MOLECULAR AND CELLULAR BIOLOGY, Feb. 1993, p. 1078–1092
0270-7306/93/021078-15$02.00/0

Vol. 13, No. 2

P Nucleotides in V(D)J Recombination: A Fine-Structure Analysis

JOSEPH T. MEIER AND SUSANNA M. LEWIS*

Division of Biology, 156-29, California Institute of Technology, Pasadena, California 91125

Received 26 August 1992/Returned for modification 15 October 1992/Accepted 2 November 1992

Antigen receptor genes acquire junctional inserts upon assembly from their component, germ line-encoded V, D, and J segments. Inserts are generally of random sequence, but a small number of V-D, D-J, or V-J junctions are exceptional. In such junctions, one or two added base pairs inversely repeat the sequence of the abutting germ line DNA. (For example, a gene segment ending AG might acquire an insert beginning with the residues CT upon joining.) It has been proposed that the nonrandom residues, termed "P nucleotides," are a consequence of an obligatory end-modification step in V(D)J recombination. P insertion in normal, unselected V(D)J joining products, however, has not been rigorously established. Here, we use an experimentally manipulable system, isolated from immune selection of any kind, to examine the fine structure of V(D)J junctions formed in wild-type lymphoid cells. Our results, according to statistical tests, show the following. (i) The frequency of P insertion is influenced by the DNA sequence of the joined ends. (ii) P inserts may be longer than two residues in length. (iii) P inserts are associated with coding ends only. Additionally, a systematic survey of published P nucleotide data shows no evidence for variation in P insertion as a function of genetic locus and ontogeny. Together, these analyses establish the generality of the P nucleotide pattern within inserts but do not fully support previous conjectures as to their origin and centrality in the joining reaction.

When immunoglobulin (Ig) and T-cell receptor (TCR) genes are assembled from their germ line components, variable truncation of the coding segments and, often, the introduction of a small number of non-germ line residues occur at the recombinant junctions. These sequence alterations fall within the binding domain of the encoded antigen receptor and constitute an important source of diversity in the immune system. Until fairly recently, it was thought that junctional insertion was essentially random, being both variable in length and unpredictable in sequence. Several groups have documented a correlation between the presence of random N regions and terminal deoxynucleotidyl transferase activity (reference 18 and references cited therein). Nevertheless, some junctions exhibited a recurrent pattern of base addition which, as such, was inconsistent with the N region paradigm. A recent proposal integrated these exceptional cases into a consistent formulation (21). The term "P nucleotide" ("P" for "palindrome") was coined to describe insertions occurring exclusively within a subset of junctions that contain at least one nontruncated coding end. ("Coding end" and related terms are defined in Materials and Methods). In such junctions, one or two residues contiguous to the full-length coding segment might conform to the inverse-complementary sequence of that segment (an example is shown in Fig. 1A, bottom). P nucleotide residues often occur together with N regions within a single junction.

The distinctive character of P inserts prompted a fresh round of speculation about how V(D)J recombination might work. Several explanations for their origin have been offered. An implicit assumption of these models is that P nucleotide addition is the manifestation of an intrinsic, essential aspect of the V(D)J joining reaction, either as a direct consequence of a site-specific cleavage step (27, 31) or a necessary step for subsequent end joining (21, 40). Nevertheless, despite an impressive accumulation of junctional data obtained through rearranging loci in both humans and mice over the last several years, the documentation pertaining to the existence of P nucleotides is fragmentary and is essentially anecdotal. Furthermore, because endogenously generated V(D)J junctions are subjected to several levels of selection within the immune system, the actual frequency of P inserts among junctions when they first arise is difficult to assess. According to the sequences of endogenous junctions, the possibility exists that P nucleotide addition occurs only at certain coding end sequences, at particular loci, or even might be attributable in some cases to immune selection of randomly generated N region inserts. In the present work, we had two aims: to verify the existence of P nucleotides as a general feature of V(D)J junctions and, through their characterization, to make deductions about the joining mechanism itself.

MATERIALS AND METHODS

Terminology. The substrates used throughout these studies contain restriction sites in place of V, D, or J coding segments: for clarity, (and in keeping with previous studies [16, 17, 22–30]) the following terms are defined. "V(D)J joining" is used in a generic sense to indicate recombination events that were mediated by site-specific recognition of 12- and 23-bp spacer joining signals. "Coding" end (designated Sal, Bam, Spe, or Xho) refers to the sequences that abut these 12 and 23 signals prior to recombination (i.e., in the present case, such sequences are noncoding). "Coding joint" refers to the site-specific connection between two coding ends, having the base pair loss and addition characteristic of V(D)J junctions. "Signal joint," as always, refers to the site-specific junction between a 12 and a 23 signal.

