MOLECULAR AND CELLULAR BIOLOGY, July 1993, p. 3900–3906
0270-7306/93/073900-07$02.00/0

Vol. 13, No. 7

Unequal Signal and Coding Joint Formation in Human V(D)J Recombination

GEORGE H. GAUSS AND MICHAEL R. LIEBER*

Laboratory of Experimental Oncology, Department of Pathology, Stanford University School of Medicine, Stanford, California 94305-5324

Received 4 February 1993/Returned for modification 17 March 1993/Accepted 6 April 1993

Substrates for studying V(D)J recombination in human cells and two human pre-B-cell lines that have active V(D)J recombination activity are described. Using these substrates, we have been able to analyze the relative efficiency of signal joint and coding joint formation. Coding joint formation was five- to sixfold less efficient than signal joint formation in both cell lines. This imbalance between the two halves of the reaction was demonstrated on deletional substrates, where each joint is assayed individually. In both cell lines, the inversional reaction (which requires formation of both a signal and a coding joint) was more than 20-fold less efficient than signal joint formation alone. The signal and coding sequences are identical in all of these substrates. Hence, the basis for these differential reaction ratios appears to be that coding joint and signal joint formation are both inefficient and their combined effects are such that inversions (two-joint reactions) reflect the product of these inefficiencies. Physiologically, these results have two implications. First, they show how signal and coding joint formation efficiencies can affect the ratio of deletional to inversional products at endogenous loci. Second, the fact that not all signal and coding joints go to completion implies that the recombinase is generating numerous broken ends. Such unresolved ends may participate in pathologic chromosomal rearrangements even when the other half of the same reaction may have proceeded to resolution.

The exons encoding the immunoglobulin and T-cell receptor variable region domains are assembled during lymphocyte development by V(D)J recombination (43). In this reaction, two site-specific cuts are made to create four DNA ends, two of which (the coding ends) are joined to generate the exon that encodes the variable domain of the antigen receptor (Fig. 1). The two signal ends are joined to form a signal joint. Modification of the coding ends occurs by terminal deoxynucleotide transferase (TdT)-associated nucleotide addition (3, 6, 19, 24), by P-element addition (18, 28), and by nucleolytic chewback of the exposed ends prior to joint resolution (22, 37). These modifications are important to the diversification of the antigen receptor repertoire.

Most of what we know about the mechanism of V(D)J recombination is from analyses of murine cells. Some information can be gathered from analysis of human endogenous antigen receptor loci (31, 33, 38, 40, 44). The difficulty in drawing conclusions from the endogenous loci is that these events can be highly skewed by immunologic selection, cellular selection, clonal expansion, polymerase chain reaction amplification preferences, and biases due to large numbers of polymerase chain reaction sister products.

For those loci in the genome that undergo recombination by inversion, the necessity for both coding joint and signal joint resolution is obvious: the failure of either joint to form results in a chromosomal break. At least 20% of all childhood malignancies appear to involve translocations derived from site-directed lymphoid recombinations (23). Chromosomal translocations are detectable in the lymphoid tissues of approximately 50% of the human population (1, 27). Determining the degree to which steps of the V(D)J recombination reaction are tightly coupled is essential for understanding the extent of V(D)J recombinase-mediated genomic instability.

* Corresponding author.

Four lines of work indicate that signal joints and coding joints are not always completed and are uncoupled. First, it has been known for some time that there is independent processing of coding and signal ends. End modifications commonly observed at coding joints are infrequently detected at signal joints (24, 32). Second, data that may be considered relevant to the uncoupling of signal joint and coding joint formation are obtained from the disruption of V(D)J recombination by the murine *scid* defect. In mice expressing the *scid* mutation, signal joint formation occurs in the absence of normal coding joint formation (4, 5, 25, 39). Third, free signal ends have been detected in normal thymic T cells, whereas unresolved coding ends were not detected (36). It is unclear whether this finding indicates differential survival of free coding and signal ends; uncoupling of the formation of these two products; or mechanistic coupling, but with different resolution half-times. Fourth, and most directly thus far, coding joint resolution and signal joint resolution have been measured on plasmid substrates under circumstances in which one of the two joint resolution reactions was markedly constrained (41). Disruption of normal joint resolution was achieved by reducing the distance between recombination signals. The data indicated that coding joint formation and signal joint formation are separable events in wild-type cells.

In this report, we describe a system for examining the human V(D)J recombination reaction, using extrachromosomal V(D)J recombination substrates transfected into lymphoid cell lines. We have also identified two human pre-B-cell lines, Reh and Nalm-6, which permit analysis of the recombination reaction. Three types of plasmid recombination substrates were constructed; two different deletion substrates were used to analyze signal and coding joint formation individually (pSJ and pCJ, respectively), and one inversion substrate was used to analyze signal and coding joint formation together (pINV). In both of the human pre-B-cell lines studied, signal joint formation was higher

