Implications of transposition mediated by *V(D)J*-recombination proteins RAG1 and RAG2 for origins of antigen-specific immunity

Alka Agrawal*, Quinn M. Eastman† & David G. Schatz‡

** Department of Pharmacology, † Department of Molecular Biophysics and Biochemistry, and ‡ Howard Hughes Medical Institute, Section of Immunobiology, Yale University School of Medicine, New Haven, Connecticut 06510, USA*

Immunoglobulin and T-cell-receptor genes are assembled from component gene segments in developing lymphocytes by a site-specific recombination reaction, *V(D)J* recombination. The proteins encoded by the recombination-activating genes, *RAG1* and *RAG2*, are essential in this reaction, mediating sequence-specific DNA recognition of well-defined recombination signals and DNA cleavage next to these signals. Here we show that RAG1 and RAG2 together form a transposase capable of excising a piece of DNA containing recombination signals from a donor site and inserting it into a target DNA molecule. The products formed contain a short duplication of target DNA immediately flanking the transposed fragment, a structure like that created by retroviral integration and all known transposition reactions. The results support the theory that *RAG1* and *RAG2* were once components of a transposable element, and that the split nature of immunoglobulin and T-cell-receptor genes derives from germline insertion of this element into an ancestral receptor gene soon after the evolutionary divergence of jawed and jawless vertebrates.

Lymphocytes of the vertebrate adaptive system rely on a diverse array of immunoglobulins and T-cell antigen receptors (TCRs) for specific recognition of antigens. In the germ line, the genes encoding the variable portions of these receptors are typically split into component *V* (variable), *J* (joining) and, in some cases, *D* (diversity) gene segments. One of each type of gene segment is joined together in a site-specific recombination reaction to form the exon that encodes the antigen-binding portion of the polypeptide[1]. This reaction, known as *V(D)J* recombination, occurs only in lymphocytes, and in some vertebrate species is responsible for generating much of the diversity seen in antigen receptors. *V*, *D* and *J* gene segments are flanked by recombination signals that consist of conserved heptamer and nonamer sequences separated by a poorly conserved spacer sequence whose length is usually 12 or 23 base pairs (bp) (referred to hereafter as 12-signals and 23-signals, respectively). The signals target the reaction, and efficient recombination only occurs between a 12-signal and a 23-signal, a restriction referred to as the 12/23 rule[2].

The first phase of the recombination reaction is achieved by the products of the recombination-activating genes, *RAG1* and *RAG2* (refs 3, 4). Together, the RAG1 and RAG2 proteins bind two recombination signals, bring them into close juxtaposition (this process is termed synapsis), and cleave the DNA, thereby separating the signals from the flanking coding segments[5–7]. The DNA-bending high-mobility group proteins HMG1 and HMG2 substantially enhance the efficiency of coordinate cleavage[8,9], in part by improving binding to the 23-signal[9]. Cleavage occurs in two steps, with a nick first introduced adjacent to the heptamer to expose a 3′-hydroxyl group on the coding flank, followed by direct nucleophilic attack of the 3′-hydroxyl on the opposite DNA strand[5,10]. The products are blunt, 5′-phosphorylated signal ends and covalently sealed hairpin coding ends (Fig. 1a). In the second phase of the reaction, the two signal ends are joined precisely to form a signal joint, whereas the coding ends are processed to form a coding joint that typically exhibits both nucleotide loss and addition[2]. An array of ubiquitously expressed DNA double-strand break repair proteins is essential for the joining phase of the reaction[11–13].

There may be several parallels between *V(D)J* recombination and transposition[14–16]. First, the *RAG1* and *RAG2* genes have a compact genomic organization, as would be expected for components of a transposable element. The two genes lie immediately adjacent to one another and are convergently transcribed in all vertebrate species examined so far, and, in most cases, the open reading frames are not interrupted by introns[17]. Second, *V(D)J* DNA cleavage, retroviral integration and transposition proceed through a common pathway that involves exposure of a 3′-hydroxyl group and its attack on a target phosphodiester bond in a magnesium-dependent reaction[10,18]. Third, the RAG proteins remain stably associated with a synapsed pair of recognition elements after DNA cleavage[19], as is common in transposition[20]. Finally, RAG1 and RAG2 can join signal ends to coding ends in a reversal of the cleavage reaction[21]. Here we provide a unifying explanation for these observations and for the split configuration of antigen-receptor genes by showing that the RAG proteins catalyse a transposition reaction with striking similarities to those catalysed by other transposases.

An unexpected cleavage product

The cleavage substrates used here were ^{32}P-body-labelled DNA fragments containing a 12-signal and a 23-signal orientated such that cleavage releases a signal end/signal end (SE/SE) fragment of roughly 320 bp, as well as smaller fragments containing the hairpin coding ends (Fig. 1a). Reactions were done with truncated RAG1 and RAG2 proteins substantially purified from the B-lymphoma cell line F2A1 and with purified recombinant HMG2 in a buffer containing magnesium but not energy source. Cleavage products were detectable after 4 min and accumulated for at least 2 h (Fig. 1b), concurring with previous results[5,6]. In addition, species with a mobility slower than that of the input substrate were visible from 15 min onwards (Fig. 1b). These were not a result of proteins remaining associated with the DNA, because proteins were elimi-

nated with protease digestion and organic extraction before loading of the gel. Two closely spaced bands were visible, with the upper predominating at longer time points. The upper band, hereafter referred to as X, is shown below to be a circular product produced by intramolecular transposition of the SE/SE fragment. The lower band may be a more highly knotted topo-isomer or the result of a one-end strand-transfer event.

Restriction mapping of purified band X DNA indicated that it contained sequences from the SE/SE fragment but not from the left or right arms, and that it was not the result of end-to-end ligation of the SE/SE fragment (data not shown). To begin to study the structure of band X, we incubated cleavage products with *Bal*31 nuclease, which is a double-stranded-DNA exonuclease and a single-stranded-DNA exo- and endonuclease. In the presence of the RAG and HMG2 proteins, band X and the SE/SE fragment were substantially more resistant to *Bal*31 than were other DNA fragments, whereas all bands were highly nuclease-sensitive after removal of the proteins by protease treatment and organic extraction (Fig. 2a). The protein-dependent nuclease resistance of band X indicated that it, like the SE/SE fragment[19], might be stably associated with the RAG and HMG2 proteins. Indeed, band X DNA could be immunoprecipitated with anti-RAG1, anti-RAG2, and anti-HMG2 antibodies, but not with several control antibodies (Fig. 2b). Thus band X DNA is stably bound by RAG1, RAG2 and HMG2, and this association appears to render the DNA resistant to nuclease digestion.

