MOLECULAR AND CELLULAR BIOLOGY, June 1997, p. 3125–3136 Vol. 17, No. 6
0270-7306/97/$04.00+0


# Cryptic Signals and the Fidelity of V(D)J Joining

SUSANNA M. LEWIS,[1,2,3]* EMILY AGARD,[2] SUSIE SUH,[1] AND LINDA CZYZYK[1]

*Division of Biology, California Institute of Technology, Pasadena, California 91125[1]; Immunology Department, University of Toronto, Toronto, Ontario, Canada M5S 1A8[2]; and Division of Immunology and Cancer, Hospital for Sick Children Research Institute, Toronto, Ontario, Canada M5G 1X8[3]*

Received 12 September 1996/Returned for modification 7 November 1996/Accepted 1 March 1997

**V(D)J recombination is responsible for the de novo creation of antigen receptor genes in T- and B-cell precursors. To the extent that lymphopoiesis takes place throughout an animal's lifetime, recombination errors present an ongoing problem. One type of aberrant rearrangement ensues when DNA sequences resembling a V(D)J joining signal are targeted by mistake. This study investigates the type of sequence likely to be subject to mistargeting, the level of joining-signal function associated with these sequences, and the number of such cryptic joining signals in the genome.**

In a vertebrate, many millions of B and T cells are produced daily. A pivotal event in B- and T-cell differentiation is the assembly, through DNA recombination, of antigen receptor genes (reviewed in reference 43). Recombination is mediated by two key proteins, called RAG1 and RAG2, as well as an incompletely defined collection of other enzymes and factors (reviewed in references 28 and 56). During pre-B-cell and pre-T-cell development, DNA segments termed V, D, and J are assembled to form the variable exon of the expressed immunoglobulin (Ig) and T-cell receptor (TCR) genes. The joined gene segments encode the highly variable antigen-binding domain of the antigen receptor protein.

V(D)J assembly is accomplished through DNA rearrangements that can encompass up to several megabases. The sequence motifs that serve as recognition elements for this process (joining signals) are surprisingly compact. A joining signal, found adjacent to every V, D, or J segment, is comprised of a heptamer (CACAGTG), a spacer (of 12 or 23 bp), and a nonamer (ACAAAAACC) (Fig. 1A). No sequence apart from the 28 (or 39)-bp joining signal motif is required for site-specific recognition and recombination (reference 33 and references cited therein).

Any V(D)J recombination event is normally an interaction between two gene segments, one of which possesses a 12-bp spacer (12-spacer) signal, and the other of which possesses a 23-bp spacer (23-spacer) signal. The molecular signature of V(D)J recombination is production of two types of junctions. One of these, the coding joint, is formed from the fusion of the two coding sequences (e.g., V and J), and the other, a signal joint, is formed from the corresponding joining signals. It is usual for a coding joint to have a small and variable degree of base loss and insertion at the point where the two coding elements connect, whereas the signal joint is typified by a precise fusion, heptamer to heptamer, of the two signals.

Not surprisingly, DNA sequences that resemble an authentic joining signal (cryptic signals) are sometimes rearranged by mistake (first described in reference 36). The result is the production of recombinant junctions that are structurally analogous to coding joints and signal joints (Fig. 1) (17, 36). In this report, we take production of the characteristic junctions as the defining, diagnostic feature of a cryptic signal.

Cryptic signals are seen at several of the Ig and TCR loci in both mice and humans and are active in recombination with authentic V, D, or J joining signals. Most of these sequences make no positive contribution to antigen receptor gene assembly because they are not associated with any V, D, or J coding sequence. An earlier suggestion that certain cryptic signals played a role in T- or B-cell development, through the recombinational silencing of an Ig or TCR locus, is now considered unlikely (20, 53). Although recombination of a cryptic signal located within the coding sequence of many Vh genes may play a role in receptor editing (for a recent discussion, see reference 14), it is probable that most observed cryptic rearrangements are simply a by-product of the activation of V(D)J recombination enzymes. In the context of the early T- and B-cell developmental programs, strong selections for a functional antigen receptor (for reviews, see references 51 and 74) mean that the consequences of an occasional cryptic signal rearrangement at an Ig or TCR locus ought to be negligible.

At the same time, selective gates that sample the state of the antigen receptor would not be expected to remove those pre-B and pre-T cells that have acquired cryptic rearrangements at non-Ig/TCR loci. The fact that cryptic rearrangement can occur completely apart from the Ig and TCR loci has been well established. Site-specific, recurrent rearrangements at two different locations are seen in human T cells (2, 11, 26, 27). Both cases involve deletion and take place between a pair of relatively closely linked cryptic sequences. In one instance, a 90-kb deletion between the *sil* and *scl* genes on chromosome 1 (2, 11) was detected in patients with T-cell acute lymphocytic leukemia (T-ALL). In the other, specific deletions in the hypoxanthine-guanine phosphoribosyltransferase (*hprt*) gene (on the X chromosome) were found in peripheral blood T cells of healthy individuals (26, 27). In each example, the deletion junctions exhibited all hallmarks of V(D)J coding joints (in particular, some joints bore a very diagnostic type of sequence termed a P insertion, and multiple independent deletions with nearly identical endpoints could be documented [see examples in references 2, 7, 10, 11, 26, 27, and 37]). The deletion endpoints were consistently found adjacent to DNA sequences bearing similarity to a V(D)J joining signal. These data have provided the most conclusive evidence to date for V(D)J targeting errors involving sequences outside the Ig and TCR loci.

To more fully understand both the molecular and biological imperatives associated with the problem of fidelity in V(D)J

* Corresponding author. Mailing address: Immunology Department, University of Toronto, 1 King's College Circle, Toronto, Ontario, Canada M5S 1A8. Phone: (416) 978-4230. Fax: (416) 978-1938. E-mail: Lewis@immune.med.utoronto.ca.

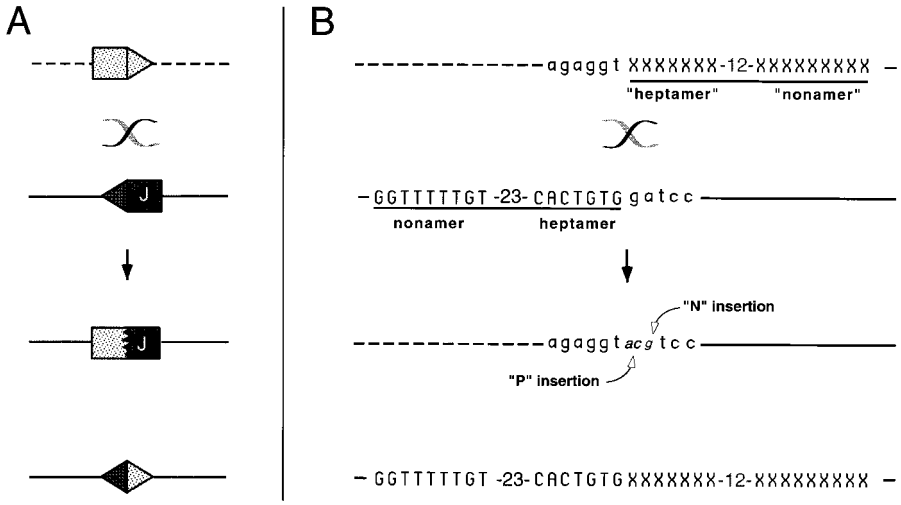



FIG. 1. Functional definition of a cryptic signal. (A) Cartoon of a cryptic rearrangement. A cryptic joining signal (dotted triangle) acts as a 12-spacer signal in a recombination with a canonical 23-spacer joining signal (shaded triangle shown attached to the J segment). The sequence flanking the cryptic signal is shown as a box for convenience but does not represent any specific element. Two products are formed: a cryptic coding joint (dotted and shaded boxes) and a cryptic signal joint (dotted and shaded triangles). Recombination between a pair of cryptic sequences is also possible but is not illustrated. (B) Sequence characteristics of recombinant junctions. For illustration, an arbitrary sequence is shown in lowercase letters flanking an unspecified cryptic signal (uppercase X's); likewise, an arbitrary sequence in lowercase represents the segment adjacent to the canonical signal (uppercase). The cryptic coding joint is variant and exhibits an irregular loss and addition of the sequences formerly flanking the two signals. One hypothetical junction is shown, with P and N insertions as indicated. P insertions occur adjacent to untrimmed coding flanks only and, as illustrated, comprise an inverted repeat of the terminal one to several base pairs of sequence. A precise heptamer-to-heptamer fusion of the cryptic signal to the canonical signal forms the cryptic signal joint.

recombination, we have attempted to learn more about the basic parameters governing cryptic signal rearrangement. We have arrived at a minimum estimate for the number of cryptic signals in the genome and have measured the relative joining function intrinsic to several such sequences. Our results indicate that cryptic signals are extremely numerous and exhibit significant V(D)J target potential and thus must present a profound challenge to joining fidelity.

## MATERIALS AND METHODS

**Cells.** The Abelson murine leukemia virus transformant 204-1-8 (49) was used in all experiments as previously described (42, 45, 50).

**Transfections.** Transfections were carried out as detailed by Lieber et al. (47). Briefly, $3 \times 10^6$ cells were transfected with 150 ng of DNA in a protocol including DEAE-dextran and osmotic shock. Plasmid DNAs were recovered from the transfected cells at about 48 h by an alkaline lysis procedure and introduced into *Escherichia coli* (33).

**Identification of spontaneously arising cryptic recombinants.** Some cryptic recombinants were deliberately isolated during surveys for rare V(D)J recombination products (42, 45, 50). A general method is summarized below. In other investigations, cryptic recombinants were discovered by chance among chloramphenicol-resistant colonies. In still other studies, chloramphenicol (rather than ampicillin)-resistant colonies were screened as described below (reference 50 and unpublished data).

In outline, cryptic signals were identified in two steps: first, putative cryptic recombinants were isolated; and second, the DNA sequence of the junction was inspected. Attributes described below determined whether a particular rearrangement was related to V(D)J joining and established the identity of the involved cryptic signal.

Many cryptic recombinants were isolated by using extrachromosomal substrates in which the last base pair of the joining signal constituted the first base pair of an adjacent, unique restriction site (Fig. 2). To isolate 23-spacer signal-specific rearrangements in a way that would include rare classes of site-specific recombinant (as illustrated in Fig. 2), plasmid DNA was treated posttransfection with *Bam*HI. The digested DNA was used to transform *E. coli*, and filter lifts of ampicillin-resistant colonies were hybridized to an oligonucleotide spanning the signal/flank border of an unrearranged 23-spacer signal (Fig. 2) (42, 45). All hybridizing colonies were disregarded, and the nonhybridizing colonies (having a deleted or interrupted region corresponding to the specific oligonucleotide probe) were picked and regrown in replicate grid arrays. These isolates were crudely mapped by testing their hybridization to the panel of oligonucleotides diagrammed in Fig. 2. Candidate cryptic recombinants were any that did not exhibit the expected hybridization pattern for a standard, hybrid, or open-and-shut joint (Fig. 2).

