nicking, on long DNA fragments associated with an inability to induce conformational changes in the intersignal sequence. This outcome was interpreted to suggest that after nicking, but before hairpin formation, there is an activation step that involves a significant conformational change to orient the DNA strands in a configuration suitable for catalyzing hairpin formation. The fact that the DNA structural distortion occurs in the intersignal sequence distal to the nonamer that is contacted by the RAG1 NBD, yet apparently triggered by C-terminal RAG1 interactions with the coding flank, suggests that DNA strand reorientation involves both the N- and C-terminal portions of RAG1. It is tempting to speculate that the activation step involves base flipping at C1b, as this step occurs after nicking, and is negatively affected by C-terminal RAG1 mutations proximal to K980 (16, 102).

What could account for differences in how short and long DNA fragments are recognized and cleaved by the RAG proteins? One possibility is that synapsis of long DNA fragments requires a larger complement of RAG proteins than is found in PCs assembled on short oligonucleotide substrates. In support of this idea, Landree et al. compared the kinetics of RAG-mediated cleavage between wild-type and catalytically inactive RAG1 heterodimers on long DNA fragments containing RSSs paired in cis and concluded the synaptic complex formed on this substrate contains a pair of RAG1 dimers (81). Consistent with this possibility, recent studies show that when precleavage synaptic complexes are assembled with long DNA fragments containing embedded RSSs in cis or in trans, discrete RAG-RSS complexes are visualized by EMSA that support 12/23-regulated cleavage activity but exhibit much slower electrophoretic mobility than comparable complexes assembled using oligonucleotide substrates (80, 119).

Additional support for this possibility comes from a recent study using AFM to directly visualize and characterize RAG precleavage synaptic complexes assembled on long DNA

---

**Figure 4**

Integrated model of RAG synaptic complex organization, assembly, and RSS cleavage mechanisms. A RAG1-RAG2 heterotetramer core initially binds an RSS (e.g., the 23RSS here) via the NBD of RAG1, and RAG1 and RAG2 collaborate to survey the sequence spanning the spacer and heptamer region for a suitable binding site. If identified, the RAG proteins and DNA undergo conformational changes to enhance protein-DNA interactions and to assume geometries that facilitate nicking. This process is repeated to capture an appropriate partner RSS. Nicks may or may not be introduced prior to synapsis, but upon synapsis, RSS nicking is completed, and the complex undergoes a further activation step associated with base flipping at C1b (see inset). Hairpin formation proceeds in a concerted manner at both RSSs to complete the cleavage reaction. Based on structural and biochemical studies on a given RSS, one subunit of the RAG1 dimer interacts with the nonamer, whereas the other RAG1 subunit interacts with the spacer and heptamer and catalyzes both steps of the cleavage reaction (cleavage in trans).
fragments in trans in the presence of HMGB1 and Mg$^{2+}$ (119). In these experiments, the estimated mass of the RAG synaptic complex, based on volume measurements of these complexes, was found to be close to the mass predicted for a RAG1/2 heterooctamer (estimated: ~760 kDa; predicted: 816 kDa). In these complexes, analysis of DNA path lengths and the distance from the center of the synaptic complex to the ends of the protruding DNA further suggests that (a) the RAG proteins specifically bind the RSS on each DNA fragment, (b) the two DNA fragments are preferentially bound in a side-by-side configuration, rather than crossing over in the complex, and (c) the two DNA fragments favor a parallel over an antiparallel arrangement of RSSs by a factor of ~2:1. This outcome is consistent with the greater efficiency of in vitro coupled cleavage and in vivo V(D)J recombination observed with DNA substrates that more readily support parallel synapsis than antiparallel synapsis (39, 116). Why might a larger RAG complex be necessary to support synapsis of long DNA fragments? One possibility is that a higher-order (heterooctameric) RAG complex is required to overcome electrostatic or steric repulsion between long DNA strands that otherwise pose a significant barrier to synapsis in vitro. These repulsive forces would be less severe for short DNA fragments, and in this case, a RAG1/2 heterotetramer might provide sufficient shielding to overcome this barrier and permit synapsis to occur.

The RAG protein content estimated in synaptic complexes studied by Shlyakhtenko et al. (119) contrasts with results obtained by Grundy et al. (54), who found that RAG pre-cleavage synaptic complexes, although more heterogeneous in size than SECs, contained no more than a RAG1/2 tetramer on average. One important difference between these two studies is that Grundy et al. assembled RAG pre-cleavage complexes in Ca$^{2+}$ on DNA fragments that contained a short coding flank (16–26 bp). Shlyakhtenko et al. reported that RAG-RSS synaptic complex size depends on the composition of the metal ion cofactor and the length of the coding flank (the transition to a higher-order complex requires ~100 bp of flanking sequence) (119). Thus, it is plausible the conditions used by Grundy et al. favored assembling the lower-order RAG1/2 tetramer complex. Alternatively, it is possible that the RAG protein stoichiometry changes during the course of the cleavage reaction from a RAG1/2 heterooctamer in the pre-cleavage synaptic complex to a RAG1/2 heterotetramer in the post-cleavage SEC. This model has an attractive feature of enabling a RAG1/2 tetramer to accompany both sets of DNA ends after cleavage and provides a potential mechanism to explain how the RAG proteins help direct end joining of both signal and coding ends.

The observation that synaptic complexes assembled on long DNA fragments favor a side-by-side configuration of DNA strands with the RSSs arranged in a parallel orientation contrasts with an earlier study by Ciubotaru et al. (22) that used FRET to interrogate the distance between the ends of oligonucleotide 12RSS and 23RSS substrates bound in a PC. In that study, the data favored a model in which the RSSs adopt a bent and crossed configuration in the PC. Although this apparent discrepancy may partly be attributed to DNA length-dependent differences in how the RAG proteins bind short oligonucleotides and long DNA fragments, the DNA strands in synaptic complexes analyzed by EM and AFM were not sufficiently resolved to determine the precise alignment of the DNA strands. Some degree of DNA bending and strand crossover can easily be accommodated in structural models of the RAG1/2 heterotetramer complex developed from EM analysis of SECs, especially along the shank of the anchorlike projection, yet still remain consistent with data obtained by FRET.

**RULES GOVERNING RSS RECOGNITION AND PAIRING**

V(D)J recombination is guided by the 12/23 rule and existing evidence favors a capture model for synaptic complex assembly. However, the extent to which the capture model applies in vivo remains an open issue, and there
is also uncertainty concerning which RSS (12 or 23) might be recognized first, how the choice of a partner RSS is made, and when during PC assembly the nicking reactions occur. In this section, we consider current evidence bearing on these and related questions.

**THE 12/23 RULE AND THE CAPTURE MODEL**

V(D)J recombination in vivo exhibits a strong (>30-fold) preference for a 12/23 RSS pair over 12/12 and 23/23 pairs, and this preference can be recapitulated in DNA cleavage assays using the core RAG proteins and HMGB1 under certain conditions (reviewed in 47, 51, 84). Hence, it is likely that the 12/23 rule is imposed at or prior to the generation of DSBs. Indeed, there is evidence that both RSS synopsis and hairpin formation are more efficient with 12/23 than with 12/12 or 23/23 RSS pairs (reviewed in 47, 51). The mechanisms underlying preferred 12/23 synopsis and cleavage remain unknown.

