The RAG Proteins and V(D)J Recombination: Complexes, Ends, and Transposition

Article *in* Annual Review of Immunology · February 2000

DOI: 10.1146/annurev.immunol.18.1.495 · Source: PubMed

5 authors, including:

Sebastian D. Fugmann
Chang Gung University
88 PUBLICATIONS **4,824** CITATIONS

SEE PROFILE

Penny E Shockett
Southeastern Louisiana University
30 PUBLICATIONS **2,047** CITATIONS

SEE PROFILE

Annu. Rev. Immunol. 2000. 18:495–527

THE RAG PROTEINS AND V(D)J RECOMBINATION: Complexes, Ends, and Transposition

Sebastian D. Fugmann[1], Alfred Ian Lee,
Penny E. Shockett, Isabelle J. Villey, and David G. Schatz
Howard Hughes Medical Institute, Section of Immunobiology, Yale University School of Medicine, New Haven, Connecticut 06520–8011; e-mail: david.schatz@yale.edu

Key Words antigen receptor, site-specific recombination, transposition, RAG1, RAG2

■ **Abstract** V(D)J recombination proceeds through a series of protein:DNA complexes mediated in part by the RAG1 and RAG2 proteins. These proteins are responsible for sequence-specific DNA recognition and DNA cleavage, and they appear to perform multiple postcleavage roles in the reaction as well. Here we review the interaction of the RAG proteins with DNA, the chemistry of the cleavage reaction, and the higher order complexes in which these events take place. We also discuss postcleavage functions of the RAG proteins, including recent evidence indicating that they initiate the process of coding end processing by nicking hairpin DNA termini. Finally, we discuss the evolutionary and functional implications of the finding that RAG1 and RAG2 constitute a transposase, and we consider RAG protein biochemistry in the context of several bacterial transposition systems. This suggests a model of the RAG protein active site in which two divalent metal ions serve alternating and opposite roles as activators of attacking hydroxyl groups and stabilizers of oxyanion leaving groups.

INTRODUCTION

The genes encoding immunoglobulin and T cell receptor proteins are unique in being split into multiple gene segments in the germline that are then made contiguous by recombination in somatic tissues. The assembly process, known as V(D)J recombination, is named for the V (variable), D (diversity), and J (joining) gene segments that are the targets of the reaction. For assembly of an antigen receptor gene, one V, one J, and in some cases one D gene segment are joined by recombination to create an exon that encodes the antigen binding portion of

[1]The first four authors are listed alphabetically and contributed equally to this work.

0732–0582/00/0410–0495$14.00

the receptor chain. Following transcription, the V(D)J exon is spliced to one or more exons encoding the constant region to produce the mature mRNA and subsequently the receptor polypeptide.

In many species, including primates and rodents, V(D)J recombination is responsible for generating a great deal of the diversity found in these receptors. Such diversity arises from two sources. The first, combinatorial diversity, is a consequence of the fact that there are typically many different V, D, and J gene segments and that each different V(D)J combination yields a different receptor specificity. The second, junctional diversity, arises from imprecise joining of the V, D, and J gene segments.

The sites of recombination are specified by recombination signal sequences (RSSs) that immediately flank each gene segment. Each RSS consists of a highly conserved 7-bp sequence (the heptamer; consensus 5′-CACAGTG) and an AT-rich 9-bp sequence (the nonamer; consensus 5′-ACAAAAACC) that are separated by a poorly conserved spacer whose length is either 12 ± 1 or 23 ± 1 bp. Spacer length therefore defines two types of RSSs, termed the 12-RSS and the 23-RSS. Efficient recombination occurs only between a 12-RSS and a 23-RSS, a restriction known as the 12/23 rule.

V(D)J recombination of the endogenous antigen receptor gene loci is a complex and highly regulated process. Events can span more than a megabase and can occur in lineage and developmental stage-specific patterns. Factors such as nucleosomal positioning and higher order chromatin structure are likely to play critical roles in regulating the reaction (for reviews, see 1, 2). Relatively little is understood about how these forces regulate V(D)J recombination, and they are not discussed further here. Instead, we consider the reaction at a much simplified level, that of artificial DNA substrates containing one or two RSSs, typically analyzed in cell-free reaction systems.

At this level, the reaction can be considered to occur in two phases (Figure 1). In the first, the two RSSs are recognized by the recombination machinery, brought into close juxtaposition (synapsis), and the DNA is cleaved precisely between the RSSs and their flanking coding elements. This generates four free ends: two blunt, 5′-phosphorylated signal ends and two covalently sealed, hairpin coding ends. In the second phase of the reaction, the coding ends are processed, often with the loss and addition of a small number of nucleotides, and joined to form a coding joint (CJ), while the signal ends are joined, typically precisely, to form a signal joint. It is also possible, though less frequent, for a signal end to become joined to a coding end. If the signal end is joined to the coding end to which it was previously connected, an open/shut (O/S) joint results, whereas if it is joined to the opposite coding end, a hybrid joint results.

The first phase of V(D)J recombination can be performed in its entirety (3, 4) by the proteins encoded by the recombination activating genes *RAG1* and *RAG2* (5, 6). The high mobility group proteins –1 and –2 (HMG1/2) also make an important contribution to this phase of the reaction (7, 8). The RAG1 and RAG2 proteins are coexpressed only in cells of the B and T lymphocyte lineages and

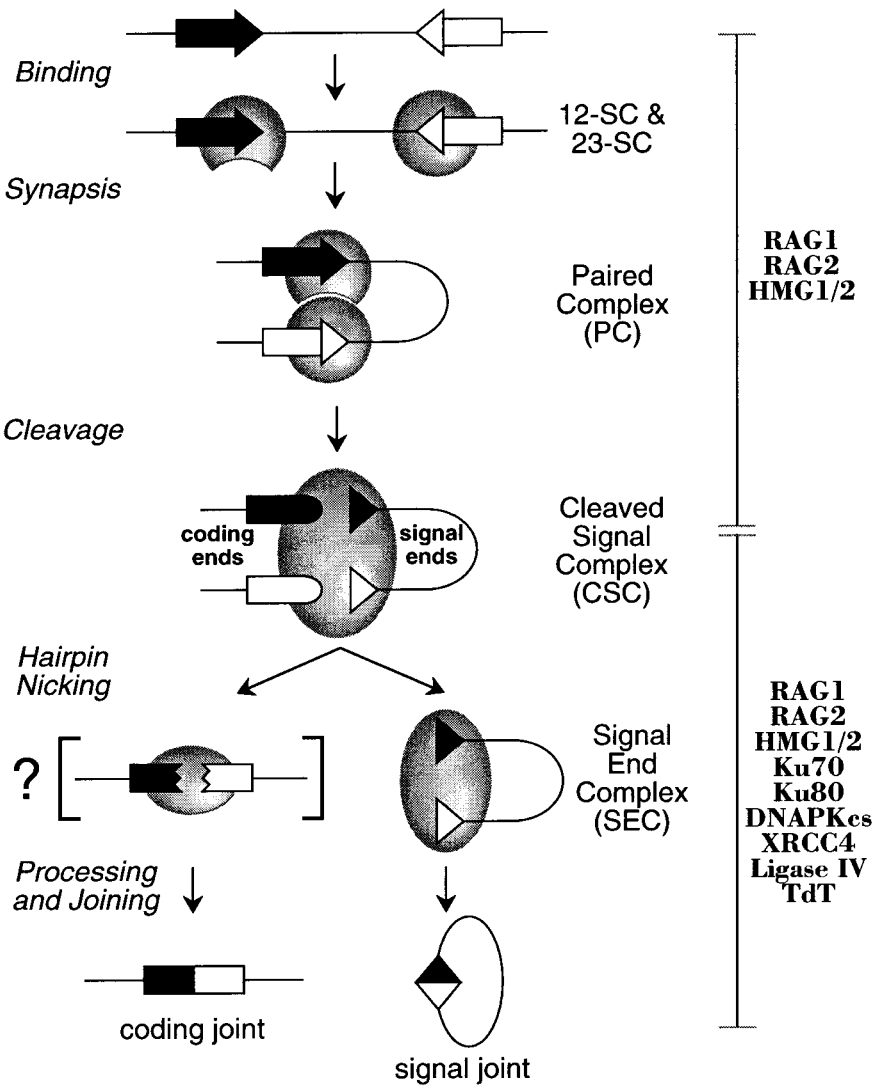

Figure 1 Schematic model of the protein-DNA complexes in V(D)J recombination. See text for details. The 12-RSS and 23-RSS are represented as black and white triangles, respectively, coding segments as rectangles and proteins as shaded ovals. Several aspects of the reaction are not depicted, including nicking adjacent to RSSs (which may occur before or after synapsis), asymmetric opening of the hairpin coding ends to generate P-nucleotides, and nucleotide addition by TdT. Coding end processing likely occurs in the context of the CSC, but the existence of a complex containing just the coding ends (brackets) cannot be ruled out. Figure adapted from (24) and relies also on data from (7, 31, 65).

are each essential for cleavage activity (9–11). Mutagenesis studies have defined minimal ''core'' regions of each RAG protein required for catalytic activity (12–15). While it is clear that the ''nonessential'' regions of the RAG proteins make important contributions to their activity (16–19), these roles are not well understood. One striking feature of the *RAG* genes is that they are located immediately adjacent to one another in the genomes of all jawed vertebrate species examined to date, and in most species they lack introns in their coding regions (20).

The second phase of V(D)J recombination is not well understood but appears to require RAG1, RAG2, and a group of ubiquitous DNA repair proteins. The enzyme terminal deoxynucleotidyl transferase (TdT) is not essential, but when present it contributes substantially to receptor diversity by adding nontemplated (N) nucleotides to coding junctions (21, 22). The essential DNA repair proteins include the three components of the DNA-dependent protein kinase (DNAPK), Ku70, Ku80, and the kinase catalytic subunit DNAPKcs, and the XRCC4 and DNA ligase IV proteins (23 and references therein). A deficiency in any of these factors results in defective V(D)J recombination, hypersensitivity to ionizing radiation, and an early block in lymphocyte development. A full consideration of these proteins is beyond the scope of this review.

Recent biochemical studies have demonstrated that V(D)J recombination probably proceeds through a series of well-defined protein:DNA complexes (Figure 1). The RAG proteins first recognize the 12-RSS and 23-RSS to form stable complexes referred to as the 12-SC and 23-SC (nomenclature adapted from 24). Synapsis of these complexes leads to formation of the paired complex (PC), within which DNA cleavage is completed. After cleavage, the four ends are held in a cleaved signal complex (CSC), which is likely the complex within which much or all of CJ formation occurs. This leaves behind the quite stable signal end complex (SEC), which is the precursor for signal joint formation. In this review, we focus on the properties and activities of the RAG proteins, particularly the protein:DNA complexes in which they participate. We also discuss the recent finding that RAG1 and RAG2 constitute a transposase (25, 26) and compare the properties of this transposase-become-recombinase with several well-studied bacterial transposases.

RECOGNITION OF THE RSS

RAG-RSS Interaction

The first step in V(D)J recombination generates a stable complex of the RAG proteins and the RSS (12-SC or 23-SC; Figure 1), and both the heptamer and nonamer make important contributions to complex formation. An in vivo one-hybrid system and a surface plasmon resonance analysis revealed an essential role of the nonamer for detectable binding (27, 28). This is mediated by an interaction with the NBD (''nonamer-binding-domain'', aa 390–460) of RAG1. RAG1 alone

binds the 12- and 23-RSSs with approximately equal affinities, and the heptamer makes a small but clearly detectable contribution to this interaction (27, 29). RAG2 by itself displays no detectable binding activity, but the 12-SC and 23-SC, containing RAG1 and RAG2, are much more stable and sequence-specific than complexes containing RAG1 alone (30–32). Overall, these results suggest that the RAG-RSS interaction may involve two steps: a ''recruitment'' step in which the nonamer acts as an anchoring motif; and a ''stabilization'' step, in which the heptamer, in the presence of RAG2, now supports a larger part of the interactions, closer to the coding flank and the site of cleavage. In all cases, formation of the 12-SC or 23-SC requires a proper spacer length and the presence of a divalent metal cation (31, 33).