Cells and transfections. The cell line 204-1-8 (referred to here as 1-8) is an Abelson murine leukemia virus transformed line derived from adult BALB/c mouse bone marrow. Transfections and tissue-culture conditions were as previously described (26, 29).

Plasmids. pSal-Bam is pJH288 (30). The other three substrates in the series were derived by the substitution of synthetic cassettes containing either the 12- or 23-bp spacer signals for those present in pJH288. In all cases, cassettes

* Corresponding author.

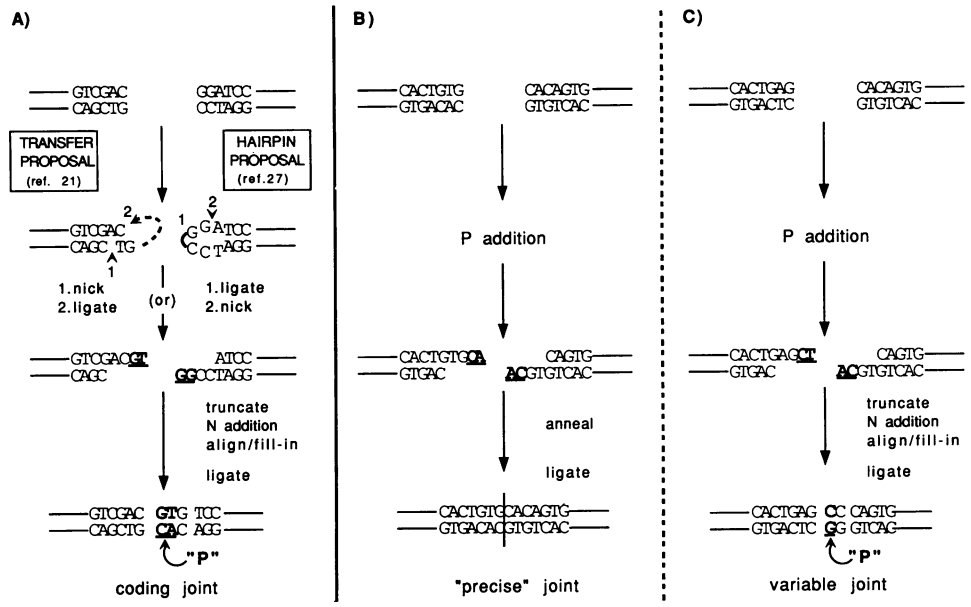

FIG. 1. Hypothetical origin of P nucleotide inserts. (A) P inserts at coding joints. Two models for coding joint formation are shown. The top portion shows coding ends, derived as an example from pSal-Bam (the corresponding signal ends are shown in panel B). Blunt-ended termini are indicated; however, the initial structure of these cleavage products is unknown. The upper middle portion shows P nucleotide addition. To the left is the proposal of Lafaille et al. (21), and to the right is that of Lieber (27). The Lafaille proposition is that a dinucleotide is obligatorily removed from one strand of the coding end and then ligated to the other. The Lieber model suggests that the two strands of the coding end become connected to form a hairpin, after which a nick opens the hairpin up. Both coding termini are proposed to be modified prior to joining. In the lower middle portion, overhang intermediates are resolved, giving rise to P inserts. The ends of the overhangs are altered by an unknown number of activities in an unknown order. Modifications include N insertion, truncation, and filling-in of recessed termini. At the bottom, a product coding joint is shown. (As an example, the sequence shown is that of isolate 10.1.11 from pSal-Bam [Fig. 3].) The junctional insert is offset by spaces on either side: P nucleotides are in boldface type (underlined), while a single N nucleotide is shown in plain type. For simplicity, two aspects of the Lieber model are not explicitly represented; as proposed, the hairpin is created after only one of two strands has been cleaved. Also, after hairpin formation, the location of the nick that opens the hairpin is variable. (For details, see Discussion and reference 27). (B and C) P inserts at signal joints (a case of covert addition?). P nucleotides may figure in the formation of precise signal joints. The hypothetical fate of signal ends (if handled in the same fashion as coding ends) is indicated in panel B. The top portion shows two signal ends after cleavage. The middle portion shows overhang intermediates produced after P addition. Note the complementary relationship between the single-strand protrusions. The bottom portion shows a precise junction formed by direct annealing of the complementary overhangs followed by ligation. The vertical bar represents the apparent crossover site. No P inserts are evident despite active P nucleotide addition. (C) A test of covert P nucleotide addition, showing the predicted fate of noncomplementary signal ends. The top portion shows non-inverse-complementary signal ends (pMut-2,27). The middle portion shows overhang intermediates. Note that the extensions are no longer complementary, and direct annealing is not possible. The bottom portion shows the imprecise outcome. An invented sequence exhibiting truncation and N and P insertion is shown. Panel B is also a representation of the predicted fate of inverse-complementary coding ends found in pDSJ. By this model, the coding joints in pDSJ ought to have a precise structure.

