Genomic instability due to V(D)J recombination-associated transposition

Yeturu V.R. Reddy,[1] Eric J. Perkins,[2,4] and Dale A. Ramsden[1,2,3,5]

[1]Lineberger Comprehensive Cancer Center, [2]Curriculum in Genetics and Molecular Biology, and [3]Department of Biochemistry and Biophysics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA

The first step in assembling immunoglobulin and T-cell receptors by V(D)J recombination has similarities to transposon excision. The excised transposon-like element then integrates into DNA targets at random in vitro, but whether this activity significantly threatens the genomic integrity of its host has been unclear. Here, we recover examples where the putative transposon associated with V(D)J recombination integrated into the genome of a pre-B-cell line. Transposition accounted for a surprisingly high proportion (one-third) of integrations, while most of the remaining events had parallels to other aberrant V(D)J recombination pathways linked to oncogenic translocation. In total, transposition occurred approximately once every 50,000 V(D)J recombinations. Transposition may thus contribute significantly to genomic instability.

[*Keywords*: DNA repair; V(D)J recombination; double-strand break repair; transposition]

Supplemental material is available at http://www.genesdev.org.

Received March 22, 2006; revised version accepted April 19, 2006.

The mammalian immune system's mature antigen-specific receptor genes are assembled during lymphocyte development by a rearrangement of separate coding segments [V(D)J recombination]. V(D)J recombination is initiated when the RAG1 and RAG2 proteins cleave chromosomal DNA at a pair of conserved recombination targeting signals that flank coding segments. The nonhomologous end-joining (NHEJ) pathway for repair of chromosome breaks then joins the ends of coding segments (coding ends) together to form the mature receptor gene, as well as typically joining signal-containing ends (signal ends) together to generate an extrachromosomal circle (for review, see Gellert 2002). RAG proteins use a transposase-like mechanism to cleave chromosomal DNA at recombination signals (van Gent et al. 1996). Importantly, the signal end fragment excised by RAG protein cleavage activity can also be likened to a transposon. It has been demonstrated that RAG proteins can direct the integration of this fragment into DNA targets approximately at random (e.g., Fig. 1A) in vitro (Agrawal et al. 1998; Hiom et al. 1998), in yeast cells (Clatworthy et al. 2003), and even in mammalian cells (Chatterji et al. 2006).

Association of transposition activity with V(D)J recombination suggests a possible source for the oncogenic translocations observed in lymphoid malignancies. Previously, these translocations were attributed in part to aberrant joining of a broken intermediate in V(D)J recombination to a spontaneous break within an oncogene locus. More recently, RAG protein activities that are not targeted by canonical recombination signals have been proposed as a less arbitrary source for the breakpoint in the oncogene locus (Hiom et al. 1998; Raghavan et al. 2004). With respect to transposition activity, for example, it has been suggested that RAG proteins could initiate recombination at a site within a receptor locus, but then "transpose" one end of the receptor locus double-strand break into a target site near an oncogene (one-ended transposition) (Hiom et al. 1998).

However, whether V(D)J recombination-associated transposition activity could be a significant source of genomic instability is not yet clear. Studies of transposition activity in cellular contexts indicate it is infrequent (Clatworthy et al. 2003; Chatterji et al. 2006), and have been limited to measuring targeting of transposition into artificial episomes: As yet, there is only one clear example where a transposition-event targeted its host genome (Messier et al. 2003). Therefore, we address here whether or not the transposon-like fragment excised during V(D)J recombination can significantly target its host genome. Moreover, to more closely mimic V(D)J recombination in the whole animal, we used a mouse pre-B-cell line as host, and a chromosomally resident recombination substrate. The substrate was further designed to determine the frequency of genomic integration of the excised fragment as a function of each excision: This is the key measure of the danger posed by V(D)J recombination-associated transposition, as the ex-

[4]Present address: Department of Molecular and Cellular Biology, Harvard University, 16 Divinity Avenue, Biolabs Room 2025, Cambridge, MA 02453, USA.
[5]Corrresponding author.
EMAIL dale_ramsden@med.unc.edu; FAX (919) 966-3015.
Article is online at http://www.genesdev.org/cgi/doi/10.1101/gad.1432706.

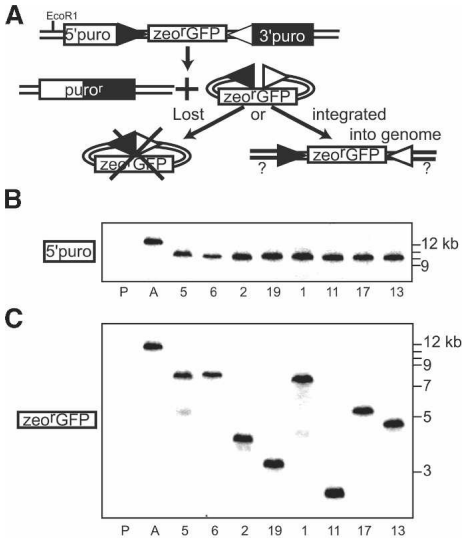

Figure 1. Recovery of integrations of a transposon-like fragment. (*A*) The assay for recovering potential transposon integrations is described. (Boxes) Coding sequence; (filled triangle) 12-type recombination signal (12-RS); (open triangle) 23-type recombination signal (23-RS); (puro^r) intact gene for puromycin resistance; (zeo^{r}GFP) fusion gene that confers zeocin resistance and green fluorescence. In *B* and *C*, DNA from subclones resistant to puromycin and zeocin after induction of V(*D*)J recombination (numbered as in Table 2) were digested with EcoRI (cuts once 5′ of the puro^r gene) and PstI (cuts only in flanking chromosomal DNA), Southern blotted, and analyzed with a probe from the puro^r gene (*B*) or zeo^{r}GFP gene (*C*). DNA from the parental line (P) and the initial clone with unrearranged substrate (clone A) are included for comparison. The mobility of molecular weight markers (in kilobase pairs) are shown at the *right* of each panel.