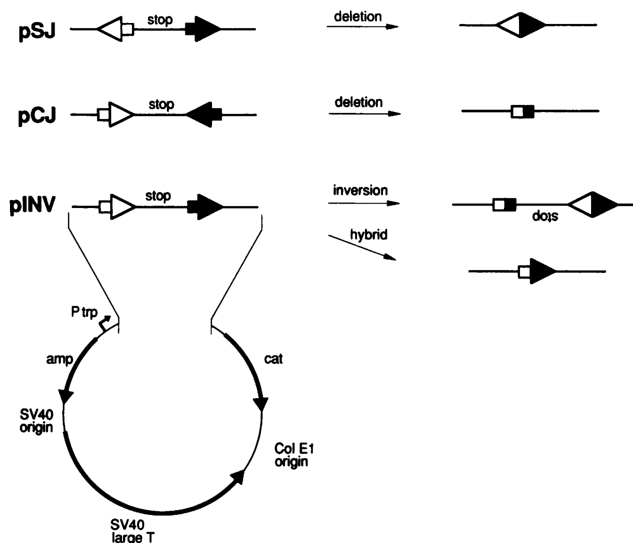

FIG. 1. Human V(D)J recombination substrates. The ColE1-SV40 shuttle vector backbone is shown at the lower left. Eukaryotic replication of transfected substrates is directed by the SV40 origin and large T antigen, indicated by SV40 origin and SV40 large T, respectively. Col E1 origin and amp represent the prokaryotic replication origin and β-lactamase (ampicillin resistance) gene, respectively; P trp and cat indicate the prokaryotic tryptophan promoter and a promoterless *cat* gene, respectively. Inserted as a cassette between P *trp* and *cat* is the OOP transcription terminator from bacteriophage λ, represented by stop in the three cassettes indicated at the upper left. The terminator is flanked by 12- and 23-bp spacer V(D)J recombination signal sequences, indicated by open and closed triangles. Upon transfection into human lymphoid cells, V(D)J recombination results in the deletion or inversion of the terminator, generating the recombinant products shown at the upper right. Substrates are recovered from the lymphoid cells and transformed into bacteria, where recombinant substrates confer both chloramphenicol resistance and ampicillin resistance. Nonrecombinant substrates confer ampicillin resistance only.

than coding joint formation. Inversions formed at much lower frequencies than did either the signal or coding deletional reaction. Inversions require both signal and coding joint formation, whereas deletions merely require one (pSJ requires a signal joint, and pCJ requires a coding joint). Signal and coding joint formation at less than 100% efficiency appears to explain the lower frequency of inversional products. The two-joint reaction (inversion) is limited by the combined inefficiencies of signal and coding joint formation.

MATERIALS AND METHODS

Substrates. The three types of V(D)J recombination substrates used in this study share a common ColE1-simian virus 40 (SV40) shuttle vector backbone and differ from one another only in the orientation of their respective V(D)J recombination signal sequences (RSS). The substrates were constructed in a stepwise fashion starting with pBR327 (42). pBR327 bears the prokaryotic ColE1 replication origin and β-lactamase gene conferring ampicillin resistance. In the first step of the construction, the SV40 replication origin and large T-antigen coding sequence (3.1-kb *Hpa*II-*Bam*HI fragment from the SV40 genome) were cloned into the *Dra*I site of pBR327. Second, the gene conferring tetracycline resistance was deleted from pBR327 by converting the pBR327 *Ava*I site into an *Eco*RI site with a linker and removing the resulting 1.4-kb *Eco*RI-*Eco*RI fragment. (This pBR327 derivative was a gift from Michael Seidman.) Third, a 29-bp synthetic oligonucleotide (ML99/100) containing *Sal*I and *Bgl*II restriction sites was inserted into the pBR327 *Eco*RI site. Fourth, the 1.2-kb *Sal*I-*Bgl*II fragment of pJH298 (25) was cloned into the *Sal*I and *Bgl*II sites of ML99/100. This 1.2-kb fragment contains a prokaryotic transcription terminator, a 23-bp spacer RSS, and the chloramphenicol acetyltransferase (*cat*) gene. Fifth, a synthetic oligonucleotide (ML81/82 or ML81.1/82.1) containing the prokaryotic tryptophan operon promoter (P_{trp}) and a 12-bp spacer RSS was inserted into the *Sal*I site of ML99/100. ML81/82 was used for pSJ, pCJ, and pINV; ML81.1/82.1 was used for pCJ1 and pINV1. Finally, the five different substrates were made by reversing the orientations of the 12- and 23-bp spacer RSS, which are flanked by *Sal*I and *Bam*HI sites, respectively.

The sequences of the oligonucleotides used in the construction of our substrates are as follows: ML81/82 (*Sal*I-compatible ends), 5′-TCGATGAATTCCCCTGTTGACAAT TAATCATCGAACTAGTTAACTATACGCAGCTTGGCT GCAGGTCGACCACAGTGCTACAGACTGGAACAAAA ACCCTGCAG-3′; ML81.1/82.1 (*Sal*I-compatible ends), 5′-TCGATGAATTCCCCTGTTGACAATTAATCATCGAAC TAGTTAACTATACGCAGCTTGGCTGCAGGTCGACA CAGTGCTACAGACTGGAACAAAAACAG-3′; ML99/100 (*Eco*RI-compatible ends), 5′-AATTGTCGACGGGGGAGA TCTGGCGCCGG-3′; and *Eco*RI linker (blunt ends), 5′-GGAATTCC-3′. These oligonucleotides and their complementary sequences were synthesized (Applied Biosystems 380A synthesizer) and then annealed, leaving 5′ overhanging, single-stranded ends to facilitate cloning.

Cell lines and tissue culture. The human cell lines Reh and Nalm-6 were obtained from the American Type Culture Collection (Rockville, Md.) and are pre-B-cell lines that were originally derived from patients with acute lymphoblastic leukemia (ALL) (16, 34). Reh and Nalm-6 were maintained in RPMI 1640 supplemented with 10% fetal bovine serum, 100 U of penicillin G per ml, 100 U of streptomycin sulfate per ml, and 50 μM 2-mercaptoethanol.