Two studies have provided evidence for a 12RSS-first model in which PC formation occurs by initial RAG binding to the 12RSS followed by capture of a 23RSS. Biochemical experiments by Jones & Gellert (71) using core RAG proteins and a truncated form of HMGB1 lacking the C-terminal acidic tail, found that adherence to the 12/23 rule was influenced by the order in which the proteins were allowed to bind to the RSSs. When RAG and HMGB1 were bound first to the 12RSS, a very strong preference for synopsis and cleavage of a 23RSS versus a 12RSS partner was observed, but if the order of RSS addition was reversed, the preference for a 12RSS versus a 23RSS partner was much smaller (5–6 fold). In the second study, Curty et al. (28) examined RSS nicking at IgH and Igk in pre-B cell lines and at Tera in primary thymocytes and observed nicks at 12RSSs (flanking Dh, Vκ, and Jκ gene segments) but not 23RSSs (flanking Vh, Jα, and Vα gene segments). These data are consistent with a model in which RAG binds and nicks a 12RSS prior to 23RSS capture, and argue against initial binding to a 23RSS because, if so, 23RSS nicking should have been detected.

There are several considerations to take into account when interpreting the findings of these two studies. First, when HMGB1 is present, the core RAG proteins bind equally well to 12RSS and 23RSS substrates (reviewed in 51, 131). Hence, if 12RSSs are to be bound first in vivo, a mechanism would have to exist to make them the preferred binding sites for RAG. Second, there is no evidence that two 23RSSs are better able to support V(D)J recombination or DNA cleavage than two 12RSSs (59, 127), as might have been expected based on biochemical studies by Jones & Gellert (71). As discussed previously (131), it is conceivable that use of truncated HMGB1 in this study allowed for enhanced synopsis and cleavage between two 23RSSs. And third, nicking probably does not provide a reliable measure of RAG binding because the rate of nicking is strongly influenced by the DNA sequence of the RSS and the flanking coding nucleotides (37, 148).

Two recent studies have cast doubt on the general applicability of the 12RSS-first model for PC formation. In the first, nicking was assessed at Tcrb gene segments in early developing thymocytes and in biochemical experiments; in both assays, nicks were more abundant at the 23RSS 3′ of Dβ1 than at Jβ1 12RSSs (see Figure 1b for diagram of Tcrb), suggesting that during Dβ1-to-Jβ1 recombination, RAG binds first to the 3′ Dβ1 23RSS and then captures a Jβ1 12RSS (45). In the second, direct assessment of RAG1 and RAG2 binding to antigen receptor loci in primary developing lymphocytes by chromatin immunoprecipitation (ChIP) revealed no preference for RAG binding to 12RSSs over 23RSSs (67). Both RAG1 and RAG2 were found to bind selectively to small regions of highly active chromatin in the IgH, Igk, Tcrb, and Tera loci. These regions, referred to as recombination centers, span some or all of the J gene segments in all four loci as well as J-proximal D gene segments in IgH and Tcrb. Importantly, whereas J gene segments in Tcrb and Tera are flanked by 12RSSs, those in IgH and Igk are flanked...
by 23RSSs. RAG binding correlated well with elevated levels of histone 3 acetylation and H3K4me3 (both markers of transcriptionally active chromatin), and there was no evidence that 12RSSs were selectively located in open, accessible chromatin. It remains unclear why Curry et al. (28) observed RSS nicking at Vκ but not Jκ gene segments, whereas ChIP revealed RAG binding at Jκ but not Vκ gene segments (67). One possibility is that Igκ recombination is initiated when RAG binds to a small number of Vκ gene segments in each pre-B cell, but evidence of this went undetected in the ChIP analysis because of the large number of Vκ gene segments (>100). Alternatively, RAG binding might occur first at Jκ 23RSSs but is not accompanied by efficient nicking, as proposed in a recent, speculative model (114).

An appealing model arising from the RAG ChIP data is that RAG1 and RAG2 bind to RSSs in accessible chromatin regardless of whether they are 12RSSs or 23RSSs (67, 114) and hence that initial RAG binding is dictated by RSS accessibility. It seems likely, however, that other mechanisms in addition to accessibility influence the in vivo pattern of RAG-RSS binding. For example, endogenous RSSs exhibit considerable sequence variation and some of these variations, particularly in the spacer and nonamer, are expected to substantially affect RAG binding affinity (44, 47, 131). In addition, particular RSSs might have special mechanisms designed to facilitate RAG binding. For example, in the Tcrb locus (Figure 1b), the 23RSSs 3′ of Dβ1 and Dβ2 contain an embedded binding site for the AP1 transcription factor (Figure 1b,c), and the c-Fos protein, which is a component of AP1, is able to interact with RAG and facilitate RAG loading onto the 23RSSs of Dβ1 and Dβ2 (143). The AP1 and RAG binding sites in Dβ 23RSSs overlap (Figure 1c), raising interesting questions about the mechanism by which AP1 facilitates RAG binding (discussed in 134). RAG binding to Dβ 23RSSs would selectively support Dβ-to-Jβ recombination and could interfere sterically with RAG binding to Dβ 12RSSs (Dβ1 and Dβ2 are only 12 bp or 14 bp long, respectively). This might contribute to developmentally ordered recombination of Tcrb (Dβ-to-Jβ before Vβ-to-Dβ). This model is consistent with the finding that mutation of the 3′ Dβ 23RSS allows for readily detectable direct Vβ-to-Dβ recombination (123). Additional evidence for this model was recently obtained by two groups, who showed that c-Fos and the Dβ1 23RSS inhibit DNA cleavage and recombination at the 5′ Dβ1 12RSS (45, 143), and that direct Vβ-to-Dβ recombination occurs at an increased frequency in Fos−/− as compared to wild-type thymocytes (143). AP1-directed binding of RAG to 3′ Dβ 23RSSs might help explain the observation, discussed above, that nicks are detected at higher levels at the Dβ1 23RSS than Jβ 12RSSs (45).

A similar mechanism has been proposed to facilitate RAG binding to Vh 23RSSs. The transcription factor PAX5 has binding sites in the body of many Vh gene segments, interacts with RAG, and can increase RAG-mediated cleavage at Vh RSSs in biochemical assays and recombination of Vh gene segments on episomal recombination substrates in vivo (152). This is intriguing given the importance of PAX5 for recombination of Djh-distal Vh gene segments and for Igk locus contraction, which is thought to bring distal Vh gene segments into close proximity of Dh-Jh and allow for synapsis and Vh-to-DJh recombination (50, 65). It is possible that PAX5 binding to these sites helps deliver RAG to Vh RSSs prior to synapsis and a Vh-to-DJh recombination (123). Alternatively, since Vh gene segments lie outside of the Igk recombination center (67), PAX5 may help to deliver Vh RSSs to RAG within the recombination center (134).

In summary, initial RSS binding by RAG is likely determined by a superposition of several mechanisms, including accessibility within chromatin and the action of RAG-interacting targeting factors, but a consistent preference for the 12RSS is not supported by current evidence.

**PATHWAYS OF RAG RECRUITMENT IN VIVO**

DNA binding, protein-protein interaction, and other types of biochemical experiments had led
to the expectation that RSS recognition in vivo would be mediated by a complex containing both RAG1 and RAG2 (reviewed in 47, 51, 131). It was surprising, therefore, when ChIP experiments revealed that RAG1 and RAG2 each exhibited highly specific patterns of chromatin binding in vivo, including recruitment into recombination centers, in the absence of the other (67). This led to the proposal that there are three distinct pathways for RAG recruitment to specific sites in chromatin (discussed in 67, 114): (a) RAG1 binds to the RSS and RAG2 is recruited later by virtue of its ability to interact with RAG1 and H3K4me3; (b) RAG2 uses its PHD finger to bind to nucleosomes containing H3K4me3, and RAG1 is recruited afterwards by binding to RAG2, perhaps without involving a direct interaction with DNA; and (c) a preformed RAG1-RAG2 complex binds to an accessible RSS, to H3K4me3, or to both. The relative contributions of these three pathways are not known, nor is it known what fraction of the RAG1 and RAG2 molecules in the nucleus are associated with one another or how dynamic this association is (114).