cision step is implicit in development of each mature lymphocyte. Our results implicate V(D)J recombination-associated transposition activity as an important possible source of oncogenic rearrangements.

Results

The substrate (Fig. 1A) was arranged such that a gene for puromycin resistance (puro^r) was interrupted by the putative transposable fragment: an intact gene for zeocin resistance (a zeocin-binding protein–green fluorescent protein fusion, or zeo^{r}GFP), flanked by one each of the pair of recombination targeting signals required for a V(D)J recombination event (a 12-RS [12-type recombination signal] and a 23-RS [23-type recombination signal]). The puromycin coding sequences flanking the signals were further adjusted to reduce the frequency of junctions that fail to confer puromycin resistance (see Materials and Methods for details). Puromycin resistance identifies cells that have successfully initiated V(D)J recombination, joined the ends of the puromycin coding region together (analogous to assembling the mature receptor gene), and have excised the potential transposon. The ends of the excised fragment are normally ligated together to form a circle, but this extrachromosomal circle is not maintained as cells continue to grow. Subcloning of cells in both puromycin and zeocin thus enriches for cells where the fragment instead reintegrated into the genome. Additional screening by flow analysis for GFP expression, PCR analysis of DNA, and finally Southern blotting was used to definitively identify clones that had reintegrated the putative transposon associated with substrate V(D)J recombination (Fig. 1B,C, Supplementary Fig. 1).

The temperature-sensitive Abelson Murine Leukemia virus (ts-AMuLV) line used as a host can be induced in culture to undergo multiple developmental steps analogous to the in vivo transition from pre-B cells to immature B cells (Muljo and Schlissel 2003), including initiation of V(D)J recombination at its endogenous immunoglobulin κ (*Igκ*) light-chain locus (Chen et al. 1994). Three different recombination substrate-containing clones were generated from the initial ts-AMuLV line, and multiple experiments were performed with each clone. We were unable to detect any significant differences between different clones or experiments in our current sample; thus, only the pooled results are discussed (see Table 1; Materials and Methods for differences in clones and experimental conditions). In total, 21 subclones were identified that had reintegrated the putative transposon, from a total of 2.2×10^7 cells screened (Table 1). Parallel assessments of cells resistant only to puromycin determined that 2.8×10^5 of the screened cells had undergone productive V(D)J recombination. The transposon-like fragment associated with V(D)J recombination thus reintegrates into the genome approximately once every 1.3×10^4 recombinations.

The selection scheme used does not distinguish between different integration pathways. However, authentic transposition products have a characteristic structure; thus, the location and junction structure for each integration was determined using inverse PCR (Table 2; Figs. 2–6). In RAG-mediated transposition in vitro, the 3′ ends of the two signals are joined to opposite strands of

Table 1. *Frequencies of reintegration*

Clone[a]	Cells screened × 10^6	Puro^r cells[b] × 10^4	No. reintegrations (transpositions)	Efficiency[c] × 10^{-5}
A	3.0	7.8	6 (3)	7.7 (3.8)
A*	0.6	0.19	2 (1)	11 (5.3)
B	11	8.8	2 (0)	2.3 (0)
C	7.5	11	11 (3)	10 (2.7)
Total	22	28	21 (7)	7.5 (2.5)

[a]Three clones were used, varying as described in Materials and Methods. In one experiment using clone A*, cells were exposed to 0.5 Gy γ radiation immediately after induction of recombination.
[b]The number of cells screened expected to be at least puromycin resistant, based on frequencies of resistance to puromycin only determined in parallel assays.
[c] Efficiencies are the reintegration frequency or, in parentheses, transposition frequency, as a fraction of the total puro^r cells screened.

Table 2. *Location of integration sites*

Clone[a]	Integration #	12-RSS flank		23-RS flank	
		Accession no.	Position	Accession no.	Position
A	1	AC026478	13041	AC026478	13037
A	2	AL606914	41953 rc[b]	AL606914	41956 rc[b]
C	3	AC141471	184523	AC141471	184519
C	4	AC126023	31830 rc[b]	AC126023	31833 rc[b]
A	5	AC073553	137778	AC073553	137775
A	6	AC073553	~139000[c]	AC073553	~139000[c]
C	7	AC115293	76224[d]	AC115293	76232[4]
C	8	AC109232	109804	M29209 (*env*)[e]	591
C	9	AC126608	103234 rc[b]	AF010170 (*gag*)[e]	2773
C	10	AL604022	31535 rc[b]	AL604022	31525 rc[b]
A	11	AC109138	164160	AC109138	164242
C	12	AY591724	2131	AY591724	2134
A	13	AC079273	45819	AC079273	45821
C	14	AC073561	160089	AC073553	133793
B	15	AY591618	2521	L80040	1098
B	16	AY591667	1805	L80040	2066
A	17	AC153615	59540 rc[b]	AY591668	1735 rc[b]
C	18	L80040	744 rc[b]	AY591682	1788 rc[b]
A	19	AY591727	1648	AY591727	383
C	20	L80040	1731 rc[b]	AY591725	1629
C	21	AY591620	1769 rc[b]	AY591682	1737 rc[b]