Possible cryptic recombinants were further analyzed by treating miniprep DNA with diagnostic restriction enzymes and categorized according to whether they contained a signal joint (as in either a deletional signal joint or inversional rearrangement) or only a coding joint.

**(i) Signal joint-containing isolates (with inversion or deletion).** Either *Apa*L1 or *Hgi*A1 can cleave the sequence GTGCAC, which is created upon the formation of most signal joints. Two distinct bands are observed for a recombinant containing a joint between the authentic 12-spacer signal and the authentic 23-spacer signal. The presence of one of the two bands, in the absence of the other, was taken to indicate an isolate containing a cryptic signal joint. If a variant pattern had not previously been noted, isolates were subjected to DNA sequence analysis to verify that a novel cryptic element had become site-specifically attached to the heptamer of an authentic joining signal (as shown in Fig. 1B). For inversional recombinants, the DNA sequence of at least one example of a coding joint was determined in addition to the signal joint.

**(ii) Coding-joint-containing isolates.** The *Apa*L1 or *Hgi*A1 banding pattern is not informative if a coding joint has been formed by deletion (or if a signal joint has been formed from a cryptic signal lacking CAC at its heptamer border). For isolates whose recurrent restriction patterns lacked the *Apa*L1 (or *Hgi*A1) bands diagnostic of a signal joint, DNA sequence analysis was performed. Junctions that met the following criteria were considered to be cryptic coding joints: (i) inspection of the nonrecombinant precursor indicated that a properly oriented CAC sequence (deleted in the recombinant) lay, at most, several base pairs from the recombination site; (ii) for the canonical partner signal, the crossover site was positioned normally, that is, at most a few base pairs displaced away from the heptamer edge to the side opposite from the nonamer; and (iii) at least two different coding joints involving the same putative cryptic sequence were isolated. In one case, a single joint fulfilling the first two criteria was accepted because it contained a 2-bp P nucleotide insertion (see the legend to Fig. 1 for definition; reviewed in reference 44). P insertions of greater than one nucleotide are sufficiently rare (50) to be considered diagnostic.

Any signal joint recombinant in which the cryptic signal lacked CAC at its heptamer edge would have been among the isolates subjected to DNA sequence analysis. No examples were found.

**(iii) Parental plasmids and specific protocols.** The approximately 8-kb plasmids from which the cryptic signals were isolated all had a 7,637-bp region in common. The plasmids varied in the following features: the presence or absence of the *oop* transcription terminator of bacteriophage λ, the presence of an overlapping restriction site at one or both joining signals, the presence or identity of a nonoverlapping restriction site flanking one or both joining signals, the presence, identity, or orientation of the 12-spacer signal, and the identity of 2 bp at positions 6130 and 6131 (34). All of the differences between plasmids were

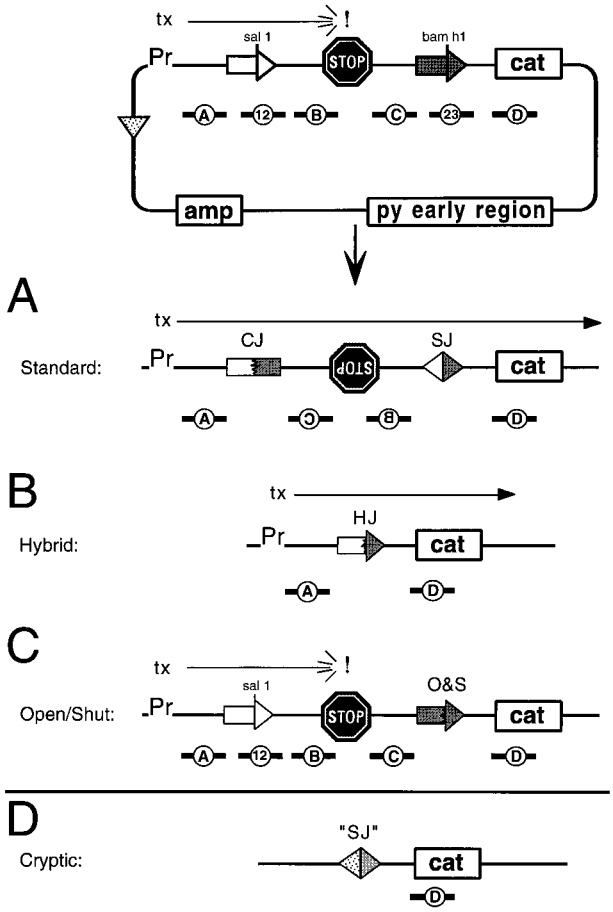



FIG. 2. Detection of cryptic signals within extrachromosomal V(D)J joining substrates. At the top is a diagram of a representative extrachromosomal V(D)J joining substrate, p12×23 (45). An example of a cryptic signal embedded in the vector backbone is represented as a dot-filled triangle. The DNA segments adjacent to the canonical 12- and 23-spacer joining signals are shown as boxes but do not represent any specific element. Unique restriction sites for *Sal*I and *Bam*HI overlap the joining signals by one nucleotide. Bars labeled A to D, 12, and 23 represent oligonucleotides used as hybridization probes (not to scale). Pr, stop, cat, and amp are, respectively, the bacterial promoter, a transcription terminator, the chloramphenicol acetyltransferase coding sequence, and a complete, constitutively expressed β-lactamase gene (33). The eukaryotic polyomavirus (py) early region is also indicated. Panels A to D show different V(D)J joining products along with their hybridization properties. Outcomes A to C all arise when canonical joining signals are targeted. The standard V(D)J recombination products (outcome A) are a coding joint (CJ) and a signal joint (SJ). Alternative products are a hybrid joint (HJ [B]) or an open-and-shut joint (O&S [C]). Panel D gives an example of the outcome when a cryptic signal is targeted and shows a cryptic signal joint (analogous to the signal joint shown in panel A). Cryptic recombinants are distinguished by their anomalous hybridization properties, which, though varied, are inconsistent with those of the recombinants shown in panels A to C; for details, see Materials and Methods.

confined to an approximately 415-bp region; the remaining 7,637 kb of the recombination substrates were identical.

Specific cryptic recombinants used to establish the DNA sequences of individual cryptic signals were derived as detailed below; however, most cryptic signals were recovered in more than one experiment: 7451, 6131 (p12×23 [45]); 1632, 3762, 6912, and 7451 (hybrid joint) (pMut2 [50]); 5145 (pMut27 [50]); 1318, 2213, and 5865 (pMut7 [41a], which is identical to pJH288 except for the 12-spacer signal, which is identical to that of pMut2; cryptic signals were isolated as described for pMut2); 77 and 6069 (pD16 [45, 68a]); 1947 (pSILSJD [this report]); 1318 and 5145 (p6131inv [identical to p6130SJΔ save for the orientation of the 6131 sequence in site 1] [unpublished]); and 2213 (p6130SJΔ [this report; see below]). The signals 6131 and 1318 were originally identified by Joanne Hesse (National Institutes of Health). It was found upon DNA sequence analysis that cryptic sites 2213, 5145, and 5865 were rearranged in recombinants independently isolated by Joanne Hesse and kindly provided by her.

**Oligonucleotide probes.** The oligonucleotides indicated in Fig. 2 were as follows: A is LAC-1, B is TER-1, C is TER-2, D is JH33 (42, 45), and 12 and 23 are oligonucleotides that hybridize to 15 nucleotides spanning the joining-signal borders as shown (45). An additional probe sometimes used in mapping was 7453 (CGTTGTGTAGGTACC [not shown]).

Other oligonucleotides, indicated in Fig. 6, were the X oligonucleotides specific for each construct plus E and F. E (6130SIG in reference 45) spanned the signal/flank border of the 6131 signal as situated in site 2. The F oligonucleotide hybridized to a sequence in the site 1-to-site 2 interval and had the sequence TCCTAACAGCTATGACCATG. The construct-specific X oligonucleotides were O/6131SJΔ (ATGTTGTGTCGACTC), O/WTSJΔ (CCACTGTGTCGACTC), O/6912SJΔ (CTGGAGTGTCGACTC), O/SILSJΔ (GCAGAGTGTCGACTC), and O/HPRSJΔ (TGTGTGTGTCGACTC).

**Standardized recombination substrates and assay.** Extrachromosomal V(D)J recombination substrates for the standardized joining assay (Fig. 6 and 7) were constructed by ligation of a double-stranded oligonucleotide cassette bearing *Sal*I-compatible termini to *Sal*I-digested p23-only DNA (45). For each sequence listed below, G on the left and TCGAC on the right of the insert constitute the vector-derived, interrupted *Sal* site. For clarity, nonamers and heptamers are underlined. The recombination substrates were as follows: pWTSJΔ (canonical; GTCGA GGTTTTTGTTCCAGTCTGTACCACTGTGTCGAC); p6131SJΔ (6131; GTCG AACTTTATGCTTCCGGCTCGTATGTTGTGTCGAC); p6912SJΔ (6912; GTC GAACTGAAACGTTTTCATCGCTCTGGAGTGTCGAC); p5145SJΔ (5145; GTCGATCTTTTCTACGGGGTCTGACGCTCAGTGTCGAC); pHPRSJD (Hpr-3a; GTCGATGTATTTGTGTGTGTGTGTGTGTGTGTGTGTCGAC); pSILSJΔ (Sil-1; GTCGAAACTGATCCCTTACTACGTAGCAGAGTGTCGA C); and pCASJΔ (CA repeat; GTCGATGTGTGTGTGTGTGTGTGTGTGTGTG TGTGTCGAC). For each construction, the insert was confirmed by DNA sequence analysis.

Recombination substrates were transfected into the cell line 204-1-8 as described above except that caffeine was routinely included in the medium posttransfection. As shown by others (52), 1 mM caffeine (ICN Biomedicals, Inc.) increased the recovery of recombinants in the extrachromosomal assay. Caffeine did not affect the ratios between various classes of recombinants (see the appendix).

In the standardized recombination assay, at least 100 and, as necessary, more than 600 chloramphenicol-resistant colonies were picked to replicate grids. The ratio of site 1 recombinants to site 2 recombinants was then determined. Site 1 recombinants were positive for probes A, E, F, and D; site 2 recombinants were positive for A and D only (42, 45). Routinely, probes A, E, F, and C were used to determine site 1 and site 2 recombinants. In early experiments, probe D and an oligonucleotide specific for the unrecombined signal (X) of the given construct were used as well but proved redundant (chloramphenicol-resistant colonies were always positive for probe D, and use of the other probes in combination never gave results that were discrepant with X). When randomly selected colonies designated site 1 and site 2 recombinants according to hybridization analyses were subjected to *Hgi*A1 or *Apa*L1 DNA digestion, all proved to be of the predicted structure. Additionally, DNA preparations from any candidate colony that could not be unambiguously assigned to the site 1 or site 2 category on the basis of hybridization analyses were subjected to diagnostic restriction analysis. For each construct, the DNA sequence of at least one site 1 recombinant was determined.