Nonspecific DNA binding by RAG1 or by the RAG1-RAG2 complex might also influence the distribution of the RAG proteins in the genome. As noted above, RAG1 can recognize the RSS in a sequence-specific manner, whereas inclusion of RAG2 increases the affinity and specificity of the interaction (131). Zhao et al. (153) recently reported that core RAG1 binds strongly to nonspecific DNA and that this interaction can be dramatically reduced by RAG2. They also observed a relatively stable RAG1-12RSS complex (half-life of 558 s), leading to the proposal that there are two distinct modes of DNA binding by RAG1: a nonspecific mode that relies heavily on electrostatic interactions and a sequence-specific mode that is less dependent on electrostatic interactions and might involve conformational changes in the protein and the DNA (153). Previous studies had not observed such strong interactions of RAG1 with nonspecific DNA (6, 23, 126); the reasons for this discrepancy are not clear. Strong nonspecific DNA binding by RAG1, as observed by Zhao et al., predicts the possibility of RAG1 binding to many open sites in the genome where DNA is relatively exposed (such as promoters and other nuclease hypersensitive sites; such binding should not only be RAG2-independent, but might actually be inhibited by RAG2. In addition, the ability of RAG1 to form relatively stable, sequence-specific complexes could help explain the RAG2-independent accumulation of RAG1 in recombination centers, which contain multiple, accessible RSSs.

It is plausible that HMGB1 further enhances RSS binding in vivo by RAG1 because RAG1 and HMGB1 form a specific complex with the 12RSS (66) and mediate synapsis of two RSSs in solution in a manner that is enhanced only modestly by RAG2 (146). Genome-wide analyses of RAG1 binding patterns should provide substantial insight into some of these issues.

**THE BEYOND 12/23 RESTRICTION**

The 12/23 rule dictates that recombination will be inefficient between two 12RSSs or between two 23RSSs, but does it also mean that recombination will occur efficiently between any 12/23 RSS pair? Analysis of the Tcrb locus has revealed that the answer to this question is no. Vβ gene segments invariably recombine with a Dβ 12RSS rather than with one of the multiple Jβ 12RSSs that typically remain in the locus (Figure 1b). Why is this? Elegant in vivo experiments involving a transgenic Tcrb minilocus or modified endogenous Tcrb loci (reviewed in 18, 139) have demonstrated that Vβ gene segments display a very strong preference to recombine with the 12RSS 5' of Dβ1 over any of the Jβ1 12RSSs, and this is not due to the positioning of the various RSSs or to competition between Dβ and Jβ 12RSSs (11, 123). Importantly, this B12/23 restriction can be reconstituted with episomal recombination substrates in nonlymphoid cell lines, demonstrating that neither lymphoid-specific factors nor residence within
the genome is essential for enforcement of the restriction (74, 103, 142). Furthermore, using the core RAG proteins, HMG1 or HMG2, and long DNA substrates, Vβ 23RSSs exhibit a five- to tenfold preference for cleavage with the Dβ1 12RSS over a Jβ 12RSS (74, 103). Notably, this preference can be as high as 50- to 500-fold in vivo with episomal recombination substrates (74) and is likely much greater than 100-fold during endogenous Tcrb recombination (11, 123). Hence, although the B12/23 restriction is substantially determined by the intrinsic properties of interactions between RAG and Tcrb DNA substrates, additional mechanisms might also contribute.

One way to explain preferential recombination of Vβ with DJβ as opposed to Jβ would be if RAG binds to Dβ 12RSSs better than to Jβ 12RSSs. EMSA experiments utilizing the core RAG proteins reveal no such preference, and suggest that Dβ and Jβ 12RSSs are bound much less efficiently by RAG than the consensus 12RSS (37, 103). Nor is there any indication from ChIP experiments that RAG binding is stronger at Dβ as opposed to Jβ gene segments (67). It is not yet clear, however, how well RAG ChIP patterns reflect RSS occupancy, or whether the pattern of RAG binding to Tcrb or Igb changes after D-to-J recombination has occurred. It is therefore possible that, after Dβ-to-Jβ recombination, RAG binding occurs preferentially to the 12RSS 5′ of the DJβ junction as opposed to the remaining Jβ 12RSSs. This would be consistent with the detection of nicks 5′ of Dβ1 but not at Jβ gene segments in vivo (45). Perhaps the close proximity of the 12RSS 5′ of DJβ to the recent D-to-J joining event creates a favorable context for RSS engagement by RAG.

Alternatively, a plausible interpretation of existing data is that DJβ and Jβ 12RSSs are occupied by RAG with roughly equal efficiency: If this is the case, how is the B12/23 restriction enforced? One possibility, suggested by a biochemical analysis using core RAG proteins, truncated HMG2, and oligonucleotide RSS substrates, is that Jβ gene segments are crippled for recombination with Vβ gene segments (37). Jβ substrates were shown to nick slowly and/or to synapse weakly with Vβ substrates, whereas the Dβ1 12RSS exhibited more efficient synapsis and more rapid nicking than the Jβ 12RSSs tested. This leads to a model in which initial RAG binding to a Jβ 12RSS fails to support recombination with Vβ either because slow nicking allows RAG to dissociate from Jβ prior to synapsis or hairpin formation, or because, in the case where Jβ nicking does occur, the synaptic complex does not form or is unstable. The Dβ1 12RSS exhibits neither defect and hence, once bound by RAG, nicks rapidly (helping to retain the bound RAG) and synapses with Vβ efficiently (37). This in turn might increase the efficiency of Vβ nicking (45), allowing for base flipping at both RSSs and subsequent hairpin formation. It will be important to confirm the crippled nature of Jβ gene segments in experiments using full-length HMG1 protein and highly active preparations of the RAG proteins because in the published analysis, coupled cleavage between Vβ14 and either of the Dβ 12RSSs was inefficient (37).

If indeed Jβ gene segments are crippled for recombination, how does Dβ-to-Jβ recombination occur with sufficient efficiency? As discussed above, AP1 may help promote efficient loading of RAG onto the Dβ 23RSSs. In addition, the Dβ1 23RSS forms a stable synaptic complex with Jβ gene segments 2- to 3.5-fold more efficiently than does the Vβ14 23RSS (37). Recent data also raise the possibility that nicking of Jβ is enhanced by synapsis with the Dβ1 23RSS (45). It might also be relevant that Dβ and Jβ gene segments coexist in relatively close proximity to one another within recombination centers, where most or all of the RAG protein associated with Tcrb appears to be located (67).

Overall, the B12/23 restriction appears to be controlled in large part by intrinsic features of RAG interactions with Tcrb gene segment substrates, perhaps complemented by mechanisms that have yet to be recapitulated biochemically, such as nucleosome positioning, histone modifications, and the consequences of DNA repair of a prior recombination event. It will be
important to determine the extent and stability of RSS occupancy and synopsis by RAG both before and after Dβ-to-Jβ recombination.

**CAPTURE AND RELEASE: SAMPLING MULTIPLE PARTNERS PRIOR TO RECOMBINATION?**

The half life of the PC has not been determined in solution; even more uncertain is how stable the SC or PC is in vivo, where the RAG proteins must compete with histones and other DNA binding proteins for RSSs. It is unknown whether SC formation involves multiple cycles of RSS binding and release by RAG, or how many times the SC samples partner RSSs prior to formation of a PC that is sufficiently stable to support hairpin formation. Studies of the Tcrb locus have suggested the idea that stable PC formation and recombination do indeed involve cycles of capture and release of a partner RSS. The initial evidence for this came from the analysis of mice harboring a Tcrb allele in which the 23RSS flanking Vβ14 was replaced with the strong Dβ1 23RSS (144). This replacement resulted in a large (sevenfold) increase in the frequency of Vβ14 recombination events and of Vβ14-positive thymocytes and peripheral T cells, formally demonstrating that RSS sequence can influence recombination efficiency. On the assumption that the RSS swap did not increase the frequency with which an accessible Vβ14 comes into close proximity of DJβ, these data support the conclusion that in normal thymocytes, Vβ14 has multiple encounters with DJβ, most of which do not lead to recombination—that is, recombination typically occurs after multiple cycles of partner RSS capture and release (144).