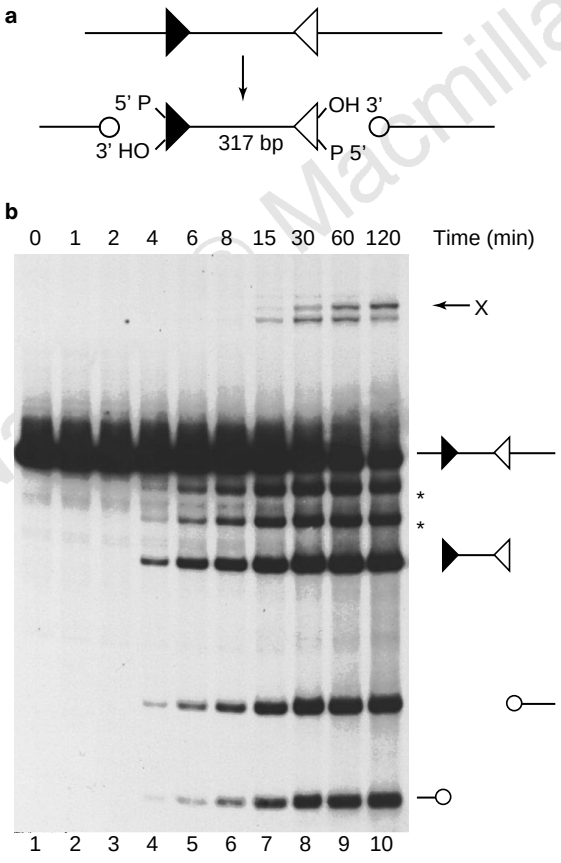


Figure 1 Time course of cleavage and band X formation. **a**, The cleavage substrate was a ^{32}P-body-labelled DNA fragment, generated by PCR, containing a 12-signal (black triangle) and a 23-signal (white triangle). Cleavage by the RAG proteins generates a 5′-phosphorylated SE/SE fragment and two smaller fragments containing the hairpin coding ends (circles). **b**, Cleavage using RAG and HMG2 proteins was done for different lengths of time at 37 °C and aliquots of the reaction were analysed on a native polyacrylamide gel. Asterisks indicate bands resulting from single cleavage at the 12- or 23-signals.

As band X contained only DNA derived from the SE/SE fragment, we wondered whether band X could be generated using the SE/SE fragment as the starting substrate. Incubation of an SE/SE fragment generated by the polymerase chain reaction (PCR) (Fig. 3a, lane 3) with the RAG and HMG2 proteins generated a low-mobility species

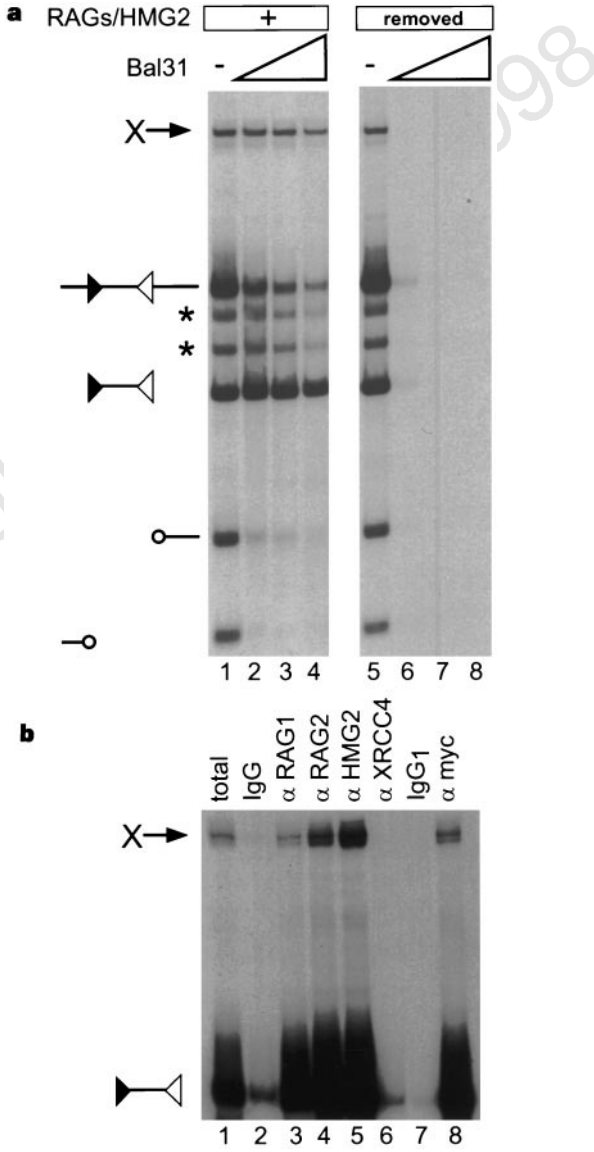


Figure 2 Stable association of the RAG and HMG2 proteins with band X confers nuclease resistance. **a**, Nuclease sensitivity of the cleavage reaction products. After 2 h of cleavage, reaction products were incubated directly with 0, 0.125, 0.5, or 1.0 U *Bal*31 nuclease for 1 h at 30 °C in the presence of 2 mM $CaCl_2$ (lanes 1–4, respectively), or were treated with proteinase K, precipitated and then subject to identical *Bal*31 digestions (lanes 5–8). Products were analysed on a native polyacrylamide gel. Asterisks indicate bands resulting from single cleavage at the 12- or 23-signals. **b**, Co-immunoprecipitation of band X DNA with RAG1, RAG2 and HMG2. Band X was generated from the 329-bp SE/SE DNA substrate and immunoprecipitations were done with polyclonal rabbit antibodies (indicated by α) specific for RAG1, RAG2 or HMG2 (lanes 2–4), an immunoglobulin (Ig)G1 monoclonal antibody specific for the Myc epitope tag (lane 8; both RAG1 and RAG2 contain the epitope tag), an IgG1 isotype control monoclonal antibody (lane 7), control polyclonal rabbit IgG (lane 2), or polyclonal rabbit antibodies specific for an irrelevant protein (XRCC4, lane 6). Co-immunoprecipitated DNA was analysed on a native polyacrylamide gel. Lane 1 represents the reaction before immunoprecipitation. As expected[19], the SE/SE DNA substrate is immunoprecipitated by the same antibodies as is band X, with a small amount of nonspecific precipitation seen in lanes 2 and 6.

(Fig. 3a, lane 6) that migrated together with gel-purified band X DNA (Fig. 3a, lane 1) and which was exonuclease-resistant in the presence of the RAG and HMG2 proteins (data not shown). Identical results were obtained using as the substrate an SE/SE fragment purified from a cleavage reaction (Fig. 3b, lane 1). If, however, the substrate lacked one or both of the signals, no product was produced (Fig. 3b, lanes 5, 7, 9). Generation of band X was more rapid when an SE/SE substrate was used rather than a substrate that required cleavage, with products first visible after 8 min of reaction with an SE/SE substrate (data not shown). We conclude that the formation of band X can be dissociated from the process of cleavage and requires a substrate with two signal ends.

The requirement for two signal ends indicated that the RAG proteins, instead of contaminating proteins in the RAG preparation, were probably responsible for formation of band X. To confirm this, we prepared highly purified glutathione *S*-transferase (GST)–RAG fusion proteins[22] and showed that these proteins generate band X (Fig. 3a, lane 8). The weaker activity of this preparation may be explained by its lower specific activity for DNA cleavage (data not shown). These results also showed that formation of band X depended on the addition of HMG2 protein (Fig. 3a, lanes 6, 8), and that HMG2 by itself had no activity (Fig. 3a, lane 9).