Interference and Footprinting Studies Interactions of the RAG proteins with the RSS have been studied at the nucleotide level by methylation and ethylation interference and footprinting assays. Interaction of RAG1 with the RSS is clearly evident (30, 34, 35), and ethylation interference studies (34) revealed that this interaction is contributed by nonspecific interactions with the DNA backbone, especially in the spacer region, as well as by base-specific interactions, restricted to the nonamer. Strong overall nonamer occupancy is also observed within the RAG1-RAG2-RSS complex, and in this situation the protection extends through the spacer to the spacer-proximal side of the heptamer (Figure 2, see color insert). In contrast with what was observed with RAG1 alone, interactions outside the nonamer are then more base-specific, consistent with the observation that RAG2 and the heptamer favor a more stable complex. For the 23-RSS, the interaction with the nonamer is evident in a RAG1-RAG2 complex, but the protein:DNA contacts do not propagate as far toward the heptamer as with the 12-RSS. This fits well with the observation that the 23-SC is less stable than the 12-SC (31) and implies that additional factor(s) are required to stabilize the 23-SC (see below). Strikingly, in any situation, the interactions are biased toward one face of the helix throughout the RSS (34) (Figure 2). This could explain why cleavage requires appropriate phasing (an integral number of helical turns) between the heptamer and the nonamer. Indeed, changing the length of the spacer by more than ± 1 bp dramatically reduces cleavage and recombination (36–40).

UV Cross-Linking Footprinting studies have failed to reveal any interaction of the RAG proteins with nucleotides near the coding flank/heptamer border. Since this is the site at which cleavage occurs, there must be protein:DNA interactions in this region, but they are likely to be transient or weak. Indeed, several studies have suggested that RAG1 interacts with the coding flank (31, 41, 42). To detect such weak interactions, we designed a UV cross-linking assay that involved substituting 5-iodouridine in place of thymidine at several positions of the coding flank and the heptamer in RSS oligonucleotides (43). RAG1 was shown to be cross-linked at positions flanking the site of cleavage (Figure 2): the first position of the coding flank on the bottom strand (C-1B), the second position of the heptamer on the bottom strand (H2B), and the third position of the heptamer on the

top strand (H3T). In the presence of RAG1, RAG2 could be cross-linked at the C-1B and H3T positions, although much more weakly than RAG1 (Figure 2). Differences in intensity of cross-linking may simply reflect the selectivity of the assay. Such biases include the facts that the 5-position of pyrimidines lies in the major groove of the DNA helix, and that iodo-groups preferentially cross-link to aromatic residues (44). Cross-linking of RAG1 and RAG2 to the RSS was also observed using aryl-azide-modified bases (45). However, these results are more difficult to interpret because the azido group allows cross-linking to more distant residues. Another study using iodo-groups confirmed the RAG1-heptamer interaction but did not reveal cross-linking to RAG2 (32).

Interestingly, the C-1T position (first position of the coding flank on the top strand) was also strongly cross-linked to RAG1 in our assay, but only if the DNA of the coding flank was unpaired (which may mimic the DNA melting that is thought to occur during cleavage). It is reasonable that RAG1 would establish additional DNA contacts near the site of cleavage to stabilize the unpaired state. Severe distortion of the DNA at the coding flank/heptamer border favors binding (34), and unpairing of the coding flank has been shown to enhance cleavage (36, 39). Taken together, these studies suggest that single-stranded coding flank/heptamer borders are recognized as well as, if not better than, double-stranded borders, which implies that the RAG proteins are able to bind and cleave single-stranded DNA. Consistent with this, cleavage of single-stranded RSSs occurs (36, 39), and our unpublished data show that RAG1 and perhaps RAG2 can be cross-linked by UV light to a single-stranded RSS containing an iodo group at position C-1B (IJVilley and DG Schatz, unpublished). Together, the results demonstrate that the RAG proteins contact and surround the site of cleavage.

Stoichiometry and Configuration of Subunits The composition of the RAG-RSS complex has been investigated recently, revealing that core RAG1 exists as a homodimer in solution and retains its dimeric form upon binding to the RSS (29, 45, 46), as was suggested by a previous study (33). One study found that RAG2 forms multimers, particularly dimers, in solution, and that the 12-SC consists of a tetramer of two molecules of each RAG protein bound to a 12-RSS (46). This conflicts with other data which suggest that only a monomer of RAG2 is found in the RAG1-RAG2-RSS complex (45).

The configuration of the different subunits within the complex remains hypothetical. In a first model, one RAG1 molecule contacts both the nonamer and the heptamer/coding flank. Alternatively, two molecules of RAG1 might interact with one RSS, with one contacting the nonamer and the other the heptamer/coding flank (29, 45).

Role of HMG1/2

How are the RAG proteins able to recognize the 12-RSS and 23-RSS given the extra helical turn found in the latter? Parallels with other recombination systems suggested the involvement of a DNA-bending accessory factor, and it has been

shown that HMG1/2, very abundant and ubiquitous DNA-binding and bending proteins, enhance binding and cleavage by the RAG proteins (7, 8, 47, 48). Their effect, however, is different with the two types of RSSs. Formation of the 12-SC with purified RAG proteins is relatively efficient and only slightly improved by the addition of HMG1/2. In contrast, formation of the 23-SC is stimulated over tenfold by HMG1/2 and is then formed as efficiently as the 12-SC (7). The interaction of HMG1/2 with RAG1 (see below) can only account for a portion of the effect on 23-SC formation (29).

How might HMG proteins stimulate formation of the 23-SC? One possibility is that they act to bend the DNA in the spacer region, thereby bringing the nonamer and heptamer elements closer together and allowing the RAG proteins to more easily contact both elements simultaneously (7). However, a recent study has shown that RAG1 and RAG2 by themselves induce a bend in the RSS, whose magnitude (about 60°) is not increased by the addition of HMG1/2 (49). It was therefore proposed that HMG1/2 are incorporated into the complex to stabilize it through the stabilization of the bend. In this model, the RAG proteins would bind and bend the DNA, and the HMG1/2 protein would behave like a clamp to ensure a durable and favorable bending of the RSS.

HMG1/2 contain two HMG boxes and an acidic C-terminal tail, with the HMG boxes involved in both protein:DNA and protein:protein interactions (50). Interestingly, the protein:protein interactions invariably involve the DNA-binding-domain of the partner protein as well. In keeping with this, it is the NBD of RAG1 that interacts with HMG1, and both HMG boxes are required for the binding (49). Based on sequence and functional similarities with the Hin recombinase, the NBD of RAG1 has been postulated to consist of a homeobox-like domain containing a GGRPR motif and three α-helices (27, 28).

HMG1/2 display no sequence-specific DNA-binding properties. Instead, they recognize unusual structures, such as DNA bends, and themselves induce sharp bends (80°) in the helix upon binding. It is now clear that HMG1/2 interact with the DNA binding domains of many transcription factors (51). Although these factors are able to bind their target sequence by themselves, the addition of HMG proteins typically greatly strengthens their binding. It is interesting that steroid hormone receptors, and not other hormone receptors, are stimulated by HMG1/2 (52). The DNA binding domain of nonsteroid nuclear receptors is composed of three α-helices that contact both major and minor grooves in the target sequence, whereas for the steroid hormone receptors, only two α-helices are involved, with contacts restricted to the major groove. Because the HMG1/2 proteins bind DNA in the minor groove, it is plausible that they provide an additional interaction surface for the steroid hormone receptors, but they would compete with the third helix of the nonsteroid hormone receptors (52). RAG1 interacts extensively with both the major and minor grooves of the nonamer (34, 35). In contrast, in the RAG1-RAG2-RSS complex, the interaction with the heptamer/spacer border appears to involve the major groove only. It is tempting to think that HMG1/2

may stabilize this complex by providing minor groove DNA contacts in vicinity of the heptamer.

SYNAPSIS, CLEAVAGE, AND THE 12/23 RULE

Chemistry of the Cleavage Reaction

The cleavage reaction occurs in two steps and introduces a DNA double-strand break between the coding gene element and the flanking RSS. In the first step (nicking), a single-strand break is introduced on the top strand between the gene element and the first nucleotide of the heptamer. In the subsequent hairpin formation step, the bottom strand is cleaved, creating a hairpin structure at the coding end and a blunt, 5′-phosphorylated signal end.

Nicking The phosphate ester between the last nucleotide of the coding element and the first nucleotide of the heptamer on the top strand is hydrolyzed, creating a nick characterized by a 3′-OH group at the end of the coding element and a 5′-phosphorylated end at the heptamer (Figure 3, see color insert) (3). The exact reaction mechanism has not been characterized yet, but the analysis of nicking reactions catalyzed by other site-specific recombinases suggests two possibilities. The first is a one-step reaction in which the recombinase proteins catalyze the direct hydrolysis of the phosphate ester. The second is a two-step reaction in which a serine or tyrosine residue of the recombinase protein acts as a nucleophile in a transesterification reaction, creating a covalent protein-DNA linkage, as occurs with bacteriophage lambda integrase, resolvases, and invertases (reviewed in 53). Subsequent hydrolysis of this ester generates the nicked product.

The comparison of the absolute configuration of the central phosphor atom before and after the reaction would distinguish between the two different mechanisms. One-step hydrolysis is a single S_N2 reaction leading to inversion of the configuration of the phosphor atom, whereas the two-step mechanism consists of two S_N2 reactions leading to retention of the configuration. For HIV-1 integrase it was shown, by substituting one of the nonbridging oxygen atoms of the phosphate with sulfur (creating a chiral phosphorothioate), that the configuration of the phosphor atom was inverted after the nicking reaction (54). But it is important to note that, for the analyzed reaction products, the 3′-OH group at the end of the top strand was used as the nucleophile instead of water. Therefore, strictly speaking, the direct hydrolysis mechanism still remains a conclusion drawn from indirect evidence.

The use of a phosphorothioate ester to determine the mechanism of RAG-mediated nicking was not successful because the ester was not cleaved by the RAG proteins (33). The failure, thus far, to isolate a covalently linked RAG-DNA complex suggests that the RAG proteins catalyze direct hydrolysis analogous to

HIV-1 integrase. It is conceivable, however, that the covalently linked intermediate is very unstable and thus quickly hydrolyzed.

Hairpin Formation Hairpin formation by the RAG proteins occurs by direct transesterification (55). The top strand 3′-OH group acts as a nucleophile attacking the central phosphor atom of the phosphate on the lower strand in a S_N2 reaction (Figure 3), with the lower strand of the heptamer serving as the leaving group. Hairpin formation also occurs starting from a pre-nicked substrate, indicating that the nicked product is a real intermediate of the cleavage reaction (3, 56).

Attack of the bottom strand by the 3′-OH of the top strand could not occur without a significant bend in one or both strands of the DNA. Some coding flank sequences inhibit cleavage at an isolated RSS in Mn^{2+} in vitro, selectively at the hairpin formation step (36, 39). Unpairing of the first two positions of the coding flank (C-2 and C-1) usually restores hairpin formation on these substrates, suggesting that flexibility of the DNA at the site of cleavage is mandatory. Because the RAG proteins have been shown to contact the site of cleavage, one can imagine that they are directly involved in its physical modification. In other recombination systems, unpairing of the flanks facilitates the strand transfer reaction (57, 58).

The Precleavage Synaptic Complex (Paired Complex)

In general, V(D)J recombination occurs effectively only on 12/23 RSS pairs (59). Given that gene segments in the antigen receptor loci are separated by up to a megabase, synapsis of the two RSS to generate the PC is critical for reaction fidelity and coordinate cleavage. RAG1 and RAG2 form multimeric complexes in vivo (60–62) and in vitro (45, 46), and it is likely that the two RSSs in the PC are held together by the RAG proteins (24, 46). Recently the PC has been detected and isolated in vitro by blocking the nuclease activity of the RAG proteins using Ca^{2+} as the divalent metal ion (24). This complex was only detected in the presence of HMG1. The catalytic activity of the RAG proteins in this ''captured'' PC was restored by adding an excess of Mg^{2+}, suggesting that the isolated complex had a structure similar to that of the PC formed in the presence of Mg^{2+} ions.

Alignment of RSSs and Stoichiometry In the PC, RSSs may be aligned in a parallel, antiparallel, or some intermediate orientation. Thus far, this issue has not been investigated using imaging techniques such as electron microscopy or atomic force microscopy. However, the influence of the distance between two RSSs on the efficiency of the reaction has been studied in vitro (37) and in vivo (63). The steric constraint imposed by a parallel alignment of the RSS is stronger for inversional than deletional substrates. Indeed, with inversional substrates, cleavage and recombination were more sensitive to shorter inter-RSS distances than with dele-

tional substrates (37, 63), suggesting a parallel alignment (or something close to it) of the RSSs in the PC.

Except for the presence of two RSSs, the stoichiometry of the PC is unknown. As described above, the complex assembled on a single RSS contains two molecules of RAG1 and either one (45) or two (46) molecules of RAG2. From symmetry considerations, it seems likely that the minimum protein content of the PC is two molecules of each RAG protein, although this requires each molecule of RAG1 to be able to interact with a different RSS. Another reasonable possibility is that the PC contains four molecules each of RAG1 and RAG2. A similar situation exists for the PC of the Mu transposon, which contains two MuA homodimers, and only one subunit of each MuA dimer uses its active site during the transposition reaction (64). HMG1/2 is likely to be a component of the PC, since it is a stable component of at least two postcleavage complexes, the SEC (65) and the strand transfer complex that arises from RAG-mediated transposition (25). In addition, HMG2 is stably incorporated into an RSS-RAG1 complex (29).