When the Vβ14 RSS was replaced with an unrecombined Dβ1-Jβ1 cassette, the cassette was observed to undergo recombination in almost 90% of developing thymocytes, far higher than the frequency of primary Vβ14 recombination events (7%) (108). The simplest interpretation is that the unmanipulated Vβ14 gene segment resides in accessible chromatin in a high percentage of developing thymocytes and that recombination of this gene segment is limited by its ability to couple productively with DJβ rather than by its accessibility. This would argue against the possibility that replacing the Vβ14 RSS with the Dβ 23RSS usage primarily by increasing Vβ14 accessibility, and provides support for the interpretation that Vβ14 recombination involves cycles of capture and release.

It is plausible that Vβ14 is not unique and that many or all of the Vβ gene segments engage in cycles of capture and release with the DJβ region. If correct, the different Vβ gene segments might compete with one another for stable PC formation and recombination with DJβ. Competition of this sort predicts that elimination of some Vβ gene segments would increase the recombination frequency of the remaining ones. Analysis of heterozygous thymocytes containing a Tcrb allele lacking 18 of the 21 functional Vβ gene segments yielded findings consistent with this prediction: Two of the three Vβ gene segments on the deleted allele showed significantly elevated recombination frequencies (13). Importantly, the deleted allele recombined somewhat less efficiently than the intact allele, ruling out the possibility that the outcome could be attributed to a generally elevated recombination efficiency on the Vβ-deleted allele. These results led the authors to propose that Vβ gene segments compete with one another for productive synopsis with DJβ (13).

Together with the finding of focal RAG binding in recombination centers, these studies suggest a model (described in 67, 114) in which V gene segments in looped/contracted chromatin cluster around a recombination center and compete with one another to be captured by RAG bound to J or DJ. This competition would typically involve multiple cycles of V segment capture and release, during which the V RSS is sampled by RAG within the recombination center. A better and/or more accessible RSS increases the probability that any given capture event would lead to a stable PC and recombination. This is consistent with data suggesting that RSS sequence variations help
explain primary recombination frequencies (reviewed in 44). Because RAG1-RSS interactions are largely centered on the nonamer, whereas RAG1/2-RSS interactions extend through the spacer and heptamer, it is tempting to speculate that under normal circumstances, the sampling process involves primary identification of a nonamer-like sequence by RAG1, followed by surveillance of the heptamer and spacer sequence for suitable sites of interaction in collaboration with RAG2. Stable complex formation occurs when an appropriate sequence is identified and is expected to be accompanied by significant RAG and DNA conformational changes that provide the required geometry to catalyze the biochemical steps of the cleavage reaction. We envision this sampling process could operate at both the initial RSS binding and subsequent RSS capture steps of synaptic complex assembly. A simplified model of this process that is based on the RAG1/2 organization and subunit stoichiometry suggested from EM and AFM studies of the SEC is shown in Figure 4.

There is, however, a caveat associated with the data that support the concept of capture and release. Specifically, the sequences used to replace the Vβ14 RSS included the Dβ1 23RSS and its embedded AP1 binding site (108, 144). This arrangement offers the possibility of cFos-mediated RAG recruitment at Vβ14, which might in turn alter properties of Vβ14, including its accessibility and propensity to become juxtaposed with the bona fide recombination center at DJβ, thereby increasing Vβ14 recombination. It will be important to determine whether RAG binding can be detected at Vβ14 in these mutant alleles because no binding at Vβ14 is detected in wild-type alleles (67). A more definitive resolution of this issue will likely require the creation of additional RSS replacement alleles.

**RAG NONCORE REGIONS**

**RAG1 Noncore Regions**

As discussed above, the N-terminal third of RAG1 is dispensable for the catalytic activity of the RAG complex but required for efficient and high-fidelity V(D)J recombination. This region contains several conserved cysteine-containing elements (C1, C2, C3, CH), four basic motifs (BI, BIIa, BIIb, BIII), and a C3HC4 RING domain flanked by a classical C2H2 zinc finger domain (Figure 3). Mutation of the cysteine-containing elements causes modest (C3, CH) to severe (C1, C2) reduction in the V(D)J recombination activity of RAG1 (112). Full-length RAG1 is expressed at lower levels than core RAG1; mapping studies (55, 93, 109) show little or no improvement in expression by truncating the first 134–150 residues, whereas further truncation to residue 211 results in a modest increase in expression. Additional truncation between residues 238–265 results in expression levels at or near core RAG1 levels. Thus, the RING and zinc finger domains between residues 265 and 383 are not directly responsible for the diminished RAG1 expression levels associated with the full-length protein. The region between residues 216 and 238 contains the BIIa basic motif that mediates interactions with KPNA1 (see below) (125) and is necessary to maintain the high recombination levels of full-length RAG1 (93), as well as a lysine residue (K233) that undergoes auto-ubiquitination by an isolated RAG1218–389 fragment (72). However, alanine or arginine substitution at K233 alone or in combination with K234 and K236 in the context of full-length RAG1 does not fully restore expression to core RAG1 levels (55), suggesting that ubiquitination of K233, if it occurs at this site in vivo, plays at most a partial role in regulating full-length RAG1 expression.

The RING domain defines a family of E3 ubiquitin ligases known for their ability to facilitate the attachment of ubiquitin (Ub), a small (76-amino-acid) protein modifier, to other proteins to alter their function or stability (for review, see 34). E3 Ub ligases promote transfer of Ub from an E2 Ub carrier protein to an internal lysine residue on the target substrate and dictate target specificity, either alone or in combination with the E2 protein. Through this process, targets proteins may
become monoubiquitinated at one or several sites, or through multiple iterations become polyubiquitinated. The Ub chain length and composition largely determine the biological outcome of ubiquitination, with monoubiquitination and some types of polyubiquitin chains generally serving to alter protein function, whereas other polyubiquitin chains (e.g., those assembled on Lys48) typically direct the target to the 26S proteasome for degradation.

The presence of a RING domain in RAG1 suggests that it functions as an E3 Ub ligase, and its importance in regulating V(D)J recombination is suggested by the identification of RING domain mutations associated with Omenn syndrome (hC328Y; see Figure 3). In support of this contention, two reports showed that isolated RAG1 fragments containing the RING domain and all (RAG11–381) or a portion (RAG1218–389) of the upstream sequence support ubiquitination of a test substrate (S protein) (150), and RAG1 autoubiquitination primarily at K233 (72). A fragment containing only the RING domain (GST-RAG1250–350) reportedly promotes polyubiquitin chain formation (150). Subsequently, RAG1218–389 was found to mediate ubiquitination of karyopherin alpha 1 (KPNA1; also called importin alpha 5, SRP1, and RAG cohort protein 2) (122). KPNA1 is a protein implicated in nuclear trafficking that was originally identified as a RAG1-interacting factor using a yeast two-hybrid screen (25) and subsequently shown to bind the BIIa motif upstream of the RING domain (125). RAG1 autoubiquitination and RAG1-mediated ubiquitination of KPNA1 have yet to be reconstituted in vitro in the context of full-length RAG1.