[a]The initial substrate-containing clone (A, B, or C; see also Table 1) that was used to generate integration-containing subclones.
[b]The position of flanking nucleotides are reported as "rc" (reverse complement) when the 12-RS and 23-RS are inserted in the opposite orientation relative to that of the published sequence.
[c]The flanking sequences recovered for clone #6 by inverse PCR contained only canonical pentanucleotide switch μ repeats, and thus they could be placed in multiple different locations within AC073553. Additionally, Southern analysis indicated germline fragment sizes were ~2 kbp larger than that predicted from the sequence in AC073553, in accordance with reports that this region is highly polymorphic in different mouse strains (Marcu et al. 1980). Further characterization of this clone by Southern analysis allowed us to determine that both 12-RS and 23-RS flanks were located ~1 kbp further away from the heavy-chain intron enhancer than the corresponding flanks of subclone #5, and that the subclone #6 integration was not accompanied by significant (>200 bp) deletion of genomic DNA flanking the integration site. Additionally, we note that at least in the reported switch μ region, the sequence TGGGG is never tandemly repeated; thus, together with the lack of significant deletion as determined by Southern analysis, we argue the structure of the subclone #6 integration is best explained by a transposition event.
[d]Sequences of the microsatellite targeted for integration from the parent clone (clone C) argued for polymorphism in this region relative to accession AC115293, and possibly within the population of clone C as well. The indicated positions were thus determined from a best-fit alignment to sequence from AC115293.
[e]For subclones #8 and #9, the 23-RS flanks contain only sequence from Moloney Murine leukemia virus genes (MoMuLV). There are no pre-existing MoMuLV proviruses near the regions identified by the12-RS flanks in published sequence, suggesting viral DNA may have "cointegrated" with the zeoʳGFP fragment. However, we cannot exclude the possibility that deletion of flanking sequence during integration was sufficient to extend into a distally located provirus, or even that there is a pre-existing proviral integration in this region that is novel to the SP9 line's strain background.

a target DNA duplex 4–5 base pairs (bp) apart, resulting in the precise insertion of the signal end fragment and a characteristic 4–5 bp direct repeat of flanking target DNA (target site duplication) (Fig. 2A,B). Six of the 21 integrations had such a structure (Fig. 2A). In addition, target sequences typically had a locally high GC content, in agreement with previously characterized RAG-mediated transpositions (Hiom et al. 1998; Clatworthy et al. 2003; Tsai et al. 2003; Chatterji et al. 2006). Target sites were otherwise distributed throughout the genome, although, interestingly, two of the six transpositions (integrations #5 and #6) (Fig. 2A) were located within a region of GC-rich pentanucleotide repeats (Sμ repeat region) that targets immunoglobulin H (*IgH*) isotype switch recombination. Southern blotting indicated there were no rearrangements of the Sμ region in the other induced subclones, arguing switch recombination is not highly active in this pre-B-cell line (Y.V.R. Reddy and D.A. Ramsden, unpubl.). It is tempting to speculate instead that the richness of secondary structure-forming sequences in this region (e.g., G quartets, hairpins) (Gellert et al. 1962; Tashiro et al. 2001) might make this region a transposition "hot spot" (see also integration #7, below).

One additional event can be attributed to an alternative transposition mechanism (integration #7) (Fig. 3A,B). Strikingly, both signal ends integrated into chromosome 5, in the middle of an ~100-bp $(CA)_N$–$(TG)_N$ inverted repeat. Hairpins that form at inverted repeat sequences are preferred targets for RAG-mediated transposition in vitro (Lee et al. 2002), and hairpin-targeted transposition might partly explain a complex rearrange-

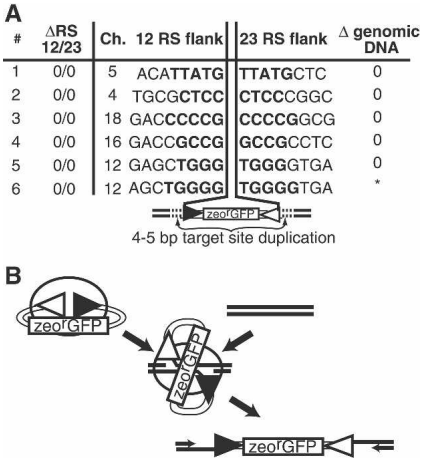

Figure 2. Integrations consistent with transpositions. (*A*) For each of the identified integration events (#, numbered as in Table 2) we report the number of base pairs deleted from 12-RS and 23-RS ends of the zeo^{r}GFP fragment (ΔRS 12/23), the chromosome where the fragment integrated (Ch.), a brief description of genomic DNA immediately flanking 12-RS ends and 23-RS ends, and the amount of genomic DNA between integration flanks (Δ genomic DNA). The sequences of flanking target site duplications are in bold. At the *bottom* is a graphic summarizing the typical integration structure. (*) As described in detail in footnote c of Table 2, the repetitive nature of sequences flanking integration #6 did not allow for unambiguous location of flanks within the region. (*B*) Pathway for integration by transposition. Ovals represent RAG proteins.