Cleavage Within the Paired Complex In vitro analyses of the nicking reaction in Mg^{2+} (the physiological divalent cation) show that nicking occurs readily on substrates containing a single RSS (31, 66), or on substrates containing two RSSs separated by short distances that prevent synapsis within the same molecule (66). In these latter substrates, the fraction of doubly-nicked RSSs is equal to the product of the fractions of singly-nicked RSSs. However, as the distance between the two RSSs increases, the fraction of doubly-nicked RSSs also increases. Thus, nicking at two RSSs is a largely uncoupled phenomenon, although synapsis of 12- and 23-RSSs substantially stimulates nicking at both RSSs.

By contrast, in vitro studies on hairpin formation show that products in which both RSSs are fully cleaved appear before products in which only one RSS is cleaved (37), suggesting that single-end cleavage products result from off-pathway reactions. Moreover, although mutations in one RSS do not significantly affect nicking at the partner RSS, such mutations can impair hairpin formation at both RSSs (24, 37, 66). Hence, it appears that hairpin formation at the two coding flank/heptamer junctions is both physically and temporally coupled.

This conclusion has recently gained strong support from in vitro cleavage experiments in Mg^{2+} using RSS oligonucleotide substrates in which the sequence of the coding flank was varied (66a). Certain coding flank sequences can dramatically inhibit recombination by wild-type RAG proteins in vivo (66b) and cleavage by purified RAG proteins in vitro (66a). These coding flanks greatly slow the nicking step of the reaction but have no effect on hairpin formation if a prenicked substrate is used. Interestingly, if nicking at one RSS is inhibited hairpin formation (but not nicking) is inhibited at the partner RSS (66a). This indicates that nicking of both RSSs is required for hairpin formation at either one and that the proteins engaging one RSS can sense the status (nicked or unnicked) of the partner RSS.

A question that remains to be answered concerns whether RAG1 and RAG2 in the PC cleave the RSS to which they are bound (*cis* cleavage) or the opposite

RSS (*trans* cleavage). At first glance, the fact that nicking occurs in the absence of a proper PC and the results of cleavage experiments in Mn^{2+} (see below) seem to suggest that RAG proteins can perform both nicking and hairpin formation in *cis.* However, the arrangement of the RAG proteins and stoichiometry of RSSs within ''single-RSS'' complexes are unknown. Such complexes may represent half of the true synaptic complex, or they may represent the full synaptic complex with one RSS lacking. An alternative, albeit less likely, scenario is that they may be improper synaptic complexes generated between RSSs on different DNA molecules.

The 12/23 Rule The specific requirements for enforcement of the 12/23 rule, the precise step at which enforcement occurs, and the molecular mechanism that mediates it have all been subjects of considerable debate.

Initial in vitro studies demonstrated the importance of the divalent cation in conferring 12/23-coordinated cleavage. In the presence of Mn^{2+}, both RAG-containing lymphoid extracts (37, 56) and purified core RAG proteins (3) catalyze nicking and hairpin formation on substrates containing either 12/23 pairs of RSSs or isolated 12- or 23-RSSs; in substrates containing 12/23 pairs, single-RSS cleavages predominate. In Mn^{2+}, therefore, cleavage events at two RSSs are uncoupled. By contrast, in the presence of Mg^{2+}, lymphoid extracts show a 25- to 50-fold preference for double cleavage of 12/23 substrates over cleavage of 12/12 or 23/23 substrates (37), in good agreement with in vivo data (38). The traditional explanation for this ion dependency is that Mn^{2+} allows RAG1 and RAG2 to cleave at isolated RSSs, perhaps by altering the geometry of the active site in relation to the substrate. Indeed, Mn^{2+} skews both the temperature and pH activity profiles of the RAG proteins (33), suggesting that the architecture of the RAG-RSS complex in Mn^{2+} is distorted. Interestingly, Tn10 cleavage in the presence of low levels of Mn^{2+} yields an accumulation of single-end events (67), although cleavage is still believed to occur within the context of a PC.

Purified core RAG proteins in the presence of Mg^{2+} show only a three- to fivefold preference for 12/23 double cleavage versus 12/12 double cleavage (4); single cleavage events are still seen at low frequency and mainly on the 12-RSS. Addition of HMG1 or HMG2 has a moderate stimulatory effect on 23-RSS single cleavage and a large stimulatory effect on 12/23 double cleavage (7, 8), but this addition does not result in strict conformity to the 12/23 rule because substantial cleavage of 12/12 and 23/23 substrates still occurs. Full restoration of the 12/23 rule can be achieved by adding crude cellular extracts (8). Thus, it appears that both Mg^{2+} as the divalent cation and additional nonlymphoid factors including HMG are involved in full enforcement of 12/23-coordinated cleavage.

Recent studies have suggested that purified core RAG and HMG proteins may be the crucial protein factors required to fully recreate the 12/23 rule. Using short double-stranded oligonucleotides containing single RSSs, Hiom & Gellert showed that RAG1, RAG2, and HMG1 in Mg^{2+} cleave only 12/23 pairs of oligonucleotides and not 12/12 or 23/23 pairs (24). Kim & Oettinger constructed an

oligonucleotide substrate with a single-stranded region between the RSSs; using purified core RAG proteins, HMG1, and nonspecific double-stranded competitor DNA in the presence of Mg^{2+}, strict 12/23-coordinated cleavage is seen, with suppression of 12/12 or 23/23 events, suggesting that the effects of cell extracts might be at least partly due to nonspecific DNA (48). West & Lieber constructed a double-stranded oligonucleotide substrate with a nick between the RSSs; while purified core RAG proteins in Mg^{2+} have no cleavage activity on this substrate, strict 12/23-regulated double cleavage is seen upon addition of HMG1 (47).

The precise step at which the 12/23 rule is imposed remains to be determined. Using EMSA, Hiom & Gellert showed that in the presence of Ca^{2+}, purified core RAG proteins with HMG1 synapse 12/23 oligonucleotide pairs more efficiently than either 12/12 or 23/23 pairs, although both 12/12 and 23/23 oligonucleotides show a significant background level of synapsis (24). In contrast, West & Lieber showed that preincubation of their labeled 12/23 substrate with unlabeled 12/23, 12/12, or 23/23 competitor substrate reduces double hairpin formation in the labeled substrate 10- to 20-fold in all cases (47). This suggests that RAG and HMG1 proteins can synapse 12/23, 12/12, and 23/23 substrates, all with similar efficiencies, and that enforcement of the 12/23 rule occurs at the level of double hairpin formation.

Both studies raise important issues. It is possible that, in the study by Hiom & Gellert, the use of Ca^{2+} as the divalent cation alters the events that take place under normal physiological conditions. Also, assaying RAG-mediated synapsis by using isolated RSSs on two separate DNA molecules may be a less sensitive means of study, as lack of connectivity between the two RSSs automatically decreases the probability of synapsis. On the other hand, the nicked substrate used in West & Lieber's study may allow for artifactual synapsis. The RSSs in their substrate are separated by less than 70 bp, and the tight tethering of two RSSs combined with increased flexibility between the RSSs may allow for noncanonical synapsis to occur at increased efficiencies. Finally, it is possible that the 12/23 rule is imposed at both synapsis ($\geq$threefold) and double hairpin formation ($\geq$tenfold), generating the $\geq$30-fold preference for 12/23 substrates seen in vivo (38).

POSTCLEAVAGE SYNAPTIC COMPLEXES

Before they are joined, hairpin coding ends must be nicked open and nucleotides may be inserted or deleted. Nucleotide insertion usually results either from the action of TdT or from asymmetric opening of the hairpin to generate short stretches of palindromic (P) nucleotides (reviewed in 59). While N nucleotide addition occurs predominantly in postnatal animals, P nucleotides are found in V(D)J coding junctions at most loci in fetal and adult animals. The mechanism of nucleotide deletion from coding ends is unknown.

The Cleaved Signal Complex (CSC)

Immediately after cleavage, the RAG proteins are bound to the signal ends and coding ends in a four-end complex, the CSC (Figure 1). One model of coding end processing predicts that coding ends are held in this complex long enough for modifying enzymes, ligase, and necessary cofactors to be recruited and to act, and hence that CJ formation occurs in the CSC. Most of the available data is consistent with this model. Alternatively, it is possible that the two signal ends and two coding ends are associated in the CSC immediately after cleavage, but that the complex falls apart rapidly and coding ends are processed or joined independent of the SEC (Figure 1). Although the stability of association of coding ends in the CSC has rightly been questioned, the available experimental data argue against this model.

The earliest hint that coding ends and signal ends must exist at least transiently within a common protein:DNA complex came from studies that identified hybrid joints and O/S joints with high efficiency in transfected cells (68, 69). The ability of coding ends to be joined to signal ends implies that all four ends must exist within a common postcleavage complex. Additionally, the occurrence of a similar spectrum of coding end processing events in hybrid, O/S, and coding junctions supports the idea that coding ends and signal ends exist within a common complex during coding end processing (68–70).

Recently, direct physical evidence has been provided for the CSC (24). RAG-mediated cleavage in *trans* of RSS oligonucleotide substrates (one labeled with biotin and the other with ^{32}P) allowed streptavidin capture of ^{32}P-labeled signal ends and coding ends. Coding ends were captured less efficiently than signal ends, consistent with the failure to detect stable coding end/coding end or coding end/signal end complexes in an earlier study (65). Definitive proof of the four-end complex awaits experiments demonstrating that both RSSs have been cleaved in the captured complexes.

Hairpin Opening and the Processing of Coding Ends

Signal ends were the first V(D)J recombination-specific DNA intermediates to be detected in normal lymphoid precursors, and they are considerably more abundant than coding ends (71–75). A pre–B cell line with inducible RAG expression was used to show that coding ends are present at 10- to 100-fold lower levels than signal ends at the Ig Jκ locus (76). After RAG protein induction, CJ formation correlates with the appearance of signal ends, suggesting that while signal ends persist, coding ends are processed and joined rapidly after cleavage.

The RAG proteins appear to be necessary for coupling the cleavage and joining stages of V(D)J recombination. In vitro, coding ends generated by RAG-mediated cleavage cannot be joined if they are first deproteinized (77, 78). There are several possible postcleavage roles for the RAG proteins. First, they might help to maintain the structure or nuclear location of the postcleavage complex. Second, they

may be required for the recruitment of DNA processing or repair proteins. Finally, as discussed below, they may play a catalytic role in coding end processing or joining.

Ubiquitous Hairpin Opening Activities Activities have been detected that can resolve and join transiently transfected synthetic hairpin DNA substrates in a variety of cell lines (79, 80, and discussed in 81). These observations indicated that ubiquitous hairpin opening activities exist in vivo and raised the possibility that hairpin opening during V(D)J recombination is performed by a ubiquitous factor.

Recent work has suggested that the RAD50/Mre11/NBS DNA repair complex could potentially serve in this capacity (82, 83). In vitro, the human Mre11/RAD50/NBS complex possesses activities that can nick synthetic DNA hairpins, remove 3′ overhangs, and mediate 3′ to 5′ resection of DNA termini (83). This latter activity promotes homology-mediated ligation of DNA ends and could conceivably resect coding ends to promote joining stimulated by short homologies during V(D)J recombination (82). As yet, no direct involvement of RAD50/Mre11/NBS in V(D)J recombination has been demonstrated.

Hairpin Opening by the RAG Proteins Recent studies have suggested that hairpin opening during V(D)J recombination might be performed by the RAG proteins (84, 85). In vitro, the RAG proteins nick synthetic hairpins a few nucleotides 5′ of the tip, and the presence of both RAG1 and RAG2 is required. The hairpin structure is not required for this nuclease activity because a homologous DNA duplex substrate is nicked at the same position. With synthetic hairpins, HMG2 focuses the nicking activity to the vicinity of the hairpin tip, presumably by interacting with the RAG proteins and promoting their interaction with the altered DNA structure at the tip (85). HMG1/2 may perform a similar function in vivo. Like the disregulated cleavage of isolated RSS oligonucleotides, nicking of synthetic hairpins occurs in Mn^{2+} and not Mg^{2+} (84, 85).