were designed to introduce changes only in the sequences adjacent to the signals, not in the signals themselves. For pSpe-Bam, the sequence TCGACTAGTCACAGTGCTACA GACTGGAACAAAAACCG (*Sal*I-compatible overhangs, *Spe*I site underlined) replaced the 12 signal of pJH288 at its *Sal*I site. Likewise, GATCCTCGAGCACAGTGGTAGTAC TCCACTGTCTGGCTGTACAAAAACCCTCGG (*Bam*HI-compatible overhangs; *Xho*I site underlined) was inserted into the *Bam*HI site of pJH288, to create pSal-Xho. pSpe-Xho was constructed by replacing the 23 signal of pSpe-Bam with the *Xho*I 23 signal cassette as described above.

pDSJ was constructed by replacing the small *Sal*I fragment of pJH200 (16) with that derived from a pJH200 recombinant containing a precise signal joint (kindly provided by J. Hesse, National Institutes of Health). The inserted fragment was oriented with the included chloramphenicol gene in reverse transcriptional orientation relative to the pJH200 *lac* promoter. A precise signal joint in an *Alu*I-*Dde*I fragment originating from recombinant N (24) was then introduced at a *Sal*I site of the intermediate. (Here, as in all following steps, recessed ends were filled in as necessary with the Klenow fragment of DNA polymerase 1.) The signal joint within the *Alu*I-*Dde*I fragment was composed of the joining signals of V_κ21-C and J_κ2 and was positioned as indicated in Fig. 2B. The other signal joint was composed of the joining signals of V_κL-8 and J_κ1 (Fig. 2B). To prevent background chloramphenicol acetyltransferase (CAT) gene expression, the *Cla*I fragment from pJH288, containing the *oop* terminator (28), was introduced at the remaining *Sal*I site 5′ to the inverted CAT gene. The particular combination of signal joints was chosen with care to minimize the possibility that inversional homologous recombination might occur. Such events would produce inversional rearrangement with two apparently precise junctions, confounding the analysis. Thus all four signals have different spacers; additionally, the J_κ2 heptamer has a nonconsensus residue in the fifth position (CACACTG).

pMut-2 was derived from p23 (26) by insertion of a *Sal*I cassette containing a change in the second position of the heptamer (CTCAGTG). The sequence of pMut-2 is otherwise identical to that of p12X23 (26). pMut-27 is identical to pMut-2, except for the absence of a cryptic site called