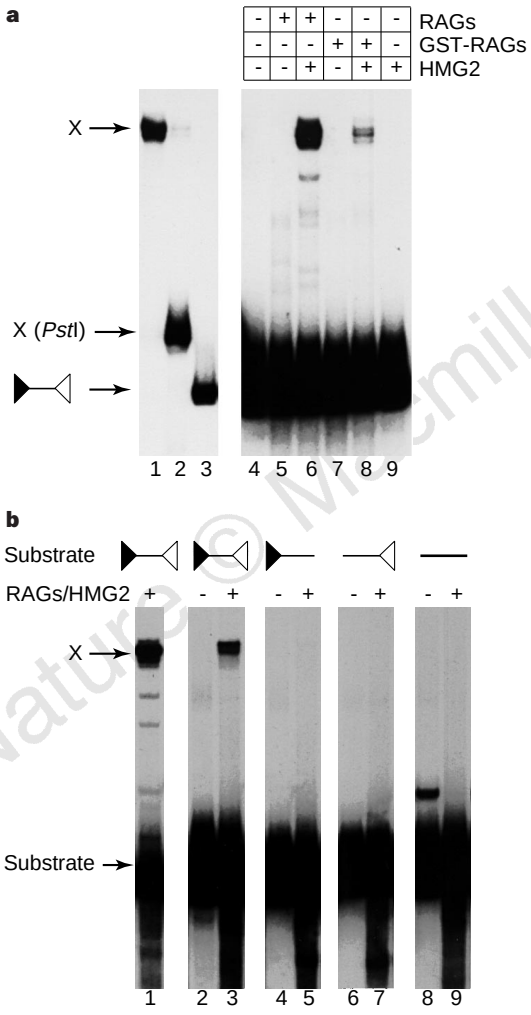


Figure 3 Conditions for band X formation. **a**, A ^{32}P-end-labelled 329-bp SE/SE DNA substrate was generated by PCR and incubated with combinations of RAG, GST-RAG and HMG proteins, as indicated above the lanes, for 2 h at 37 °C. Products were analysed on a native polyacrylamide gel. Lanes 1–3 show gel-purified band X, band X digested with *Pst*I, and the SE/SE fragment, respectively. The reduced mobility of *Pst*I-digested band X relative to the SE/SE fragment may be due to the single-stranded gaps in the band X DNA. **b**, ^{32}P-body-labelled substrates terminating in 12- and 23-signals, a 12-signal only, a 23-signal only, or no signals were generated by PCR (lanes 2–9) or by RAG-mediated cleavage (lane 1) and incubated in the presence or absence of RAG and HMG2 proteins for 2 h at 37 °C. The structure of the substrates is indicated above the lanes. Products were analysed on a native polyacrylamide gel.

Intramolecular transposition

The results above show that a low-mobility DNA species could be generated by RAG1, RAG2 and HMG2 from an SE/SE fragment, but that the reaction does not involve end-to-end ligation to form circles or concatamers. An alternative process that would generate circular DNA molecules without requiring end joining is intramolecular transposition, a strand-transfer reaction in which the free 3′-hydroxyl groups of the signal ends attack phosphodiester bonds in the DNA backbone (Fig. 4a). Such a reaction is a well-characterized feature of bacterial transposases[23–25] and retroviral integrases[26,27], and generates two distinct types of product depending on the topology of the strand-transfer reaction. Attack of the 3′-hydroxyl groups on the opposite strands from which they derive results in an inversion circle in which one portion of the molecule is inverted with respect to the other (Fig. 4a, bottom). In contrast, attack of the 3′-hydroxyl groups on the same strands from which they derive generates two deletion circles, each of which contains one of the recombination signals (Fig. 4a, top). If the attacks occur with a 5′ staggered polarity (as is typical for transposition and retroviral integration[24,28,29]), the two recombination signals should come to be flanked by gaps whose length reflects the staggered positions of strand transfer. When filled in, these gaps should yield target-site duplications, a hallmark of mobile DNA elements[30].

On the basis of its discrete mobility in a gel, we predicted that band X would be made up primarily of inversion circles, which should be of uniform size. If this were the case, digestion of band X with a restriction enzyme that cuts once in the SE/SE fragment (for example, *Pst*I) should generate single product with a distinctive structure (Fig. 4a). It should contain all of the sequences from the SE/SE fragment, with one recombination signal and variable amounts of flanking DNA inverted with respect to the rest of the molecule. The two recombination signals should be flanked by gaps derived from the target site. To test these predictions, we cloned *Pst*I-digested band X (Fig. 3a, lane 2) and sequenced the inserts in their entirety. Of the 50 randomly selected inserts analysed, 36 had precisely the structure expected for a transpositionally derived inversion circle and 5 had a structure consistent with a deletion circle (a fragment containing the 12-signal and variable amounts of flanking DNA). The remaining nine clones contained more complex rearrangements or deletions and are not described further here. All 36 inversion-circle clones contained a target-site duplication immediately flanking the recombination signals: 25 (69%) were 5 bp in length, 10 (28%) were 4 bp in length, and 1 (3%) was 3 bp in length (Fig. 4b).

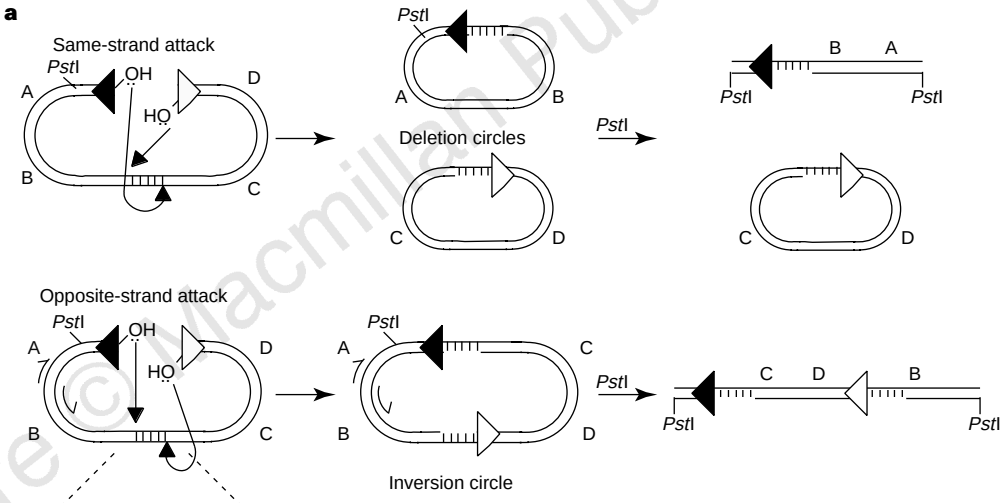
Among the 36 inversion-circle clones, strand transfer occurred at 17 different target sites scattered over the central one-third of the SE/SE fragment (Fig. 4c; target sites for deletion-circle formation are indicated below the line). The paucity of events near the ends may reflect topological constraints or interference from proteins bound to the terminal sequences. The sequences of the target-site duplications and the regions immediately flanking them in the SE/SE fragment are heterogeneous (Fig. 4b), indicating that strand transfer can occur at several different sequences. In addition, the results show that there is one transposition hot spot at which 13 (10 with a 5-bp duplication and 3 with a 4-bp duplication) out of 36 (36%) of the events occurred (Fig. 4c). This indicates that certain target sequences may be preferred over others.