Strikingly, hairpin nicking can occur in Mg^{2+}, either in *trans* on oligonucleotide substrates or in *cis* with plasmid substrates, but only in the context of 12/23 regulated cleavage. Using oligonucleotide substrates, coding end hairpins are opened at the tip (84). After coupled cleavage of plasmid substrates, however, coding end hairpins are opened predominantly 1 to 2 nt 5′ of the tip, generating palindromic extensions similar in length to the short P regions found in CJs in vivo (Figure 4) (85). Hairpin nicking in Mg^{2+} has also been observed using a large DNA fragment containing an RSS at one end and a hairpin at the other (PE Shockett, DG Schatz, unpublished). Since nicking of synthetic hairpin substrates (lacking a signal end) does not occur in Mg^{2+}, it is possible that transient synapsis occurs between the signal end bound by the RAG proteins and the coding end, to some extent mimicking the postcleavage complex. Additionally, synthetic hairpin nicking in Mg^{2+} appears to be stimulated by short signal end oligonucleotides

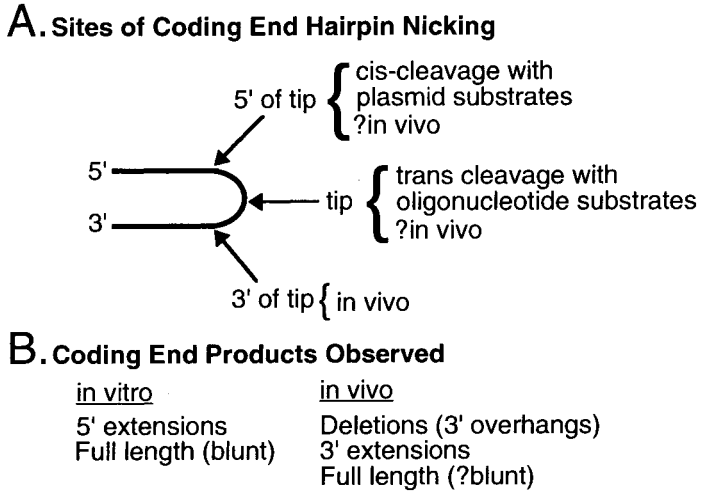

Figure 4 Summary of sites of hairpin nicking by the RAG proteins within the postcleavage complex (A) and types of coding ends detected in vivo (B).

(PE Shockett, DG Schatz, unpublished). Thus, in Mg^{2+}, organization of the RAG protein active site for hairpin nicking probably requires signal end binding.

These findings suggest that in vitro, the RAG proteins nick coding end hairpins within the CSC, generating full-length coding ends, P nucleotides, and conceivably deletions at coding ends. Similarities between mechanistic aspects of V(D)J recombination and bacterial transposition (see below) further support the hypothesis that the RAG proteins initiate both RSS cleavage and subsequent hairpin opening (86, 87). These similarities were further extended by the recent finding that both the RAG proteins and Tn10 transposase are sequence-nonspecific 3′ flap endonucleases (87a). Both 3′ extensions and 3′ flaps can be removed by endonucleolytic cleavage at or near the double strand/single strand junction. In addition to raising the possibility of an additional role for RAG1 and RAG2 in creating junctional diversity, the results suggest that sequence-specific nicking by the RAG proteins may proceed through a flap-like structure at the RSS-coding flank border (87a).

Questions Raised by RAG-Mediated Hairpin Opening One unresolved question is why oligonucleotide and plasmid substrates yielded different patterns of coding end hairpin opening (Figure 4) in the two studies cited above (84, 85). This difference may result from a combination of factors including the sequence, structure, and single-stranded character of the hairpin coding end, the substrate orientation, and the preparation of RAG proteins (88 and references therein). Preliminary data suggest that the differences observed in the two studies can be

accounted for by differences in both the substrates and the protein preparations used (85; P Cortes, personal communication).

A second question arises from the fact that while coding end hairpins are opened at, or 5′ of, the tip in vitro, coding ends in vivo exhibit deletions and extensions consistent with 3′ overhangs and nicking of the hairpin 3′ of the tip (Figure 4) (74–76). Whether the ends detected in vivo represent products of the initial coding end hairpin processing event and a true intermediate in V(D)J recombination remains to be shown. A possible precursor-product relationship between some of the deleted ends detected in vivo and deletions found within actual CJs has been pointed out (75). In vivo, both deletion and P nucleotide formation are influenced by the coding end sequence (88, 89 and references therein). It has been proposed that double-stranded hairpin nicking at AT-rich sequences serves as a mechanism of nucleotide loss from coding ends (88). The low recovery of coding ends with nucleotides deleted on both strands in vitro after RAG-mediated cleavage might be a function of the coding end sequences used or the need for a more stable and prolonged hairpin:RAG association for double-stranded nicking of the hairpin. Additionally, the studies of RAG-mediated hairpin opening in vitro have been performed with truncated RAG proteins, and it is possible that the full-length proteins will behave differently. We have observed synthetic hairpin nicking in the presence of full-length RAG2, but full-length RAG1 has not yet been examined (PE Shockett, DG Schatz, unpublished).

Finally, in lymphocytes of DNAPKcs- and Ku80-deficient mice, hairpin coding ends accumulate despite RAG protein expression (90–92). This presents an interesting paradox: In vivo, DNAPKcs and Ku are necessary for RAG-mediated hairpin opening, while in vitro, the RAG proteins can nick coding end hairpins without DNAPKcs or Ku. One resolution is suggested by the observation that coding ends are retained inefficiently in the CSC in vitro (24, 65). It is plausible that in vivo other factors are required to retain coding ends in the postcleavage complex long enough to be nicked by the RAG proteins. Ku and DNAPKcs are good candidates for factors that would act to stabilize coding ends in the CSC. Ku has been proposed to regulate remodeling or disassembly of the postcleavage complex, thereby facilitating coding end processing (91). Furthermore, the Ku proteins have been shown to facilitate the ligation of DNA ends similar to those generated during V(D)J recombination, suggesting that they might bridge two coding ends to promote joining (93). DNAPKcs has been shown to bind to DNA hairpins (94), and furthermore, in the presence of Ku but in the absence of DNA ends, it can be activated in vitro by the putative nuclear matrix protein C1D (95). By virtue of this interaction, DNAPKcs could tether the postcleavage complex to the nuclear matrix, thereby stabilizing coding end association with the complex and enhancing RAG-mediated hairpin nicking. Another possibility is that DNAPKcs may regulate the RAG proteins or other proteins in the postcleavage complex by phosphorylation. DNAPKcs activity might be required in vivo to

move DNA end-binding proteins out of the way so that processing enzymes, including the RAG proteins, have access to coding ends (91, 96). Coding end blocking proteins might include Ku and DNAPKcs themselves, or PARP, which binds and is activated by DNA ends and which may stimulate DNAPKcs by ADP-ribosylation (96–98). Finally, it is possible that a complex of Ku and possibly DNAPKcs must move along the coding end hairpin altering its conformation before it can be nicked by the RAG proteins. It has been proposed that Ku might unwind the hairpin coding end to allow nicking (99).

The Signal End Complex

In vitro, signal ends remain associated with each other in the signal end complex (SEC) (Figure 1), a nuclease resistant protein:DNA complex containing the RAG proteins and HMG1/2 (24, 65). This complex has not been directly demonstrated in vivo; however, several observations suggest that it exists. First, signal ends are usually not subject to deletion and N-nucleotide insertion as are coding ends, and this could be explained if RAG proteins protect signal ends from the processing activities that act on coding ends (65). Second, in normal lymphoid precursors, signal ends are easily detected (71–73) and persist before being joined (76). This persistence causes no obvious increase in p53 in normal animals (100). Thus, signal ends might not trigger conventional DNA damage responses in the cell because of their association with the RAG proteins. Finally, the formation of signal joints correlates with the downregulation of RAG gene expression, consistent with a need to remove RAG proteins from the SEC before joining (76).

Recent experiments demonstrated that removal of short terminal portions of RAG1 and RAG2, previously thought to be ''nonessential'' for signal joint formation, is associated with an accumulation of signal ends in transient recombination assays, consistent with decreased efficiency of signal joint formation (19). Requirements for chaperone-mediated disassembly of Mu transposase protein:DNA complexes have prompted speculations that signal joint formation requires the binding of chaperones for disassembly of the SEC and subsequent recruitment of repair proteins (19, 65). It has been suggested that these RAG terminal domains may interact with such a factor (19). Additionally, it has been proposed that the requirement for SEC disassembly before signal joint formation directs processing and joining activities to coding ends first, thereby minimizing hybrid joint formation (99).

While it is accepted that Ku proteins are required for signal joint formation, the role of DNAPKcs is more controversial (92, 101, 102). After RAG-mediated cleavage in cell extracts, associations between signal ends and either Ku or DNAPKcs have been detected (65). The finding that signal joints on transfected substrates and at TCR loci can exhibit N-nucleotide addition and deletion suggests that in some circumstances signal ends encounter some of the same processing enzymes as coding ends (103, 104, and references therein).

RAG-MEDIATED TRANSPOSITION

From the beginning, it had been postulated that V(D)J recombination must have evolved from an ancient transposition or site-specific recombination system; both the compact genomic structure of the RAG locus and the organization of the RSSs were compatible with this notion (6, 105). Subsequent elucidation of the reaction pathway (3), the observation that alcohols could serve as nucleophiles in the nicking reaction, and the demonstration that hairpin formation proceeded via a direct S_N2 transesterification mechanism (4) all provided strong correlations with transposition. RAG-mediated hybrid joint formation in vitro appeared to be analogous to the disintegration reactions catalyzed in vitro by retroviral integrase (70, 106). Finally, the demonstration that purified RAG proteins in vitro could catalyze transpositional insertion confirmed the link between V(D)J recombination and transposition (25, 26).

Basic Features of RAG-Mediated Transposition

RAG-mediated transposition was demonstrated in two different studies using either plasmid (25) or oligonucleotide (26) substrates containing RSSs. These studies showed that RAG proteins catalyze efficient insertion of signal end substrates into a target vector, generating a strand transfer product. The target vector can be the same molecule as the signal end substrate, resulting in an intramolecular transposition product (25); or the target vector can be a different molecule from the signal end substrate, resulting in an intermolecular transposition product. As with all transposition reactions, RAG-mediated coupled strand transfer of two signal ends results in a target site duplication at the site of insertion. The duplication is generally 5-bp long for RAG-catalyzed transposition, although duplications arising from intramolecular reactions show more variation in length. The sites at which insertions occur vary in sequence, and GC-rich sequences are somewhat preferred; a GC-rich hotspot was observed in one instance (25). RAG-mediated transposition is strongly dependent on the presence of HMG1/2 (25, 26) and typically requires both a 12- and a 23-RSS, although a low level of strand transfer products can be seen using only a 12-RSS substrate (26). The reaction can occur using either Mg^{2+} or Ca^{2+} as the divalent cation.

Comparison to Other Transposition Systems

Many aspects of RAG behavior are shared by composite transposons, large elements whose internal sequences encode drug resistance and accessory factors, and whose ends are functionally dissociable insertion sequences (IS) that encode transposase (for reviews of different classes of transposable elements, see 107–109 and references therein). In particular, many parallels are seen with the composite transposon Tn10/IS10, the first transposon in which formation of an obligatory hairpin intermediate was conclusively demonstrated (87).

Divalent Cation As a general rule, it appears that Mg^{2+} is the physiological divalent cation for proper functioning of transposases (107–110). In the absence of a divalent cation, RAG1 and RAG2 fail to stably bind isolated RSSs (31); by contrast, MuA can bind its cognate RSSs but is incapable of performing synapsis (107), whereas Tn10 can perform both binding and synapsis (108, 111). Without a divalent cation, transposases have no catalytic activity presumably because, as discussed below, divalent cations are essential components of the active site.

In the presence of Ca^{2+}, transposase binding and synapsis are quite stable. However, enzymatic activities are impaired, with Tn10 showing no activity (67) and RAG proteins and MuA capable of only strand transfer (24, 57). In the presence of Mn^{2+}, transposases behave strangely. In various Mn^{2+}-based assays, both RAG and Tn10 show (*a*) uncoordinated single-end cleavage (47, 56, 67), (*b*) inability to complete cleavage, resulting in accumulation of nicked intermediates (36, 39, 112), and (*c*) relaxed sequence specificities (33, 67). Importantly, Mn^{2+}-induced relaxation of sequence specificity is also seen in other enzymes that catalyze phosphoryl transfer reactions, such as polymerases (113) and restriction endonucleases (114).

Signal Sequences The organization of the RSS separates the primary site of RAG protein binding (the nonamer) from the site of cleavage (the heptamer). The dissociation of these two sites appears to be a general property among transposons (108, 115, 116).

It has been suggested that the RSS resembles the sites at the ends of Tc family transposons (117). The RSS heptamer motif is almost identical to the Tc end sequence; the two sequences differ primarily in the first nucleotide (C for the heptamer; T for Tc transposons). However, the cleavage products generated by the two reactions differ (115). The RSS nonamer motif resembles an AT-rich sequence in the Tc1 transposase binding site immediately downstream of the heptamer-like motif; however, this sequence is not well conserved among the Tc family (115, 117). Thus, the functional significance of the parallels between RSSs and Tc end sequences are uncertain.

