V(D)J Recombination: Mechanism, Errors, and Fidelity

DAVID B. ROTH
Department of Pathology and Laboratory Medicine and the Center for Personalized Diagnostics, Perelman School of Medicine of the University of Pennsylvania, Philadelphia, Pennsylvania 19104

ABSTRACT V(D)J recombination, the mechanism responsible for generating antigen receptor diversity, has the potential to generate aberrant DNA rearrangements in developing lymphocytes. Indeed, the recombinase has been implicated in several different kinds of errors leading to oncogenic transformation. Here we review the basic aspects of V(D)J recombination, mechanisms underlying aberrant DNA rearrangements, and the types of aberrant events uncovered in recent genomewide analyses of lymphoid neoplasms.

GENERATION OF ANTIGEN-RECEPTOR DIVERSITY: A DOUBLE-EDGED SWORD
The realization, now more than half a century ago, that B cells can generate antibodies to an astounding variety of chemical structures sparked intense interest in the “generation of diversity” question (reviewed in reference 1). The correct solution to this puzzle turned out to be both surprising and simple: the exons encoding the antigen-binding portions of the receptor (the so-called variable regions) are assembled by chromosomal breakage and rejoicing in developing lymphocytes (2). Immunoglobulins and T-cell receptors are composed of two polypeptide chains, each of which contributes to the antigen-binding domain. The exons encoding the antigen-binding domains are assembled from so-called V (variable), D (diversity), and J (joining) gene segments by “cut and paste” DNA rearrangements. This process, termed V(D)J recombination, chooses a pair of segments, introduces double-strand breaks adjacent to each segment, deletes (or, in selected cases, inverts) the intervening DNA, and ligates the segments together (Fig. 1). Rearrangements occur in an ordered fashion, with D-to-J joining proceeding before a V segment is joined to the rearranged D-J segments. This process of combinatorial assembly—choosing one segment of each type from several (sometimes many) possibilities—is the fundamental engine driving antigen-receptor diversity in mammals. Diversity is tremendously amplified by the characteristic variability at the junctions (loss or gain of small numbers of nucleotides) between the various segments. This process leverages a relatively small investment in germline coding capacity into an almost limitless repertoire of potential antigen-binding specificities.

This elegant process does, however, have a potential downside. A system that must break chromosomal DNA several times in order to generate a functional antigen receptor gene—many millions of times over the lifetime of an organism—creates significant opportunities for error. The necessity of enforcing a high degree of fidelity in V(D)J recombination has been recognized for decades (reviewed in reference 3). Aberrant V(D)J recombination events do occur, and they can be life-threatening, underlying the genesis of common lymphoid neoplasms (4, 5, 6, 7), as discussed below. Recent genome-wide analyses of lymphoid neoplasms have revealed V(D)J recombination-driven oncogenic events and have intensified interest in the regulatory mechanisms responsible for ensuring fidelity during V(D)J recombination. This chapter reviews the basic aspects of V(D)J recombination, the mechanisms...
responsible for aberrant rearrangements, and the types of events uncovered in recent analyses of tumor genomes. Recent advances in understanding the mechanisms responsible for safeguarding genomic integrity during V(D)J recombination are also discussed.

**THE NORMAL V(D)J RECOMBINATION MECHANISM**

This section briefly overviews the normal mechanism of V(D)J recombination (for more details, see references 8 and 9). The V(D)J recombinase recognizes conserved DNA sequence elements, termed recombination signal sequences (RSSs), located adjacent to each V, D, and J coding segment. RSSs consist of conserved heptamer and nonamer elements, separated by 12 or 23 nucleotides of less conserved “spacer” sequence (Fig. 2). Efficient recombination only occurs between RSSs with different spacer lengths (the “12/23 rule”). Additional restrictions are imparted at some antigen-receptor loci by other DNA sequence features (the so-called “beyond 12/23 rule”) (10). The RSSs are the only DNA segments required to allow V(D)J recombination to occur on artificial substrates, and their relative orientation determines whether the reaction proceeds by inversion or by deletion (Fig. 3) (11). An additional outcome, occasionally observed at antigen receptor loci, is formation of a “hybrid joint,” in which a coding segment is joined to an RSS (12) (Fig. 3). Hybrid joints do not contribute to antigen-receptor diversity, nor do they appear to play a role in oncogenic transformation.