Recently, Grazini et al. (52) reported that full-length, but not core, RAG1, expressed in 293 cells, associates with histone H3 (variant unidentified), and that full-length RAG1 and RAG2 support monoubiquitination of H3 in vitro (using RAG immunoprecipitates) and in 293 cells. However, in these assays, RAG-mediated ubiquitination was not particularly robust and was only mildly impaired by mutations in the RING domain (H307A or C325G). This result raised some question about whether RAG1 directly mediates H3 ubiquitination and whether RAG1 selectively ubiquitinates different variants or posttranslationally modified forms of H3. Very recently, Jones et al. reported that a bacterially expressed RAG1 fragment containing the entire RAG1 N terminus (RAG11–189), but not a RAG1218–389 fragment, supports mainly monoubiquitination of full-length histone H3 contained within a preparation of mixed histones purified from a RAG1−/− pro-B cell line (69). By analyzing the monoubiquitinated H3 protein by mass spectrometry, it was defined as histone variant H3.3 that was posttranslationally modified by acetylation (K14, K18, and K23) and Ser31 phosphorylation. Consistent with this finding, recombinant histone H3.1 and H3.3 proteins were poor substrates for in vitro ubiquitination, and antibodies specific for H3K23a and H3.3S31p inhibited in vitro ubiquitination (anti-H3, H3K9a, and H3K18a antibodies had little or no effect). Whether these findings are recapitulated in the context of full-length RAG1 remains untested.

To establish a relationship between RAG-mediated ubiquitination and V(D)J recombination, several groups have analyzed the V(D)J recombination activity of RAG1 RING mutants. Grazini et al. (52) showed that RAG1 H307A and C325G single or double mutants exhibit impaired V(D)J recombination activity on extrachromosomal and endogenous substrates, with the defect being milder for coding joint formation (40% to 60% reduction versus wild type) than signal joint formation (~80% reduction versus wild type), but in this study, RAG-mediated DNA break induction on extrachromosomal substrates was unaltered by RING domain mutations. Another RAG1 RING domain mutant, P326G, characterized by Simkus et al. (120), showed similar defects in signal and coding joint formation to those investigated by Grazini et al. Interestingly, however, these authors found that a C325Y mutant supports much less recombination activity (>50-fold reduction versus wild type) than reported for the C325G mutant, which
they attributed in part to increased protein instability (no folding defects were observed with the P326G mutant). Whether a RAG1 C325Y mutant initiates cleavage as well as wild-type RAG1 was not tested. However, both the C325Y and P326G mutants fail to support autoubiquitination in an in vitro assay. Analysis of an additional panel of RAG1 RING domain mutants (N265A, H270A, I292A, E294A, S327F) suggests there is a positive correlation between ubiquitin ligase activity in vitro and cellular V(D)J recombination activity in vivo (121). However, a formal link between these processes remains to be established because it requires showing that V(D)J recombination is impaired by the loss of the substrate targeted for RAG1-mediated ubiquitination and more specifically mutation of the target ubiquitination site(s).

Studies of RAG1-mediated ubiquitination conducted to date must be interpreted with some caution because although some RING domain–containing proteins function as an E3 Ub ligase as a single subunit, other RING domain proteins mediate ubiquitination in the context of multisubunit assemblies (34). For example, in the cullin family of RING Ub ligases (CRL), the RING domain subunit associates with a cullin scaffold protein and uses different adaptor and substrate receptors to assemble unique E3 Ub ligase complexes. It is notable in this regard that the CRL complex Cul4-DDB1-Roc1 has been shown to mediate histone H3 and H4 ubiquitination in response to UV irradiation (142). Whether the E3 Ub ligase activity and target specificity of RAG1 are modulated through its association with other factors remains unclear. It is also possible that full-length RAG1 directly or indirectly associates with other E3 Ub ligases, a possibility offered by Jones et al. (69) to explain the relative insensitivity of histone H3 ubiquitination to RAG1 RING domain mutations observed by Grazini et al. In addition, the E2 pairings with RAG1 profoundly influence its E3 ligase activity in vitro: UbcH3 (CDC34) supports RAG1 autoubiquitination (72), UbcH5b supports polyubiquitin chain formation (150), and histone H3 ubiquitination (52), and UbcH2 supports ubiquitination of KPN1 and acetylated/phosphorylated histone H3.3 (69, 122). The relevant E2 carrier proteins supporting physiological RAG1-dependent ubiquitination in vivo remain to be established. Thus, the E3 Ub ligase activity of RAG1 is likely to be complicated and context dependent. Nevertheless, there is good reason to suspect ubiquitination plays an important role in V(D)J recombination, as it is implicated in regulating DNA repair processes in response to other forms of DNA damage (7, 138). How might it do so? Grazini et al. (52) have speculated that histone H3 ubiquitination by RAG1 enhances the efficiency of the repair steps of V(D)J recombination by serving to mark nucleosomes at RAG-mediated DNA breaks to facilitate detection and joining of these intermediates. Jones et al. (69) refine this model by suggesting that modified histones associated with recombinationally active loci are selectively targeted for RAG1-mediated ubiquitination. Whether there is a mechanistic link between DNA break induction by the RAG proteins and the activation of RAG1 E3 Ub ligase activity remains unclear, but it is tempting to speculate that the two events are coupled in some manner, possibly triggered by remodeling or modification of the RAG complex following DNA cleavage.

**RAG2 NONCORE REGION**

Like the noncore regions of RAG1, the noncore region of RAG2, which constitutes the last 140 aa of the protein, plays important roles in V(D)J recombination efficiency and fidelity (discussed below under End Joining and Postcleavage Complex Stability). This region includes a portion of a flexible acidic hinge, a PHD finger, and the T490 residue that is important for directing degradation of RAG2 in the S, G2, and M phases of the cell cycle (Figure 3). The function of the very C-terminal residues of RAG2 (residues 500–527) is not well understood, although the last seven residues of the protein contribute to phosphotidylinositol phosphate binding
and decrease protein stability (42, 55). The PHD finger is discussed here; readers are directed to a recent review for a more complete discussion of the RAG2 noncore region (73).

The RAG2 PHD finger (residues 414–487) has been shown to interact with two quite distinct entities: phosphotidylinositol phosphate (42) and covalently modified forms of the N-terminal tail of histone H3, particularly H3K4me3 (87, 91). Interaction with phosphoinositides was suggested to modulate an equilibrium between permissive and inhibitory states of the PHD finger (42), but the functional significance of phosphoinositide binding is not well understood at this point.

The discovery that the RAG2 PHD finger allows RAG to read the histone code represented a paradigm shift for the field of V(D)J recombination by suggesting new pathways for controlling the targeting and activity of RAG (87, 91). The RAG2 PHD finger interacts with the N-terminal tail of histone 3 when lysine 4 is methylated, with the binding preference $H3K4me3 \gg H3K4me2 > H3K4me1 > H3K4me0$. Binding to H3K4me3 ($Kd \approx 4 \mu M$) is at least an order of magnitude stronger than to H3K4me2 ($Kd \approx 60 \mu M$) (91). Structural analysis revealed that the PHD finger forms an aromatic channel into which the K4me3 side chain fits, with residues Y415, M443, and W453 constituting critical components of the channel (91, 106). Mutation of any one of these residues eliminates RAG2 binding to a histone 3 N-terminal peptide containing K4me3 (hereafter, H3K4me3 peptide) and significantly reduces V(D)J recombination activity in cell based assays (87, 91). An unusual feature of the RAG2 PHD finger is that it can accommodate methylation of H3 arginine 2 (R2), and asymmetric dimethylation of R2 on the H3K4me3 peptide increases the binding affinity somewhat (106). Curiously, asymmetric dimethylation of H3R2 actually inhibits trimethylation of H3K4 (56, 64). The relevance of H3R2 dimethylation for RAG2 binding in vivo or for V(D)J recombination remains unknown. Interestingly, mutations of several PHD finger residues, including W453, have been implicated in Omenn syndrome (Figure 3) (124). Overall, current data indicate that the structural integrity of the PHD finger as well as its ability to bind H3K4me3 are critical for efficient V(D)J recombination (73).