Synapsis Transposons such as Mu or Tn7 have strict requirements for synapsis (107, 109). In Mu, under physiological conditions, topological requirements restrict PC formation to supercoiled substrates in which the two end sequences have specific orientations along the DNA (107). By contrast, Tn10 shows considerable versatility in synapsis; in vitro and in vivo, transposition can occur using many different arrangements of IS10 ends on the same or on different molecules (108). The behavior of the RAG proteins in synapsis appears to be similar to that of Tn10 transposase.

Protein-DNA interactions within PCs are quite strong. Mu and Tn10 cleavage from supercoiled donor substrates retains supercoils within the transposon elements; similarly, insertion of excised Tn10 into a supercoiled target molecule

retains the supercoils in the target (107, 108). Thus we may expect protein:DNA interactions within the RAG PC to be similarly strong (see below).

The 12/23 Rule The 12/23 rule provides a means by which the two ends of the primordial RAG transposon can be distinguished. This asymmetry in end usage is echoed in varying degrees in other transposition systems, particularly in more complicated systems such as Mu and Tn7 (107, 109). Perhaps the most striking parallel is seen in the apparent ''9–21'' rule of P-element transposition (116). The two ends of P-elements contain a 31-bp inverted repeat and a 10-bp transposase binding site, separated by a spacer whose length is either 9 bp (3′ end) or 21 bp (5′ end). Cleavage occurs within the 31-bp inverted repeat and typically requires both 3′ and 5′ ends, although in vitro, uncoupled cleavage can be observed at low levels on substrates containing a single 3′ end.

Cleavage In addition to restrictions governing PC formation, Mu and Tn7 transposons have another control feature to regulate initiation of cleavage. In the absence of target DNA, MuA performs the initial nicking reaction somewhat slowly; addition of target DNA greatly stimulates the nicking rate (107, 108). Tn7 is even more rigid in that cleavage is not performed at all unless a target DNA sequence is present (109). In contrast, both RAG and Tn10 transposases readily cleave synapsed ends in the absence of target DNA (108). This dissociation between cleavage and target interaction may have been an important evolutionary factor in allowing the RAG transposase to function in recombination.

As described previously, RAG1 and RAG2 appear to catalyze coupled cleavage at two RSSs almost simultaneously. However, in many other systems (e.g., Tn7 and Tn10), the cleavage reactions at the two ends are temporally separable (108, 109).

The cleavage reactions catalyzed by RAG, Tn5, and Tn10 transposases all result in the formation of hairpin intermediates (87, 118). In Tn5 and Tn10 cleavage, however, the hairpin resides on the end of the transposon, whereas in V(D)J cleavage, the hairpin resides on the end of the flanking sequence. This seemingly minor distinction may have important functional and evolutionary ramifications (see below). It has been suggested that excision of certain elements in plants (*Ac/Ds, Tam3,* and *Slide*), *Drosophila* (*hobo*), and fungi (*Ascot-1*) may resemble RAG cleavage in this regard (119). Interestingly, the ends of *Ascot-1* bear the sequence CAGTG that is found at the ends of Tc family transposons (117, 119) and that represents the last five bases of the heptamer.

Postcleavage Complex As discussed above, current evidence indicates that after RAG-mediated cleavage has occurred, the CSC retains all four cleaved ends in a single postcleavage complex (24). Available evidence on Tn10 cleavage, however, suggests that the Tn10 transpososome releases each flanking sequence as it is cleaved, resulting in a two-end postcleavage complex (111). This release may

be necessary to ensure that efficient hairpin opening occurs during Tn10 transposition (see below).

Target Immunity Transposons such as Mu or Tn7 typically do not transpose into the same donor molecule from which they originate, a phenomenon known as target immunity (107, 109, 120). The mechanistic basis for this resides in the fact that under physiological conditions, strand transfer requires additional proteins (MuB; TnsC) that bind to target molecules other than the transposon donor. By contrast, RAG and Tn10 transposons do not appear to be subject to target immunity; in both systems, intramolecular transposition occurs at relatively high frequencies (25, 108). The rationale for this discrepancy is that, while intramolecular transposition would be deleterious to a transposon such as Mu or Tn7, such intramolecular events within a composite transposon are advantageous for an insertion sequence whose primary aim is to disseminate as widely as possible (108).

Thermodynamics of Transposition In transposition reactions, the stability of each complex along the reaction pathway increases as the reaction progresses (107, 108). Consistent with this notion, RAG complexes with cleaved signal end–signal end molecules are extremely stable (65), and RAG proteins appear to remain tightly associated with strand transfer complexes resulting from intramolecular transposition (25).

THE RAG1-RAG2 ACTIVE SITE

One of the most interesting and still open questions is the structure and functionality of the active site of the RAG complex. No structure has been reported for the core regions of the RAG proteins thus far, and hence this section attempts to address this issue based primarily on analogies to other site-specific recombinases, together with evidence from in vitro experiments.

How Many Active Sites Participate in Cleavage?

Comparison of active site stoichiometries in other transposition systems suggests that in V(D)J cleavage, a single active site may catalyze both nicking and hairpin formation at a given coding end–signal end junction (55). Both Tn5 and Tn10 transposition generate double-strand break intermediates that are formed via a nicking-hairpin approach, and both transposases appear to use one active site for cleavage at each transposon end (86, 108, 118). By contrast, Tn7 transposition generates double-strand break intermediates that are formed by nicking each of the two complementary strands; two separate proteins are required to catalyze the two single-strand cleavages, apparently to accommodate the two substrate strands with opposite polarities (109, 121).

Neither RAG1 nor RAG2 alone exhibits any catalytic activity (3, 25, 26, 70, 84, 85), suggesting either that both proteins contribute amino acid side chains to the active site, or that interaction between them leads to a structural change in one of them, bringing the catalytic site into its active conformation. In keeping with the latter model, transposases often have cofactors that bind to and activate the catalytic subunit, e.g. MuA-MuB and TnsA/B-TnsC (reviewed in 107, 109).

A Two-Divalent Metal Ion Model for the Active Site

HIV-1/ASV integrase, MuA, Tn5, and Tn10 transposases all have catalytic activities similar to those of the RAG complex. Initial hydrolysis of a phosphate ester bond on one strand releases a 3′-OH group that is subsequently used in a direct transesterification reaction, creating either a strand transfer product or a DNA double-strand break. Therefore it is possible that all these proteins contain a similar active site. Although at the amino acid level the homology between the core regions of ASV integrase, HIV-1 integrase, MuA, and Tn5 is only 9–15% (118), X-ray crystal structure analyses have revealed that these four proteins, together with RuvC and RNaseH, belong to a polynucleotidyl-transferase family defined by structural similarities (118, reviewed in 122, 123). The active site of all four proteins contains a triplet of acidic residues, known as the DDE motif (124), that is shared among many other transposases. These residues chelate divalent metal ions and are important for all chemical steps of transposition and integration. Given the mechanistic parallels between transposition and V(D)J recombination, and the strong influence that divalent metal ions have on the activity and specificity of the RAG proteins, it is appealing to think that the active site of the RAG proteins contains one or more divalent metal ions that are directly involved in the catalytic process. Indeed, recent experiments (SD Fugmann, IJ Villey, LM Ptaszek, DG Schatz, unpublished) have identified two aspartic acid residues within RAG1 (murine amino acids D600 and D708) that function specifically in catalysis. Mutation of either of these residues abolishes all cleavage and strand transfer activities, but results in properly folded proteins that are able to bind to the RSS and form synaptic complexes with pairs of RSSs. Additional data indicate that D708 directly coordinates a divalent metal ion. The results suggest that these two amino acids and at least one divalent metal ion are critical, catalytic components of the RAG active site.

A model describing how two divalent metal ions catalyze phosphoryl-transfer reactions was originally proposed for the DNA polymerase 3′-5′-exonuclease domain (126, 127), and that model is now thought to be a general principle in many phosphoryl-transfer enzymes (128). Applied to the RAG proteins, the model offers a simple explanation for all their activities reported so far: nicking, hairpin formation, hybrid joint and O/S joint formation, hairpin opening, and strand transfer.

The general principle is that the two metal ions stabilize the trigonal bipyramidal transition state of the S_N2 reactions and either activate hydroxyl nucleo-

philes or act as a Lewis acid stabilizing oxyanion leaving groups for the S_N2 reactions. The two metal ions switch between these two functions after each step of the reaction, i.e. the metal ion that activates the nucleophile in the first step stabilizes the leaving group in the next step and vice versa (Figure 3).

The nucleophile and the target phosphor atom are well defined for the two steps of the V(D)J cleavage reaction. First, water is activated as a nucleophile by a Mg^{2+} ion (designated number 1 in Figure 3), whereas Mg^{2+} ion number 2 stabilizes the 3′-OH leaving group on the top strand of the coding end. Second, the hairpin is formed by the attack of the previous 3′-OH leaving group, now activated by Mg^{2+} number 2 as a nucleophile, at the phosphate ester on the lower strand. Mg^{2+} number 1 enhances the leaving group properties of the 3′-OH group on the resulting signal end.

The final outcome of the recombination event has not been established at this point. It will be determined by the nucleophile and the target phosphate ester in the third reaction step. If the signal end 3′-OH is used as the nucleophile, either hybrid joint (or O/S joint) formation, a reversal of the second step or transposition occurs, depending on whether the hairpin coding end or a different DNA molecule, respectively, is attacked (25, 26, 70). However, if a water molecule is activated to perform the nucleophilic attack, hairpin opening can occur. The model therefore predicts that hairpin opening and transposition (and also hybrid joint and O/S joint formation) are mutually exclusive. Once hairpin hydrolysis occurs (which starts the reaction down the pathway leading to CJ formation), the signal ends lose their ability to perform strand transfer reactions.

EVOLUTION AND CONTROL OF RAG-MEDIATED TRANSPOSITION

According to the proposed two-divalent metal ion model (Figure 3), the fate of the SEC is determined in the CSC shortly after the cleavage. Either coding ends and signal ends get rejoined to form hybrid joints or O/S joints, or the hairpins are opened, or the hairpins fall out of the complex before they get opened and the 3′-OH groups at the signal ends are active for transposition. If the hairpins get opened, the signal end 3′-OH groups are inactive for strand transfer and the signal ends can be ligated to form a signal joint. Since inversional recombination and subsequent genome integrity require the formation of signal joints, regulatory mechanisms should exist to control which pathway is chosen.

Does RAG-Mediated Transposition Occur In Vivo?

In general, transposition activity is maintained at a low level, since highly active transposons would be detrimental to the host cell due to the mutagenic effect of genome rearrangements. It has been proposed that chromosomal translocations involving antigen receptor gene loci arise from incomplete RAG-mediated trans-

position events (26), but no experimental evidence for these events has been presented so far. To date, no RAG-mediated transposition has been observed in vivo, although SECs are not rapidly joined or degraded after the cleavage reaction (see above). So, why don't we see transposition? First, transposition events might be too rare to be detected with the current methods. Second, the transposase activity might be suppressed in vivo either by the RAG proteins themselves, or by other factors.

Thus far, no active transposon has been isolated from vertebrates. The putative vertebrate transposase genes isolated to date contain multiple, inactivating mutations (129). In one case, clever reversion of these mutations resulted in an active transposase (130). The RAG proteins might also have acquired mutations affecting their ability to perform transposition in vivo, but these would have had to spare all of their activities necessary for V(D)J recombination. Perhaps these mutations affected target capture or strand transfer, although both activities are robust in vitro. As mentioned above, RAG-mediated hairpin opening might be detrimental to transposition, and an alteration of the active site might have caused a switch in the preferred nucleophile for the third transesterification (Figure 3) from the 3′-OH group on the signal end to a water molecule. Additionally, since the in vitro experiments were performed using the core regions of RAG1 and RAG2 (25, 26), it is possible that the amino-terminus of RAG1 and/or the carboxy-terminus of RAG2, missing in the core proteins, might inhibit transposition in vivo, by facilitating the disassembly of the SEC (19).

Additional factors that are present in the nucleus of the cell but not in the in vitro reactions might reduce or abolish the capacity of the RAG proteins to catalyze a complete transposition reaction. These factors could cause conformational changes in the SEC either by performing modifications (e.g. phosphorylation) or by interacting directly with the RAG proteins or the DNA in the SEC. Such changes could destabilize or inactivate this complex and therefore inhibit transposition. The DNA double-strand break repair factors XRCC4, DNA Ligase IV, Ku70, Ku80 are candidates because they are important for the formation of signal joints, a process that is mutually exclusive of the strand transfer reaction. Additionally, the Ku70/Ku80 heterodimer and DNAPKcs could facilitate disassembly of the SEC by competing with the RAG complex for the binding to the signal ends. This mechanism might at least prevent reassembly of SEC in the event that the RAG proteins dissociated from the signal ends.