## RESULTS

**Cryptic signals within extrachromosomal V(D)J recombination substrates.** To assess the frequency with which potential V(D)J joining targets arise by chance within a defined DNA, we counted cryptic joining-signal sequences in an extrachromosomal V(D)J recombination substrate. A representative construct as originally developed by Hesse et al. (33, 34) and Lieber et al. (47) is shown in Fig. 2. The extrachromosomal plasmid and related derivatives contain joining signals positioned so that site-specific rearrangement will disable a prokaryotic transcription terminator through deletion or inversion (Fig. 2). This DNA is transfected into immortalized mouse pre-B cells that are active for V(D)J recombination. After 2 days, the plasmid DNA is reisolated and is introduced into *E. coli*. A chloramphenicol selection allows recovery of the recombinant molecules in the population (33, 34).

Unintended rearrangements are also recovered with this assay (34). We found that cryptic recombinants were particularly evident in studies where the assay was adapted to retrieve low-level hybrid and open-and-shut junctions (Fig. 2) (42, 45). We took advantage of this circumstance to collect all examples where a joining signal had recombined with a cryptic sequence

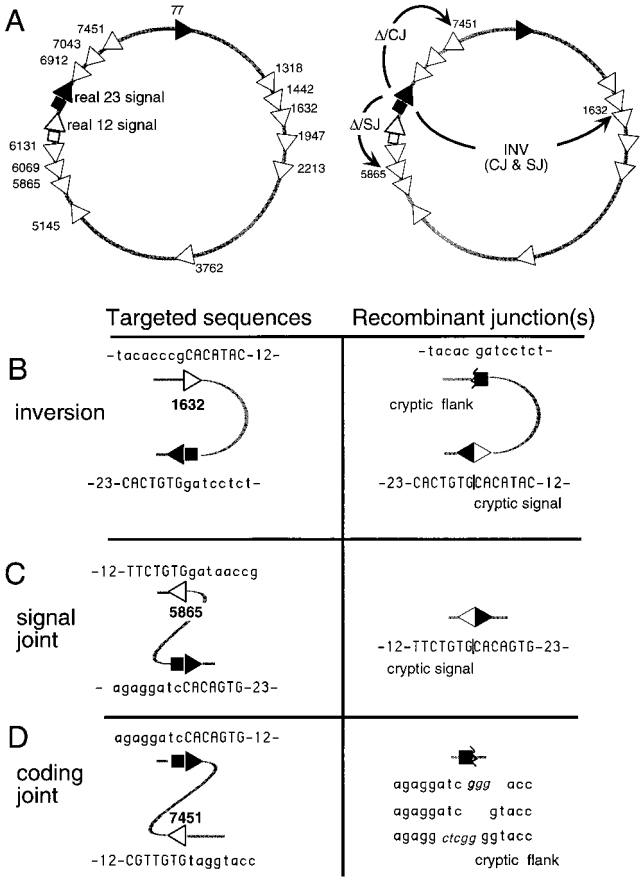



FIG. 3. Distribution and orientation of cryptic signals. (A) Cryptic signals are named according to their locations on plasmid pJH288 (46) (roughly to scale). Sequences that functioned as 12-bp joining signals are shown as white triangles; those that functioned as 23-bp signals are shaded. The canonical 12- and 23-bp signals are located at 6210 and 6496. To the right in panel A, examples of the three categories of cryptic rearrangement are indicated. Δ/SJ, deletion with retained signal joint; Δ/CJ, deletion with retained coding joint; INV, inversion. In panels B to D, examples of the data used to establish the identity of cryptic signals are shown; e.g., the coding joints for signal 7451 are shown in panel D. Junction sequences, with insertions, are indicated as in Fig. 1.

over several years' time. Details are given in Materials and Methods.

Where a recombinant plasmid appeared to have been rearranged by V(D)J mistargeting events, closer inspection determined whether defined structural and functional criteria were met. As shown in Fig. 3, if an authentic joining signal recombines with a cryptic sequence, there are three possible outcomes. The resulting plasmid may contain either a signal joint, a coding joint, or both. If (as shown in Fig. 3B and C) rearrangement results in retention of a signal joint, it is a fairly simple matter to determine the DNA sequence of the novel junction and by this means to directly identify the cryptic signal (Fig. 1, 3B, and 3C). However, in cases where a coding joint, having been formed by deletion, is the only plasmid-retained V(D)J joining product (Fig. 3D), the identity of the cryptic signal must be deduced indirectly from the recombinants. In such cases, it is crucial to distinguish putative cryptic coding joints from illegitimate rearrangements arising through unrelated mechanisms.

To maximize confidence in coding joint data, it is almost always necessary to have multiple, slightly variant examples available for analysis (Fig. 3D). In a cryptic coding joint (Fig. 1B, 3B, and 3D), two coding ends, one terminating within several base pairs (and to the proper side) of an authentic joining signal and the other likewise terminating near a candidate cryptic signal, are connected. Variable truncation should be evident among cryptic coding joints arising from misrecognition of a given cryptic sequence. Junctional insertions should have the characteristics of either an N or a P region (reviewed in references 43 and 44). N insertions are short, GC-rich, but otherwise random additions; P insertions are inverted repeats of one to several base pairs of the terminal residues of a nontruncated coding end (a hypothetical example is shown in Fig. 1). In practice, we have found that putative coding joints generally satisfy either all of the above criteria or none of them. For example, short (e.g., 4- to 10-bp) direct repeats are not characteristic of V(D)J coding joints (62), and such inserts were found within junctions that occurred no more than once and had crossover sites inconsistent with a V(D)J joining product (not shown).

Altogether, 14 cryptic signals were identified (Fig. 3 and 4). Eleven of these were established on the basis of signal joint data, and three were established on the basis of coding-joint formation. Three other coding-joint-type recombinants, though otherwise acceptable, were each isolated only once, and so the corresponding cryptic signals have not been listed. The pUC-based polyomavirus shuttle vector plasmid was composed of eukaryotic, prokaryotic, and completely synthetic DNA, with an overall GC content of 48%. There are no repetitive stretches or unusual structural features that would make this DNA obviously unsuited as a generic sequence.

Thus, in this representative DNA of roughly 8 kb, the occurrence of cryptic signals is, on average, at least one per 600 bp. The number of cryptic signals that we were able to find is doubtless less than the actual number present, because the constraints of the assay system are such that, for detection, a recombinant molecule must be minimally able to confer ampicillin resistance to *E. coli* and to replicate in both *E. coli* and mouse cells.

**Features of cryptic signals derived from the extrachromosomal substrate.** To evaluate the sequence features of the cryptic signals, they were aligned according to a spacer length of exactly 12 or 23 bp. Such alignment showed that each cryptic signal varied substantially from the canonical joining signal sequence. Typically, only about 9 of the 16 canonical heptamer and nonamer nucleotides were present in a cryptic sequence (Fig. 4A, ALL). Identities were inconsistent from one sequence to the next. The only exception to this generality was that every cryptic signal had CAC at positions 1 to 3 of the heptamer. (As described in Materials and Methods, cryptic signals lacking $CAC_{1-3}$ were less likely to be detected, but were not systematically excluded).

Remarkably, despite the individual variation, a consensus DNA sequence derived from the aligned sequences very nearly reconstructed the canonical V(D)J joining signal. The consensus and canonical sequences are identical (with two ambiguities) at 13 of 16 positions (Fig. 4B). For comparison, if one were to assume an incorrect spacer length for each cryptic signal (i.e., 23 instead of 12 bp and 12 instead of 23 bp), the new consensus nonamer has no resemblance to the canonical sequence (Fig. 4C).

These observations are considered further, with respect to the targeting problem in V(D)J joining, in Discussion.

**Cryptic signals in the genome.** Using the criteria that were applied to the plasmid-derived cryptic signals, we compiled a list of naturally occurring cryptic signals from the human and mouse genomes (Fig. 5). That is, endogenously derived cryptic signals were functionally defined, as documented by the for-

**A. Cryptic signals in the plasmid backbone**

| SIGNAL | (heptamer) | (spacer) | (nonamer) | HEP | NON | ALL | IMP | $A_5$ | RRMT |
|---|---|---|---|---|---|---|---|---|---|
| | CACAGTG | CTACAGACTGGA | ACAAAAACC | 7 | 9 | 16 | 7 | 5 | |
| 77 | CACAGTA | CTGACAAACCCA<br>TACACCTCCTC | TGAAATACC | 6 | 6 | 12 | 6 | 4 | Δ/SJ |
| 1318 | CACAGCG | TGTATAATCCAA | GTAAGTATC | 6 | 4 | 10 | 5 | 3 | INV |
| 1442 | CACGATG | ACTACTGGTCAT | TCAGCTATG | 5 | 3 | 8 | 4 | 2 | INV |
| 1632 | CACATAC | TGCTGGAAGAAG | ACGAAATCC | 4 | 7 | 11 | 6 | 3 | INV |
| 1947 | CACAGAG | CTGCTCTCAGAG | CTCATTCAA | 6 | 1 | 7 | 4 | 1 | INV |
| 2213 | CACAGGG | TTTCAGCAGTTA | AGAATTATT | 6 | 4 | 10 | 5 | 3 | INV |
| 3762 | CACCGCA | TATGGTGCACTC | TCAGTACAA | 4 | 3 | 7 | 4 | 2 | INV |
| 5145 | CACTGAG | CGTCAGACCCCG | TAGAAAAGA | 5 | 4 | 9 | 6 | 4 | INV |
| 5865 | CACAGAA | TCAGGGGATAAC | GCAGGAAAG | 5 | 4 | 9 | 6 | 3 | Δ/SJ |
| 6069 | CACATIA | ATTGCGTTGCGC | TCACTGCCC | 5 | 4 | 9 | 4 | 1 | Δ/SJ |
| 6131 | CACAACA | TACGAGCCGGAA | GCATAAAGT | 4 | 5 | 9 | 7 | 4 | Δ/SJ |
| 6912 | CACTCCA | GAGCGATGAAAA | CGTTTCAGT | 3 | 1 | 4 | 4 | 1 | Δ/CJ |
| 7043 | CACCCAG | GGATTGGCTGAG | ACGAAAAAC | 4 | 7 | 11 | 6 | 4 | Δ/CJ |
| 7451 | CACAACG | AAGAGGTCCCTA | CTGTAAATA | 5 | 3 | 8 | 7 | 3 | Δ/CJ |