V(D)J recombination assay. Human lymphoid cell lines are transfected with plasmid substrates by using an electroporation/DEAE-dextran method as described previously (8) and cultured for 48 h. During this time, a fraction of the substrates are rearranged by V(D)J recombination. Plasmid DNA is subsequently recovered from the cells by alkaline lysis (13) and electrotransformed into *Escherichia coli* (15). Bacterial transformants are selected on LB plates containing ampicillin (100 μg/ml) or ampicillin and chloramphenicol (100 and 22 μg/ml, respectively). In bacteria, all substrates confer ampicillin resistance (designated A). V(D)J recombination results in the deletion (pSJ and pCJ) or inversion (pINV) of the prokaryotic transcription terminator on the substrates, allowing *cat* gene expression in bacteria. Thus, recombinant substrates confer both ampicillin and chloramphenicol resistance (designated AC). The ratio of AC colonies to A colonies reflects the fraction of recovered substrates that underwent V(D)J recombination. A more meaningful ratio can be obtained by focusing on those molecules that have replicated in the eukaryotic cells (26). The ratio of AC to A among replicated molecules is designated *R* and represents the recombination frequency of the substrate. We control for replication by digesting the recovered substrates with *Dpn*I before bacterial transformation. Eukaryotic replicated molecules lose their *dam* methylation and are *Dpn*I resistant; nonreplicated molecules are *Dpn*I sensitive, and thus only replicated (i.e., transfected) mole-

cules transform the bacteria. pCJ and pCJ1 had comparable *R* values in both cell lines, as did pINV and pINV1.

Typically, the number of AC and A colonies from a single bacterial transformation was 10 to 200 AC and 1×10^3 to 20×10^3 A colonies. AC colonies from nontransfected substrates were not detectable (<1 AC colony for every 2×10^6 unrecombined substrates transformed directly into bacteria without passage through lymphoid cells).

Recombinant and nonrecombinant substrates replicate at equal frequency in eukaryotic cells. For each substrate, a 50:50 mixture of recombinant and nonrecombinant substrates was transfected into eukaryotic cells and recovered 48 h posttransfection. Recovered molecules were found to be 50 ± 5% recombinant, indicating equivalent eukaryotic replication.

The plating efficiency of recombinant pSJ, pCJ, pCJ1, pINV, and pINV1 on AC plates was 90 ± 5% for all three substrates.

Analysis of recombinants. Individual DAC (see the footnote to Table 1 for definitions of DAC and DA) colonies were cultured overnight in LB-ampicillin, and plasmid DNA was isolated by alkaline lysis. With the two V(D)J recombination-positive cell lines used in this study, greater than 95% of the DAC colonies represented bona fide V(D)J events, as judged from high-resolution restriction analysis of all DAC colonies picked and sequencing of representative samples of each type of recombinant substrate to confirm the restriction analysis. The remaining 1 to 5% DAC colonies, which did not have the expected restriction pattern based on V(D)J rearrangement, were excluded from subsequent analysis. The level of background DAC colonies, when the substrates were transfected into human lymphoid cell lines which had no V(D)J recombination activity in our assay, was $<1.4 \times 10^{-6}$ DAC colonies per replicated substrate recovered. The following restriction digests were used to determine the structures of recombinant substrates: with the *Hgi*AI digest a precise signal junction on pSJ, pINV, and pINV1 generates a new *Hgi*AI restriction site. The *Bam*HI-*Sal*I digest, in the analysis of recombinants with imprecise signal junctions, generates an 82- to 92-bp fragment with pSJ, pINV, and pINV1 recombinants, the size of the band depending on the amount of nucleotide addition at the signal junction. These digests were run on 8% polyacrylamide gels, permitting resolution to approximately ±3 bp. The *Eco*RI-*Pvu*II digest was used to distinguish hybrid junctions from inversions with pINV and pINV1 and to analyze coding junctions with pCJ and pCJ1. This digest generates a 273 ± 10-bp band with hybrid junctions on pINV and pINV1 recombinants, again the size of the band depending on the amount of nucleotide loss or addition at the junction. In analyzing coding junctions, a 230 ± 10-bp band is generated upon *Eco*RI-*Pvu*II digestion of pCJ and pCJ1 recombinants. *Eco*RI-*Pvu*II digests were run on 5% polyacrylamide gels.

Statistics. Probabilities of inverted repeat junctional additions occurring by chance were calculated by using the binomial distribution as described in Meier and Lewis (29). The probabilities of A, C, G, and T were taken to be 0.161, 0.327, 0.344, and 0.168, based on a collection of 741 junctional nucleotides seen in recombination events. This collection includes those reported by Meier and Lewis (29) and 196 additional junctional nucleotides observed in the coding junctions of the substrates pCJ and pINV as reported in another study (10).

TABLE 1. Recombination frequencies in various human V(D)J recombination substrates[a]

Cell line and substrate	No. of molecules		R (%)	Range (%)
	DA	DAC		
Reh				
pSJ	77,565	8,890	11.5	10.6–20.6
pCJ	35,061	828	2.4	1.8–5.3
pINV	19,465	212	0.55	0.25–1.5
Nalm-6				
pSJ	28,763	1,163	4.0	1–4.6
pCJ	55,863	352	0.63	0.3–2.1
pINV	33,001	61	0.18	0.04–0.33

[a] Substrates were transfected individually into the cell lines shown, recovered 48 h posttransfection, digested with *Dpn*I, and transformed into bacteria. The recombination frequency, *R*, of the substrates was measured by the ratio of the number of replicated, recombinant substrates (designated DAC for *Dpn*I resistant, chloramphenicol and ampicillin resistant) to the total number of replicated substrates (designated DA for *Dpn*I resistant and ampicillin resistant) recovered from the cells. Data for DA and DAC represent the summations of five independent transfections for each of the substrates. The range of values of *R* among the five independent transfections is given in the last column. Hybrid joints have been excluded from the tabulation of inversion events of pINV. Inclusion of pINV hybrid joints with inversions would increase DAC and *R* by a factor of 2.0 for Reh cells and by a factor of 1.05 for Nalm-6 cells.