Evolutionary Implications

Previously, the RAG transposon was proposed to consist of the RAG genes flanked by a single RSS on each side (25, 131). To explain the separation of the *RAG* genes and the RSSs, we would suggest a modification of this idea in which the RAG transposon, like Tn10/IS10 and Tn5/IS50, was a composite transposon, consisting of a central sequence flanked by a pair of insertion sequences (IS), each flanked by a pair of RSSs (Figure 5). The RAG genes may have resided in

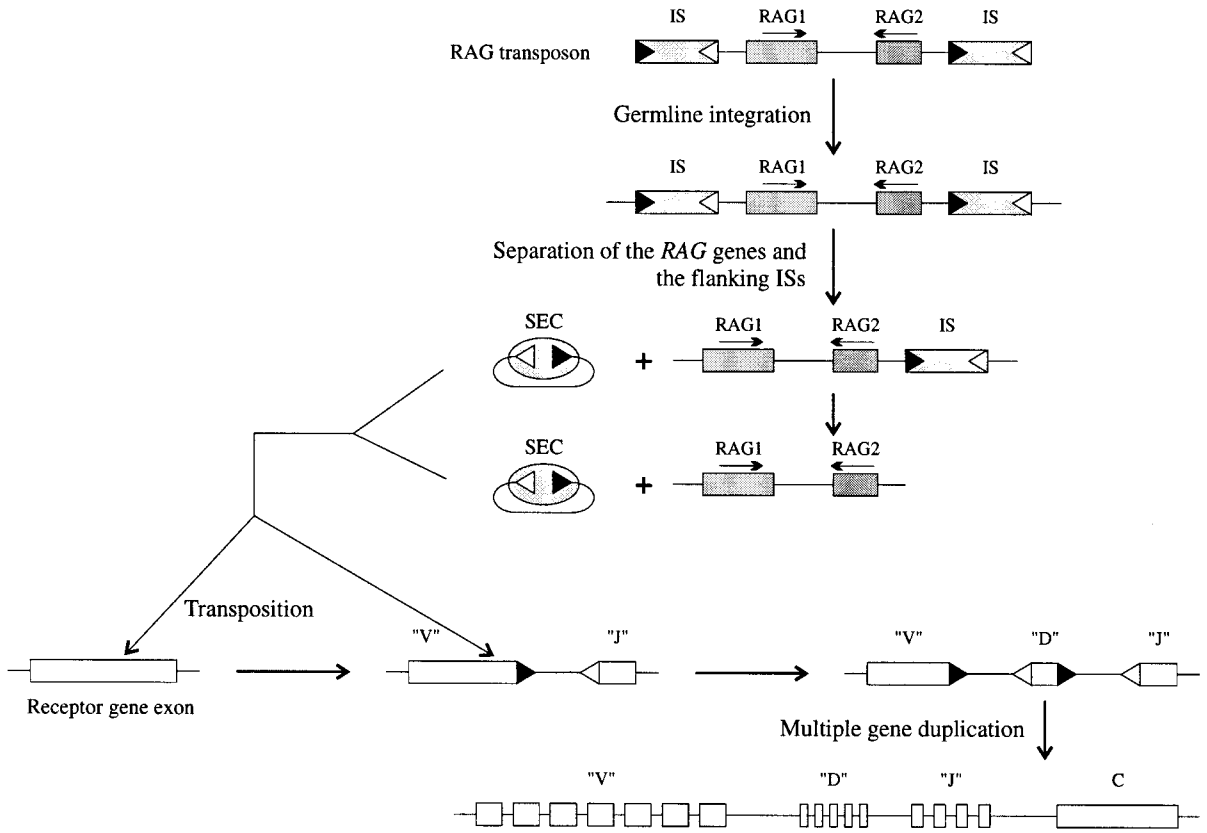

Figure 5 Schematic model of the ancient RAG transposon and the generation of the split antigen receptor genes. See text for details. The 12-RSS and 23-RSS are represented as black and white triangles, respectively. The RAG genes are drawn as boxes with arrows indicating the direction of transcription; the RAG-proteins in the SECs are shown as shaded ovals.

the central region or, as is the case for the composite bacterial transposons, inside one of the insertion sequences. After this complete transposon integrated into the germline, one or both of the IS elements was excised and integrated elsewhere in the genome. If the RAG genes resided in the central portion of the composite transposon, such an event would also have rendered them incapable of further transposition. At some point during evolution, one of the IS elements integrated into an exon of a receptor gene. The functional receptor gene could then only be produced if the RAG proteins excised the IS and the chromosomal break was repaired. All of the current split antigen receptor gene loci would thereafter have arisen by repeated gene duplications. The putative microorganism from which the RAG transposon originated has not been identified, and it is possible that it has not survived subsequent evolutionary selection processes.

ACKNOWLEDGMENTS

This review is dedicated to the memory of our friend and colleague, Eugenia Spanopoulou. We would like to thank L Ptaszek for help in preparation of Figure 2, and M Lieber, P Cortes, and M Bianchi for sharing unpublished data. This work was supported by a Howard Hughes Medical Institute predoctoral fellowship to AIL and by NIH grant AI32524 to DGS. SDF, IJV, and DGS are employees of the Howard Hughes Medical Institute.

LITERATURE CITED

1. Sleckman BP, Gorman JR, Alt FW. 1996. Accessibility control of antigen-receptor variable-region gene assembly—role of cis-acting elements. *Annu. Rev. Immunol.* 14:459–81
2. Schlissel MS, Stanhope-Baker P. 1997. Accessibility and the developmental regulation of V(D)J recombination. *Sem. Immunol.* 9:161–70
3. McBlane JF, van Gent DC, Ramsden DA, Romeo C, Cuomo CA, Gellert M, Oettinger MA. 1995. Cleavage at a V(D)J recombination signal requires only RAG1 and RAG2 proteins and occurs in two steps. *Cell* 83:387–95
4. van Gent DC, Ramsden DA, Gellert M. 1996. The RAG1 and RAG2 proteins establish the 12/23 rule in V(D)J recombination. *Cell* 85:107–13
5. Schatz DG, Oettinger MA, Baltimore D. 1989. The V(D)J recombination activating gene (RAG-1). *Cell* 59:1035–48
6. Oettinger MA, Schatz DG, Gorka C, Baltimore D. 1990. RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. *Science* 248:1517–23
7. van Gent DC, Hiom K, Paull TT, Gellert M. 1997. Stimulation of V(D)J cleavage by high mobility group proteins. *EMBO J.* 16:2665–70
8. Sawchuk DJ, Weis-Garcia F, Malik S, Besmer E, Bustin M, Nussenzweig MC, Cortes P. 1997. V(D)J recombination: modulation of RAG1 and RAG2 cleavage activity on 12/23 substrates by whole cell extract and DNA bending proteins. *J. Exp. Med.* 185:2025–32

9. Schatz DG, Oettinger MA, Schlissel MS. 1992. V(D)J recombination: molecular biology and regulation. *Annu. Rev. Immunol.* 10:359–83
10. Schatz DG. 1997. V(D)J recombination moves in vitro. *Semin. Immunol.* 9:149–59
11. Gellert M. 1996. Recent advances in understanding V(D)J recombination. *Adv. Immunol.* 64:39–64
12. Sadofsky MJ, Hesse JE, McBlane JF, Gellert M. 1993. Expression and V(D)J recombination activity of mutated RAG-1 proteins. *Nucleic Acids Res.* 22:5644–50
13. Sadofsky MJ, Hesse JE, Gellert M. 1994. Definition of a core region of RAG-2 that is functional in V(D)J recombination. *Nucleic Acids Res.* 22:1805–9
14. Cuomo CA, Oettinger MA. 1994. Analysis of regions of RAG-2 important for V(D)J recombination. *Nucleic Acids Res.* 22:1810–14
15. Silver DP, Spanopoulou E, Mulligan RC, Baltimore D. 1993. Dispensable sequence motifs in the RAG-1 and RAG-2 genes for plasmid-V(D)J recombination. *Proc. Natl. Acad. Sci. USA* 90:6100–4
16. McMahan CJ, Difilippantonio MJ, Rao N, Spanopoulou ES, Schatz DG. 1997. A basic motif in the N-terminal region of RAG1 enhances recombination activity. *Mol. Cell. Biol.* 17:4544–52
17. Roman CAJ, Cherry SR, Baltimore D. 1997. Complementation of V(D)J recombination deficiency in RAG-1(−/−) B cells reveals a requirement for novel elements in the N-terminus of RAG-1. *Immunity* 7:13–24
18. Kirch SA, Rathbun GA, Oettinger MA. 1998. Dual role of RAG2 in V(D)J recombination–catalysis and regulation of ordered Ig gene assembly. *EMBO J.* 17:4881–86
19. Steen SB, Han JO, Mundy C, Oettinger MA, Roth DB. 1999. Roles of the ''dispensable'' portions of RAG-1 and RAG-2 in V(D)J recombination. *Mol. Cell. Biol.* 19:3010–17
20. Litman GW, Anderson MK, Rast JP. 1999. Evolution of antigen binding receptors. *Annu. Rev. Immunol.* 17:109–47
21. Komori T, Okada A, Stewart V, Alt FW. 1993. Lack of N regions in antigen receptor variable region genes of TdT-deficient lymphocytes. *Science* 261:1171–75
22. Gilfillan S, Dierich A, Lemeur M, Benoist C, Mathis D. 1993. Mice lacking TdT: mature animals with an immature lymphocyte repertoire. *Science* 261:1175–78
23. Critchlow SE, Jackson SP. 1998. DNA end-joining: from yeast to man. *Trends Biochem. Sci.* 23:394–98
24. Hiom K, Gellert M. 1998. Assembly of a 12/23 paired signal complex: a critical control point in V(D)J recombination. *Mol. Cell.* 1:1011–19
25. Agrawal A, Eastman QM, Schatz DG. 1998. Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. *Nature* 394:744–51
26. Hiom K, Melek M, Gellert M. 1998. DNA transposition by the RAG1 and RAG2 proteins: a possible source of oncogenic translocations. *Cell* 94:463–70
27. Difilippantonio MJ, McMahan CJ, Eastman QM, Spanopoulou E, Schatz DG. 1996. RAG1 mediates signal sequence recognition and recruitment of RAG2 in V(D)J recombination. *Cell* 87:253–62
28. Spanopoulou E, Zaitseva F, Wang F-H, Santagata S, Baltimore D, Panayotou G. 1996. The homeodomain of Rag-1 reveals the parallel mechanisms of bacterial and V(D)J recombination. *Cell* 87:263–76
29. Rodgers KK, Villey IJ, Ptaszek L, Corbett E, Schatz DG, Coleman JE. 1999. A dimer of the lymphoid protein RAG1 recognizes the recombination signal sequence and the complex stably incor-

porates the high mobility group protein HMG2. *Nucleic Acids Res.* 27:2938–46
30. Akamatsu Y, Oettinger MA. 1998. Distinct roles of RAG1 and RAG2 in binding the V(D)J recombination signal sequences. *Mol. Cell. Biol.* 18:4670–78
31. Hiom K, Gellert M. 1997. A stable RAG1-RAG2-DNA complex that is active in V(D)J cleavage. *Cell* 88:65–72
32. Mo X, Bailin T, Sadofsky MJ. 1999. RAG1 and RAG2 cooperate in specific binding to the recombination signal sequence in vitro. *J. Biol. Chem.* 274:7025–31
33. Santagata S, Aidinis V, Spanopoulou E. 1998. The effect of Me^{2+} cofactors at the initial stages of V(D)J recombination. *J. Biol. Chem.* 273:16325–31
34. Swanson PC, Desiderio S. 1998. V(D)J recombination signal recognition–distinct, overlapping DNA-protein contacts in complexes containing RAG1 with and without RAG2. *Immunity* 9:115–25
35. Nagawa F, Ishiguro K, Tsuboi A, Yoshida T, Ishikawa A, Takemori T, Otsuka A, Sakano H. 1998. Footprint analysis of the RAG protein recombination signal sequence complex for V(D)J type recombination. *Mol. Cell. Biol.:*655–63
36. Cuomo CA, Mundy CL, Oettinger MA. 1996. DNA sequence and structure requirements for cleavage of V(D)J recombination signal sequences. *Mol. Cell. Biol.* 16:5683–90
37. Eastman QM, Leu TMJ, Schatz DG. 1996. Initiation of V(D)J recombination in vitro obeying the 12/23 rule. *Nature* 380:85–88
38. Hesse JE, Lieber MR, Mizuuchi K, Gellert M. 1989. V(D)J recombination: a functional definition of the joining signals. *Genes Dev.* 3:1053–61
39. Ramsden DA, McBlane JF, van Gent DC, Gellert M. 1996. Distinct DNA sequence and structure requirements for the two steps of V(D)J recombination signal cleavage. *EMBO J.* 15:3197–3206
40. Steen SB, Gomelsky L, Speidel SL, Roth DB. 1997. Initiation of V(D)J recombination in vivo: role of recombination signal sequences in formation of single and paired double-strand breaks. *EMBO J.* 16:2656–64
41. Roman CAJ, Baltimore D. 1996. Genetic evidence that the RAG1 protein directly participates in V(D)J recombination through substrate recognition. *Proc. Natl. Acad. Sci. USA* 93:2333–38
42. Sadofsky MJ, Hesse JE, van Gent DC, Gellert M. 1995. RAG-1 mutations that affect the target specificity of V(D)J recombination—a possible direct role of RAG-1 in site recognition. *Genes Dev.* 9:2193–99
43. Eastman QM, Villey IJ, Schatz DG. 1999. Detection of RAG protein-V(D)J recombination signal interactions near the site of DNA cleavage by UV cross-linking. *Mol. Cell. Biol.* 19:3788–97
44. Meisenheimer K, Meisenheimer P, Willis M, Koch T. 1996. High yield photocross-linking of a 5-iodocytidine substituted RNA to its associated protein. *Nucleic Acids Res.* 24:981–82
45. Swanson PC, Desiderio S. 1999. RAG-2 promotes heptamer occupancy by RAG-1 in the assembly of a V(D)J initiation complex. *Mol. Cell. Biol.* 19:3674–83
46. Bailin T, Mo X, Sadofsky MJ. 1999. A RAG1 and RAG2 tetramer complex is active in cleavage in V(D)J recombination. *Mol. Cell. Biol.* 19:4664–71
47. West RB, Lieber MR. 1998. The RAG-HMG1 complex enforces the 12/23 rule of V(D)J recombination specifically at the double-hairpin formation step. *Mol. Cell. Biol.* 18:6408–15
48. Kim DR, Oettinger MA. 1998. Functional analysis of coordinated cleavage in V(D)J recombination. *Mol. Cell. Biol.* 18:4679–88
49. Aidinis V, Bonaldi T, Beltrame M, Santagata S, Bianchi ME, Spanopoulou E. 1999. The RAG1 homeodomain recruits HMG1,2 to facilitate RSS binding and to enhance the intrinsic DNA bending