**B Heptamer, nonamer identities in the cryptic collection (plasmid)**

canonical sequence:

| | C | A | C | A | G | T | G | 12/23 | A | C | A | A | A | A | A | C | C |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | 0 | 14 | 0 | 9 | 3 | 5 | 6 | | 3 | 1 | 8 | 7 | 6 | 7 | 10 | 4 | 4 |
| C | 14 | 0 | 14 | 2 | 2 | 5 | 1 | | 3 | 7 | 1 | 1 | 1 | 1 | 3 | 3 | 5 |
| G | 0 | 0 | 0 | 1 | 7 | 1 | 7 | | 3 | 3 | 4 | 3 | 2 | 1 | 0 | 3 | 2 |
| T | 0 | 0 | 0 | 2 | 2 | 3 | 0 | | 5 | 3 | 1 | 3 | 5 | 5 | 1 | 4 | 3 |

cryptic consensus:

| | A | C | A | G | AC | G | 12/23 | T | C | A | A | A | A | A | AT | C |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|

**C. Nonamer identities, irrelevant spacer**

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| A | 5 | 7 | 3 | 4 | 3 | 2 | 2 | 4 | 1 |
| C | 2 | 2 | 3 | 4 | 4 | 5 | 3 | 3 | 5 |
| G | 3 | 2 | 2 | 4 | 3 | 5 | 5 | 4 | 4 |
| T | 4 | 3 | 6 | 2 | 4 | 2 | 4 | 3 | 4 |

| cryptic consensus: | A | A | T | - | CT | CG | G | AG | C |
|---|---|---|---|---|---|---|---|---|---|

FIG. 4. Sequences and analysis of identified cryptic signals. (A) Cryptic signals from the plasmid backbone. The canonical joining-signal sequence is shown above the line, and positions that are most critical to joining-signal function for in vivo recombination (1, 34) are shown in boldface. Cryptic signals, named as in Fig. 3, are aligned below; in each case, nucleotides that match the canonical sequence are underlined. In the six adjacent columns are scored the identities to the canonical heptamer (HEP), to the nonamer (NON), to the total (heptamer plus nonamer; ALL), to functionally important positions (IMP), and to $A_{3-7}$ in the nonamer ($A_5$). The type of rearrangement is indicated in the column marked RRMT (Fig. 3): Δ/SJ (deletion with retained signal joint), Δ/CJ (deletion with retained coding joint), and INV (inversion). (B) Consensus sequence for cryptic signals. The frequency of A, C, G, or T at each position is given. In the last line, a consensus for the cryptic sequences is shown. Matches to the canonical sequence are underlined and in boldface. Ambiguities are noted. (C) Consensus nonamer sequence derived by using an incorrect spacer length (12 bp in place of 23 bp or vice versa).

mation of junctions with the appropriate structural characteristics. The collection of sequences in Fig. 5 is, to the best of our knowledge, a comprehensive listing of all published non-V-, non-D-, or non-J-associated sequences observed to undergo site-specific rearrangement. A DNA sequence was listed only if it had been isolated in recombinant form as a signal joint [with a typically precise connection to an authentic V(D)J joining signal] or if the criteria specified above for coding joints (including the existence of multiple isolates) were met.

When plasmid-derived cryptic signals were compared to

**A. Endogenous cryptic signals (authentic joining signal partners)**

| SIGNAL | heptamer - spacer - nonamer | HEP | NON | ALL | IMP | $A_5$ | (ref) |
|---|---|---|---|---|---|---|---|
| canonical | CACAGTG -12/23- ACAAAAACC | 7 | 9 | 16 | 7 | 5 | |
| 11p13-1 (H) | CACAGTA -12- GCAATAATT | 6 | 5 | 11 | 6 | 4 | (15) |
| 1p32-1 (H) | CACACCG -23- CGAAAAAGG | 5 | 5 | 10 | 7 | 5 | (6) |
| 11P15-1 (H) | CACAGTG -23- CTCTGGCAT | 7 | 0 | 7 | 4 | 0 | (8, 48) |
| Y2-6 | CACCATG -12- AGTGTATCA | 5 | 3 | 8 | 4 | 1 | (72) |
| Y1-1 | CACATTA -12- TTCAATTCA | 5 | 3 | 8 | 5 | 2 | (72) |
| Y2-16 | CACTGTC -12- TCTTTAACA | 5 | 4 | 9 | 5 | 2 | (72) |
| Y2-12 | CACTCTT -23- CACACAGCT | 4 | 3 | 7 | 4 | 2 | (72) |
| AT8-1-12B | CACAGCC -12- GCAAGAAAG | 5 | 5 | 10 | 6 | 4 | (40) |
| f41 | CACAGTG -12- ACTTCATTA | 7 | 3 | 10 | 5 | 1 | (36) |
| RS-1 | CACAGTG -23- AAGAGAACA | 7 | 5 | 12 | 6 | 3 | (64) |
| | CACAGTG -12- CCACTAATC | | 5 | 12 | 6 | 3 | (64) |
| RS-2 | CACAAAA -12- ACTTTAATT | 4 | 4 | 8 | 6 | 2 | (64) |
| | CACAAAA -23- AAGAGAACA | | 5 | 9 | 6 | 3 | (64) |
| RS(3' Ck) | CACAGTG -23- GCAGAAACT | 7 | 6 | 13 | 7 | 4 | (64) |
| | CACAGTG -12- TGACTGCCC | | 3 | 10 | 4 | 1 | (64) |
| J-Ck (H) | CACAGTG -12- GCCACTAAG | 7 | 3 | 10 | 5 | 2 | (65) |
| | CACAGTG -23- AAGAGAACA | | 5 | 12 | 6 | 3 | (54) |
| Vhq52 | CACAGTA -12- GTCATTAGT | 6 | 2 | 8 | 5 | 2 | (38) |
| Vh7183 | CACAGTA -12- TGTGTCCTC | 6 | 1 | 7 | 1 | 0 | (59) |
| Vh10,81X | CACAGTA -12- GTCCTCAGA | 6 | 1 | 7 | 5 | 1 | (17, 72) |
| $V_{T15.42}$ | CACATAC -23- ACACACATA | 4 | 5 | 9 | 6 | 3 | (70) |
| $V_{T15.49}$ | CACTCCA -23- TACCAAGC- | 3 | (3) | (6) | 5 | 2 | (70) |
| Ψ-Jα (H) | CACAGGA -12- ACAAAAACC | 5 | 9 | 14 | 7 | 5 | (21) |
| δRec (H) | CACGGTG -23- GCCTAAACC | 6 | 6 | 12 | 6 | 3 | (22) |
| δRec1 | CACAGTG -23- GCATAAACC | 7 | 7 | 14 | 7 | 4 | (69) |
| δRec2 | CACAGAC -23- ATATAAGAA | 5 | 4 | 9 | 5 | 3 | (69) |
| δRec3 | CACAGTA -23- CCAAATCTA | 6 | 4 | 10 | 4 | 3 | (69) |
| 5'del (H) | CACAGTG -23- ATTAGAAGC | 7 | 5 | 12 | 6 | 3 | (6, 12) |
| Nalm-6a (H) | CACTGTG -12- TACTTAAAC | 6 | 3 | 9 | 5 | 2 | (31) |
| Vγdel (H) | CACAGCG -23- ACTCCAAGG | 6 | 4 | 10 | 6 | 2 | (13, 71) |

FIG. 5. Naturally occurring cryptic signals. (A) All cryptic sequences, of genomic origin, reported to recombine with an authentic V(D)J joining signal (some participate in cryptic signal-to-cryptic signal rearrangements as well). Some signals are renamed for convenience. In all cases, spacer lengths were taken as exactly 12 or 23 bp, regardless of the specified length in the original report [column marked (ref)]. Two nonamers are shown if a sequence is known to function as both a 23-spacer signal and a 12-spacer signal in rearrangement with an authentic signal. In such cases, both nonamers were scored in panel C (and the heptamer was counted twice). (H) indicates sequences from the human genome; other sequences are from the mouse. (B) Cryptic signals reported to recombine with other cryptic signals. A spacer length of 12 or 23 bp is not certain for any of these signals, and it is possible that some sequences are bifunctional (10). Only the heptamer sequences were scored in panel C. Column designations in panels A and B are as in Fig. 4; references are given in the last column. Consensus data (C) are displayed as in Fig. 4.

those isolated from a physiological context, there were no conspicuous differences between the two collections. In each case, there were examples where fewer than half of 16 heptamer and nonamer nucleotides matched an authentic joining signal. The consensus sequences derived from the endogenous collection and plasmid collections (Fig. 4B and 5C) were also similar. For both, the most strongly conserved nucleotides lay in the first four positions of the heptamer and within the A

**B. Endogenous cryptic signals (cryptic partners; all human)**

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| SIL-1 | CACTCTG | -12- | GGATCAGTT | 5 | 2 | 7 | 4 | 2 | (2, 11) |
| | | -23- | TGTTTGAAG | | 1 | 6 | 4 | 1 | |
| SCL-1 | CACAGCC | -23- | TAAGGAAAA | 5 | 3 | 8 | 6 | 3 | (2, 11) |
| | | -12- | GTATATTGC | | 3 | 8 | 5 | 2 | |
| SCL-2 | CACAGAG | -12- | GCCAAAACT | 6 | 6 | 12 | 7 | 4 | (7, 37) |
| | | -23- | CAGTAACCC | | 4 | 10 | 6 | 2 | |
| HPRT-1A | CACAGTA | -12- | GGAGATGCT | 6 | 3 | 9 | 5 | 2 | (26, 27) |
| | | -23- | TGGCCTCAT | | 0 | 6 | 4 | 0 | |
| HPRT-3A | CACACAC | -12- | ACAAATACA | 4 | 7 | 11 | 6 | 4 | (27) |
| | | -23- | TATGTGTTT | | 0 | 4 | 4 | 0 | |
| HPRT-3B | CACAGAG | -23- | AATAAAAAA | 6 | 5 | 11 | 7 | 4 | (27) |
| | | -12- | AGAATATTC | | 5 | 11 | 5 | 3 | |

**C Heptamer, nonamer identities in the cryptic collection (endogenous)**

canonical sequence:

| | C | A | C | A | G | T | G | 12/23 | A | C | A | A | A | A | A | C | C |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | 0 | 36 | 0 | 29 | 5 | 7 | 11 | | 11 | 6 | 11 | 13 | 10 | 21 | 20 | 5 | 11 |
| C | 36 | 0 | 36 | 1 | 4 | 5 | 6 | | 5 | 14 | 9 | 6 | 6 | 3 | 4 | 13 | 8 |
| G | 0 | 0 | 0 | 1 | 26 | 1 | 18 | | 8 | 4 | 3 | 3 | 4 | 2 | 3 | 5 | 4 |
| T | 0 | 0 | 0 | 5 | 2 | 23 | 1 | | 6 | 6 | 7 | 8 | 10 | 4 | 3 | 7 | 6 |

cryptic consensus:

| C | A | C | A | G | T | G | 12/23 | A | C | A | A | AT | A | A | C | A |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|

FIG. 5—*Continued.*

tract of the nonamer. (The endogenous collection overall more closely approximated the canonical sequence; this is likely a reflection of the greater number of sequences in the collection as well as the fact that some sequences, such as Ψ-Jα [21], clearly evolved from a real joining signal.) The similarity between the two groups of cryptic signals indicates that the sequences that we can identify with the extrachromosomal recombination substrate are representative of those that recombine in vivo. This observation helps to establish the defined substrate as an appropriate experimental model for V(D)J target misrecognition. Further confirmation was obtained as described below.