ment recovered from human cells (Messier et al. 2003). Additionally, RAG proteins can mediate a reversal of the initial cleavage reaction by inserting signal ends into hairpin coding end intermediates in vitro (Melek et al. 1998), and, under certain conditions, in vivo (Bogue et al. 1997; Sekiguchi et al. 2001). However, the presence of a target site duplication cannot be used to identify this type of event as a transposition, and inverted repeat sequences are generally associated with chromosome breakage.

We therefore determined if the chromosome 5 inverted repeat sequence was sufficient to target transposition activity in vitro. We incubated a purified RAG protein signal end complex with a plasmid containing a 415-bp fragment of the chromosome 5 sequence as it was prior to integration of the zeo^{r}GFP fragment. Of 16 in vitro transposition products, 12 integrated near the tip of the predicted hairpin, close (within 11 bp) to the cellular integration sites (Fig. 3C). This is a 30-fold greater frequency than would be expected by chance. For comparison, we performed a similar experiment using a chromosome 7 fragment that included the site of an integration that was neither associated with a flanking target site duplication nor an inverted repeat (#8) (Fig. 4). Using the chromosome 7 fragment as a target for in vitro transposition, integration sites were more dispersed (Fig. 3, cf. D and C), and were not generally near the site of cellular integration. Therefore, integration into the chromosome 5 inverted repeat is consistent with transposition targeted to a hairpin-forming sequence.

In sum, there are seven integrations attributable to RAG-mediated transposition, or approximately one transposition every 50,000 V(D)J recombinations (Table 1). We might have expected the remaining events that were not consistent with transposition also to be randomly distributed throughout the genome. Only four of the remaining 14 integrations fit into this category (Fig. 4). The cause of these integrations is unclear, but the most probable explanation is that the zeo^{r}GFP fragment was captured during repair of a spontaneous double-strand break. The remaining 10 events are more easily explained: All of the integration sites for these events were located <60 bp away from sites broken during V(D)J recombination of immunoglobulin (*Ig*) loci. As described below, these 10 integrations can be further classified as either intermolecular recombinations (Fig. 5) or end donations (Fig. 6), the two aberrant V(D)J recombination pathways previously proposed to cause the translocations associated with lymphoid malignancy (for review, see Roth 2003).

In three of these integrations (Fig. 5A), one of the sig-

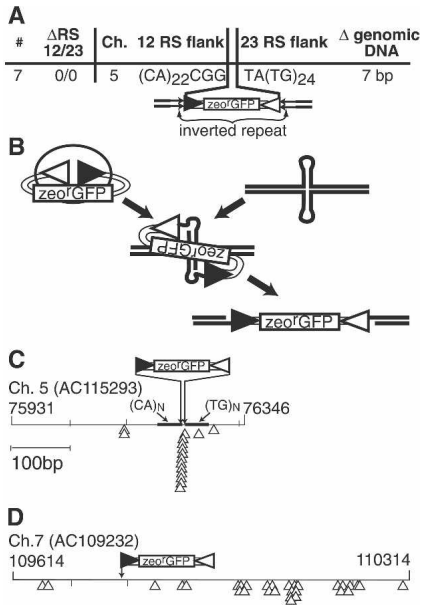

Figure 3. Integration into an inverted repeat. (*A*) The structure of integration #7 is summarized as in Figure 2A. (*B*) Pathway for integration by transposition into a hairpin-forming sequence. Ovals represent RAG proteins. (*C*) A 415-bp chromosome 5 fragment that contained the targets for cellular integration #7 is represented as a line. The termini of the cloned fragment are defined by nucleotide numbers that map to the sequence in accession AC115293. The approximate location of $(CA)_N$ and $(TG)_N$ repeats within this fragment are noted by thick lines (see also footnote d of Table 2). The locations of cellular integration sites are noted *above* the line, while the locations of in vitro-defined integrations are noted with open triangles *below* the line. (*D*) Seven-hundred base pairs of a chromosome 7 fragment (from AC109232) that contained the target for cellular integration #8 are represented as in *C*. The site of integration of the 12-RS of the zeo^{r}GFP fragment is noted *above* the line. The 23-RS flank could not be located in this region (see footnote e of Table 2).