RESULTS

The human assay system for V(D)J recombination. We have constructed substrates for assaying V(D)J recombination in human cells (Fig. 1). These substrates are similar to those that we have previously described for use in the murine system except that the SV40 replicon instead of the polyoma virus replicon is used. The three types of substrates shown permit assay of all of the reaction products described for V(D)J recombination. Signal joint formation and coding joint formation are the two standard products that can be assayed independently with the deletional recombination substrates pSJ and pCJ, respectively. Inversional recombination is measured by using pINV. Hybrid joints, an alternative product observed in murine cells (21), can also be measured with these substrates (Fig. 1). Open/shut joints, another alternative product (21), can most readily be assessed with pCJ1 and pINV1, which have a 1-bp overlap between a restriction site (*Sal*I) and the 12-signal heptamer. We have used these substrates to define the characteristics of the V(D)J recombination reaction in human pre-B-cell lines.

The human V(D)J recombination reaction: an imbalance of signal and coding joint formation. The recombination substrates were tested on a developmental and lineage array of human hematopoietic cell lines. Included in this array were erythroid, myeloid, monocytic, early and mature T-cell, and numerous pre-B- and B-cell lines. Two human pre-B-cell lines, Reh and Nalm-6, were found to have V(D)J recombination activity (Table 1). These two human pre-B-cell lines have TdT activity (30, 34), have rearrangements at their heavy, kappa, and lambda immunoglobulin loci (17), and have early rearrangements at their T-cell receptor alpha/delta loci (11).

Our analysis of V(D)J recombination in these two human pre-B lines indicated an interesting difference from the corresponding analyses of murine early lymphoid cells. In murine lymphoid cells transfected with analogous substrates, we find that the recombination efficiencies of inversion substrates and of deletion substrates that retain the coding joint or the signal joint are similar (23). In contrast,

the human substrates presented here gave a consistent imbalance of signal and coding joint formation, even though they have the same signal and coding end sequences as do the murine substrates. When transfected into Reh cells, the deletion substrate pSJ, which retains the signal joint on the plasmid after recombination, had an average *R* value of 11.5% (Table 1). That is, 11.5% of the plasmids that replicated in the human lymphoid cells had undergone V(D)J recombination (see Materials and Methods). The substrate pCJ, which retains the coding joint after deletional recombination, recombined in Reh cells at an average level of 2.4% (Table 1), approximately fivefold lower than the level of signal joint formation. Analysis of the efficiency of signal and coding joint formation in Nalm-6 cells, the other human pre-B-lymphoid cell line, demonstrated an approximately sixfold deficit in coding joint formation (Table 1). This finding suggests that only one coding joint is being formed for every five to six signal joints.

Inversional recombination in human cells. Inversional recombination, analyzed by using pINV, requires the formation of both a signal joint and a coding joint. If only coding joint formation were limiting the reaction, then inversion should occur at approximately the same efficiency as with pCJ. If signal joint formation were also inefficient, then inversional recombination would occur at efficiencies that are lower than levels for either signal or coding joint formation. We find that in Reh and Nalm-6 cells, the inversion frequency is 21- and 22-fold below the signal joint frequency and 4- and 3.5-fold below the coding joint frequency, respectively (Table 1). Hence, inversion reflects the combined inefficiency of both signal and coding joint formation.

Hybrid joints. When the inversion substrate is transfected into Reh cells, approximately 50% of the recombinant products are the alternative product, called hybrid joints (footnote to Table 1; Fig. 1). Hybrid joints arise when a coding end is joined not to the other coding end but to the other signal end (21). The hybrid formation frequency in Reh cells contrasts markedly with a 5% frequency of hybrid joints observed in Nalm-6 cells (footnote to Table 1). Hence, the relative frequency of hybrid joints to inversions can vary over at least a 10-fold range. This is noteworthy because these two lines are similar in their relative rates of formation of signal and coding joints. The finding that the incidence of hybrid joint formation can be so markedly different raises the issue of whether a component of the recombinase varies the magnitude of the hybrid joint product.

Junctional addition and nucleotide loss. Nucleotide addition can occur at signal joints, coding joints, hybrid joints, and open/shut joints (22). Nucleotide addition at signal joints appears to be the simplest because it correlates well with TdT levels (24) and does not appear to include inverted repeats that would suggest more complex mechanisms of addition (29). In Reh and Nalm-6 cells, nucleotide addition was observed at approximately 50% and at approximately 25%, respectively, of the signal joints. Representative signal joint sequences from recombination events with pINV1 are GC rich (Fig. 2). This is consistent with the known GC richness of signal joint inserts in murine cells and with data indicating that TdT may add G and C nucleotides to DNA ends more often than it adds A or T (2).