We then studied the structure of the two strands of the inversion circles. The 5′ ends of the SE/SE fragment should remain unjoined, resulting in two gaps on one strand. To test this, band X was

generated from SE/SE fragments labelled at the 5′ end of either the 12-signal or the 23-signal (Fig. 5a) and gel-purified. This material was digested with restriction enzymes that cut close to the ends of the SE/SE fragment (Fig. 5a), and the products were resolved on a denaturing polyacrylamide gel. For each enzyme, band X gave rise to a single band that migrated together with the band produced by digestion of the SE/SE fragment with the same enzyme (Fig. 5b). Therefore, the 5′ ends of the SE/SE fragment are unjoined in the band X material, and hence one strand must contain two nicks or gaps. In addition, we confirmed that the other strand was covalently sealed by PCR amplification using divergent primers (depicted as half arrows in Fig. 4a) and gel-purified band X as the template. This reaction yielded a predominant product whose structure was identical to that of inversion-circle clones described above (data not shown). We conclude that the RAG proteins constitute a transposase. Together with HMG2, they can perform intramolecular transposition to yield products that are identical in all major respects to those produced by transposases and retroviral integrases.

Intermolecular transposition

The ability to move to a new genomic location is a defining feature of transposable elements. To determine whether the RAG proteins could carry out intermolecular transposition, we performed reactions with a 1.4-kilobase (kb) SE/SE substrate containing a tetracycline-resistance gene and a 6.3-kb circular target plasmid containing an amplicillin-resistance gene (Fig. 6a). Transposition should result in a 7.7-kb relaxed circle carrying both antibiotic-resistance genes (Fig. 5a). Direct analysis of the reaction products by Southern blotting showed that substrate sequences were transferred into a product that migrated at the position expected for the relaxed circle (Fig. 6b, lane 5). Formation of this product required the target plasmid and the RAG and HMG2 proteins (Fig. 6b, lanes 2–4). Experiments described below show that at least a portion of this band consists of the expected transposition product; we cannot rule out that it also contains a 'panhandle' product generated by strand transfer of only one end. In addition, substrate sequences also appeared in a band migrating at roughly the position of a linear 7.7-kb product. This band might be generated from relaxed circular molecules by nonspecific nuclease activity, or by concerted transposition of two signal ends derived from different substrate molecules (in which case the product would consist of the linear target plasmid and two copies of the substrate). We ruled out the latter



b

5′ flank	Target	3′ flank	N
GTTGC	CGCCG	GGCGT	10*
CCACT	CGGCA	GGTCT	2
AGAAT	CGCAG	CAACT	2
TTTTA	TTGGT	GAGAA	2
AGTCC	CTCTT	CCACC	2
AATGA	CCTCA	GAACT	1
CGGCA	GGTCT	CCAGT	1
TTTAT	TGGTG	AGAAT	1
ACTCC	ATCTG	GATTT	1
TCTCC	AGTAA	TGACC	1
AATCC	AGTCC	CTCTT	1
GTTCA	GAACG	CTCGG	1
GTTGC	CGCC	GGGCG	3*
GAACG	CTCG	GTTGC	2
AGGTC	TCCA	GTAAT	2
GGTTG	CCGC	CGGGC	1
AGTCC	CTCT	TCCAC	1
GGGCG	TTTT	TTATT	1
AATCG	CAG	CAACT	1

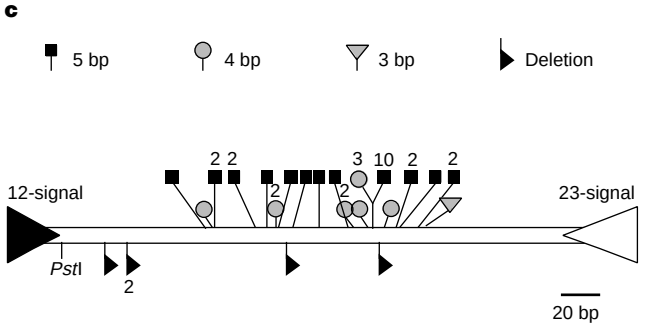


Figure 4 Intramolecular transposition events using the 329-bp SE/SE substrate. **a**, Depending on the topology of the strand attacks, two deletion circles (same-strand attack, top) or a single inversion circle (opposite-strand attack, bottom) are generated during intramolecular transposition. Deletion circles may be catenated and are represented as uncatenated circles for simplicity. Staggered nicks at a 5-bp target site (five short vertical lines) are depicted. This produces 5-bp gaps adjacent to the recombination signals and a 5-bp duplication after repair. The half-arrow indicate the positions of PCR primers used to confirm that one strand of the inversion circle is intact. **b**, Target-site sequences. *Pst*I-digested band X was cloned and the inserts sequenced. Only inversion-circle target sites and the 5 bp flanking them on either side in the starting substrate are shown. **c**, Location of target sites within the 329-bp SE/SE fragment. Multiple events at a single position are indicated by numbers above or below the symbols. Inversion-circle events are marked above the line (with different symbols used to indicate the length of the target-site duplication) and deletion-circle events below.

explanation by repeating the experiment with smaller target plasmids (<3 kb); here, the linear product contained only one copy of the 1.4-kb substrate (data not shown).

Bacteria were transformed with the reaction products and plated on agar containing ampicillin and tetracycline, or ampicillin alone, and the ratio of the numbers of colonies obtained on these two media provided a measure of the transposition frequency. This indicated that, in the presence of the RAG and HMG2 proteins, 0.016%, or 1 in 6,300, target plasmids were transposition recipients (average of two experiments). A fivefold molar excess of target-to-donor molecules was used in the reaction, indicating that about 1 in 1,300 SE/SE substrates was transposed. The transposition frequency dropped by at least tenfold in the absence of HMG2, and no ampicillin and tetramycin resistant colonies were obtained in the absence of the RAG proteins. Roughly equal numbers of ampicillin-resistant colonies were obtained from the different reactions.

Restriction mapping of plasmids prepared from 38 ampicillin and tetramycin resistant colonies confirmed the presence of the SE/SE substrate in all of them, and PCR reactions showed that the element was full-length in the ten representative plasmids analysed (data not shown). Sequencing of the two ends of the SE/SE element and flanking target DNA showed that all 38 plasmids contained the expected target-site duplication immediately flanking the recombi-