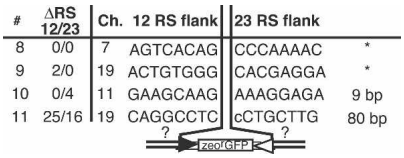

Figure 4. Other nontargeted integrations. The structures of integration #8–11 are summarized as in Figures 2A and 3A. A possible inserted nucleotide (nontemplated) is noted in lowercase. (*) Sequences flanking the 23-RS for #8 and #9 were derived from Moloney Murine leukemia virus genes, and could not be located near the 12-RS flanks. (See also footnote e of Table 2.)

nal ends of the zeo^{r}GFP fragment was joined to an *Igκ* or *IgH* signal end of the opposite type, generating a precise 12-RS/23-RS-type signal junction. The remaining signal end of the zeo^{r}GFP fragment was imprecisely joined to a *Igκ*/*IgH* coding end. These integrations can be attributed to intermolecular recombination. The RAG proteins mistakenly juxtaposed a signal from the *Igκ* or *IgH* locus with a signal of opposite type from the zeo^{r}GFP fragment (intermolecular synapsis) (Fig. 5B), triggering cleavage at the *Igκ*/*IgH* signal. The zeo^{r}GFP signal may have been part of a signal junction circle that is recleaved as a consequence of the intermolecular synapsis (as shown in Fig. 5B), as there is precedent for secondary recombinations involving a signal within a signal junction (Lewis et al. 1985). However, we cannot exclude the possibility that signal junction fomation is unnecessary, and a zeo^{r}GFP signal end participated in synapsis with, and triggered cleavage of, the *IgH* signal. In any event, integration is the consequence of the subsequent resolution of all four broken ends by NHEJ, as described in Figure 5B. Importantly, the intermolecular recombinations described here can be likened to those that occur between a signal within an antigen receptor locus and a cryptic signal within an oncogene locus (Fig. 5C), leading to a subset of cancer-causing translocations. In both cases, the critical aberrant step is the inappropriate intermolecular synapsis of signals by RAG proteins.

Seven other examples also involve targeting of integration to *Ig* loci (Fig. 6A). However, only one of the 14 junctions from these integrations involves two signals (integration #21), and joining was never precise (numbers of nucleotides deleted from *Igκ* ends are noted in parentheses in Fig. 6A). Instead, zeo^{r}GFP signal ends were often perfectly retained and usually joined to ends of *Igκ* coding segments (coding ends) with short deletions, a pattern consistent with previous examples of NHEJ-dependent joining of comparable ends (Sekiguchi et al. 2001). Broken intermediates from a single V(D)J recombination event are normally joined by NHEJ to each other (joining in *cis*): In these seven events, joining in *cis* apparently failed, both for the signal end intermediates from recombination at the substrate locus, as well as for the intermediates from an independent recombination occurring in parallel within the *Igκ* locus. Instead, the intermediates from these two independent recombinations were joined to each other in *trans* (Fig. 6B). Such events are comparable to oncogenic "end donations," where intermediates from an antigen receptor recombination are joined to the ends of a double-strand break generated independently within an oncogene locus (Fig. 6C). The critical aberrant step in both examples is the failure of NHEJ to join broken ends in *cis*, leading instead to inappropriate joining of broken ends in *trans*.

Discussion

Integrations of the transposon-like fragment could thus often be explained by one or the other of the two aberrant recombination pathways mechanisms previously associated with oncogenic translocation (intermolecular recombination or end donation). Does the high proportion (one-third) of transpositions thus implicate transposition activity as an important additional source of oncogenic translocations? The assay described here is specific for integrations involving a signal end pair, a restriction that was essential for definitive identification of transpositions. Recovery of transpositions will thus be favored relative to other integration pathways. However, we will also not score the kinds of transpositions (Hiom et al. 1998) best reconciled with oncogenic translocations (e.g., one-ended transpositions, cycles of integration, and excision). Indeed, these alternate transposition pathways were initially proposed because they can generate aber-

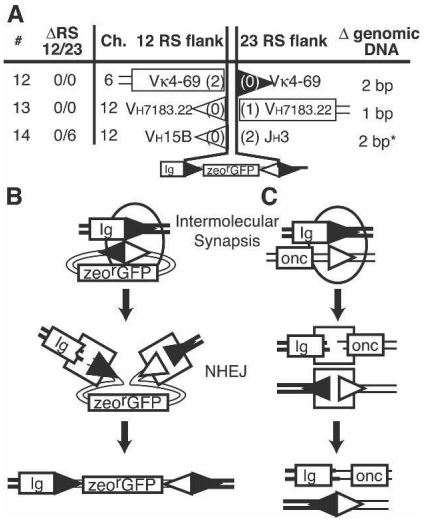

Figure 5. Integration by intermolecular V(D)J recombination. (*A*) The structure of integrations #12–14 are summarized as in Figures 2–4. Rectangles represent flanking immunoglobulin (Ig) locus coding segments, and triangles represent flanking *Ig* locus recombination signals. The numbers in parentheses within these symbols refer to the base pairs of the genomic sequence deleted relative to the site of RAG protein cleavage. (*) Integration #14 is the product of two recombinations, as described in Supplementary Figure 2. *B* and *C* describe how intermolecular recombination can similarly lead to both integration of the zeo^{r}GFP fragment (*B*) as well as translocations between receptor loci and oncogene (*onc*) loci (*C*). Ovals represent RAG proteins, while boxes represent NHEJ proteins.