Nucleotide sequences of natural RSSs display considerable variability. Those with sequences closest to the consensus support the most efficient recombination (13). The first three nucleotides of the heptamer (closest to the coding flank) show the highest sequence conservation and are critical for recombination, whereas the remaining heptamer positions are much less important (13) (Fig. 2). The nonamer sequence conforms less closely to the consensus, with only a few highly conserved positions (particularly the A/T tract), and the nonamer is dispensable under certain conditions in vitro.
Spacer length is critical and can be changed successfully only in increments that preserve the helical spacing of the nonamer and heptamer elements (13). Other nucleotide sequence features can influence recombination efficiency, most notably the sequence of the coding segment adjacent to the heptamer (the coding flank) (14, 15). This reflects a requirement for DNA distortion during DNA cleavage (14, 15, 16, 17).

The V(D)J recombinase consists of two lymphoid-specific proteins, RAG1 and RAG2 (18), which work together with a non-lymphoid-specific DNA bending factor, HMG1A or HMG1B (19), to carry out DNA cleavage. The RAG1 and RAG2 genes are located quite close to each other in all species examined, and their open reading frames are generally encoded in single exons. These observations led to speculation that the V(D)J recombinase may have evolved from an ancestral prokaryotic transposase (20). Indeed, the mechanism of DNA cleavage by RAG1/2 (one-step transesterification) (21) is shared with a class of bacterial transposases, and the RAG proteins can catalyze bona fide transposition events (22, 23). Definitive evidence that RAG1/2 indeed evolved from an ancestral transposase remains elusive (24).

The functional anatomy of the RAG1 and RAG2 proteins has been reviewed recently (8). Mutational studies have defined the minimally functional regions of both proteins. The “core” region of murine RAG1 is comprised of amino acids 384 to 1008 of the 1040-amino acid protein (25, 26) and is sufficient to catalyze V(D)J recombination, albeit with some abnormal features (27). Core RAG1 contains elements important for binding to the nonamer, as well as amino acids required for catalysis of cleavage. Neither specific DNA binding nor catalytic activities have been attributed to RAG2, leading to the view that RAG1 is the catalytic component of the recombinase, with RAG2 serving as an essential cofactor with some regulatory activities (described below). The essential “core” region of RAG2 historically has been defined as amino acids 1 to 383 (of 527) (28, 29). Recent work has shown that the minimal region extends only to amino acid 360 (30), closely coinciding with the predicted six-bladed β-propeller structure (31). This core domain is connected to the C-terminal domain via a flexible acidic hinge. The C terminus, while dispensable for recombination, is important for optimal recombination (32) and for enforcing the proper order of recombination events in developing lymphocytes (33). In its absence, aberrant recombination events are observed (33, 34, 35). The C terminus is also important for maintaining genomic stability (36, 37, 38), as is the acidic hinge (30). Within the
C terminus reside a plant homeodomain (PHD) capable of recognizing histone H3K4 trimethylation (39, 40) and a cell-cycle-regulated protein degradation signal (41). These elements are discussed in more detail below.

Efficient cleavage of a DNA substrate requires only RAG1, RAG2, a divalent metal ion, and HMGB1 or HMGB2 (19, 42). Cleavage proceeds via a two-step mechanism (Fig. 4). First, a nick is introduced between the RSS and the coding flank, and the resulting 3’OH group is then used to attack the opposite strand by transesterification, forming a hairpin coding end and a blunt signal end. This second step is similar to transesterification reactions catalyzed by the HIV integrase and by bacterial transposases (21). Whereas nicking can occur independently on either RSS, in the presence of the physiological divalent metal ion Mg$^{2+}$, transesterification requires assembly of a synaptic complex including both a 12- and a 23-RSS (43), providing a molecular basis for the 12/23 rule. After cleavage, the four DNA ends remain associated with the RAG proteins in a post-cleavage complex (Fig. 5), which retains the signal ends more stably than the coding ends (43, 45, 46). This complex is important for the proper rejoining of the broken DNA ends (47, 48, 49) and shepherds the ends to the classical nonhomologous end-joining (cNHEJ) pathway (45). This function, which prevents access of the ends to the low fidelity, translocation-prone alternative nonhomologous end-joining machinery (aNHEJ) joining pathway (50), is hypothesized to be important for maintenance of genomic stability during V(D)J recombination (30, 36), as discussed below. The C terminus of RAG2 contributes to the stability of the post-cleavage signal end complex (30, 50). Flexibility of the acidic hinge is also important: mutations (including some nucleotide sequence polymorphisms identified in humans) that reduce the negative charge destabilize the RAG-signal end complex and reduce genomic stability in pre-B-cell lines (30).