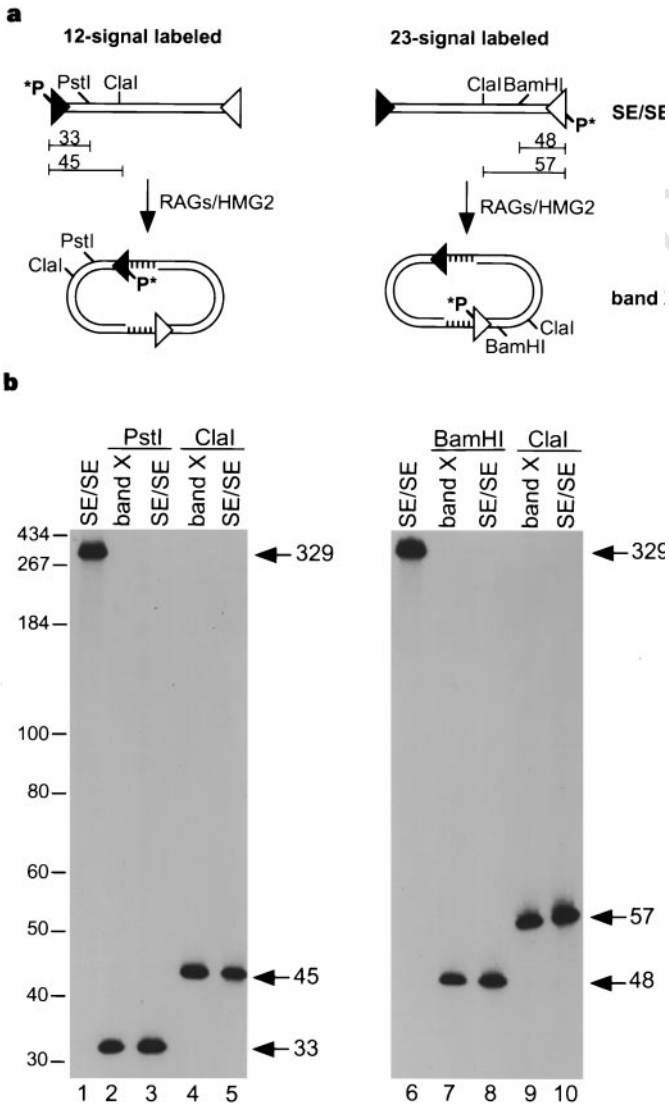


Figure 5 Mapping of the 5′ ends of the SE/SE fragment and band X. **a**, A 329-bp SE/SE fragment, 5′-end-labelled with ^{32}P (P*) at either the 12-signal or the 23-signal, was incubated with RAG and HMG2 proteins. The resulting band X products were gel-purified and digested with restriction enzymes that cut close to the labelled ends (distances are indicated in nucleotides). **b**, Digested products were analysed on an 8% denaturing polyacrylamide formamide/urea gel. The sizes of the molecular weight markers are given in nucleotides.

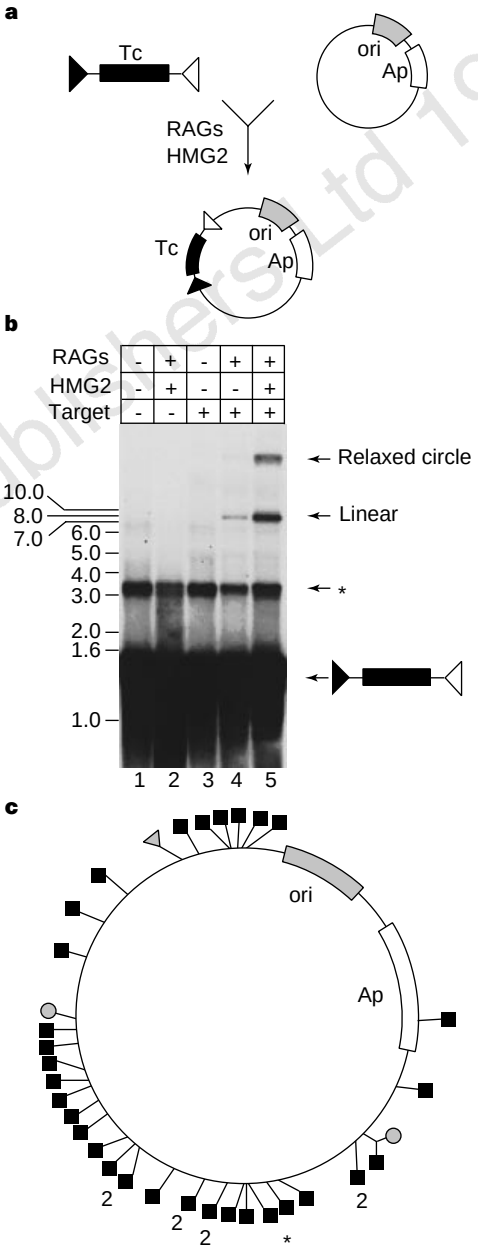


Figure 6 Intermolecular transposition events. **a**, A 1.4-kb substrate containing a tetracycline-resistance gene (Tc) flanked by 12- and 23-signals, and 6.3-kb plasmid target (a mixture of supercoiled and relaxed circular forms) containing an ampicillin-resistance gene (Ap) and a bacterial origin of replication (ori), was incubated with RAG and HMG2 proteins. Intermolecular transposition should generate a relaxed circular plasmid containing both antibiotic-resistance genes. **b**, Southern blot analysis of intermolecular transposition. Reactions were electrophoresed on a 0.7% agarose gel, transferred to a membrane, and hybridized with a ^{32}P-labelled probe generated form the 1.4-kb substrate by random priming. Arrows indicate the positions of the 7.7-kb relaxed circle and linear product. An asterisk indicates the position of a background band present in the substrate preparation. The positions of the molecular weight markers are given in kilobases. **c**, Location of the 38 integration sites in the target plasmid. An asterisk indicates the one event in which target sequences flanking the site of integration had become rearranged. Symbols are as in Fig. 4c.

nation signals. In comparison to the intramolecular reaction, a larger proportion (35 of 38, 92%) of target-site duplications were 5 bp in length, only 2 (5%) were 4 bp in length, and 1 (3%) was 3 bp in length (Fig. 6c). Strand transfer occurred at 33 different sites distributed over much of the target plasmid (Fig. 6c). Target-site sequences were heterogeneous and included GC- and AT-rich regions (data not shown). No hot spot or consensus target sequence was apparent.

We also detected intermolecular transposition using 329-bp or 139-bp SE/SE substrates and a 1.1-kb linear target (data not shown), showing that intermolecular transposition does not require a supercoiled target. In addition, a substrate consisting of an equimolar mixture of double-stranded 12- and 23-signal end oligonucleotides could be transposed into a circular target to generate a linear product (data not shown), indicating that the two signal ends need not reside on the same donor DNA molecule for the reaction to occur. Our results show that the RAG and HMG2 proteins can mediate intermolecular transposition into diverse target sequences, usually generating a target-site duplication of 5 bp.

Discussion

The transposition reaction mediated by RAG1 and RAG2 is mechanistically similar to that mediated by other transposases. The transposed segment of DNA has terminal inverted repeats (recombination signals), and is excised from a donor site and inserted into a target site on another DNA molecule or within the element itself. Insertion (strand transfer) generates short gaps flanking each end of the transposed element which become target-site duplications when repaired. No energy source is required for the reaction, in keeping with other transpositional strand-transfer reactions[31].

RAG-protein-mediated transposition may proceed through a series of higher-order protein–DNA complexes analogous to those identified in well-studied bacterial transposition systems[20,32].