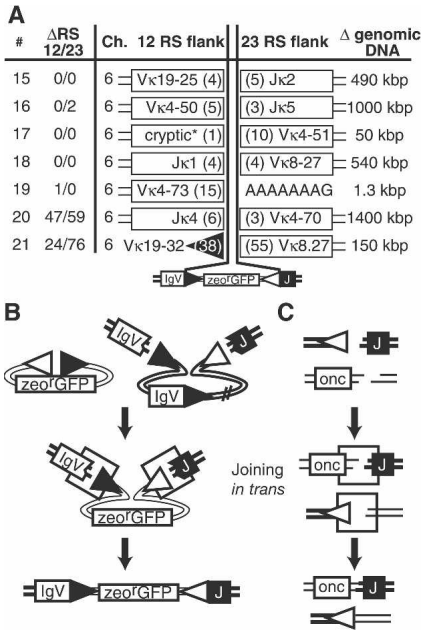

Figure 6. Integration by end donation. (*A*) The structure of integrations #15–21 are summarized as in Figures 2–5. Rectangles represent flanking *Ig* locus coding segments, and triangles represent flanking *Ig* locus recombination signals. The numbers in parentheses within these symbols refer to the base pairs of the genomic sequence deleted relative to the site of RAG protein cleavage. (*) Sequence flanks a cryptic 23-type signal within the *V*κ region: CACtGTG (23 bp) AgAAAAACC (nucleotides that match consensus RS in capital letters). *B* and *C* describe how end donation can similarly lead to both integration of the zeo^{r}GFP fragment (*B*) as well as translocations between receptor loci and *onc* loci (*C*). Boxes represent NHEJ proteins.

rant rearrangement structures that are mostly indistinguishable from those previously attributed to end donation. Therefore, with respect to the oncogenic potential of transposition, we consider it noteworthy that at least when transpositions can be definitively identified, transposition competes effectively with the two pathways previously used to explain oncogenic translocation. Perhaps more importantly, transposition was the primary means (7/11 events) by which signal end intermediates were resolved in a near-random, and thus potentially oncogenic, manner. In contrast, the probable intermolecular recombinations and end donations we recovered were mostly targeted to a recombining locus, and consequently made "safe."

Several characteristics of our assay argue for its biological relevance. The use of a substrate where the transposon donor was chromosomally integrated was probably critical. It has not been possible to recover transpositions in cellular genomes from an episome-based donor, and transposition between episomes appears to be much less efficient (Chatterji et al. 2006). Additionally, we used a cell line of lymphocyte lineage, and consequently native RAG1 and RAG2 proteins, expressed under native transcriptional control. As noted above, concurrent recombination activity at *Ig* loci also led to other biologically relevant aberrant resolutions, allowing us to assess how well transposition competes with such events. Recombination activity was also not excessive, as ~15% of light chain loci (and ~3% of substrate loci) recombined in the 24-h period during which recombination was active. Nevertheless, transposition frequencies in a whole animal could still be significantly different. For example, we might be overestimating transposition in vivo due to characteristics of this transformed cell line model (e.g., probable disruption of DNA damage checkpoints). Alternatively, our assay might underestimate transposition in vivo, as it is unlikely we recover all integrations (the ability to express sufficient zeo^{r}GFP to survive selection is probably highly dependent on integration context).

Is V(D)J recombination-associated transposition an important threat to genomic integrity? At least under conditions where transpositions can be definitively identified, we show it accounts for a high proportion of aberrant V(D)J recombinations. We estimate there are 2.5×10^{-5} events associated with each V(D)J recombination. For comparison, the well-characterized sleeping beauty transposon, which has been engineered to mutagenize the mouse genome, transposes in embryonic stem cells with similar or lower frequency (3.5×10^{-5} to 2×10^{-7} events/cell/generation) (Luo et al. 1998). Humans generate several hundred million new lymphocytes/day, with each new lymphocyte requiring at least three V(D)J recombination events. Our data would thus suggest this daily output of lymphocytes is accompanied by as many as 10,000 transpositions: in all liklihood, an important potential source of oncogenic genome rearrangements.

Materials and methods

Constructs and cell lines

The substrate was derived from amplified fragments of pPUR and pTRACER-EF plasmids (Clontech), and assembled as described in Figure 1 within a murine stem cell virus-based retrovirus vector (Cheng et al. 1998). Diversity in coding junction formation could potentially lead to infrequent puromycin resistance, reducing the sensitivity of our assay. Therefore, since NHEJ often precisely deletes repeated sequence at ends during coding junction formation, we designed the substrate such that the 4 bp of the puromycin coding sequence immediately flanking recombination signals was repeated. Sequencing of coding junctions formed in the absence of selection indicated the puromycin ORF was precisely assembled in three out of five V(D)J recombinations.

Retrovirus was prepared by cotransfection of the substrate plasmid together with plasmids expressing *gag, pol,* and *vsv-g* genes into 293T cells (Pear et al. 1993). Virus recovered from the supernatant was used to infect the ts-AMuLV line SP-9 (gift of Y. Chang, University of New Mexico, Albuquerque, NM) (Chang and Brown 1999). Clones were obtained by limiting dilution plating in media with 100 μg/mL zeocin, and those with unique, single-copy integrations were identified by Southern blotting of genomic DNA from zeo^r colonies. The construct used to generate clones A and B possessed a deletion of 1 nucleo-

tide (nt) within the 23-RS nonconserved spacer. This is within the ±1-bp natural variation of spacer size used in vivo; thus, we continued study of lines made with this construct, but additionally corrected the substrate and generated a line with a consensus spacer size 23-RS (Clone C). The integrated version of the substrate in Clone A also has a 400-bp deletion between the 12-bp spacer RS and zeo^{r}GFP reading frame that apparently had no significant impact on zeo^{r}GFP expression.