activity of RAG1/2. *Mol. Cell. Biol.* 19:6532–42
50. Bustin M, Reeves R. 1996. High-mobility-group chromosomal proteins: architectural components that facilitate chromatin function. *Prog. Nucleic Acid Res. Mol. Biol.* 54:35–100
51. Bianchi ME, Beltrame M. 1998. Flexing DNA: HMG-box proteins and their partners. *Am. J. Hum. Genet.* 63:1573–77
52. Boonyaratanakornkit V, Melvin V, Prendergast P, Altmann M, Ronfani L, Bianchi ME, Taraseviciene L, Nordeen SK, Allegretto EA, Edwards DP. 1998. High-mobility group chromatin proteins 1 and 2 functionally interact with steroid hormone receptors to enhance their DNA binding in vitro and transcriptional activity in mammalian cells. *Mol. Cell. Biol.* 18:4471–87
53. Craig NL. 1988. The mechanism of conservative site-specific recombination. *Annu. Rev. Genet.* 22:77–105
54. Engelman A, Mizuuchi K, Craigie R. 1991. HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer. *Cell* 67:1211–21
55. van Gent DC, Mizuuchi K, Gellert M. 1996. Similarities between initiation of V(D)J recombination and retroviral integration. *Science* 271:1592–94
56. van Gent DC, McBlane JF, Ramsden DA, Sadofsky MJ, Hesse JE, Gellert M. 1995. Initiation of V(D)J recombination in a cell-free system. *Cell* 81:925–34
57. Savilahti H, Rice PA, Mizuuchi K. 1995. The phage mu transpososome core–DNA requirements for assembly and function. *EMBO J.* 14:4893–4903
58. Zhu XD, Sadowski PD. 1998. Selection of novel, specific single-stranded DNA sequences by Flp, a duplex-specific DNA binding protein. *Nucleic Acids Res.* 26:1329–36
59. Lewis SM. 1994. The mechanism of V(D)J joining: lessons from molecular, immunological, and comparative analyses. *Adv. Immunol.* 56:27–150
60. Spanopoulou E, Cortes P, Shih C, Huang CM, Silver DP, Svec P, Baltimore D. 1995. Localization, interaction, and RNA binding properties of the V(D)J recombination-activating proteins RAG1 and RAG2. *Immunity* 3:715–26
61. Leu TMJ, Schatz DG. 1995. Rag-1 and rag-2 are components of a high-molecular-weight complex, and association of rag-2 with this complex is rag-1 dependent. *Mol. Cell. Biol.* 15:5657–70
62. Grawunder U, Schatz DG, Leu TMJ, Rolink A, Melchers F. 1996. The half-life of RAG-1 protein in precursor B cells is increased in the absence of RAG-2 expression. *J. Exp. Med.* 183:1731–37
63. Sheehan KM, Lieber MR. 1993. V(D)J recombination: signal and coding joint resolution are uncoupled and depend on parallel synapsis of the sites. *Mol. Cell. Biol.* 13:1363–70
64. Namgoong SY, Harshey RM. 1998. The same two monomers within a MuA tetramer provide the DDE domains for the strand cleavage and strand transfer steps of transposition. *EMBO J.* 17:3775–85
65. Agrawal A, Schatz DG. 1997. RAG1 and RAG2 form a stable post-cleavage synaptic complex with DNA containing signal ends in V(D)J recombination. *Cell* 89:43–53
66. Eastman QM, Schatz DG. 1997. Nicking is asynchronous and stimulated by synapsis in 12/23 rule-regulated V(D)J cleavage. *Nucleic Acids Res.* 25:4370–78
66a. Yu K, Lieber MR. 1999. Mechanistic basis for coding end sequence effects in the initiation of V(D)J recombination. *Mol. Cell. Biol.* 19:8094–8102
66b. Gerstein RM, Lieber MR. 1993. Coding end sequence can markedly affect the initiation of V(D)J recombination. *Genes Dev.* 7:1459–69
67. Junop MA, Haniford DB. 1996. Multiple roles for divalent metal ions in DNA transposition: distinct stages of Tn*10* transposition have different Mg^{2+} requirements. *EMBO J.* 15:2547–55
68. Lewis SM, Hesse JE, Mizuuchi K, Gel-

lert M. 1988. Novel strand exchanges in V(D)J recombination. *Cell* 55:1099–107
69. Lewis SM, Hesse JE. 1991. Cutting and closing without recombination in V(D)J joining. *EMBO J.* 10:3631–39
70. Melek M, Gellert M, van Gent DC. 1998. Rejoining of DNA by the RAG1 and RAG2 proteins. *Science* 280:301–3
71. Roth DB, Nakajima PB, Menetski JP, Bosma MJ, Gellert M. 1992. V(D)J recombination in mouse thymocytes: double-stranded breaks near T-cell receptor delta rearrangement signals. *Cell.* 69:41–53
72. Schlissel M, Constantinescu A, Morrow T, Baxter M, Peng A. 1993. Double-strand signal sequence breaks in V(D)J recombination are blunt, 5'-phosphorylated, RAG-dependent, and cell cycle regulated. *Genes Dev.* 7:2520–32
73. Livàk F, Schatz DG. 1996. T-cell receptor alpha locus V(D)J recombination by-products are abundant in thymocytes and mature T cells. *Mol. Cell. Biol.* 16:609–18
74. Livàk F, Schatz DG. 1997. Identification of V(D)J recombination coding end intermediates in normal thymocytes. *J. Mol. Biol.* 267:1–9
75. Schlissel MS. 1998. Structure of non-hairpin coding-end DNA breaks in cells undergoing V(D)J recombination. *Mol. Cell. Biol.* 18:2029–37
76. Ramsden DA, Gellert M. 1995. Formation and resolution of double-strand break intermediates in V(D)J rearrangement. *Genes Dev.* 9:2409–20
77. Ramsden DA, Paull TT, Gellert M. 1997. Cell-free V(D)J recombination. *Nature* 388:488–91
78. Leu TMJ, Eastman QM, Schatz DG. 1997. Coding joint formation in a cell free V(D)J recombination system. *Immunity* 7:303–14
79. Lewis SM. 1994. P nucleotide insertions and the resolution of hairpin DNA structures in mammalian cells. *Proc. Natl. Acad. Sci. USA* 91:1332–36
80. Staunton JE, Weaver DT. 1994. Scid cells efficiently integrate hairpin and linear DNA substrates. *Mol. Cell. Biol.* 14:3876–83
81. Lewis SM. 1999. Palindromy is eliminated through a structure-specific recombination process in rodent cells. *Nucleic Acids Res.* 27:2521–28
82. Paull TT, Gellert M. 1998. The 3' to 5' exonuclease activity of Mre11 facilitates repair of DNA double-strand breaks. *Mol. Cell.* 1:969–79
83. Paull TT, Gellert M. 1999. Nbs1 potentiates ATP-driven DNA unwinding and endonuclease cleavage by the Mre11/Rad50 complex. *Genes Dev.* 13:1276–88
84. Besmer E, Mansilla-Soto J, Cassard S, Sawchuk DJ, Brown G, Sadofsky M, Lewis SM, Nussenzweig MC, Cortes P. 1998. Hairpin coding end opening is mediated by RAG1 and RAG2 proteins. *Mol. Cell.* 2:817–28
85. Shockett PE, Schatz DG. 1999. DNA hairpin opening mediated by the RAG1 and RAG2 proteins. *Mol. Cell. Biol.* 9:4159–66
86. Bolland S, Kleckner N. 1996. The three chemical steps of Tn10/IS10 transposition involve repeated utilization of a single active site. *Cell* 84:223–33
87. Kennedy A, Guhathakurta A, Kleckner N, Haniford D. 1998. Tn10 transposition via a DNA hairpin intermediate. *Cell* 95:125–34
87a. Santagata S, Besmer E, Villa A, Bozzi F, Allingham JS, Sobacchi C, Haniford DB, Vezzoni P, Nussenzweig MC, Pan Z-Q, Cortes P. 1999. The RAG1/RAG2 complex constitutes a 3' flap endonuclease: implications for junctional diversity in V(D)J and transpositional recombination. *Mol. Cell. Biol.* In press
88. Nadel B, Feeney AJ. 1997. Nucleotide deletion and P addition in V(D)J recombination—a determinant role of the coding-end sequence. *Mol. Cell. Biol.* 17:3768–78
89. Nadel B, Feeney AJ. 1995. Influence of

coding-end sequence on coding-end processing in V(D)J recombination. *J. Immunol.* 155:4322–29
90. Roth DB, Menetski JP, Nakajima PB, Bosma MJ, Gellert M. 1992. V(D)J recombination: broken DNA molecules with covalently sealed (hairpin) coding ends in scid mouse thymocytes. *Cell* 70:983–91
91. Zhu CM, Bogue MA, Lim DS, Hasty P, Roth DB. 1996. Ku86-deficient mice exhibit severe combined immunodeficiency and defective processing of V(D)J recombination intermediates. *Cell* 86: 379–89
92. Gao YJ, Chaudhuri J, Zhu CM, Davidson L, Weaver DT, Alt FW. 1998. A targeted DNA-PKcs-null mutation reveals DNA-PK-independent functions for ku in V(D)J recombination. *Immunity* 9:367–76
93. Ramsden DA, Gellert M. 1998. Ku protein stimulates DNA end joining by mammalian DNA ligases: a direct role for Ku in repair of DNA double-strand breaks. *EMBO J.* 17:609–14
94. Smider V, Rathmell WK, Brown G, Lewis S, Chu G. 1998. Failure of hairpin-ended and nicked DNA to activate DNA-dependent protein kinase—implications for V(D)J recombination. *Mol. Cell. Biol.* 18:6853–58
95. Yavuzer U, Smith GCM, Bliss T, Werner D, Jackson SP. 1998. DNA end-independent activation of DNA-PK mediated via association with the DNA-binding protein C1d. *Genes Dev.* 12:2188–99
96. Calsou P, Frit P, Humbert O, Muller C, Chen DJ, Salles B. 1999. The DNA-dependent protein kinase catalytic activity regulates DNA end processing by means of Ku entry into DNA. *J. Biol. Chem.* 274:7848–56
97. Satoh MS, Lindahl T. 1992. Role of poly(ADP-ribose) formation in DNA repair. *Nature* 356:356–58
98. Ruscetti T, Lehnert BE, Halbrook J, Le Trong H, Hoekstra MF, Chen DJ, Peterson SR. 1998. Stimulation of the DNA-dependent protein kinase by poly(ADP-ribose) polymerase. *J. Biol. Chem.* 273:14461–67
99. Smider V, Chu G. 1997. The end-joining reaction in V(D)J recombination. *Semin. Immunol.* 9:189–97
100. Guidos CJ, Williams CJ, Ildiko G, Knowles G, Huang MTF, Danska JS. 1996. V(D)J recombination activates a p53-dependent DNA damage checkpoint in *scid* lymphocyte precursors. *Genes Dev.* 10:2038–54
101. Bogue MA, Jhappan C, Roth DB. 1998. Analysis of variable (diversity) joining recombination in DNA dependent protein kinase (DNA-PK)-deficient mice reveals DNA-PK-independent pathways for both signal and coding joint formation. *Proc. Natl. Acad. Sci. USA* 95:15559–64
102. Taccioli GE, Amatucci AG, Beamish HJ, Gell D, Xiang XH, Arzayus MIT, Priestley A, Jackson SP, Rothstein AM, Jeggo PA, Herrera VLM. 1998. Targeted disruption of the catalytic subunit of the DNA-PK gene in mice confers severe combined immunodeficiency and radiosensitivity. *Immunity* 9:355–66
103. Lieber MR, Hesse JE, Mizuuchi K, Gellert M. 1988. Lymphoid V(D)J recombination: nucleotide insertion at signal joints as well as coding joints. *Proc. Natl. Acad. Sci. USA* 85:8588–92
104. Candeias S, Muegge K, Durum SK. 1996. Junctional diversity in signal joints from T cell receptor beta and delta loci via terminal deoxynucleotidyl transferase and exonucleolytic activity. *J. Exp. Med.* 184:1919–26
105. Sakano H, Hüppi K, Heinrich G, Tonegawa S. 1979. Sequences at the somatic recombination sites of immunoglobulin light-chain genes. *Nature* 280:288–94
106. Chow SA, Vincent KA, Ellison V, Brown PO. 1992. Reversal of integration and DNA splicing mediated by integrase of