Nucleotide addition at hybrid joints is more complex because nucleotides consistent with inverted repeats, in addition to nucleotides that are consistent with TdT addition, have been noted in hybrid joints (21). The inverted repeats have previously been described at full-length coding ends, and it has been proposed that these are due to the

A. Reh x pINV1 Signal Junctions

CCAGTCTGTAGCACTGTG		**CACAGTGGTAGTACTC**
1. CCAGTCTGTAGCACTGTG	GGC	CACAGTGGTAGTACTC
2. CCAGTCTGTAGCACTGTG	GGGAGG	CACAGTGGTAGTACTC
3. CCAGTCTGTAGCACTGTG	GAT	CACAGTGGTAGTACTC
4. CCAGTCTGTAGCACTGTG	GCCCCCTG	CACAGTGGTAGTACTC
5. CCAGTCTGTAGCACTGTG	AGGCCT	CACAGTGGTAGTACTC

B. Reh x pINV1 Hybrid Junctions

AGCTTGGCTGCAGGTCGA		**CACAGTGGTAGTACTC**
6. AGCTTGGCTGCAGGTCG.	GGTAG	.ACAGTGGTAGTACTC
7. AGCTTGGCTGCAGGTCGA	GAAAC	CACAGTGGTAGTACTC
8. AGCTTGGCTGCAGGTCG.	CC	CACAGTGGTAGTACTC
9. AGCTTGGCTGCAGGTCG.	CA	CACAGTGGTAGTACTC
10. AGCTTGGCTGCAGGTCGA	TCC	CACAGTGGTAGTACTC

C. Nalm-6 pINV1 Hybrid Junctions

AGCTTGGCTGCAGGTCGA		**CACAGTGGTAGTACTC**
11. AGCTTGGCTGCA......	CGCGCGGGC	CACAGTGGTAGTACTC
12. AGCTTGGCTGCAGGTCGA	TCTG	CACAGTGGTAGTACTC

FIG. 2. Signal and hybrid junction sequences. pINV1 was transfected into Reh and Nalm-6 cells, and recombinant substrates were recovered and sequenced. The sequences of full-length signal and coding ends, as they are in the substrate, are shown in bold type above the sequences of the recombinant junctions. Each numbered line represents the sequence of a signal or hybrid junction recovered from the cells. Nucleotides lost from the recombinant ends are indicated by dots. Nucleotides added to each junction are shown between the columns.

processing of hairpin intermediates at the coding ends (22, 29, 35). Two of the seven human hybrid joint sequences show a 2-nucleotide inverted repeat (P insert); this is the TC in sequences 10 and 12 of Fig. 2. One would expect this incidence (two of seven inserts) of this particular dinucleotide by chance with only 5% probability. These short inverted repeats may be P nucleotides. The remaining nucleotide inserts are GC rich, as is consistent with TdT additions (19, 24).

Nucleotide addition at coding joints involves complex processing. In human cells, these complexities are particularly marked and have been the subject of a separate study (10).

Nucleotide loss does not usually occur at either signal end in signal joints (24), though it often occurs in hybrid joints (21). The human signal and hybrid joints are similar, in these respects, to their corresponding joints in murine cells (Fig. 2).

The endpoints of nucleotide loss at the coding ends, plotted as a function of position (Fig. 3), are very similar to those determined in a corresponding study in murine cells (24). Both the human and murine data indicate that most coding ends lose from one to five nucleotides.

DISCUSSION

In two human pre-B-cell lines, only one coding joint is detected for every five to six signal joints, even though the overall recombinase activity level is equivalent to that in murine cells (26). This study makes two points: (i) signal and coding joint formation may not be balanced and (ii) efficiencies of joint formation can be well below 100%. This obser-

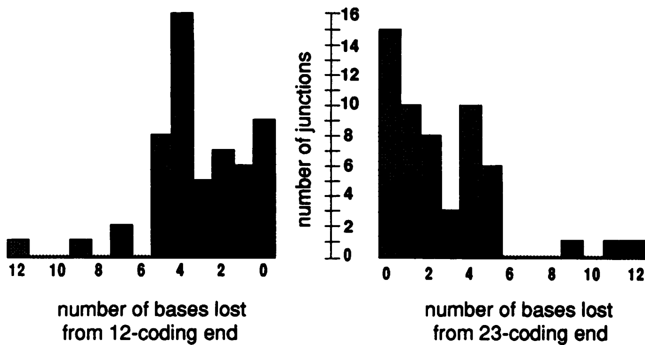

FIG. 3. Distribution of nucleotide loss from the coding ends. pINV1 and pCJ1 were transfected into Reh and Nalm-6 cells, recombinant substrates were recovered, and 56 coding junctions were sequenced. Distribution of nucleotide loss from the coding ends is shown. 12-coding end and 23-coding end refer to the coding ends that are adjacent to the 12 signal and the 23 signal in the substrate.

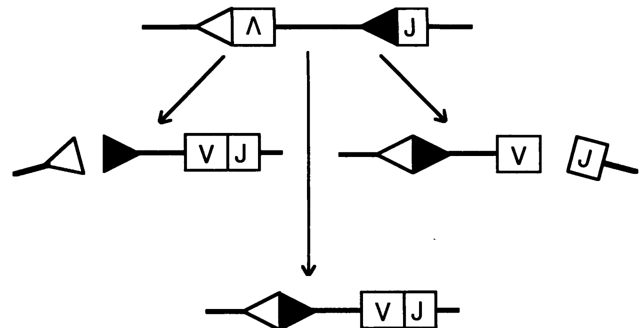

FIG. 4. Inversion reaction generating a VJ junction, indicating that inversional V(D)J recombination may be inefficient because of failure to complete signal and coding joints. The starting substrate is shown at the upper center. V and J coding regions are indicated by boxes; V(D)J recombination signals are represented by open and closed triangles. The completed inversion product is shown at the lower center, where the V and J coding regions are joined together forming a coding joint and the recombination signals are joined in a signal joint. The unequal recombination frequencies of the three substrates, pINV, pCJ, and pSJ, suggest that both coding joint formation and signal joint formation are not 100% efficient. Failure of signal joint formation is shown at the lower left; failure of coding joint formation is shown at the lower right. Neither of these hypothesized failed inversion products is recovered in our cellular assay, hence the observed lower recombination frequency of inversion relative to deletions.

vation agrees with previous indications that the efficiency of signal and coding joint formation is not always 100% (7, 36).