Screening colonies

Cells were washed once and V(D)J recombination induced in the absence of both puromycin and zeocin by transfer to 40°C for 24 h. Induced cultures were then returned to preinduction conditions to recover for 24 h, still without antibiotics. Cells were then subcloned by plating in 96-well plates at 2×10^3 cells/plate in media with 5 μg/mL puromycin alone [to determine the frequency of productive V(D)J recombination], or at 10^5 cells/plate in 5 μg/mL puromycin and 100 μg/mL zeocin. Screening by flow analysis indicated a small proportion of subclones that survived in zeocin nevertheless had green fluorescence indistinguishable from background (within 10% of the median fluorescence intensity of the parental line). This indicates cells can infrequently acquire zeocin resistance by means other than maintenance of the zeo^{r}GFP cassette: Later analysis of DNA from selected zeo^r, GFP negative clones by PCR (Supplementary Fig. 3A, P4/P7) confirmed the absence of the zeo^{r}GFP cassette. DNA was prepared from the remaining clones and analyzed by PCR to confirm coding junction formation (assembly of a complete puromoycin gene) (Supplementary Fig. 3A, P2/P9), as well as loss of the "germline" 5′ puro fragment–recombination signal–zeo^{r}GFP configuration (Supplementary Fig. 3A, P2/P5).

The majority of puro^r–zeo^r clones were eliminated from further analysis when they tested positive for both coding junction and "germline" PCR products. Southern analysis of selected clones of this type determined this was due to apparent duplication of the substrate in an estimated 1% of the parental line prior to induction of recombination: Upon induction of recombination in these cells, one copy recombined, conferring puromycin resistance to the cell, while the remaining copy remained in the unrecombined configuration, allowing the cell to also retain resistance to zeocin. Such duplications may have been caused by mobilization of our proviral substrate by a helper virus in *trans*, or possibly unequal sister chromaid exchange at proviral long terminal repeats.

The 21 puro^r–zeo^r subclones remaining after screening by PCR analysis were then analyzed by Southern blotting, and confirmed to have a pattern consistent with transposition; that is, relative to the clone before recombination, a 5′ puro hybridizing species showed evidence of transposon excision (~2 kb smaller) (Fig. 1B; Supplementary Fig. 1A,C,E), while zeo^{r}GFP hybridizing species were of a distinct size, but varied essentially at random (Fig. 1C; Supplementary Fig. 1B,D,F). Zeo^{r}GFP (543 bp) and Puro (423 bp) probe fragments for Southern blot analysis were generated by PCR amplification of substrate with primer pairs P3/P8 and P1/P6, respectively (Supplementary Fig. 3). Each probe fragment was radiolabeled with α^{32}P-dCTP by two rounds of annealing with appropriate primers and extension with klenow polymerase. Hybridizations were performed in buffer with 50% formamide at 43°C for the zeo^{r}GFP probe and 49°C for the Puromycin probe using standard protocols.

In vitro transposition

The in vitro transposition reactions were performed as previously described, using the same oligonucleotide recombination signals and PCR primers (Lee et al. 2002), except we used as target plasmid cloned chromosomal sequences that acted as cellular targets for integrations #7 and #8. Briefly, a complex of recombinant core RAG1 and RAG2 proteins, HMG1, and oligonucleotide recombination signals was first formed by preincubation in the presence of $CaCl_2$. Cleavage of recombination signals and transposition was then initiated by addition of $MgCl_2$ and 0.8 μg of plasmid target. Integrations of cleaved signals were then PCR-amplified from deproteinized reactions using a primer in the recombination signal and a primer in the plasmid target. Target plasmid carried through the transposition and PCR reactions was then inactivated by digestion with DpnI before PCR products were cloned (TOPO-TA, Invitrogen). Consistent with integration by a transposition mechanism, we recovered only integrations of accurately cleaved recombination signals, despite using oligonucleotide signals with flanking "coding" sequence.

Characterization of junctions

Junction sequences were cloned by inverse PCR. Briefly, genomic DNA was first enriched for fragments with integrations by separating EcoRI-digested genomic DNA by electrophoresis on a 0.8% agarose gel, followed by recovery of DNA of molecular weight appropriate to species with integrations as determined by prior Southern analysis. This DNA was then digested with HaeIII or MspI, ligated with T4-DNA ligase at 0.25–0.5 μg/mL, amplified by 25 cycles using the 12R1/12F1 and 23F1/23R1 primer pairs (Supplementary Fig. 3B), diluted 50-fold, and amplified again for 25 cycles with 12R2/12F2 and 23F2/23R2 (Supplementary Fig. 3B). PCR products were then cloned using a TOPO-TA cloning kit (Invitrogen) and sequenced. We then used the sequences derived from inverse PCR to design primers that permitted direct amplification of integration junctions.