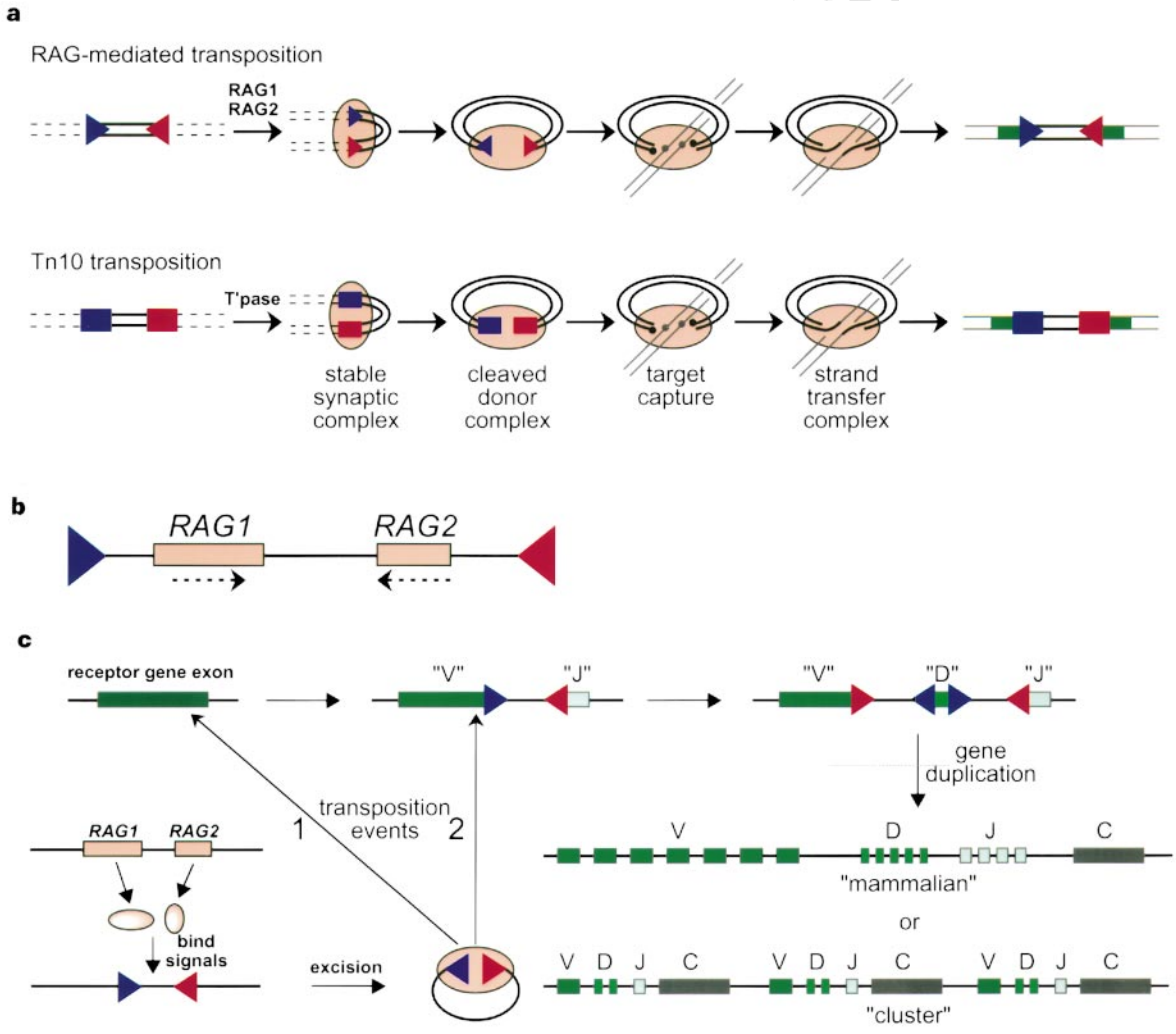


Figure 7 RAG-mediated transposition and a model for the origins of split antigen-receptor genes. **a**, Parallels between RAG- and Tn10-mediated transposition. Tn10 transposase (T'pase) or the RAG proteins (shaded ovals) recognize and synapse the terminal sequences of the transposable element (blue and red triangles or rectangles) and then excise it from donor sequences (dashed lines). Target capture, strand transfer, and gap repair result in target-site duplications (green rectangles) flanking the elements (see text). **b**, Possible structure of the original transposable element that integrated into the germ line of an ancestral vertebrate. Dashed arrows indicate the direction of transcription of the RAG genes. **c**, The present 'split' nature of immunoglobulin and TCR genes is proposed to have arisen from transposition mediated by RAG (bottom left) of one or two SE/SE elements into a primordial receptor gene exon (top, dark green rectangle), thereby dividing the exon into two or three gene segments, each flanked by one or two recombination signals (blue or red triangles). These gene segments would represent the evolutionary precursors of current *V*, *D* and *J* gene segments (top). Other models for the generation of *D* segments can also be envisioned. Different patterns of gene duplication (right) would result in the 'mammalian' or 'cluster' configurations of gene segments characteristic of the heavy chain locus of mammals or cartilaginous fishes, respectively. The constant region exons (*C*) are represented as single grey rectangles. Similar models have been proposed previously[14,41].

The first such complex contains two uncleaved recombination signals, the RAG proteins and HMG1 or HMG2 (HMG1/2)[33] ('stable synaptic complex', Fig. 7a). After DNA cleavage, the RAG proteins and HMG1/2 remain tightly associated with the signal ends in a 'cleaved donor complex'[19,33]. This complex is then capable of target capture and strand transfer to form a 'strand-transfer complex'. The existence of such a complex is suggested by the stable association of the RAG and HMG2 proteins with inversion-circle DNA and by the protein-dependent exonuclease resistance of this DNA species (Fig. 2). Analogous stable strand-transfer complexes have been characterized in bacteriophage Mu and Tn10 transposition[20,23]. These mechanistic similarities suggest a possible common evolutionary origin for the RAG proteins and other transposases. Although no significant sequence similarities have been found yet to support this idea, we note that the MuA and retroviral integrase catalytic domains adopt similar topologies but have little extended sequence similarity[34–36]. We emphasize that some aspects of this model for RAG-mediated transposition require experimental confirmation.

Our results extend the range of activities and DNA interactions known to be associated with the RAG proteins. They bind in a sequence-specific way to the recombination signals, and also interact with the flanking coding sequences and with coding ends after cleavage[33]. Our results show that RAG proteins bound to a pair of signal ends can capture duplex DNA and transfer the signal end 3′-hydroxyl groups to staggered positions on the two strands of the target DNA. It will be interesting to determine whether interactions with coding flanks, coding ends and transposition target sequences use a single, flexible active site, or whether multiple active sites are involved.

HMG2 is vital for both intramolecular and intermolecular transposition, and may have several distinct functions in the reaction. It probably enhances binding of the RAG proteins to the 23-signal and subsequent formation of a functional 12/23-signal synaptic complex[9]. With short SE/SE substrates, it may stabilize the severe bends in the DNA required for signal-end synapsis and subsequent intramolecular transposition, and, by analogy with the role of integration host factor in Tn10 transposition, may influence target-site selection and the topology of the intramolecular reaction[32,37]. Retroviral integration occurs preferentially into bent DNA[38] and HMG2 may enhance RAG-mediated transposition by modifying the local structure of the target DNA.

Many mobile DNA elements exhibit only limited target-site selectivity[38], and, in keeping with this, the target sites of RAG-mediated transposition are heterogeneous in location and sequence. An exception is the hot spot observed in the intramolecular reaction (Fig. 4), which may represent a preferred target-site sequence or may instead be a consequence of the constrained topology of this reaction. Two aspects of target-site selection in RAG-mediated transposition resemble the transposition mediated by Tn10 (refs 32, 39), but not by Tn7 or bacteriophage Mu. First, target-site capture appears to occur after cleavage (Fig. 7a). Second, target-site immunity is not seen (at least under the conditions used), as shown by the robust intramolecular transposition reaction.