The recombination frequencies of pINV, pCJ, and pSJ can be analyzed to determine the efficiency with which signal and coding joints go to completion once the reaction has been initiated. In Reh cells, the recombination frequencies of the signal joint, coding joint, and inversion substrates are 11.5, 2.4, and 0.55%, respectively (Table 1). Hybrid joints have been eliminated from consideration in this case. The relative ratio of these recombination frequencies is approximately 21:4.4:1. Hence, there appears to be a 21-fold difference between the yield of signal joint events (pSJ) and the yield of inversional events (pINV). Coding joints form at an intermediate efficiency. Because the signal sequences of all three substrates are identical, it seems reasonable to assume that the recombinase binds to each of the three substrates with equal efficiency and that the recombination reaction is initiated on the three substrates with equal efficiency. Under this assumption, one possible explanation for the unequal recombination ratios is that all reactions initiated do not go to completion. Assume, for example, that once the reaction is initiated, the probability of a signal joint going to completion is 0.23 and that the probability of a coding joint going to completion is 0.048. The probability of forming both joints together (as is required for inversion) would be the product of the two independent probabilities: (0.23)(0.048) = 0.011. This is the unique combination of probabilities (0.23:0.048:0.011) that would result in the observed ratio of recombination frequencies (11.5%:2.4%:0.55%) for pSJ, pCJ, and pINV, respectively. Hence, inversions are less frequent than deletions because inversional recombination requires the completion of two inefficient junctions, whereas deletional recombination requires only one (Fig. 4).

Analysis of the recombination frequencies from Nalm-6 cells yields similar conclusions. In Nalm-6 cells, the frequencies are 4.0, 0.63, and 0.18% for pSJ, pCJ, and pINV, respectively (Table 1). Here, the ratio of frequencies is 22:3.5:1, indicating that the yield of signal joints is 22-fold more than the yield of inversions. If the probability of the signal joint going to completion is 0.29 and the probability of the coding joint going to completion is 0.045, then the expected probability of inversion would be (0.29)(0.045) = 0.013. Again, the ratio of probabilities (0.29:0.045:0.013) is similar to the ratio of frequencies (4.0%:0.63%:0.18%).

The inequality of signal and coding joint formation implies that numerous broken ends result from aborted V(D)J recombination. Recently, thymocytes from normal cells have been shown to contain broken chromosomal ends at heptamer and nonamer signal sites (36). These free signal ends seem likely to be the physical result of the inefficiency that we are measuring. In *scid* thymocytes, free unresolved coding ends can also be identified (35). In the *scid* defect, coding end joining occurs in less than one reaction in several thousand (12). Our results indicate that a less marked inefficiency may be operative for the wild-type reaction in coding joint formation.

What does the inequality that we observed here indicate about the V(D)J recombination reaction in vivo? Are only 5% of coding ends resolved to form coding joints? There are several possible interpretations. First, the two halves of the reaction in normal human pre-B and pre-T cells may be intrinsically inefficient and unbalanced. There is reason to believe that the high frequency of chromosomal translocations that would be expected from the kind of inefficiency observed here does indeed occur in human lymphoid cells during V(D)J recombination. Specific translocations into the immunoglobulin locus can be detected in a large percentage of routine human tonsillectomy specimens (1, 27). The picture that emerges is that failed V(D)J recombination reactions may be much more common than previously appreciated. The normal immune system must have sufficient numbers of cells to expend to tolerate an inefficient joining in order to achieve repertoire diversity. It is important to note that measures to counter this intrinsic joining inefficiency may exist at the level of the chromosome. Though these minichromosomes acquire phased nucleosomes (14), they may lack unidentified chromatin features. Nuclear scaffolding, matrix attachment sites, CpG methyla-

tion, and other features of fully chromatinized DNA may contribute to the more efficient resolution of a chromosomal recombination event. Hence, 5% may be a low estimate for the efficiency with which initiated reactions result in coding joints in the chromosome.

A second possibility is that the imbalance observed here is an artifact of human cell lines in culture. This possibility must always be kept in mind in working with cell lines. Such a concern has always been attendant in analysis of murine lymphoid cell line studies of the V(D)J recombination mechanism. Cell lines often show karyotypic abnormalities. Reh cells, which have 45 chromosomes, and Nalm-6 cells, which have 46 chromosomes, both show a deletion on the long arm of chromosome 5 (16, 34). In no case thus far has a mechanistic difference been identified between studies of murine lymphoid cell lines and primary murine lymphoid rearrangement at endogenous loci. In fact, many mechanistic insights initially described in murine lymphoid cell lines were later verified with observations of rearrangement events that occurred within mice (21). The concern about cell lines is impossible to rule out until methods are devised to do reaction efficiency analyses in primary human cells; even then, primary human cells may also be subject to cell culture artifacts.

A third possibility is that intrinsic to the development of human ALL is a variant V(D)J recombinase that is defective in efficiency and product balance. Both of the cell lines used here were derived from patients with ALL. Until we are able to compare the reaction in these cells with primary human lymphoid cells, we can only speculate. We have surveyed a large number of human hematopoietic cell lines, and these two are the only ones thus far that demonstrate human recombinase activity (9).

Among these possibilities, we favor the view that the joining inefficiencies seen in these two human pre-B ALL lines reflect the wild-type reaction. There is physical evidence of inefficient signal joint formation in murine thymocytes (36). Moreover, certain physiologic biases in recombination of the murine endogenous loci can be recapitulated on minichromosomes, and part of the basis for such physiologic biases appears to be inefficiency in joint resolution (7).