**Joining potential of cryptic signal sequences.** We wished to determine the relative target functions of various cryptic sequences, compared to a canonical sequence, in a standardized recombination assay. To this end, we prepared a set of comparably modified extrachromosomal V(D)J recombination substrates that had been substituted with selected cryptic signals (Fig. 6). In the parental construct, an authentic 23-spacer signal could join to either of two 12-spacer signals, positioned at site 1 or site 2. Both site 1 and site 2 recombinants lead to chloramphenicol resistance (Fig. 6, bottom).

Seven recombination substrates (pWTSJΔ, p6131SJΔ, p6912SJΔ, p5145SJΔ, pHPR-3aSJΔ, pSIL-1SJΔ, and pCASJΔ) differed from one another only in the identity of the site 1 sequence (as designated by the underlining). The signal at site 2 was invariant and served as an internal standard. The different constructs were singly transfected into cells. The site 1/site 2 ratio for each was determined by picking large numbers of chloramphenicol-resistant colonies to replicate grids and testing hybridization to the indicated oligonucleotide probes (Fig. 7 and Materials and Methods).

Two aspects of the assay maximized sensitivity. One was that the sequence serving as the invariant site 2 standard was itself a cryptic signal, 6131 (Fig. 3 and 4). By measuring joining frequencies relative to a weak signal, roughly 100-fold fewer colonies had to be picked for each determination. This effectively extended the practical range of the assay 2 orders of magnitude. In addition, very low levels of joining-signal function could be detected because the hybridization analyses allowed us to screen out all background chloramphenicol-resistant colonies (as arise from illegitimate rearrangements as well as cryptic recombination to the sites in the vector backbone). The hybridization screen, though time-consuming, allowed measurement of joining activity well below previous limits of 1% of wild-type activity (34). The site 2 standard has previously been used as an internal control in other studies, and the frequency with which site 2 recombinants are recovered is independent of the presence or absence of a joining signal at site 1 (45).

To establish the absence of any gross bias in site 1 and site 2 recombination, we first constructed a plasmid in which 6131 was present at both positions. As expected, the site 1/site 2 ratio was close to 1:1 (p6131SJΔ [Fig. 7]). The site 1/site 2 ratio was then measured for the construct containing the canonical 12-spacer signal sequence at site 1. In this construct, the site 1/site 2 ratio was 94:1. Thus, the 6131 sequence, here used as our standard, had approximately 1% of the activity of an authentic joining signal.

With this information, we compared recombination function for three plasmid-derived cryptic signals, two human-derived cryptic signals and one synthetic sequence (a CA repeat) to that of the canonical signal.

Cryptic signals exhibited widely different V(D)J joining function (Fig. 7). While the 6131 sequence functioned at about 1% of wild-type levels, and a second plasmid sequence (5145)

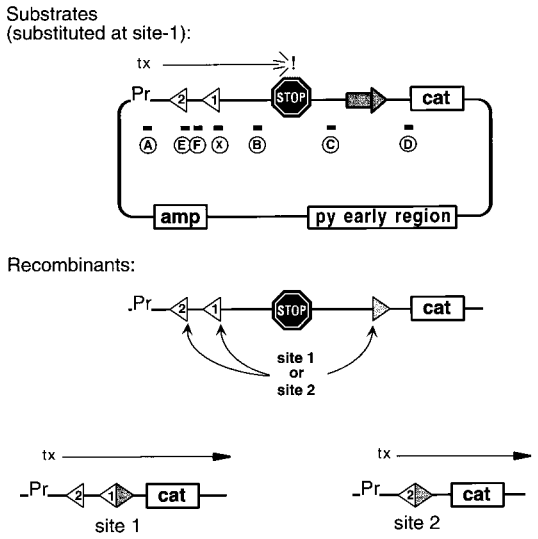



FIG. 6. The standardized test for joining-signal function. Seven versions of the substrate shown were constructed for which the site 1 signal was the canonical 12-spacer signal (as in pWTSJD) or was substituted with the 6131, 5145, 6912, HPR-3a, or SIL-1 sequence or a CA repeat sequence. The sole alteration in these constructs was in the site 1 sequence. Signal joints (formed via deletion) were measured. The heptamer-to-heptamer distance between site 1 and the 23-spacer signal is 253 bp, and that between site 1 and site 2 is 359. Both types of recombinant are chloramphenicol resistant and are distinguished from one another, and from other products, according to their hybridization properties. Oligonucleotide probes are as in Fig. 1, with additional probes described in Materials and Methods.

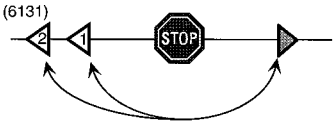
scored about fivefold lower, a third (6912) functioned below the practical limits of the assay at <0.003%. The relative joining activities of the human genome-derived sequences showed similar spreads: HPR-3a was very active at 3% of wild-type levels, but the activity SIL-1 was far lower, at 0.002%. In fact, as noted in Fig. 7, we recovered only one SIL-1 recombinant, which was confirmed by DNA sequence analysis, upon screening over 600 site 2 recombinants from four transfections. This result demonstrated that cryptic sequences scoring on the order of $10^{-5}$ in our assay can have an impact in a physiological context (2, 11).

We also tested the joining activity of a sequence composed of 14 CA repeats. CA repeats are preferentially found in euchromatin and are present in an estimated 50,000 to 100,000 copies in the genome (67). This simple repeat had significant activity in the assay, functioning relative to a canonical joining signal at 0.6%.

In summary, the plasmid-derived cryptic sequences are both structurally and functionally like sequences that are subject to mistargeting by the V(D)J joining machinery in the human genome. As mentioned above, the 14 cryptic signals that we have been able to identify are doubtless only a subset of the cryptic signals that exist within the 7.6-kb DNA. By simple extrapolation from the plasmid data, at least 10 million sequences with intrinsic V(D)J target function exist in a mammalian genome of roughly $6 \times 10^9$ bp. By comparison to a canonical joining signal in a standardized assay, the joining potential of a significant fraction of these sites is likely within severalfold of the 1% level.

## DISCUSSION

V(D)J recombination is an obligatory part of T- and B-cell development and is apparently ongoing throughout life (55). As such, the impact of V(D)J recombination errors can be considered on three levels: the single cell, the immune system, and the animal. At the level of the single cell, those T- and B-cell precursors in which cryptic signals rearrange are likely culled (or not culled) by the same mechanisms that eliminate T- and B-cell precursors where V(D)J joining has created




| Sequence at site 1 | | | | Site 1 | Site 2 | Relative joining function |
|---|---|---|---|---|---|---|
| Canonical: | CACAGTG | (12) | ACAAAAACC | 942 | 10 | 1 |
| Plasmid origin: | | | | | | |
| 6131 | CACAACA | (12) | GCATAAAGT | 177 | 209 | $1 \times 10^{-2}$ |
| 5145 | CACTGAG | (12) | TAGAAAAGA | 20 | 134 | $2 \times 10^{-3}$ |
| 6912 | CACTCCA | (12) | CGTTTCAGT | 0 | 388 | $<3 \times 10^{-5}$ |
| Genomic: | | | | | | |
| HPR-3a | CACACAC | (12) | ACAAATACA | 185 | 73 | $3 \times 10^{-2}$ |
| SIL-1 | CACTCTG | (12) | GGATCAGTT | 1 | 602 | $2 \times 10^{-5}$ |
| Synthetic: | | | | | | |
| CA repeat | CACACAC | (12) | ACACACACA | 151 | 274 | $6 \times 10^{-3}$ |

FIG. 7. Results with the standardized joining assay. Site 1 and site 2 recombinants were determined as described in Materials and Methods. At least four different transfections were sampled for each determination. In the last column, results were normalized according to site 2 recombinants to give the joining function relative to the canonical signal.

nonfunctional out-of-frame variable exons (25, 51, 74). Unless cryptic recombination is rampant, one would imagine that there could be little negative impact on lymphopoiesis, because the immune system is already set up to cope with the fact that two-thirds of recombination attempts are out of frame. Mistargeting in V(D)J recombination is an altogether different sort of problem, however, when viewed from the perspective of the whole animal. A single mistake that activates an oncogene in just one of the T- or B-cell precursors generated during the animal's lifetime can have a lethal outcome. There are qualitatively different constraints in force for this site-specific recombination system than for recombination systems (such as lox/CRE or phage λ integration) that operate in single-celled organisms.

The type of V(D)J joining error investigated here is one in which defined, characteristic recombinant junctions are formed following misrecognition of a cryptic signal. Throughout, we designate a DNA sequence as a cryptic signal on the basis of functional criteria: it must be recognized, bound, brought together with a partner joining signal, cleaved, and successfully reconnected by the V(D)J joining machinery.

Information provided by recent studies indicates that different phases of recombination (binding, cutting, and complete in vivo rearrangement) may require an overlapping but nonidentical set of DNA sequence features in a canonical joining signal. With respect to binding, RAG1-joining signal interactions have been measured, and more recently, a stable complex that involves RAG1, RAG2, and the target DNA has been described (23, 35, 66). The former studies highlighted nonamer positions $A_5$ and $A_6$ (23, 66), whereas stable complex formation was affected by the identity of the last four heptamer nucleotides ($AGTG_{1-4}$) as well as the presence of the nonamer (35). Somewhat different features govern the cleavage of joining signals in vitro (19, 58). Here it has been shown that the identity of the first three heptamer nucleotides ($CAC_{1-3}$) greatly affects the efficiency with which double-strand breaks can be introduced. The analysis showed that the nonamer could be eliminated, but it clearly augmented in vitro cleavage by purified RAG1 and RAG2. A single change in $A_6$ of the nonamer has the same effect as removing the nonamer identity entirely (58). Finally, the importance of different positions within the canonical joining signal for complete in vivo recombination has been assessed in two large studies (1, 34). These analyses showed that in the context of a wild-type sequence, single base changes within $CAC_{1-3}$ of the heptamer, or $A_5$, $A_6$, or $A_7$ of the nonamer (depending on the study), had the greatest impact. As noted, the nucleotides that were found to be functionally important for in vivo recombination corresponded well to the best-conserved positions in authentic joining signals in nature.