Evolution of the vertebrate immune system

All jawed vertebrates studied so far possess adjacent *RAG1* and *RAG2* genes as well as immunoglobulin and TCR genes, which, with a few notable exceptions, must be assembled by somatic recombination before they can be expressed[16,40]. There is no evidence that any of these molecules, or antigen-specific lymphocytes, are found in jawless vertebrates (hagfish and lamprey) or invertebrates. This indicates that split antigen-receptor genes and the enzymatic machinery necessary for their assembly into functional units arose in the ~100 million years between the divergence of jawless and jawed vertebrates and the divergence of cartilaginous and bony fishes[16,40]. Our results are evidence in favour of the theory that a vital event in the evolution of the antigen-specific immune system was the insertion of a 'RAG transposon' into the germ line of a vertebrate ancestor[14,41]. The RAG transposon would presumably have consisted of recombination signals flanking the *RAG1* and *RAG2* open reading frames (Fig. 7b). Subsequently, the RAG1/RAG2 transposase mobilized the element in the germline and inserted it into an exon of a receptor gene (Fig. 7c). The gene could then only be expressed if the inserted transposon was excised by the RAG proteins and the two ends of the exon rejoined, presumably by ubiquitous DNA double strand break repair factors. This split gene has a structure analogous to that of immunoglobulin light chain and TCR α and γ chain genes. Insertion of a second transposon into the same exon would have resulted in a tripartite structure analogous to that of immunoglobulin heavy chain and TCR β and δ chain genes (Fig. 7c). Subsequent duplication of individual gene segments or of the entire gene would have resulted in the 'mammalian-type' or 'cluster-type' configurations of gene segments observed currently at the heavy-chain locus in mammals and cartilaginous fishes, respectively (Fig. 7c). Continued activity of RAG1 and RAG2 in the germ line could have led to the formation of 'preassembled' genes observed in the immunoglobulin loci of cartilaginous fishes[16].

Our results show that the RAG proteins can perform transposition *in vitro*, and, if the above hypothesis proves correct, then they were once active *in vivo* as well, where they played the determining role in the evolution of the antigen-specific immune system. Do they retain the ability to perform transposition *in vivo*? To our knowledge, no such events have been described, and it is possible that changes in the RAG proteins or some aspect of the cellular environment has led to a suppression of transposition *in vivo*. In this case, it will be interesting to determine how transposition in the cell is regulated. This may reveal some of the evolutionary forces that have resulted in the metamorphosis of a transposase into central components of a site-specific recombinase. □

Methods

Proteins. The RAG1 and RAG2 proteins were partially purified from the cell line F2A1, a mouse B-lymphoma cell line expressing active, truncated forms of the murine RAG1 (amino acids (aa) 264–1,008) and RAG2 (aa 1–387) proteins, both of which contain carboxy-terminal polyhistidine and Myc monoclonal antibody epitope tags. RAG proteins were purified over anion-exchange, Ni^{2+}, and heparin columns (Q.M.E., I. Villey and D.G.S., manuscript in preparation). Purity was >60% as judged by silver staining of polyacrylamide gels. GST–RAG1 (aa 330–1,040) and GST–RAG2 (aa 1–388) proteins[22] were purified over glutathione–agarose after transient co-transfection of 293-T cells[22]. Purity was >90% as judged by silver staining of polyacrylamide gels. Murine HMG2 lacking the C-terminal acidic domain and containing an amino-terminal polyhistidine region and protein kinase A phosphorylation site[42] was expressed in bacteria and purified by thermal denaturation and Ni^{2+} chromatography. Most antibodies used have been described[19]. The anti-RAG1 antibody used was P7. The anti-HMG2 antibodies were generated by immunizing rabbits with the peptide CKAGKKGPG-RPTGSK, spanning murine HMG2 residues 167–182. Antibodies were affinity-purified against the peptide and exhibit no cross-reactivity towards other proteins, including HMG1 (data not shown). The anti-XRCC4 antibodies were generated by immunizing rabbits with a *Pseudomonas* exotoxin fusion protein containing XRCC4 amino acids 5–263, and were affinity-purified. Immunoprecipitations were done nearly as described[19], except that the buffer used was the transposition reaction buffer plus 0.1% NP40.

DNA substrates. ^{32}P-body-labelled DNA-cleavage substrates were generated from plasmid pC329 or pC317 by PCR as described[6,19]. pC317 was generated by inserting the 23-signal of p12×23 (ref. 43) into the *Bam*HI site of p12 (ref. 43). The Tc gene of pBR322 was PCR-amplified using primers that introduced flanking 12- and 23-signals, and cloned to generate pTetRSS. The SE/SE fragment containing the Tc gene was generated from pTetRSS by PCR. All SE/SE fragments were gel-purified before use. The target for intermolecular transposition was pJH104 (ref. 44). Sequences of primers and details of plasmid construction are available on request.

Transposition reactions. Transposition reactions (25 μl) were performed for 2 h at 37 °C with ~25 ng RAG1 and RAG2 (or 200 ng GST–RAG proteins), 75 ng HMG2 and 0.05 pmol SE/SE fragment or cleavage substrate in a buffer containing 20 mM HEPES–Na^+, pH 7.5, 10 mM magnesium acetate, 50 mM sodium acetate, 2 mM dithiothreitol, 10 μM zinc sulphate and 100 μg ml^{-1} bovine serum albumin. Reactions were stopped by adding 175 μl 50 mM Tris-Cl, pH 8, 0.5% SDS, 10 mM EDTA and 20 μg proteinase K and incubating for at least 1 h at 55 °C. DNA was precipitated, separated on a 4% 29:1 acrylamide:bisacrylamide native polyacrylamide gel in 1× TBE buffer, and visualized by autoradiography. Intermolecular reactions were performed for 4 h at 37 °C in the same reaction buffer using 0.05 pmol unlabelled SE/SE fragment and 0.25 pmol target plasmid and included a 15 min preincubation at 37 °C in the absence of target DNA. Reactions were terminated and DNA precipitated as above, and 0.1–1% of the reaction was transformed into electrocompetent MC1061 bacteria and spread on agar plates containing ampicillin (100 μg ml^{-1}) or ampicillin plus tetracycline (10 μg ml^{-1}). This yielded an average of 500,000 ampicillin-resistant colonies from 0.1% of the reaction.

Received 6 July; accepted 30 July 1998.