human immunodeficiency virus. *Science* 255:723–26
107. Mizuuchi K. 1992. Transpositional recombination: mechanistic insights from studies of Mu and other elements. *Annu. Rev. Biochem.* 61:1011–51
108. Kleckner N, Chalmers RM, Kwon D, Sakai J, Bolland S. 1996. Tn10 and IS10 transposition and chromosome rearrangements: mechanism and regulation in vivo and in vitro. *Curr. Top. Microbiol. Immunol.* 204:49–82
109. Craig NL. 1996. Transposon Tn7. *Curr. Top. Microbiol. Immunol.* 204:27–48
110. Asante-Appiah E, Skalka AM. 1997. Molecular mechanisms in retrovirus DNA integration. *Antiviral Res.* 36:139–56
111. Sakai J, Chalmers RH, Kleckner N. 1995. Identification and characterization of a pre-cleavage synaptic complex that is an early intermediate in Tn*10* transposition. *EMBO J.* 14:4374–83
112. Bolland S, Kleckner N. 1995. The two single-strand cleavages at each end of Tn10 occur in a specific order during transposition. *Proc. Natl. Acad. Sci. USA* 92:7814–18
113. Beckman RA. 1985. On the fidelity of DNA replication: manganese mutagenesis in vitro. *Biochemistry* 24:5810–17
114. Hsu M, Berg P. 1978. Altering the specificity of restriction endonuclease: effect of replacing Mg^{2+} with Mn^{2+}. *Biochemistry* 17:131–38
115. Plasterk RH. 1996. The Tc1/mariner transposon family. *Curr. Top. Microbiol. Immunol.* 204:125–43
116. Beall EL, Rio DC. 1997. Drosophila P-element transposase is a novel site-specific endonuclease. *Genes Dev.* 11:2137–51
117. Dreyfus DH. 1992. Evidence suggesting an evolutionary relationship between transposable elements and immune system recombination sequences. *Mol. Immunol.* 29:807–19
118. Davies DR, Braam LM, Reznikoff WS, Rayment I. 1999. The three-dimensional structure of a Tn5 transposase-related protein determined to 2.9-Å resolution. *J. Biol. Chem.* 274:11904–13
119. Colot V, Haedens V, Rossignol JL. 1998. Extensive, nonrandom diversity of excision footprints generated by ds-like transposon Ascot-1 suggests new parallels with V(D)J recombination. *Mol. Cell. Biol.* 18:4337–46
120. Stellwagen AE, Craig NL. 1997. Avoiding self–two Tn7-encoded proteins mediate target immunity in Tn7 transposition. *EMBO J.* 16:6823–34
121. Gary PA, Biery MC, Bainton RJ, Craig NL. 1996. Multiple DNA processing reactions underlie Tn7 transposition. *J. Mol. Biol.* 257:301–16
122. Yang W, Steitz TA. 1995. Recombining the structures of HIV integrase, RuvC and RNaseH. *Structure* 3:131–34
123. Rice P, Craigie R, Davies DR. 1996. Retroviral integrases and their cousins. *Curr. Opin. Struct. Biol.* 6:76–83
124. Kulkosky J, Jones KS, Katz RA, Mack JP, Skalka AM. 1992. Residues critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposon integrases and bacterial insertion sequence transposases. *Mol. Cell. Biol.* 12:2331–38
125. Polard P, Chandler M. 1995. Bacterial transposases and retroviral integrases. *Mol. Microbiol.* 15:13–23
126. Freemont PS, Friedman JM, Beese LS, Sanderson MR, Steitz TA. 1988. Cocrystal structure of an editing complex of Klenow fragment with DNA. *Proc. Natl. Acad. Sci. USA* 85:8924–28
127. Beese LS, Steitz TA. 1991. Structural basis for the 3'-5' exonuclease activity of *Escherichia coli* DNA polymerase I: a two metal ion mechanism. *EMBO J.* 10:25–33

128. Steitz TA, Steitz JA. 1993. A general two-metal-ion mechanism for catalytic RNA. *Proc. Natl. Acad. Sci. USA* 90:6498–502
129. Lohe AR, Moriyama EN, Lidholm DA, Hartl DL. 1995. Horizontal transmission, vertical inactivation, and stochastic loss of mariner-like transposable elements. *Mol. Biol. Evol.* 12:62–72
130. Ivics Z, Hackett PB, Plasterk RH, Izsvak Z. 1997. Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells. *Cell* 91:501–10
131. Thompson CB. 1995. New insights into V(D)J recombination and its role in the evolution of the immune system. *Immunity* 3:531–39

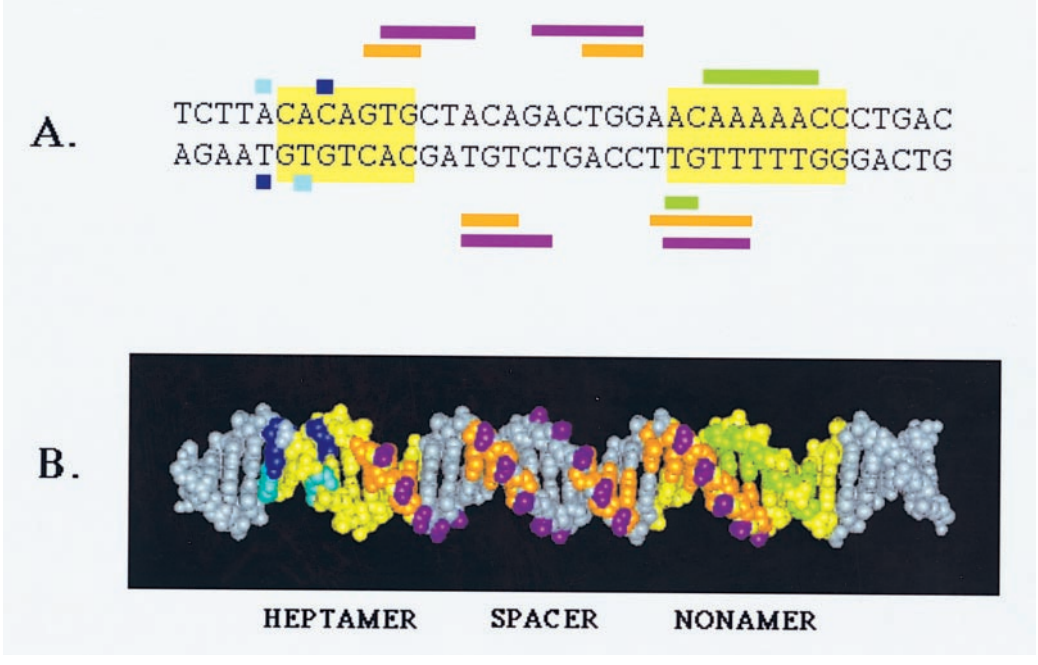

Figure 2 Map of the sites of contact of the RAG proteins on a 12-RSS. A. Linear model of the 12-RSS and coding flank DNA. B. Three-dimensional depiction as standard B-form DNA. Colors: yellow, the heptamer and nonamer; green, positions protected from cleavage in footprinting assays; light purple (phosphates); and orange (bases), sites of ethylation/methylation interference; dark purple and blue, sites of UV cross-linking to RAG1 only, or to both RAG1 and RAG2, respectively. Data derived from (30, 34, 35, 43).

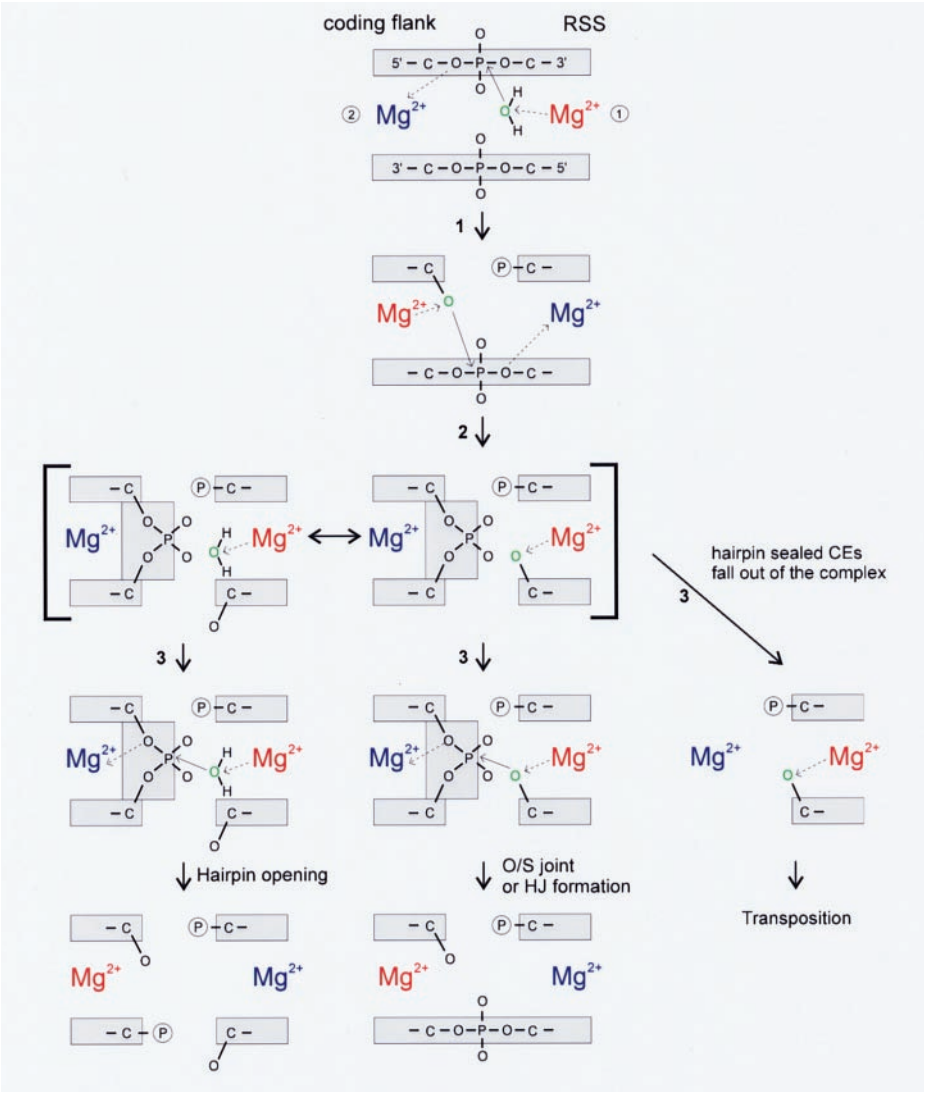

Figure 3 Schematic model of the two divalent metal ion active site model. The figure is focused on the phosphate ester bridges between the coding flank on left side and the RSS on the right side. The three steps, nicking (1), hairpin formation (2), and hairpin opening, transposition, hybrid joint and O/S joint formation (3) are proposed to be catalyzed by a single active site containing two Mg^{2+} ions. The Mg^{2+} ion that activates the nucleophile is shown in red; the ion that stabilizes the leaving group is shown in blue; and the nucleophile is shown in green. Solid arrows represent nucleophilic attacks, and dashed arrows indicate the activation of nucleophiles and the stabilization of leaving groups. Phosphate groups are drawn as a circled "P". See text for additional details.