ACKNOWLEDGMENTS

We thank M. Gallo, R. Gerstein, K. Sheehan, and A. Islas for critically reading the manuscript.

This work was supported by NIH grant CA51105 and in part by NIH grant GM43236 to M.R.L. M.R.L. is a Lucille P. Markey Scholar, and this work was supported in part by a grant from the Lucille P. Markey Charitable Trust.

REFERENCES

1. **Aster, J. C., Y. Kobayashi, M. Shiota, S. Mori, and J. Sklar.** 1992. Detection of the t(14;18) at similar frequencies in hyperplastic lymphoid tissues from American and Japanese patients. Am. J. Pathol. 141:291–299.
2. **Basu, M., M. Hedge, and M. Modak.** 1983. Synthesis of compositionally unique DNA by terminal deoxynucleotidyl transferase. Biochem. Biophys. Res. Commun. 111:1105–1112.
3. **Blackwell, T. K., and F. W. Alt.** 1988. Immunoglobulin genes, p. 1–6. *In* B. D. Hames and D. M. Glover (ed.), Molecular immunology. IRL Press, Oxford.
4. **Blackwell, T. K., B. A. Malynn, R. R. Pollock, P. Ferrier, L. R. Covey, G. M. Fulop, R. A. Phillips, G. D. Yancopoulos, and F. W. Alt.** 1989. Isolation of *scid* pre-B cells that rearrange kappa light chain genes: formation of normal signal and abnormal coding joins. EMBO J. 8:735–742.
5. **Bosma, M. J., and A. M. Carroll.** 1991. The scid mouse mutant: definition, characterization, and potential uses. Annu. Rev. Immunol. 9:323–350.
6. **Desiderio, S. V., G. D. Yancopoulos, M. Paskind, E. Thomas, M. A. Boss, N. Landau, F. W. Alt, and D. Baltimore.** 1984. Insertion of N regions into heavy-chain genes is correlated with expression of terminal deoxytransferase in B cells. Nature (London) 331:627–631.
7. **Gauss, G. H., and M. R. Lieber.** 1992. The basis for the mechanistic bias for deletional over inversional V(D)J recombination. Genes Dev. 6:1553–1561.
8. **Gauss, G. H., and M. R. Lieber.** 1992. DEAE-dextran enhances electroporation of mammalian cells. Nucleic Acids Res. 20: 6739–6740.
9. **Gauss, G. H., and M. R. Lieber.** Unpublished data.
10. **Gauss, G. H., and M. R. Lieber.** Submitted for publication.
11. **Hansen-Hagge, T. E., S. Yokota, H. J. Reuter, K. Schwartz, and C. R. Bartram.** 1992. Human common acute lymphoblastic leukemia-derived cell lines are competent to recombine their T-cell receptor δ/α regions along a hierarchically ordered pathway. Blood 80:2353–2362.
12. **Harrington, J., C.-L. Hsieh, J. Gerton, G. Bosma, and M. R. Lieber.** 1992. Analysis of the defect in DNA end joining in the murine *scid* mutation. Mol. Cell. Biol. 12:4758–4768.
13. **Hesse, J. E., M. R. Lieber, M. Gellert, and K. Mizuuchi.** 1987. Extrachromosomal substrates undergo inversion or deletion at immunoglobulin VDJ joining signals. Cell 49:775–783.
14. **Hsieh, C.-L., R. P. McCloskey, and M. R. Lieber.** 1992. V(D)J recombination on minichromosomes is not affected by transcription. J. Biol. Chem. 267:15613–15619.
15. **Hsieh, C.-L., R. P. McCloskey, E. Radany, and M. R. Lieber.** 1991. V(D)J recombination: evidence that a replicative mechanism is not required. Mol. Cell. Biol. 11:3972–3977.
16. **Hurwitz, R., J. Hozier, T. LeBein, J. Minowada, K. Gajl-Peczalska, I. Kubonishi, and J. Kersey.** 1979. Characterization of a leukemic cell line of the pre-B phenotype. Int. J. Cancer 23:174–180.
17. **Korsmeyer, S., A. Arnold, A. Bakshi, J. Ravetch, U. Siebenlist, P. Hieter, S. Sharrow, T. LeBien, J. Kersey, D. Poplack, P. Leder, and T. Waldmann.** 1983. Immunoglobulin gene rearrangement and cell surface antigen expression in acute lymphocytic leukemias of T cell and B cell precursor origins. J. Clin. Invest. 71:301–313.
18. **Lafaille, J. J., A. DeCloux, M. Bonneville, Y. Takagaki, and S. Tonegawa.** 1990. Junctional sequences of T cell receptor γ/δ genes: implications for γδ T cell lineages and for novel intermediate of V-(D)-J joining. Cell 59:859–870.
19. **Landau, N. R., D. G. Schatz, M. Rosa, and D. Baltimore.** 1987. Increased frequency of N-region insertion in a murine pre-B-cell line infected with a terminal deoxynucleotidyl transferase retroviral expression vector. Mol. Cell. Biol. 7:3237–3243.
20. **Lewis, S. M., and J. E. Hesse.** 1991. Cutting and closing without recombination in V(D)J joining. EMBO J. 10:3631–3639.
21. **Lewis, S. M., J. E. Hesse, K. Mizuuchi, and M. Gellert.** 1988. Novel strand exchanges in V(D)J recombination. Cell 55:1099–1107.
22. **Lieber, M. R.** 1991. Site-specific recombination in the immune system. FASEB J. 5(14):2934–2944.
23. **Lieber, M. R.** 1993. The role of site-directed recombinases in physiologic and pathologic chromosomal rearrangements, p. 239–275. *In* I. Kirsch (ed.), The causes and consequences of chromosomal translocations. CRC Press, Boca Raton, Fla.
24. **Lieber, M. R., J. Hesse, K. Mizuuchi, and M. Gellert.** 1988. Lymphoid V(D)J recombination: nucleotide insertion at signal joints as well as coding joints. Proc. Natl. Acad. Sci. USA 85:8588–8592.
25. **Lieber, M. R., J. E. Hesse, S. Lewis, G. Bosma, N. Rosenberg, K. Mizuuchi, M. Bosma, and M. Gellert.** 1988. The defect in murine severe combined immune deficiency: joining of signal sequences but not coding segments in V(D)J recombination. Cell 55:7–16.
26. **Lieber, M. R., J. E. Hesse, K. Mizuuchi, and M. Gellert.** 1987. Developmental stage specificity of the lymphoid V(D)J recombination activity. Genes Dev. 1:751–761.
27. **Limpens, J., D. de Jong, J. van Krieken, C. Price, B. Young, G.**