A characteristic feature of V(D)J recombination is the asymmetric processing of the signal and coding ends. Coding ends are joined with slight variations (small deletions, short insertions), whereas signal ends are generally joined with little or no end processing, so that the majority of the signal joints are perfect heptamer-to-heptamer fusions (51). This asymmetry may partly be explained by the requirement for additional processing of the coding ends, which are covalently sealed. Other factors may also contribute, including the differential stability of the RAG-coding end and RAG-signal end post-cleavage complexes. Hairpin opening occurs...
through the action of the Artemis endonuclease, which often cuts off-axis, resulting in short, single-stranded extensions that can give rise to palindromic insertions (P nucleotides) at the coding joint. These are never observed in signal joints, which are formed by blunt end-to-end joining. Formation of ends with single-stranded extensions may also increase opportunities for loss of nucleotides from the coding ends. Another source of extra nucleotides is provided by terminal deoxynucleotidyl transferase, which adds short, nontemplated, GC-rich inserts (N regions) to coding joints (and occasionally to signal joints). The frequent presence of such “microscopic” junctional alterations at coding joints provides a powerful mechanism for amplifying the diversity of antigen-binding sites in T-cell receptor and immunoglobulin molecules.

RAG-generated DNA ends are normally joined by the cNHEJ (reviewed in reference 52). Inactivation of any of the key components of cNHEJ (e.g., Ku70/80, DNA ligase IV, XRCC4) results in a severe impairment of joining. The few junctions formed under these conditions are often (but not always) abnormal, showing excessive deletions, the frequent presence of microhomologies and the occasional presence of abnormally long stretches of extra nucleotides. These features have been considered characteristic of alternative joining pathways, collectively termed aNHEJ (52), although, as discussed below, they are not always observed. aNHEJ is error-prone in two senses: the tendency toward formation of abnormal junctions and an increased propensity for forming gross genomic rearrangements such as chromosome translocations (53, 54, 55).

As noted above, the “shepherding” function of the RAG post-cleavage complex prevents the coding and signal ends from accessing aNHEJ. This was demonstrated by the observation that certain RAG mutants allow much higher levels of aNHEJ with artificial substrates in cultured cells, in both the presence and the absence of functional cNHEJ (45, 50). This may be important for preserving genomic stability, as discussed below.

**V(D)J RECOMBINATION ERRORS**

As noted above, the rearranging gene strategy that so successfully generates antigen-receptor diversity comes with a price: the potential for generating deleterious genomic rearrangements. Indeed, chromosomal rearrangements involving antigen-receptor loci were reported in both B- and T-cell neoplasms shortly after the discovery of V(D)J recombination (4, 56, 57). With the advent of next-generation sequencing technologies, the genomic landscapes of these malignancies are being studied with increasingly fine resolution. Aberrant events identified in lymphoid neoplasms include chromosome translocations, relatively small (kilobase to megabase) inversions and deletions (5, 6, 58), and reinsertion of excised fragments bounded by signal ends (59). Recent work has highlighted the importance of deletions that affect numerous genes implicated or suspected in tumorigenesis (7, 38, 60, 61) (M. Mijuskovic, S. Lewis, D. Roth, unpublished data). These deletions are RAG mediated, as they occur between pairs of sequences closely resembling RSSs, they often follow the 12/23 rule, and reciprocal
signal joints have been detected (M. Mijuskovic, S. Lewis, D. Roth, unpublished data). Thus, different types of V(D)J recombination errors play important roles in initiating oncogenic transformation.

Aberrant V(D)J recombination events observed in lymphoid neoplasms fall into two broad conceptual categories: errors in target recognition (Fig. 6) and errors in joining (Fig. 7). The first type consists of recognition of one authentic RSS and one DNA sequence fortuitously resembling an RSS (termed a “cryptic RSS” or cRSS). Given the relatively small size of RSS sequences and that recombination does not require strict adherence of this sequence to consensus heptamer/nonamer sequences, it is not surprising that cRSSs capable of supporting recombination are present approximately once per kilobase in random DNA sequences (62). Perhaps the first example of such events was provided by cytogenetic analyses of human lymphoid neoplasms, which revealed chromosome translocations involving authentic RSSs at antigen-receptor loci and cRSSs adjacent to proto-oncogenes (4, 56) (Fig. 6a). These events can cause inappropriate expression of the target gene due to, for example, the presence of transcriptional regulatory elements from the antigen-receptor loci. Recombination events involving a cRSS/RSS pair can also deregulate oncogenes through amplification, likely through a breakage–fusion-bridge mechanism (63).