How does the RAG2 PHD finger facilitate the recombination process? Given the tight link between H3K4me3 and active or poised RNA polymerase II (113), and between transcription and RSS accessibility (114), an appealing idea is that H3K4me3 recognition by RAG2 helps recruit RAG to the vicinity of RSSs, thereby facilitating stable RSS binding (87, 91). Several pieces of evidence are consistent with this idea. First, the RAG2 PHD finger can bind to chromatin containing methylated H3K4 (87). Second, ChIP sequencing shows a tight correlation between RAG2 binding and H3K4me3, both within antigen receptor loci and elsewhere in the genome, arguing that the PHD finger-H3K4me3 interaction plays a dominant role in determining RAG2 association with chromatin. Third, tethering of H3K4me3 peptide to an RSS oligonucleotide stimulated RSS binding by RAG only if RAG2 contained the PHD finger, demonstrating that the PHD finger-H3K4me3 interaction can enhance RSS recognition in a defined biochemical system (118).

Another nonmutually exclusive mechanism is that H3K4me3 binding by RAG2 enhances the intrinsic activity of the RAG complex. In support of this contention, PHD finger point mutations that eliminate H3K4me3 binding (RAG2-PHDmut) reduce V(D)J recombination activity more than deletion of the entire RAG2 C-terminal region (RAG2 core) (87, 91, 106). These findings suggest that the PHD finger-H3K4me3 peptide interaction relieves an inhibitory effect of the RAG2 C terminus, perhaps by triggering a conformational change in RAG2. A clue to how this might occur was revealed by structural analysis of the PHD finger without bound H3 peptide. Under these conditions, the binding channel contained a short peptide derived from expression vector sequences, with a proline residue occupying the position of the trimethyl-ammonium group of K4me3 (106). This led to the proposal that, in the absence of H3K4me3, RAG2 is in a
repressed configuration caused by occupancy of the PHD finger binding channel by a self-peptide of RAG2 (perhaps a proline-containing flexible loop from the RAG2 core); binding of H3K4me3 peptide would displace the self-peptide and relieve this inhibition (106). However, for this model to explain the poor activity of RAG2-PHDmut proteins, one must assume that the PHD finger mutations prevent binding of H3K4me3 peptide but still allow binding of the putative RAG2 self-peptide.

Two recent biochemical studies lend further support for the direct stimulation model by demonstrating that binding of H3K4me3 peptide enhances RAG catalytic activity in a manner that is dependent on an intact PHD finger (55, 118). The peptide need not be attached to the RSS DNA to exert its effect, arguing that stimulation occurs by a mechanism other than simple recruitment of RAG to the DNA. H3K4me3 peptide was found to stimulate RAG-RSS complex formation [the SC (118) and the PC (55)], nicking (118), and hairpin formation (55, 118). Grundy et al. (55) also demonstrated that the RAG2 C-terminal region (excluding only residues 521–527) leads to a suppression of hairpin formation that is strongly dependent on the presence of the last 32 residues of RAG1 (residues 1009–1040). Furthermore, they found that adding H3K4me3 peptide relieves this inhibition in a manner that is again far stronger when the RAG1 C-terminal residues are present. Grundy et al. (55) proposed that the RAG1 and RAG2 C-terminal regions collaborate to limit RSS substrate flexibility to inhibit a DNA conformational change required for hairpin formation and that H3K4me3 peptide binding by the PHD finger relieves this inhibition.

Shimazaki et al. (118), however, observed RAG stimulation by H3K4me3 peptide using core RAG1, which lacks the C-terminal residues. The discrepancy might be explained by the use of different RAG2 proteins (1–527 by Shimazaki et al. and 1–520 by Grundy et al.) and by the fact that Grundy et al. did not include HMGB1 in their DNA binding or cleavage experiments, noting that the effects they observed were reduced in the presence of HMGB1. There is also some disagreement as to whether the RAG2 C-terminal region is inhibitory: Some (55, 80, 109, 135), but not all (118), studies find that full-length or nearly full-length RAG2 is less active than core RAG2 in supporting DNA cleavage in vitro. Interestingly, the former studies used RAG1 containing residues 1009–1040, whereas the latter did not.

Overall, current evidence indicates that the RAG2 C-terminal domain and an intact PHD finger are important for proper RAG function in vivo, that the C-terminal domain can exert an inhibitory effect on DNA cleavage in vitro, that this is manifest at both the DNA binding and cleavage steps, and that inhibition by the RAG2 C-terminal region is most apparent when activity is assayed using RAG1 that contains its extreme C-terminal residues. Binding of the RAG2 PHD finger to H3K4me3 appears to stimulate V(D)J recombination by at least two mechanisms: enhanced RAG recruitment to the RSS and stimulation of DNA cleavage. The stimulation of DNA cleavage is mediated by effects on both RSS binding and catalysis, and is likely due at least in part to a reversal of the repressive effects of the RAG2 C-terminal domain. Important challenges for the future include determining the full range of functions exerted by the RAG2 PHD finger and whether conformational changes underlie the stimulatory effects of engagement of H3K4me3 by the RAG2 PHD finger.

CLEAVAGE-TO-JOINING TRANSITION

DNA DSBs are dangerous lesions with the capacity to lead to genome rearrangements such as chromosomal translocations. Lymphoid tumors frequently harbor chromosomal translocations involving antigen receptor loci, and a subset of these translocations are due to errors made during the process of V(D)J recombination (12). Given the very large number of V(D)J recombination events that occur each day in a normal human, it is exceptionally important that RAG-generated DSBs be repaired
efficiently and with high fidelity and that there exist surveillance systems to eliminate the rare cells in which repair goes awry. This is accomplished through the combined action of numerous DNA repair and damage response proteins and pathways. Here, we consider this only from the perspective of the RAG proteins and their role in orchestrating proper repair of the DSBs they generate. The reader is also directed to an excellent recent review of this topic (19).

**RAG AND THE CHOICE OF A REPAIR PATHWAY**

As noted above, repair during V(D)J recombination is mediated by the classical NHEJ pathway. A defect in any classical NHEJ repair factor virtually eliminates V(D)J recombination, demonstrating that no other repair pathway can compensate for the absence of NHEJ during V(D)J recombination. This is strikingly different from class switch recombination, where an alternative NHEJ pathway is able to generate products relatively efficiently in cells lacking one or more classical NHEJ factors (17 and references therein). What prevents RAG-generated ends from efficiently accessing repair pathways other than classical NHEJ?

Accumulating evidence strongly argues that the answer lies in the RAG proteins themselves and the postcleavage complex(es) they form with the coding and signal ends. Initial evidence for this came from the identification of RAG mutants, many by the Roth laboratory, that were able to cleave reasonably efficiently but were defective in V(D)J recombination, presumably due to a defect in joining (Figure 3) (19). Subsequent analysis revealed that two of these mutants (RAG1 K980A and RAG1 R838A/K839A/R840A) yielded elevated levels of repair of signal ends by homologous recombination and coding joints with an increased frequency of short sequence homologies and excessive deletions, hallmarks of alternative NHEJ (62, 83). Hence, mutations in RAG1 can allow other repair pathways to gain access to signal and coding ends, supporting the idea that RAG guides the cleaved ends it generates specifically into the classical NHEJ pathway (83).

Strong support for this idea came from subsequent studies from the Roth and Meek groups (24, 26). In one, it was revealed that a frameshift mutation of RAG2 at residue 361 (RAG2–361Δfs) allowed for increased access of ends to both the homologous repair and alternative NHEJ pathways compared to wild-type RAG2 (24). In addition, RAG2–361Δfs allowed for remarkably efficient signal and coding joint formation in cells lacking classical NHEJ repair factors; sequence analysis of these products showed an increased frequency of short sequence homologies and large deletions. The alternative NHEJ pathway is therefore quite robust and able to process and join RAG-generated ends, but is prevented from doing so in the context of the wild-type RAG proteins (24).