van Ommen, and P. Kluin. 1991. Bcl-2/JH rearrangements in benign lymphoid tissues with follicular hyperplasia. Oncogene **6:**2271–2276.

28. **McCormick, W. T., L. W. Tjoelker, L. M. Carlson, B. Petryniak, C. F. Barth, E. H. Humphries, and C. B. Thompson.** 1988. Chicken IgL gene rearrangement involves deletion of circular episome and addition of single nonrandom nucleotides to both coding segments. Cell **56:**785–791.
29. **Meier, J., and S. Lewis.** 1993. P nucleotides in V(D)J recombination: a fine-structure analysis. Mol. Cell. Biol. **13:**1078–1092.
30. **Minowada, J., G. Janossy, M. F. Greaves, T. Tsubota, B. I. S. Srivastava, S. Morikawa, and E. Tatsumi.** 1978. Expression of an antigen associated with acute lymphoblastic leukemia in human leukemia-lymphoma cell lines. J. Natl. Cancer Inst. **60:**1269–1277.
31. **Mortari, F., J. A. Newton, J. Y. Wang, and H. W. Schroeder.** 1992. The human cord blood antibody repertoire. Frequent usage of the V_H7 gene family. Eur. J. Immunol. **22:**241–245.
32. **Okazaki, K., D. D. Davis, and H. Sakano.** 1987. T cell receptor β gene sequences in the circular DNA of thymocyte nuclei: direct evidence for intramolecular DNA deletion in V-D-J joining. Cell **49:**477–485.
33. **Raaphorst, F. M., E. Timmers, M. J. H. Kenter, M. J. D. Van Tol, J. M. Vossen, and R. K. B. Schuurman.** 1992. Restricted utilization of germ-line V_H3 genes and short diverse third complementarity-determining regions (CDR3) in human fetal B lymphocyte immunoglobulin heavy chain rearrangements. Eur. J. Immunol. **22:**247–251.
34. **Rosenfeld, C., A. Goutner, C. Choquet, A. M. Venuat, B. Kayibanda, J. L. Pico, and M. F. Greaves.** 1977. Phenotypic characterization of a unique non-T, non-B acute lymphoblastic leukaemia cell line. Nature (London) **267:**841–843.
35. **Roth, D. B., J. P. Menetski, P. B. Nakajima, M. J. Bosma, and M. Gellert.** 1992. V(D)J recombination: broken DNA molecules with covalently sealed (hairpin) coding ends in SCID mouse thymocytes. Cell **70:**983–991.
36. **Roth, D. B., P. Nakajima, J. Menetski, M. Bosma, and M. Gellert.** 1992. V(D)J recombination in mouse thymocytes: double strand breaks near T cell receptor δ rearrangement signals. Cell **69:**41–53.
37. **Schatz, D. B., M. A. Oettinger, and M. S. Schlissel.** 1992. V(D)J recombination: molecular biology and regulation. Annu. Rev. Immunol. **10:**359–383.
38. **Schroeder, H. W., and J. Y. Wang.** 1990. Preferential utilization of conserved immunoglobulin heavy chain variable gene segments during human fetal life. Proc. Natl. Acad. Sci. USA **87:**6146–6150.
39. **Schuler, W., N. R. Ruetsch, M. Amsler, and M. J. Bosma.** 1991. Coding joint formation of endogenous T cell receptor genes in lymphoid cells from SCID mice: unusual P-nucleotide additions in VJ-coding joints. Eur. J. Immunol. **21:**589–596.
40. **Schwarz, K., T. E. Hansen-Hagge, C. Knobloch, W. Friedrich, E. Kleihauer, and C. R. Bartram.** 1991. Severe combined immunodeficiency (SCID) in man: B cell-negative (B^-) SCID patients exhibit an irregular recombination pattern at the J_H locus. J. Exp. Med. **174:**1039–1048.
41. **Sheehan, K. M., and M. R. Lieber.** 1993. V(D)J recombination: signal and coding joint resolution are [sic] uncoupled and depend on parallel synapsis of the sites. Mol. Cell. Biol. **13:**1363–1370.
42. **Soberon, X., L. Covarrubias, and F. Bolivar.** 1980. Construction and characterization of new cloning vehicles. Gene **8:**287–305.
43. **Tonegawa, S.** 1983. Somatic generation of antibody diversity. Nature (London) **302:**575–581.
44. **Yamada, M., R. Wasserman, B. A. Reichard, S. Shane, A. J. Caton, and G. Rovera.** 1991. Preferential utilization of specific immunoglobulin heavy chain diversity and joining segments in adult human peripheral blood B lymphocytes. J. Exp. Med. **173:**395–407.