The analyses summarized above point to important features in the joining signal but do not allow one to predict which features most significantly affect V(D)J target function for a degenerate, multiply mismatched sequence, nor do these studies allow an estimate of the number of cryptic signals that might exist within a defined region of DNA.

Here, in a long-term analysis, we determined the minimum number of fortuitous sites in a 7.6-kb DNA able to function as an acceptable V(D)J recombination target. We validated the approach in two ways: first by showing that the DNA sequences identified in the plasmid backbone were similar in nature to the sequences with cryptic signal function from the mouse or human genome, and second by demonstrating that the level of function of the plasmid-derived signals was comparable to cryptic signals of endogenous origin. By extrapolation from the plasmid data, we can therefore estimate that there are at least 10 million sites in the genome with a DNA sequence compatible with joining-signal function.

**What does it take to be a cryptic signal? (i) Heptamer and nonamer identities.** Comparison of the cryptic signals to the canonical sequence showed that, apart from $CAC_{1-3}$ of the heptamer, no other nucleotide is consistently present (Fig. 4). Seven positions previously highlighted as important to function, in vivo, for the canonical joining signal ($CACA_{1-4}$ of the heptamer and $A_5$, $A_6$, and $A_7$ of the nonamer [1, 34]) appear together in only two cryptic signals in the plasmid collection (Fig. 4A, IMP). The same observations—conservation of $CAC_{1-3}$, marked variability, and few examples where all seven of the functionally significant identities are present—apply to the chromosomally derived cryptic sequences in Fig. 5 as well.

A CAC sequence, of itself, is unlikely to have significant cryptic signal function, judging from the SIL-1 sequence, which had seven canonical identities ($CAC_{1-3}$ plus four others). SIL-1 was an extremely weak joining signal and recombined at approximately $2 \times 10^{-5}$ the canonical frequency (Fig. 7). As discussed below, it is possible that the SIL-1 sequence represents a lower functional limit for cryptic signals with a physiological impact.

Despite the variable identities found at positions other than CAC in the cryptic signal collections, a consensus sequence derived from this small collection closely approximated the canonical joining-signal sequence. This was not necessarily expected, and we interpret the fact that the canonical joining-signal sequence can be distilled from the cryptic sequences to mean that sporadic matches existing throughout a cryptic sequence are functionally important.

**(ii) Spacers in cryptic signals.** All cryptic signal sequences of either plasmid or genomic origin were analyzed on the basis of a spacer length of either exactly 12 bp or exactly 23 bp. Experimentally, binding, cleavage and in vivo V(D)J recombination all are sensitive to the length of the spacer separating a canonical heptamer and nonamer (1, 19, 34, 58, 73). We have no definitive evidence that spacers do not vary in the case of cryptic signals, but we think it unlikely because the nonamers defined by a +1- or −1-bp spacer length were generally not more highly matched than those defined by an exact 12- or 23-bp spacing. Among all 28 possible noncanonical spacer alignments, we found only five alignments that increased nonamer identity, and this could well have occurred by chance. It was notable that in all but one of the five better-matched spacer variants, the nonamers gained identities at less evolutionarily conserved positions at the expense of an important ($A_6$ or $A_7$) nonamer match (not shown).

**(iii) The effects of coding-flank sequence.** The nucleotides just adjacent to a joining signal (comprised of V, D, or J coding sequences in a real gene segment) are termed the coding flank. Work with canonical joining-signal sequences has shown that the identity of one to three nucleotides at the coding flank can have an impact on the efficiency with which a canonical signal becomes cleaved in vitro or recombined in vivo (9, 19, 24, 29, 58, 61, 63, 73).

According to these observations, it was of interest to examine whether coding-flank sequence might be an important extrasignal determinant of cryptic signal identity. As can be seen in Fig. 8, however, the DNA sequences adjacent to our cryptic signals revealed no obvious bias, purine/pyrimidine pattern (58), or other restriction. Both preferred and proscribed coding-flank dinucleotide sequences (as highlighted in other studies [29, 58, 63]) occurred within the small collection. Further, two coding flanks differing in quality were sampled for the 6131 sequence in the construct p6131SJΔ. In this substrate, the site 1 coding flank was GA; at site 2, it was CA. Sadofsky et al. (63)

| signal | flank | heptamer-(etc) |
|---|---|---|
| 77 | AACACAGG | CACAGTA-23-... |
| 1318 | TGGATGTG | CACAGCG-12-... |
| 1442 | GTCCACTA | CACGATG-12-... |
| 1632 | TACACCCG | CACATAC-12-... |
| 1947 | GAGCTACT | CACAGAG-12-... |
| 2213 | TGACTAAG | CACAGGG-12-... |
| 3762 | GTATTTCA | CACCGCA-12-... |
| 5145 | TTTCGTTC | CACTGAG-12-... |
| 5865 | CGGTTATC | CACAGAA-12-... |
| 6069 | AGCTAACT | CACATTA-12-... |
| 6131 | CAATTCCA | CACAACA-12-... |
| 6912 | GTGGTATT | CACTCCA-12-... |
| 7043 | GTGAAACT | CACCCAG-12-... |
| 7451 | GGTACCTA | CACAACG-12-... |

FIG. 8. Analysis of coding-flank composition in identified cryptic signals. Coding-flank sequences for each cryptic signal are given. See text for details.

have observed marked coding-flank preferences in connection with a canonical joining signal in cells transfected with the RAG1 D32 mutant (CA was relatively neutral, and GA was strongly disfavored). Regardless, for p6131SJΔ, the site 2/site 1 ratio was nearly 1:1 (Fig. 7). At a first approximation, it would appear that coding-flank sequence is not a dominant factor in determining whether a DNA sequence will function as a cryptic site.

In summary, although the consensus sequence derived from cryptic signals would suggest that certain heptamer and nonamer positions are more important than others (Fig. 4B and 5C), a wide spectrum of DNA sequences comprised of CAC with additional, nonunique matches to the canonical sequence are functional V(D)J joining targets. The existence of an enormous population of cryptic signals with highly variable combinations of identities in both heptamer and nonamer elements is indicated.

**Relative joining functions of identified cryptic signals.** Two human cryptic sequences were chosen for study because they illustrated two quite different biological effects of cryptic signal recognition and because it was anticipated that they might have very different intrinsic joining-signal functions. Each is subject to mistargeting in vivo but at very different apparent frequencies within the human population. The SIL-1 sequence is site-specifically rearranged (to one of several partner cryptic signals) in T-ALL (2, 11), which is a rare disease. Activation of the *scl* gene (as caused either by *sil*/*scl* deletion or by other mechanisms) is highly correlated with T-ALL, and only about 300 new cases of T-ALL are diagnosed per year in the United States (reviewed in reference 30). The HPR-3a sequence, in contrast, is recombined in an estimated one in $2 \times 10^7$ circulating fetal or adult T cells, so it is likely that every individual possesses some cells that are site-specifically rearranged at this cryptic sequence (26, 27). As anticipated, the SIL-1 and HPR-3a sequences functioned at very different levels in the standardized V(D)J joining assay. The SIL-1 sequence was $2 \times 10^{-5}$ times as active as a canonical signal, while the HPR-3a sequence had 3% of the activity of a canonical signal (Fig. 7). Direct demonstration of signal joint formation formally documented cryptic signal function in both cases (each had been defined previously on the basis of coding joint criteria only [2, 11]). The 1,000-fold difference in intrinsic joining-signal activity measured for SIL-1 and HPR-3a would seem to at least partially account for the large apparent difference in occurrence of *sil*/*scl* and *hprt* deletions in human T cells.

We have estimated that there are about 10 million cryptic sites in the diploid genome and here can make some statements about relative target site function. Two representative plasmid-derived cryptic signals (6131 and 5145), each possessing the average number of matches to the canonical sequence, roughly evenly distributed between heptamer and nonamer, showed significant joining-signal function, 1% in one case and 0.2% in another (Fig. 7). On this basis, it seems plausible that a substantial fraction of genomic cryptic sequences will possess a joining potential in the range of 1%. Moreover, a CA repeat sequence (present in 50,000 to 100,000 copies in the genome [67]) was shown to have a joining activy of 0.6% (Fig. 7), and a perfect heptamer (which should by chance arise in roughly 200,000 copies in the genome) devoid of any nonamer matches has been reported to function at 0.6 or 5% in other studies (1, 34). By any measure, DNA sequences with the potential to function around the 1% level are abundant. We can crudely relate the level of function measured in the extrachromosomal assay to rearrangement in a physiological context: as mentioned above, the HPR-3a sequence, with a 3% joining potential in the standardized assay, is rearranged in most (perhaps all) humans (26).

These analyses indicate that the genome is riddled with DNA sequences that have the potential to act as cryptic V(D)J recombination targets, with an intrinsic joining potential within a physiologically relevant range. The number of such sites alone compels the question of how it is possible to limit joining errors during T- and B-cell differentiation.

**How is fidelity achieved?** V(D)J recombination is controlled through a carefully programmed regulation of RAG protein levels and further by modulated changes in chromatin structure at the Ig and TCR loci (for recent discussions, see references 60 and 68). Less is known, however, about the degree to which the remainder of the genome (outside the Ig and TCR loci) is or is not exposed to the V(D)J joining enzymes. The *hprt* and *sil*/*scl* deletions (2, 11, 26, 27) have provided solid evidence that accessibility to the V(D)J joining enzymes is not strictly limited to Ig and TCR loci.

Conceivably, fidelity is governed by factors operating at a level other than chromatin structure. Studies of *trans* (interhomolog) V(D)J recombination in normal thymocytes indicate that if two V(D)J joining signals are on different homologs, recombination is demonstrably rare, occurring on the order of $10^{-4}$ times the frequency of *cis* recombination (4). *trans* rearrangements are also observed in B-cell precursors that have been transformed by a temperature-sensitive Abelson murine leukemia virus; these arise at a low frequency at the nonpermissive temperature and occur in violation of the 12/23 rule (5). It is interesting that class switch recombination, another type of developmentally regulated recombination in B-lineage cells that is mechanistically distinct from V(D)J joining, can occur quite readily in *trans*. For example, in a normal rabbit, there is evidence of interallelic rearrangement in fully 3 to 7% of IgA-expressing B cells (39). Favored interactions between linked DNAs in V(D)J recombination is consistent with the possibility that an active recombination machinery is present in only a few fixed sites, rather than freely diffusible, and perhaps is anchored prior to the onset of rearrangement. Other studies have indicated that components of the V(D)J joining machinery, particularly the *RAG1* and *RAG2* gene products, may be tethered to the nuclear matrix (16, 18, 41). If there is indeed a small arena within which both the machinery and the target for V(D)J recombination must come together, this could ensure

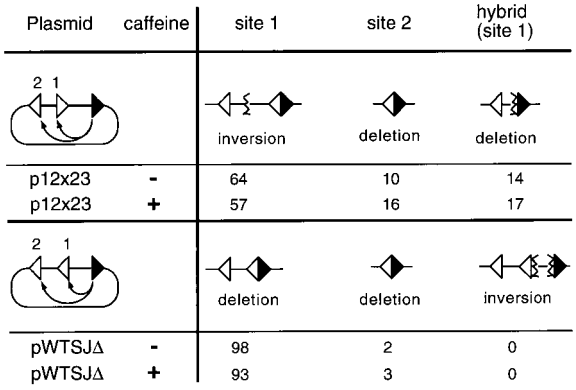



FIG. A1. Caffeine effects.

fidelity by greatly limiting the number of actively targeted cryptic signals in the genome.