Events also occur between pairs of cRSSs. These can occur in trans, generating a chromosome translocation (Fig. 6b) as in T-cell acute lymphoblastic leukemia (ALL) cases involving translocations between T-cell receptor gene segments and the SCL locus (3), or in cis, generating a deletional “coding joint” and an excised “signal joint” (Fig. 6c). Interestingly, although one might expect

---

**FIGURE 6** V(D)J recombination: recognition errors. Three types of recognition error are shown. (a) Recombination occurs between an authentic RSS (black triangle) with its associated coding flank (white box) and a cryptic RSS (cRSS; green triangle) with its associated coding flank (orange box), located on a separate DNA molecule. Recombination produces a trans rearrangement, with a pseudo-coding joint and a pseudo-signal joint. (b) The recombinase recognizes a pair of cRSSs located on separate DNA molecules. These recombine, generating a reciprocal chromosome translocation. The two products bear a pseudo-coding joint and a pseudo-signal joint. (c) The recombinase recognizes a pair of cRSSs located on the same DNA molecule and generates a deletion, forming a pseudo-coding joint (retained on the chromosome) and an excised circle containing a pseudo-signal joint.

a. **Recognition error**: RSS and cRSS in trans (translocation)

```
Diagram showing RSS and cRSS in trans configuration with pseudo-coding and pseudo-signal joints.
```

b. **Recognition error**: cRSS pair in trans (translocation)

```
Diagram showing cRSS pair in trans configuration with pseudo-coding and pseudo-signal joints.
```

c. **Recognition error**: cRSS pair in cis (deletion)

```
Diagram showing cRSS pair in cis configuration with deletional coding joint and excised signal joint.
```
events between cRSSs to also generate deletional “signal joints” retained in the chromosome, or inversion events, these are rarely observed (7) (M. Mijuskovic, unpublished data). Deletional recombination between cRSS pairs generates recurrent deletions at the SIL/SCL locus (64) and in Notch1, Izkf1, PTEN, and other critical genes in lymphoid neoplasms in humans and in mice (5, 7, 38, 58, 60, 61, 65) (Fig. 6c). These are now thought to be major drivers of oncogenic transformation in lymphocytes. Another type of target recognition error, less commonly observed, involves RAG-mediated cleavage at non-B-form DNA structures. This type of error has been implicated in oncogenic rearrangements joining the Bcl-2 major breakpoint region to an authentic RSS at the immunoglobulin heavy chain locus (66).

Errors in joining involve events that join a RAG-mediated double-strand break to a broken DNA end created by a non-RAG-mediated mechanism. The events observed in lymphoid neoplasms generally involve three DNA breaks and are referred to as “end donation” (67) or “type 2” events (68). These can involve a pair of breaks made during an apparently normal V(D)J recombination event, which are then mistakenly joined to another break generated by another mechanism (Fig. 7a). These can generate chromosome translocations or insertions of signal-ended fragments into another chromosomal location (59, 68). Similar events can involve a combination of recognition and joining errors: cleavage at a pair of cRSSs followed by joining to a non-RAG-mediated double-strand break (Fig. 7b). Although joining errors may involve the normal cNHEJ mechanism, it is thought that such events may be favored by the use of error-prone aNHEJ mechanisms (50, 63), which are known to favor the formation of translocations (53, 54).

RAG-mediated transposition events cannot be conveniently classified as errors in recognition or in joining, as in this case the initial cleavage event, occurring at a pair of authentic RSSs, is followed by RAG-mediated integration of the excised signal-ended fragment at another genomic location. These events, while observed in vitro (22, 23) in artificial systems in cultured cells (69, 70) and at the hypoxanthine artificiphosphoribosyl transferase (HPRT) locus in human peripheral T cells (71), have not yet been demonstrated definitively in lymphoid neoplasms. It should be noted, however, that certain types of transposition events generate products that would not be recognizable as having been derived from transposition (72).

HOW IS FIDELITY PRESERVED DURING V(D)J RECOMBINATION?

Given the fact that V(D)J recombination occurs in many millions of lymphocytes each day and that a variety of
V(D)J recombination errors generate oncogenic lesions in lymphoid neoplasms, it seems logical to suppose that mechanisms exist to maintain the fidelity of the process. Perhaps the most basic of these is to ensure that the recombinase is active only in the appropriate target cells and only during the appropriate developmental stages. Indeed, expression of RAG1 and RAG2 is carefully limited in a cell- and developmental stage-specific fashion. Bypassing these controls by introducing RAG1 and RAG2 transgenes under the control of strong promoters causing constitutive expression during lymphocyte development and in extralymphoid tissues results in a spectrum of phenotypes (including lymphopenia, growth retardation, and early death) reminiscent of DNA damage deficiency syndromes (73). An additional temporal control is provided by cell-cycle-specific protein degradation of the RAG2 protein, mediated by phosphorylation of a threonine (T490) located in the dispensable C terminus (41). Disabling this feature (via a T490A mutation) results in accelerated lymphomagenesis in p53-deficient mice (37). Autoubiquitylation of RAG1 may also play a regulatory role (74, 75).