1. Tonegawa, S. Somatic generation of antibody diversity. *Nature* **302,** 575–581 (1983).
2. Lewis, S. M. The mechanism of V(D)J joining: lessons from molecular, immunological, and comparative analyses. *Adv. Immunol.* **56,** 27–150 (1994).
3. Schatz, D. G., Oettinger, M. A. & Baltimore, D. The V(D)J recombination activating gene (RAG-1). *Cell* **59,** 1035–1048 (1989).
4. Oettinger, M. A., Schatz, D. G., Gorka, C. & Baltimore, D. RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. *Science* **248,** 1517–1523 (1990).
5. McBlane, J. F. *et al*. Cleavage at a V(D)J recombination signal requires only RAG1 and RAG2 proteins and occurs in two steps. *Cell* **83,** 387–395 (1995).
6. Eastman, Q. M., Leu, T. M. J. & Schatz, D. G. Initiation of V(D)J recombination *in vitro* obeying the 12/23 rule. *Nature* **380,** 85–88 (1996).
7. van Gent, D. C., Ramsden, D. A. & Gellert, M. The RAG1 and RAG2 proteins establish the 12/23 rule in V(D)J recombination. *Cell* **85,** 107–113 (1996).
8. Sawchuk, D. J. *et al*. V(D)J recombination: modulation of RAG1 and RAG2 cleavage activity on 12/23 substrates by whole cell extract and DNA bending proteins. *J. Exp. Med.* **185,** 2025–2032 (1997).
9. van Gent, D. C., Hiom, K., Paull, T. T. & Gellert, M. Stimulation of V(D)J cleavage by high mobility group proteins. *EMBO J.* **16,** 2665–2670 (1997).
10. van Gent, D. C., Mizuuchi, K. & Gellert, M. Similarities between initiation of V(D)J recombination and retroviral integration. *Science* **271,** 1592–1594 (1996).
11. Weaver, D. T. What to do at an end—DNA double-strand-break repair. *Trneds Genet.* **11,** 388–392 (1995).
12. Chu, G. Double strand break repair. *J. Biol. Chem.* **272,** 24097–24100 (1997).
13. Grawunderr, U., West, R. B. & Lieber, M. R. Antigen receptor gene rearrangement. *Curr. Opin. Immunol.* **10,** 172–180 (1998).
14. Thompson, C. B. New insights into V(D)J recombination and its role in the evolution of the immune system. *Immunity* **3,** 531–539 (1995).
15. Lewis, S. M. & Wu, G. E. The origins of V(D)J recombination. *Cell* **88,** 159–162 (1997).
16. Litman, G. W. *et al*. Phylogenetic diversification of immunoglobulin genes and the antibody repertoire. *Mol. Biol. Evol.* **10,** 60–72 (1993).
17. Gellert, M. Recent advances in understanding V(D)J recombination. *Adv. Immunol.* **64,** 39–64 (1996).
18. Craig, N. L. Unity in transposition reactions. *Science* **270,** 253–254 (1995).
19. Agrawal, A. & Schatz, D. G. RAG1 and RAG2 form a stable post-cleavage synaptic complex with DNA containing signal ends in V(D)J recombination. *Cell* **89,** 43–53 (1997).
20. Mizuuchi, K. Transpositional recombination: mechanistic insights from studies of Mu and other elements. *Annu. Rev. Biochem.* **61,** 1011–1051 (1992).
21. Melek, M., Gellert, M. & van gent, D. C. Rejoining of DNA by the RAG1 and RAG2 roteins. *Science* **280,** 301–303 (1998).
22. Spanopoulou, E. *et al*. The homeodomain of Rag-1 reveals the parallel mechanisms of bacterial and V(D)J recombination. *Cell* **87,** 263–276 (1996).
23. Weinert, T. A., Derbyshire, K. M., Hughson, F. M. & Grindley, N. D. F. Replicative and conservative transpositional recombination of insertion sequences. *Cold Spring Harb. Symp. Quant. Biol.* **49,** 251–260 (1984).
24. Benjamin, H. W. & Kleckner, N. Intramolecular transposition by Tn10. *Cell* **59,** 373–383 (1989).
25. Isberg, R. R. & Syvanen, M. Tn5 transposes independently of cointegrate resolution. Evidence for an alternative model for transposition. *J. Mol. Biol.* **182,** 69–78 (1985).
26. Shoemaker, C., Hoffmann, J., Goff, S. P. & Baltimore, D. Intramolecular integration within Moloney Murine Leukemia Virus DNA. *J. Virol.* **40,** 164–172 (1981).
27. Lee, Y. M. H. & Coffin, J. M. Efficient autointegration of avian retrovirus DNA in vitro. *J. Virol.* **64,** 5958–5965 (1990).
28. Fujiware, T. & Mizuuchi, K. Retroviral DNA integration: structure of an integration intermediate. *Cell* **54,** 497–504 (1988).
29. Brown, P. O., Bowerman, B., Varmus, H. E. & Bishop, J. M. Retroviral integration: structure of the initial covalent product and its precursor, and a role for the viral IN protein. *Proc. Natl Acad. Sci. USA* **86,** 2525–2529 (1989).
30. Berg, D. E. & Howe, M. M. *Mobile DNA* (Am. Soc. Microbiol., Washington DC, 1989).
31. Mizuuchi, K. Polynucleotidyl transfer reactions in transpositional DNA recombination. *J. Biol. Chem.* **267,** 21273–21276 (1992).
32. Kleckner, N., Chalmers, R. M., Kwon, D., Sakai, J. & Bolland, S. Tn10 and IS10 transposition and chromosome rearrangements: mechanism and regulation *in vivo* and *in vitro*. *Curr. Top. Microbiol. Immunol.* **204,** 49–82 (1996).
33. Hiom, K. & Gellert, M. Assembly of a 12/23 paired signal complex: a critical control point in V(D)J recombination. *Mol. Cell* **1,** 1011–1019 (1998).
34. Dyda, F. *et al*. Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. *Science* **266,** 1981–1986 (1994).
35. Rice, P., & Mizuuchi, K. Structure of the bacteriophage Mu transposase core: a common structural motif for DNA transposition and retroviral integration. *Cell* **82,** 209–220 (1995).
36. Bujacz, G. *et al*. High-resolution structure of the catalytic domain of avian sarcoma virus integrase. *J. Mol. Biol.* **253,** 333–346 (1995).
37. Chalmers, R., Guhathakurta, A., Benjamin, H. & Kleckner, N. IHF modulation of Tn10 transposition: sensory transduction of supercoiling status via a proposed protein/DNA molecular spring. *Cell* **93,** 897–908 (1998).
38. Craig, N. L. Target site selection in transposition. *Annu.. Rev. Biochem.* **66,** 437–474 (1997).
39. Sakai, J. & Kleckner, N. The TN10 synaptic complex can capture a target DNA only after transposon excision. *Cell* **89,** 205–214 (1997).
40. Rast, J. P. *et al*. α, β, γ and δ T cell antigen receptor genes arose early in vertebrate phylogeny. *Immunity* **6,** 1–11 (1997).
41. Sakano, H., Hüppi, K., Heinrich, G. & Tonegawa, S. Sequences at the somatic recombination sites of immunoglobulin light-chain genes. *Nature* **280,** 288–294 (1979).
42. Zwilling, S., König, H. & Wirth, T. High mobility group protein 2 functionally interacts with the POU domains of octamer transcription factors. *EMBO J.* **14,** 1198–1208 (1995).
43. Lewis, S. M. & Hesse, J. E. Cutting and closing without recombination in V(D)J joining. *EMBO J.* **10,** 3631–3639 (1991).
44. Hsieh, C., McCloskey, R. P., Radany, E. & Lieber, M. R. V(D)J recombination: evidence that a replicative mechanism is not required. *Mol. Cell. Biol.* **11,** 3972–3977 (1991).

Acknowledgements. We thank D. Hesslein for GST-RAG protein; I. Villey for HMG2 and RAG proteins; C.-L. Tsai for RAG proteins; P. Shockett for help with denaturing polyacrylamide gels; W. Stephen for bacterial plates and media; F. Livak for pTetRSS; E. Spanopoulou for the GST–RAG plasmids; T. Wirth for the HMG2 plasmid; J. Repasky and E. Corbett for help with varous procedures; N. Craig and G. Litman for comments on the manuscript; and the W. M. Keck Foundation Biotechnology Resource Laboratory at Yale University for oligonucleotides and DNA sequencing. D.G.S. is an associate investigator of the Howard Hughes Medical Institute. This work was supported by a grant from the NIH.

Correspondence and requests for materials should be addressed to D.G.S. (e-mail: david.schatz@yale.edu).