To what extent is cryptic rearrangement an unavoidable accompaniment to V(D)J recombination? At present, this is difficult to answer, because without specific molecular probes (3, 20, 21, 32, 38, 57, 72) or a strong phenotypic selection (2, 11, 26, 27), cryptic rearrangements in a normal T- and B-cell population are effectively invisible. One would expect that natural selection has long since eliminated cryptic signals with a high probability of activating an oncogene through site-specific rearrangement. However, the number and activity of cryptic sequences in regions where such sequences have little pathogenic potential (as within the *hprt* gene), or the number of weak signals near oncogenic targets (as within the *scl* gene), are for the moment, unknown.

Our results indicate that there is the possibility of cryptic V(D)J recombination for virtually any introduced DNA that may pass through the early T- and B-cell differentiation program. This ought to be a concern when it comes to evaluating the safety of certain gene therapy or DNA vaccination protocols.

## APPENDIX

Transfections of 204-1-8 cells with p12×23 (45) and pWTSJD (Fig. 5) were carried out in the presence or absence of 1 mM caffeine in the medium posttransfection (52). In our hands, as previously reported (52), caffeine increased the number of recombinants recovered per transfection by 5- to 10-fold (not shown). This was useful in the present experiments, where we were interested in quantifying function for weak joining signals. p12×23 and pWTSJD are identical except for the orientation of the 12-spacer signal in site 1. Both constructs contain the cryptic 12-spacer signal, 6131, at site 2. All pictured recombinants (Fig. A1) are chloramphenicol resistant. Transfections were carried out in parallel with and without caffeine. About 100 recombinants were picked, and the numbers of inversions, deletions, hybrid junctions, and site 2 recombinants among them were determined. The values for site 1, site 2, and hybrid junctions for a given construct were not significantly different with and without caffeine (by the $\chi^2$ test, $P > 0.3$ for values with either plasmid). For comparison between p12×23 and pWTSJΔ, values can be normalized to the number of site 2 recombinants, and it is also evident from this comparison that inversion/deletion ratios are the same with and without caffeine. The results indicated that caffeine did not affect inversion/deletion ratios, the ratio of cryptic and canonical rearrangements, or the ratio of hybrid and standard recombination.

## ACKNOWLEDGMENTS

We thank Joanne Hesse (NIH) for kindly providing cryptic recombinants isolated by her for sequence analysis, Greg Brown for expert technical assistance, and Veena Guru for her contributions in the construction and analysis of pCASJΔ. We are very grateful to J. Hesse, C. Guidos, H. Lipshitz, N. Rosenberg, M. Shulman, and G. Wu for comments on the manuscript and to M. Sadofsky for helpful discussion.

This work was supported by American Cancer Society grant IM-599C and by National Cancer Institute of Canada grant no. 007362. S.L. is a Research Scientist of the National Cancer Institute of Canada.

## REFERENCES

1. **Akamatsu, Y., N. Tsurushita, F. Nagawa, M. Matsuoka, K. Okazaki, M. Imai, and H. Sakano.** 1994. Essential residues in V(D)J recombination signals. J. Immunol. **153:**4520–4529.
2. **Aplan, P. D., D. P. Lombardi, A. M. Ginsberg, J. Cossman, V. L. Bertness, and I. R. Kirsch.** 1990. Disruption of the human SCL locus by "illegitimate" V-(D)-J recombinase activity. Science **250:**1426–1429.
3. **Arden, B.** 1992. Diversity of novel recombining elements suggests developmentally programmed expression of the T cell receptor α/δ locus. Eur. J. Immunol. **22:**1287–1291.
4. **Aster, J. C., and J. Sklar.** 1992. Interallelic V(D)J *trans*-rearrangement within the β T cell receptor gene is infrequent and occurs preferentially during attempted Dβ to Jβ gene joining. J. Exp. Med. **175:**1773–1782.
5. **Bailey, S. N., and N. Rosenberg.** 1997. Assessing the pathogenic potential of the V(D)J recombinase by interlocus immunoglobulin light-chain gene rearrangement. Mol. Cell. Biol. **17:**887–894.
6. **Bernard, O., P. Gugliemi, P. Jonveaux, D. Cherif, S. Gisselbrecht, M. Mauchauffe, R. Berger, C.-J. Larsen, and D. Mathieu-Mahul.** 1990. Two distinct mechanisms for the SCL gene activation in the t(1:14) translocation of T-cell leukemias. Genes Chromosomes Cancer **1:**194–208.
7. **Bernard, O., N. Lecointe, P. Jonveaux, M. Souyri, M. Mauchauffé, R. Berger, C. J. Larsen, and D. Mathieu-Mahul.** 1991. Two site-specific deletions and t(1;14) translocation restricted to human T-cell acute leukemias disrupt the 5′ part of the *tal-1* gene. Oncogene **6:**1477–1488.
8. **Boehm, T., R. Baer, I. Lavenir, A. Forster, J. J. Waters, E. Nacheva, and T. H. Rabbitts.** 1988. The mechanism of chromosomal translocation t(11;14) involving the T-cell receptor Cδ locus on human chromosome 14q11 and a transcribed region of chromosome 11p15. EMBO J. **7:**385–394.
9. **Boubnov, N. V., Z. P. Wills, and D. T. Weaver.** 1995. Coding sequence composition flanking either signal element alters V(D)J recombination efficiency. Nucleic Acid Res. **23:**1060–1067.
10. **Breit, T. M., E. J. Mol, I. L. M. Wolvers-Tettero, W.-D. Ludwig, E. R. van Wering, and J. J. M. van Dongen.** 1993. Site-specific deletions involving the tal-1 and sil genes are restricted to cells of the T cell receptor α/β lineage: T cell receptor δ gene deletion mechanism affects multiple genes. J. Exp. Med. **177:**965–977.
11. **Brown, L., J.-T. Cheng, Q. Chen, M. J. Siciliano, W. Crist, G. Buchanan, and R. Baer.** 1990. Site-specific recombination of the *tal*-1 gene is a common occurrence in human T cell leukemia. EMBO J. **9:**3343–3351.
12. **Burnett, R. C., R. I. Espinosa, T. B. Shows, R. L. Eddy, M. M. LeBeau, J. D. Rowley, and M. O. Diaz.** 1993. Molecular analysis of a t(11;14)(q23;q11) from a patient with null-cell acute lymphoblastic leukemia. Genes Chromosomes Cancer **7:**38–46.
13. **Carr, T. F., R. F. Stevens, C. Burton, and C. J. Harrison.** 1995. An interstitial deletion in the rearranged T-cell receptor γ chain locus in a case of T-cell acute lymphoblastic leukaemia. Br. J. Haematol. **89:**688–689.
14. **Chen, C., Z. Nagy, E. L. Prak, and M. Weigert.** 1995. Immunoglobulin heavy chain gene replacement: a mechanism of receptor editing. Immunity **3:**747–755.
15. **Cheng, J.-T., C. Y.-C. Yang, J. Hernandex, J. Embrey, and R. Baer.** 1990. The chromosome translocation (11;14)(p13-q11) associated with T cell acute leukemia: asymmetric diversification of the translocation junctions. J. Exp. Med. **171:**489–501.
16. **Cortes, P., Z.-S. Ye, and D. Baltimore.** 1994. RAG-1 interacts with the repeated amino acid motif of the human homologue of the yeast protein SRP1. Proc. Natl. Acad. Sci. USA **91:**7633–7637.
17. **Covey, L. R., P. Ferrier, and F. W. Alt.** 1990. $V_H$ to $V_HDJ_H$ rearrangement is mediated by the internal $V_H$ heptamer. Int. Immunol. **2:**579–583.
18. **Cuomo, C. A., S. A. Kirch, J. Gyuris, R. Brent, and M. A. Oettinger.** 1994. Rch1, a protein that specifically interacts with the RAG-1 recombination-activating protein. Proc. Natl. Acad. Sci. USA **91:**6156–6160.
19. **Cuomo, C. A., C. L. Mundy, and M. A. Oettinger.** 1996. DNA sequence and structure requirements for cleavage of V(D)J recombination signal sequences. Mol. Cell. Biol. **16:**5683–5690.
20. **Daitch, L. E., M. W. Moore, D. M. Persiani, J. M. Durdik, and E. Selsing.** 1992. Transcription and recombination of the murine RS element. J. Immunol. **149:**832–840.
21. **de Villartay, J.-P., R. D. Hockett, D. Coran, S. J. Korsmeyer, and D. I. Cohen.** 1988. Deletion of the human T-cell receptor δ-gene by a site-specific recombination. Nature **335:**170–174.
22. **de Villartay, J.-P., D. Lewis, R. Hockett, T. A. Waldmann, S. J. Korsmeyer, and D. I. Cohen.** 1987. Deletional rearrangement in the human T-cell re-