In the second study, joining of ends generated by the I-Sce1 endonuclease was assessed on transiently transfected plasmid substrates containing two I-Sce1 sites (26). It was first demonstrated that I-Sce1 ends were rejoined efficiently in classical NHEJ-deficient cells, demonstrating ready access of these ends to the alternative NHEJ pathway. However, when cleavage at the I-Sce1 sites was performed by a RAG2core-I-Sce1 fusion protein in the presence of core, full-length, or catalytically inactive RAG1, rejoining became strongly dependent on an intact classical NHEJ repair pathway (the effect was significantly weaker in the absence of coexpressed RAG1) (26). Interestingly, this strong dependency on classical NHEJ in the presence of RAG was not seen on a substrate containing only a single I-Sce1 site. Hence, tethering of the RAG complex to a heterologous endonuclease can prevent efficient end repair by any pathway other than classical NHEJ, and this effect is seen only at a pair of breaks (with four DNA ends), not at a single break (two ends) (26). Together, the data argue strongly that it is the RAG proteins themselves that prevent alternative NHEJ from participating in the repair phase of V(D)J recombination. Given the apparent propensity of alternative NHEJ to yield aberrant joining products, including
chromosomal translocations, this might play a significant role in ensuring the high fidelity of the V(D)J recombination reaction (19).

**END JOINING AND POSTCLEAVAGE COMPLEX STABILITY**

It is appealing to think that the ability of RAG to channel ends into the classical NHEJ repair pathway depends on the stability of postcleavage complexes (CSC, SEC, and the hypothetical complex containing the two coding ends) (Figure 2) and that premature end-release could result in the use of alternate repair pathways (19). The RAG1 K980A and RAG1 R838A/K839A/R840A mutants discussed above provide evidence for this: Their defect in constraining repair to the NHEJ pathway correlates with a severe defect in the stability of the SEC (83). Unfortunately, the weak association of coding ends with the CSC in vitro (2, 61) makes it difficult to assess the effect of RAG mutations on their retention in the complex. In addition, a recent study by Deriano et al. (33) reports that core RAG2 (residues 1–383) exhibits a defect in SEC stability and that core RAG2 knockin mice develop aggressive thymic lymphomas on a p53−/− background. The lymphomas that arise in core RAG2 × p53−/− mice are lethal much more quickly than those from p53−/− mice (mean survival approximately 12 weeks versus approximately 23 weeks) and contain recurrent translocations involving the Tcrα/Tcrδ and Igκ loci, which are not observed in thymic lymphomas from p53−/− mice. Given that core RAG2 allows for low level bypass of the requirement for classical NHEJ (24), it is plausible that the translocations in core RAG2 × p53−/− thymic lymphomas are due to a destabilized postcleavage complex that allows released ends to engage alternative repair pathways (33).

Further support for this idea comes from the analysis of the ATM (ataxia telangiectasia mutated) protein, an important DNA damage response factor and tumor suppressor. ATM is important for the stability of one or more RAG postcleavage complexes in vivo (20), and ATM-deficiency predisposes to the development of thymic lymphomas in both mice and humans (117). Interestingly, the translocations and lymphomas seen in RAG2 × p53−/− mice closely resemble those of Atm−/− mice (33). Hence, it is plausible that RAG-dependent genome instability arises from similar mechanisms in core RAG2 × p53−/− and Atm−/− mice, and that the common denominator is an unstable postcleavage complex. This idea is strengthened further by the recent characterization of a postcleavage defect during V(D)J recombination in cells lacking either histone H2AX or MDC1 (mediator of DNA damage checkpoint 1), two proteins in the ATM DNA damage response pathway. Artemis is the only nuclease normally able to process RAG-generated hairpin coding ends, but in the absence of H2AX or MDC1, a CtIP-dependent end resection pathway is able to gain access to coding ends (58). As a result, the ends suffer significant deletions and engage the alternative NHEJ repair pathway, leading to the model that ATM, H2AX, and MDC1 are important for genome stability during V(D)J recombination by restricting the access of aberrant repair pathways to the coding ends, particularly in cases where Artemis fails to open quickly the coding end hairpins (58).

It is important to recognize, however, that deletion of the entire RAG2 C-terminal region removes multiple important regulatory functions and perturbs more than just the stability of RAG postcleavage complexes. Hence, the phenotypes observed by Deriano et al. in core RAG2 × p53−/− mice might be influenced by altered RAG cell cycle regulation, targeting, activity, or cleavage site selection (29, 73, 151, and references therein) in addition to decreased postcleavage complex stability (33). Zhang et al. (151) recently revealed the importance of proper cell cycle regulation of RAG for the fidelity of V(D)J recombination through the creation and analysis of mice expressing RAG2 T490A, which cannot be destabilized in S-G2-M phase cells. As expected, these mice express RAG2 and accumulate signal end
intermediates in S-G2-M phase thymocytes, unlike thymocytes expressing wild-type RAG2. Importantly, RAG2 T490A × p53−/− and core RAG2 × p53−/− mice exhibit overlapping but distinct tumor phenotypes. Both genotypes develop thymic lymphomas containing translocations involving antigen receptor genes, but RAG2 T490A × p53−/− mice do not die more rapidly than p53−/− mice, whereas core RAG2 × p53−/− mice do (33, 151). Hence, the accelerated development of thymic lymphoma in core RAG2 × p53−/− mice cannot be explained by the loss of RAG cell cycle regulation, and can plausibly be attributed, at least in part, to destabilization of RAG postcleavage complexes.

A further link between postcleavage complex stability and strict channeling of V(D)J ends into the classical NHEJ repair pathway comes from the finding that certain alterations in the sequence of heptamer positions 4 or 5 destabilize the SEC in vitro and allow signal and coding ends to participate in alternative repair pathways in vivo (9). This is consistent with an earlier study that found that sequence variation in the RSS can influence the efficiency of coding joint formation at a postcleavage step of V(D)J recombination (1). Together, the two studies suggest that cleavage at cryptic RSSs or even certain bona fide RSSs can lead to inefficient coding end processing and breakdown of the postcleavage complex, leading either to apoptotic cell death or resolution by an alternative DNA repair pathway, which perhaps leads to a chromosomal translocation (1, 9).

Taken together, current findings suggest that stabilization of postcleavage complexes by RAG is important for proper channeling of ends into the classical NHEJ repair pathway and for genome stability, and that other factors (e.g., ATM, H2AX, MDC1) collaborate with RAG to achieve stability of postcleavage complexes and proper protection of the DNA ends. In addition, exclusion of RAG-generated ends from the S-G2 phases of the cell cycle, when homologous recombination is most active, is important for the prevention of RAG-induced genome instability.

MAKING THE CONNECTION TO CLASSICAL NONHOMOLOGOUS END JOINING

We currently have little insight into the mechanisms by which the RAG proteins direct coding and signal ends into the classical NHEJ repair pathway, and this is likely to be an area of active investigation in coming years. Several distinct types of nonmutually exclusive models can be envisioned. (a) The RAG proteins (or proteins they recruit) might exclude or inhibit the action of other repair pathways; (b) RAG might direct recombination to a compartment of the nucleus or a phase of the cell cycle in which classical NHEJ is favored; and (c) perhaps most appealingly, RAG might recruit classical NHEJ factors to the postcleavage complex. A recent study provides evidence for this third possibility by revealing an interaction between full-length RAG1 and Ku70/Ku80 (109). Under gentle conditions, the Ku complex copurifies with full-length RAG1/core RAG2 in a DNA-independent manner, and using such a protein preparation, a novel Ku-RAG-RSS complex can be detected by EMSA. Detection of this complex minimally requires residues 211–383 of the RAG1 N-terminal region. Ku does not detectably alter the DNA cleavage properties of RAG, suggesting a postcleavage function for the interaction. The authors suggest that by binding Ku, RAG1 helps coordinate the delivery of Ku to signal ends when the SEC is disassembled during signal joint formation (109). It is conceivable that the presence of Ku might also help stabilize postcleavage complexes. An interaction between RAG1 and MDC1 that involves determinants in the RAG1 N-terminal region as well as the extreme C terminus has also been detected (G. Coster, D. Chen, M. Goldberg, and D.G. Schatz, unpublished results). Although it is very tempting to think that RAG-mediated recruitment of repair factors after, or even before, DNA cleavage is important for the efficiency and fidelity of V(D)J recombination, definitive evidence for this has yet to be obtained.
SUMMARY POINTS

1. V(D)J recombination is thought to be initiated through the stepwise binding of one RSS by the RAG and HMGB1 proteins, followed by capture of an appropriate partner RSS. Both binding steps likely involve iterative sampling of DNAs to enable the RAG complex to bind, survey, and release suboptimal heptamer, spacer, and nonamer sequences.