PCR primers

The following primers were used for PCR amplifications; their approximate position within our substrate is noted in Supplementary Figure 3: (5′–3′) P1, CCAGCCGGGAACCGCTCAACT; P2, CGAGCTGCAAGAACTCTTCCTCAC; P3, CTCCAATTGGCGATGGCCCTGTCC; P4, GGTCACGCTTTTCGTTGGGATCTTTCG; P5, TACTCAGACAATGCGATGGG; P6, GAATTCTAGGCTTTTGCAAAAAGCTT; P7, AGATGACGGGAACTCCAAGACGC; P8, ATAAACAAGTTTCGAGGTCGAGTGTCAGTC; P9, CGCTCGTAGAAGGGGAGGTT; 12R1, GCCAGGCGGGCCATTTACCGTA; 12R2, TACTCAGACAATGCGATGGG; 12R3, GAAATCCCCGTGAGTCAAAC; 12F1, GACGCAAATGGGCGGTAGGC; 12F2, GTACGGTGGGAGGTCTATAT; 23R1, GTAACCATTATAAGCTGCAATAAAC; 23R2, GAGGTCGAGTGTCAGTCCTGC; 23F1, TCACTGCATTCTAGTTGTGGTTTG; 23F2, GTGGTTTGTCCAAACTCATCAATG.

PCR reactions using P2/P9 required inclusion of 1 M betaine (Sigma).

Acknowledgments

We thank Kevin Griffin, Ayyappan Nair, and Brett Weed for their assistance, all members of the Ramsden laboratory for helpful comments, and Martin Gellert, Dik van Gent, Jeff Sekelsky, and Sarah Radford for critical reading of the manuscript. This project was supported by a Searle Scholars award, ACS grant RPG-99-192-01-GMC, NIH grants 84442 and 97096, and a Leukemia and Lymphoma Society Scholar award to D.A.R.

References

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

Bogue, M.A., Wang, C., Zhu, C., and Roth, D.B. 1997. V(D)J recombination in Ku86-deficient mice: Distinct effects on coding, signal, and hybrid joint formation. *Immunity* **7:** 37–47.

Chang, Y. and Brown, M.L. 1999. Formation of coding joints in V(D)J recombination-inducible severe combined immune deficient pre-B cell lines. *Proc. Natl. Acad. Sci.* **96:** 191–196.

Chatterji, M., Tsai, C.L., and Schatz, D.G. 2006. Mobilization of RAG-generated signal ends by transposition and insertion in vivo. *Mol. Cell. Biol.* **26:** 1558–1568.

Chen, Y.Y., Wang, L.C., Huang, M.S., and Rosenberg, N. 1994. An active v-abl protein tyrosine kinase blocks immunoglobulin light-chain gene rearrangement. *Genes & Dev.* **8:** 688–697.

Cheng, L., Du, C., Lavau, C., Chen, S., Tong, J., Chen, B.P., Scollay, R., Hawley, R.G., and Hill, B. 1998. Sustained gene expression in retrovirally transduced, engrafting human hematopoietic stem cells and their lympho-myeloid progeny. *Blood* **92:** 83–92.

Clatworthy, A.E., Valencia, M.A., Haber, J.E., and Oettinger, M.A. 2003. V(D)J recombination and RAG-mediated transposition in yeast. *Mol. Cell* **12:** 489–499.

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

Gellert, M., Lipsett, M.N., and Davies, D.R. 1962. Helix formation by guanylic acid. *Proc. Natl. Acad. Sci.* **48:** 2013–2018.

Hiom, K., Melek, M., and Gellert, M. 1998. DNA transposition by the RAG1 and RAG2 proteins: A possible source of oncogenic translocations. *Cell* **94:** 463–470.

Lee, G.S., Neiditch, M.B., Sinden, R.R., and Roth, D.B. 2002. Targeted transposition by the V(D)J recombinase. *Mol. Cell. Biol.* **22:** 2068–2077.

Lewis, S., Gifford, A., and Baltimore, D. 1985. DNA elements are asymmetrically joined during the site-specific recombination of κ immunoglobulin genes. *Science* **228:** 677–685.

Luo, G., Ivics, Z., Izsvak, Z., and Bradley, A. 1998. Chromosomal transposition of a Tc1/mariner-like element in mouse embryonic stem cells. *Proc. Natl. Acad. Sci.* **95:** 10769–10773.

Marcu, K.B., Banerji, J., Penncavage, N.A., Lang, R., and Arnheim, N. 1980. 5′ Flanking region of immunoglobulin heavy chain constant region genes displays length heterogeneity in germlines of inbred mouse strains. *Cell* **22:** 187–196.

Melek, M., Gellert, M., and van Gent, D.C. 1998. Rejoining of DNA by the RAG1 and RAG2 proteins. *Science* **280:** 301–303.

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

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

Pear, W.S., Nolan, G.P., Scott, M.L., and Baltimore, D. 1993. Production of high-titer helper-free retroviruses by transient transfection. *Proc. Natl. Acad. Sci.* **90:** 8392–8396.

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

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

Sekiguchi, J.A., Whitlow, S., and Alt, F.W. 2001. Increased accumulation of hybrid V(D)J joins in cells expressing truncated versus full-length RAGs. *Mol. Cell* **8:** 1383–1390.

Tashiro, J., Kinoshita, K., and Honjo, T. 2001. Palindromic but not G-rich sequences are targets of class switch recombination. *Int. Immunol.* **13:** 495–505.

Tsai, C.L., Chatterji, M., and Schatz, D.G. 2003. DNA mismatches and GC-rich motifs target transposition by the RAG1/RAG2 transposase. *Nucleic Acids Res.* **31:** 6180–6190.

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