ceptor α-chain locus. Proc. Natl. Acad. Sci. USA **84:**8608–8612.
23. **Difilippantonio, M. J., C. J. McMahan, Q. M. Eastman, E. Spanopoulou, and D. G. Schatz.** 1996. RAG1 mediates signal sequence recognition and recruitment of RAG2 in V(D)J recombination. Cell **87:**253–262.
24. **Ezekiel, U. R., P. Engler, D. Stern, and U. Storb.** 1995. Asymmetric processing of coding ends and the effect of coding end nucleotide composition on V(D)J recombination. Immunity **2:**381–389.
25. **Fang, W., D. L. Mueller, C. A. Pennell, J. J. Rivard, Y.-S. Li, R. R. Hardy, M. S. Schlissel, and T. W. Behrens.** 1996. Frequent aberrant immunoglobulin gene rearrangements in Pro-B cells revealed by a *bcl-*$x_L$ transgene. Immunity **4:**291–299.
26. **Fuscoe, J. C., L. J. Zimmerman, K. Harrington-Brock, L. Burnette, M. M. Moore, J. A. Nicklas, J. P. O'Neill, and R. J. Albertini.** 1992. V(D)J recombinase-mediated deletion of the hprt gene in T-lymphocytes from adult humans. Mutat. Res. **283:**13–20.
27. **Fuscoe, J. C., L. J. Zimmerman, M. J. Lippert, J. A. Nicklas, J. P. O'Neill, and R. J. Albertini.** 1991. V(D)J recombinase-like activity mediates *hprt* gene deletion in human fetal T-lymphocytes. Cancer Res. **51:**6001–6005.
28. **Gellert, M.** 1996. A new view of V(D)J recombination. Genes Cells **1:**269–276.
29. **Gerstein, R. M., and M. R. Lieber.** 1993. Coding end sequence can markedly affect the initation of V(D)J recombination. Genes Dev. **7:**1459–1469.
30. **Goldfarb, A. N., and J. M. Greenberg.** 1994. T-cell acute lymphoblastic leukemia and the associated basic helix-loop-helix gene *SCL/tal*. Leukemia Lymphoma **12:**157–166.
31. **Hansen-Hagge, T. E., S. Yokota, H. F. Reuter, K. Schwarz, 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.
32. **Hara, J., Y. Takihara, K. Yumara-Yagi, S. Ishihara, A. Tawa, T. W. Mak, E. W. Gelfand, S. Okada, and K. Kawa-Ha.** 1991. Differential usage of δ recombining element and Vδ genes during T-cell ontogeny. Blood **78:**2075–2081.
33. **Hesse, J. E., M. R. Lieber, M. Gellert, and K. Mizuuchi.** 1987. Extrachromosomal DNA substrates in pre-B cells undergo inversion or deletion at immunoglobulin V-(D)-J joining signals. Cell **49:**775–783.
34. **Hesse, J. E., M. R. Lieber, K. Mizuuchi, and M. Gellert.** 1989. V(D)J recombination: a functional definition of the joining signals. Genes Dev. **3:**1053–1067.
35. **Hiom, K., and M. Gellert.** 1997. A stable RAG1-RAG2-DNA complex that is active in V(D)J cleavage. Cell **88:**65–72.
36. **Höchtl, J., and H. G. Zachau.** 1983. A novel type of aberrant recombination in immunoglobulin genes and its implication for V-J joining mechanism. Nature **302:**260–263.
37. **Janssen, J. W. G., W.-D. Ludwig, W. Sterry, and C. R. Bartram.** 1993. Sil-Tal1 deletion in T-cell acute lymphoblastic leukemia. Leukemia **7:**1204–1210.
38. **Kleinfield, R., R. R. Hardy, D. Tarlinton, J. Dangl, L. A. Herzenberg, and M. Weigert.** 1986. Recombination between an expressed immunoglobulin heavy-chain gene and a germline variable gene segment in a Ly 1+ B-cell lymphoma. Nature **322:**843–846.
39. **Knight, K. L., M. Kingzette, M. A. Crane, and S.-K. Zhai.** 1995. Transchromosomally derived Ig heavy chains. J. Immunol. **155:**684–691.
40. **Komori, T., H. Sugiyama, and S. Kishimoto.** 1989. A novel $V_HDJ_H$ to $J_H$ joining that induces H chain production in an Ig-null immature B cell line. J. Immunol. **143:**1040–1045.
41. **Leu, T. M. J., and D. G. Schatz.** 1995. Rag-1 and Rag-2 are components of a high-molecular-weight complex, and association of Rag-2 with this complex is Rag-1 dependent. Mol. Cell. Biol. **15:**5657–5670.
41a. **Lewis, S.** Unpublished data.
42. **Lewis, S., J. E. Hesse, K. Mizuuchi, and M. Gellert.** 1988. Novel strand exchanges in V(D)J recombination. Cell **55:**1099–1107.
43. **Lewis, S. M.** 1994. The mechanism of V(D)J joining: lessons from molecular, immunological, and comparative analyses. Adv. Immunol. **56:**27–150.
44. **Lewis, S. M.** 1994. P nucleotides, hairpin DNA and V(D)J joining: making the connection. Semin. Immunol. **6:**131–141.
45. **Lewis, S. M., and J. E. Hesse.** 1991. Cutting and closing without recombination in V(D)J joining. EMBO J. **10:**3631–3639.
46. **Lieber, M. R., J. E. Hesse, S. Lewis, G. C. Bosma, N. R. Rosenberg, K. Mizuuchi, M. J. 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.
47. **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.
48. **McGuire, E. A., R. D. Hockett, K. M. Pollock, M. F. Bartholdi, S. J. O'Brien, and S. J. Korsmeyer.** 1989. The t(11;14)(p15;q11) in a T-cell acute lymphoblastic leukemia cell line activates multiple transcripts, including *Ttg-1*, a gene encoding a potential zinc finger protein. Mol. Cell. Biol. **9:**2124–2132.
49. **McKearn, J. P., M. McCubrey, and B. Fagg.** 1985. Enrichment of hematopoietic precursor cells and cloning of multipotential B-lymphocyte precursors. Proc. Natl. Acad. Sci. USA **82:**7414–7418.
50. **Meier, J. T., and S. M. Lewis.** 1993. P nucleotides in V(D)J recombination: a fine-structure analysis. Mol. Cell. Biol. **13:**1078–1092.
51. **Melchers, F., H. Karasuyama, D. Haasner, S. Bauer, A. Kudo, N. Kakaguchi, J. Jameson, and A. Rolink.** 1993. The surrogate light chain in B cell development. Immunol. Today **14:**60–68.
52. **Menetski, J. P., and M. Gellert.** 1990. V(D)J recombination activity in lymphoid cell lines is increased by agents that elevate cAMP. Proc. Natl. Acad. Sci. USA **87:**9324–9328.
53. **Nakajima, P. B., J. P. Menetski, D. B. Roth, M. Gellert, and M. J. Bosma.** 1995. V-D-J rearrangements at the T cell receptor δ locus in mouse thymocytes of the αβ lineage. Immunity **3:**609–621.
54. **Nakatani, T., K. Horigome, N. Nomura, T. Kondo, H. Ohtsuka, H. Noguchi, and T. Honjo.** 1990. Deletion of human $J_k$ segments by site-specific recombination recognizing the conserved nonamer and heptamer sequences. Nucleic Acids Res. **18:**5529–5532.
55. **Nuñez, C., N. Nishimoto, G. L. Garland, L. G. Billips, P. D. Burrows, H. Kubagawa, and M. D. Cooper.** 1996. B cells are generated throughout life in humans. J. Immunol. **156:**866–872.
56. **Oettinger, M. A.** 1996. Cutting apart V(D)J recombination. Curr. Opin. Genet. Dev. **6:**141–145.
57. **Ohashi, P. S., V. A. Wallace, H. Broughton, C. T. Ohashi, D. A. Ferrick, V. Jost, T. W. Mak, H. Hengartner, and H. Pircher.** 1990. Specific deletion of the J-Cδ locus in murine α/β T cell clones and studies using transgenic mice. Eur. J. Immunol. **20:**517–522.
58. **Ramsden, D. A., J. F. McBlane, D. C. van Gent, and M. Gellert.** 1996. Distinct DNA sequence and structure requirements for the two steps of V(D)J recombination signal cleavage. EMBO J. **15:**3197–3206.
59. **Reth, M., R. Gehrmann, E. Petrac, and P. Wiese.** 1986. A novel $V_H$ to $V_HDJ_H$ joining mechanism in heavy-chain-negative (null) pre-B cells results in heavy-chain production. Nature **322:**840–846.
60. **Roberts, J. L., P. Lauzurica, and M. S. Krangel.** 1997. Developmental regulation of VDJ recombination by the core fragment of the T cell receptor α enhancer. J. Exp. Med. **185:**131–140.
61. **Roman, C. J., and D. Baltimore.** 1996. Genetic evidence that the RAG1 protein directly participates in V(D)J recombination through substrate recognition. Proc. Natl. Acad. Sci. USA **93:**2333–2338.
62. **Roth, D. B., T. N. Porter, and J. H. Wilson.** 1985. Mechanisms of nonhomologous recombination in mammalian cells. Mol. Cell. Biol. **5:**2599–2607.
63. **Sadofsky, M. J., J. E. Hesse, D. C. van Gent, and M. Gellert.** 1995. RAG-1 mutations that affect the target specificity of V(D)J recombination: a possible direct role of RAG-1 in site recognition. Genes Dev. **9:**2193–2199.
64. **Shimizu, T., T. Iwasato, and H. Yamagishi.** 1991. Deletions of immunoglobulin $C_κ$ region characterized by the circular excision products in mouse splenocytes. J. Exp. Med. **173:**1065–1072.
65. **Siminovitch, K. A., A. Bakhshi, P. Goldman, and S. J. Korsmeyer.** 1985. A uniform deleting element mediates the loss of κ genes in human B cells. Nature **316:**260–262.
66. **Spanopoulou, E., F. Zaitseva, F.-H. Wang, S. Santagata, D. Baltimore, and G. Panayotou.** 1996. The homeodomain region of Rag-1 reveals the parallel mechanisms of bacterial and V(D)J recombination. Cell **87:**263–276.
67. **Stallings, R. L., A. F. Ford, D. Nelson, D. C. Torney, C. E. Hildebrand, and R. K. Moyzis.** 1991. Evolution and distribution of (GT)n repetitive sequences in mammalian genomes. Genomics **10:**807–815.
68. **Stanhope-Baker, P., K. M. Hudson, A. L. Shaffer, A. Constantinescu, and M. S. Schlissel.** 1996. Cell type-specific chromatin structure determines the targeting of V(D)J recombinase activity *in vitro*. Cell **85:**887–897.
68a. **Suh, S., and L. Lewis.** Unpublished data.
69. **Takeshita, S., M. Toda, and H. Yamagishi.** 1989. Excision products of the T cell receptor gene support a progressive rearrangement model of the α/δ locus. EMBO J. **8:**3261–3270.
70. **Taki, S., F. Schwenk, and K. Rajewsky.** 1995. Rearrangement of upstream DH and VH genes to a rearranged immunoglobulin variable region gene inserted into the DQ52-JH region of the immunoglobulin heavy chain locus. Eur. J. Immunol. **25:**1888–1896.
71. **Taylor, J. J., D. Rowe, M. M. Reid, and P. G. Middleton.** 1993. An interstitial deletion in the rearranged T-cell receptor g chain locus in a case of T-cell acute lymphoblastic leukemia. Br. J. Haematol. **85:**193–196.
72. **Usuda, S., T. Takemori, M. Matsuoka, T. Shirasawa, K. Yoshida, A. Mori, K. Ishizaka, and H. Sakano.** 1992. Immunoglobulin V gene replacement is caused by the intramolecular DNA deletion mechanism. EMBO J. **11:**611–618.
73. **van Gent, D. C., D. A. Ramsden, and M. Gellert.** 1996. The RAG1 and RAG2 proteins establish the 12/23 rule in V(D)J recombination. Cell **85:**107–113.
74. **Zúñiga-Pflücker, J. C., and M. J. Lenardo.** 1996. Regulation of thymocyte development from immature progenitors. Curr. Opin. Immunol. **8:**215–224.