2. The RAG proteins mediate RSS cleavage through a nick-hairpin mechanism. The RAG proteins may or may not nick the first RSS upon binding it, but hairpin formation generally occurs only upon synopsis, and is likely triggered through a mechanism involving base flipping at C1b (Figure 1a and Figure 4).

3. The order in which the RSSs are assembled into a synaptic complex, the choice of which RSSs are ultimately paired together, and the likelihood that an RSS is nicked prior to synopsis are probably determined by a combination of factors, including the sequence of the RSS itself, the accessibility of the locus in which it resides, and the potential for RAG-interacting proteins to recruit or stabilize RAG complexes at particular RSSs.

4. RAG1 plays a primary role in RSS recognition and catalyzing DNA cleavage. In addition, its noncore N terminus might promote the cleavage-to-joining transition through association with Ku70/Ku80, or by mediating ubiquitylation of KPNA1 or histone H3 to influence RAG1 subcellular localization or to facilitate detection or repair of DNA breaks, respectively.

5. The RAG1 noncore C terminus might help suppress RAG complex activity through interactions with the RAG2 PHD finger. This suppression might be partly relieved upon binding of full-length RAG2 to H3K4me3.

6. RAG2 collaborates with RAG1 to mediate DNA contact near the cleavage site. Determinants around RAG2 residue 360 help stabilize postcleavage complexes to prevent alternative repair mechanisms from accessing RAG-mediated DNA breaks. The RAG2 PHD finger mediates recognition of H3K4me3, which might promote RAG recruitment to active loci and stimulate RAG DNA binding and cleavage activity.

7. The RAG proteins act as guardians of the genome and as tumor suppressors, probably by multiple mechanisms. These include the stabilization of postcleavage complexes, the channeling of coding and signal ends into the classical NHEJ repair pathway, and the restriction of RAG-generated broken ends to the G1/G0 phases of the cell cycle.

FUTURE ISSUES

1. How does RAG2 contribute to DNA binding and cleavage by the RAG complex? Does it cause a conformation change in RAG1?

2. What is the mechanistic basis of the 12/23 rule?

3. How does nicking at one RSS enable hairpin formation at the other, and how do the two active sites communicate with one another? How often does nicking occur outside of the context of the PC in vivo?
4. What are the stoichiometry and structure of the SC, and how does this structure change when a second RSS is captured to form the PC? How structurally dynamic are the SC and PC? How does the PC change configuration after cleavage?

5. How does binding of the RAG2 PHD finger to H3K4me3 increase RAG catalytic activity? What is the mechanistic basis of the inhibition mediated by the RAG2 noncore region?

6. What are the functionally important targets of RAG1 E3 ubiquitin ligase activity and how does ubiquitylation contribute to V(D)J recombination?

7. By what mechanisms does RAG ensure that coding and signal ends are joined by the classical NHEJ repair pathway? How does RAG collaborate with other factors to stabilize postcleavage complexes during V(D)J recombination?

8. What are the critical parameters that determine whether a RAG-RSS interaction will lead to a stable SC or whether partner capture will lead to a stable PC? How stable are the SC and PC in vivo?

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

The authors apologize that not all of the relevant literature could be cited due to limitations on the number of references. This work was supported in part by NIH grants R37 AI32524 to D.G.S. and 3R56 AI055599-06A1S1 to P.C.S. D.G.S. is an investigator of the Howard Hughes Medical Institute.

LITERATURE CITED

79. Kriatchko AN, Bergeron S, Swanson PC. 2008. HMG-box domain stimulation of RAG1/2 cleavage activity is metal ion dependent. BMC Mol. Biol. 9:32
82. Landree MA, Wibbenmeyer JA, Roth DB. 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.* 13:3059–69


126. Steen SB, Gomelsky L, Roth DB. 1996. The 12/23 rule is enforced at the cleavage step of V(D)J recombination in vivo. *Genes Cell.* 1:543–53


144. Wu C, Bassing CH, Jung D, Woodman BB, Foy D, Alt FW. 2003. Dramatically increased rearrangement and peripheral representation of Vβ14 driven by the 3′D1β1 recombination signal sequence. *Immunity* 18:75–85


RELATED RESOURCES


Contents

Comparative Genetics and Genomics of Nematodes: Genome Structure, Development, and Lifestyle
Ralf J. Sommer and Adrian Streit .................................................. 1

Uncovering the Mystery of Gliding Motility in the Myxobacteria
Beiyan Nan and David R. Zusman .................................................. 21

Genetics and Control of Tomato Fruit Ripening and Quality Attributes
Harry J. Klee and James J. Giovannoni ........................................... 41

Toxin-Antitoxin Systems in Bacteria and Archaea
Yoshihiro Yamaguchi, Jung-Ho Park, and Masayori Inouye .................. 61

Genetic and Epigenetic Networks in Intellectual Disabilities
Hans van Bokhoven .................................................................. 81

Axis Formation in Hydra
Hans Bode ................................................................................. 105

The Rules of Engagement in the Legume-Rhizobial Symbiosis
Giles E.D. Oldroyd, Jeremy D. Murray, Philip S. Poole, and J. Allan Downie 119

A Genetic Approach to the Transcriptional Regulation
of Hox Gene Clusters
Patrick Tschopp and Denis Duboule .............................................. 145

V(D)J Recombination: Mechanisms of Initiation
David G. Schatz and Patrick C. Swanson ........................................ 167

Human Copy Number Variation and Complex Genetic Disease
Santosh Girirajan, Catarina D. Campbell, and Evan E. Eichler .................. 203

DNA Elimination in Ciliates: Transposon Domestication
and Genome Surveillance
Douglas L. Chalker and Meng-Chao Yao ........................................ 227

Double-Strand Break End Resection and Repair Pathway Choice
Lorraine S. Symington and Jean Gautier ....................................... 247
CRISPR-Cas Systems in Bacteria and Archaea: Versatile Small RNAs for Adaptive Defense and Regulation  
_Devaki Bhaya, Michelle Davison, and Rodolphe Barrangou_ .............................. 273

Human Mitochondrial tRNAs: Biogenesis, Function, Structural Aspects, and Diseases  
_Tsutomu Suzuki, Asuteka Nagao, and Takeo Suzuki_ ................................. 299

The Genetics of Hybrid Incompatibilities  
_Shamoni Maheshwari and Daniel A. Barbash_ ................................. 331

Maternal and Zygotic Control of Zebrafish Dorsoventral Axial Patterning  
_Yvette G. Langdon and Mary C. Mullins_ ................................. 357

Genomic Imprinting: A Mammalian Epigenetic Discovery Model  
_Denise P. Barlow_ ................................. 379

Sex in Fungi  
_Min Ni, Marianna Feretzaki, Sheng Sun, Xuying Wang, and Joseph Heitman_ ...... 405

Genomic Analysis at the Single-Cell Level  
_Tomer Kalisky, Paul Blainey, and Stephen R. Quake_ ................................. 431

Uniting Germline and Stem Cells: The Function of Piwi Proteins and the piRNA Pathway in Diverse Organisms  
_Celina Juliano, Jianquan Wang, and Haifan Lin_ ................................. 447

Errata

An online log of corrections to _Annual Review of Genetics_ articles may be found at http://genet.annualreviews.org/errata.shtml