Choice of the joining pathway used to repair RAG-generated breaks also appears to be important in maintaining fidelity. Mice lacking a functional cNHEJ pathway exhibit accelerated lymphomagenesis in the absence of p53 (76, 77), with complex chromosome translocations (mediated by aNHEJ) accompanied by gene amplification (63). Choice of joining pathway appears to involve the RAG post-cleavage complex, as mutations in RAG1 or RAG2 that destabilize the complex allow the ends to be joined by alternative pathways, including homologous recombination and aNHEJ (30, 45, 50). Further work has shown that mutations in the nonessential C terminus of RAG2 lead to genomic instability, accompanied by chromosomal aberrations (30, 36). To test the hypothesis that the C-terminal mutations caused oncogenic transformation by encouraging joining errors mediated by aNHEJ, lymphomas from two mouse models lacking the RAG2 C terminus were examined by whole-genome sequencing. Scant evidence for oncogenic aberrant joining events was observed. Instead, most genomic lesions that could be linked to potentially oncogenic events were deletions between pairs of cRSSs (38, 60). These data suggest that interstitial deletions may be more important drivers of RAG-mediated oncogenesis, at least in some systems, than gross chromosomal aberrations and are in agreement with recent studies of B- and T-cell ALL in humans (7, 61).

In the case of the RAG2 C-terminal mutants, it is not yet clear whether increased access of RAG-mediated DNA breaks to aNHEJ plays a role in the observed genomic instability. A severe RAG2 C-terminal truncation that allows high levels of aNHEJ in cultured cells (50) increases access of RAG-mediated breaks to alternative joining mechanisms, as shown by rescue of joining in a Ku80/RAG2 double mutant. This observation supports the idea that the post-cleavage complex enforces pathway choice (38). Recognizing a particular junction as having arisen from aNHEJ in this system is complicated, however, because junctional features considered characteristic of aNHEJ were rarely observed, even in the double mutants (in which junctions must have formed by a cNHEJ-independent process) (38). These data are consistent with previous analysis of rare junctions isolated from Ku80-deficient mice (78). Together, these data indicate that aNHEJ is not always distinguishable on the basis of junction structures, and provide support for the suggestion that aNHEJ may actually consist of several distinct pathways (55, 79), only some of which generate junctional “signatures.” Thus, caution must be observed when inferring the involvement of aNHEJ in aberrant recombination events in cancer genomes by sequence features alone.

As noted above, the 140+ amino acids in the RAG2 C terminus, largely conserved throughout evolution, contain several known or suspected regulatory elements. These may play important roles in maintaining the fidelity of V(D)J recombination. Clear evidence implicates the cell-cycle-regulated phosphorylation of T490 in suppressing persistence of broken DNA ends through the cell cycle and in suppressing lymphomagenesis. The PHD domain, which recognizes trimethylated histone H3K4, may play a role in limiting recognition of cRSSs located outside the antigen-receptor loci and/or in downregulating RAG cleavage activity in the absence of this histone modification (80). It should be noted, however, that these potential regulatory activities do not prevent the generation of deletions between cRSS pairs in known or suspected oncogenes and tumor-suppressor genes, as these are observed in thymic lymphomas of p53-deficient mice bearing wild-type RAG2 (38, 60) and in T- and B-cell ALL genomes from patients who presumably bear wild-type RAG alleles (7, 61). Other potential regulatory mechanisms involving the C terminus include its ability to inhibit RAG-mediated transposition (81, 82) and suppression of bi-allelic cleavage at antigen-receptor loci (83).

Clearly, V(D)J recombination fidelity is also enforced by non-RAG-specific mechanisms. These include ATM (84, 85), p53 (36, 38, 76, 77), and phosphorylated histone H2AX (86), as well as other aspects of the DNA
damage response (9, 52). It is relatively straightforward to imagine how DNA damage-response factors may act to limit errors in joining, such as limiting the persistence of broken DNA ends or the survival of cells bearing persistent broken ends. How these factors might limit errors in recognition, such as deletions between pairs of cRSSs or the survival of cells bearing such events, is less obvious. Investigating these regulatory mechanisms provides an interesting focus for future research.

REFERENCES