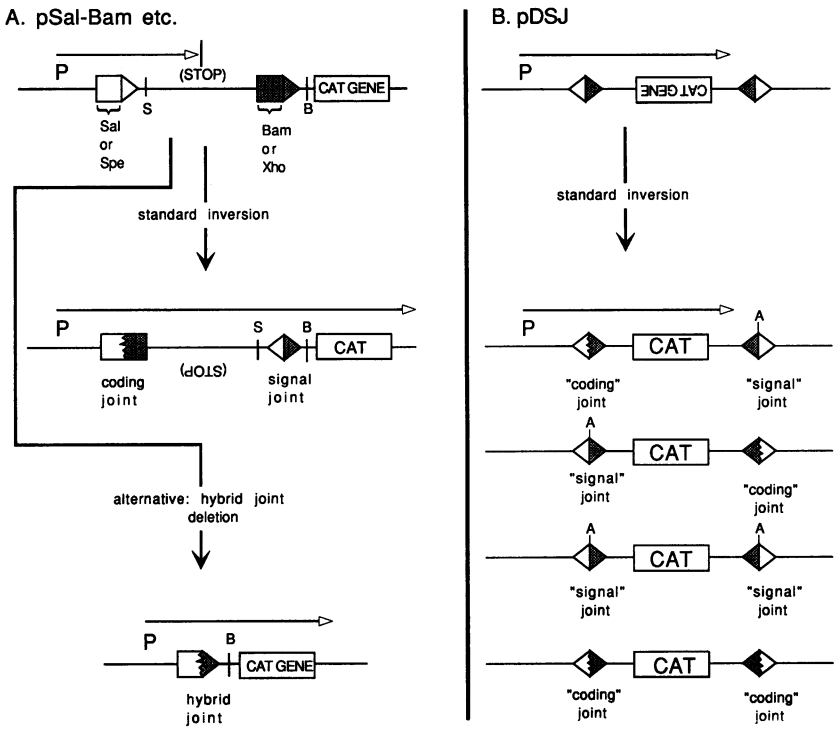

FIG. 2. Constructs and products. Relevant regions of the constructs used in this study are shown, along with possible products. Open boxes represent *Sal*I or *Spe*I sites, open triangles represent 12-bp spacer joining signals. Shaded boxes and triangles represent *Bam*HI or *Xho*I sites and 23-bp spacer signals, respectively. B, S, and A indicate *Bam*HI, *Sal*I, and *Apa*LI sites, respectively. (All sites for these enzymes are shown, except *Apa*LI, which cleaves at several positions in the plasmid outside the illustrated region.) STOP indicates the prokaryotic transcription terminator, and P indicates the promoter that drives CAT gene expression after rearrangement. (A) Structures of products obtained with the related constructs, pSal-Bam, pSpe-Bam, pSal-Xho, pSpe-Xho, pMut-2, and pMut-27. A standard reaction results in an inversion, with a coding joint at one boundary and a signal joint at the other (shown in the middle). A hybrid deletion (shown at the bottom) connects a coding end (here, Sal or Spe) to the 23 signal. Variable coding and hybrid junctions are indicated by jagged lines; the precise signal joints are shown with a straight edges. (B) pDSJ and possible products. pDSJ contains four signals arranged in two signal joints as follows from left to right: the 12 signal from $V_{\kappa 21\text{-}C}$ connected to the 23 signal from $J_{\kappa 2}$ and the 23 signal from $J_{\kappa 1}$ abutting the 12 signal from $V_{\kappa L\text{-}8}$. Four possible outcomes in which one, both, or neither product junction has the typical signal joint structure are possible. Variable and precise junctions are indicated as in panel A. The predicted sensitivity of each class of junction to *Apa*LI digestion is shown.

"6130" (described in reference 26 and the references therein).

Identification of various classes of recombination product. (i) **pSal-Bam.** For pSal-Bam, untrimmed coding joints were identified by the presence of a diagnostic fragment of approximately 245 bp after digestion of DNA samples prepared from Cam^r colonies with *Sal*I and *Bam*HI (the exact size of the fragment varied depending upon the structure of the individual junction); all other classes of recombinants, however, were linearized by this treatment. To determine which of the two coding ends were full-length in the untrimmed junctions thus identified, each was tested with *Bam*HI alone (for the presence of an approximately 330-bp fragment) or *Sal*I alone (for the presence of the approximately 245-bp fragment). Hybrid recombinants were quantified by digesting 76 randomly selected pSal-Bam isolates with *Hgi*AI. The identified hybrids were then digested with *Sal*I alone to look for linearization; this would indicate an untrimmed coding end. The *Hgi*AI digestions also revealed any standard recombinants in the sample that contained base additions or deletions at the signal joint.

(ii) **pSpe-Bam.** pSpe-Bam recombinants were tested with *Spe*I alone to test for linearization, and with *Bam*HI alone (as described above), in order to test for the presence of the approximately 330-bp fragment. Those that were linearized with *Spe*I were subjected to DNA sequence analysis in order to distinguish between standard recombinants and hybrid recombinants containing full-length ends. Those with the *Bam*HI fragment were scored as untrimmed Bam coding joints. Signal joints were analyzed as described above for pSal-Bam by screening a randomly selected sample of 75 pSpe-Bam isolates with the enzyme *Hgi*AI.

(iii) **pSal-Xho.** pSal-Xho recombinants were tested by doubly digesting with *Sal*I and *Xho*I. Those with untrimmed coding joints produced an approximately 245-bp fragment. Candidate recombinants were tested with *Sal*I alone (to test for the presence of the 245-bp fragment) or *Xho*I alone (to test for linearization) to determine which coding ends were full length in each case. Hybrid joints and signal joints were analyzed as described above for pSal-Bam by screening 63 randomly selected pSal-Xho recombinants.

(iv) **pSpe-Xho.** pSpe-Xho recombinants were tested by doubly digesting with *Spe*I or *Xho*I. Linearization indicated the presence of a full-length coding end. Candidates were then digested with *Spe*I alone. DNA sequence analysis of those that were linearized distinguished between standard and hybrid recombinants. Signal joints among 77 randomly selected pSpe-Xho isolates were analyzed as described above for pSal-Bam.

For all of the above, the number of standard inversions

TABLE 1. P nucleotides in coding joints

Summary according to:	No. of transfections/ construct	No. of Cam^r colonies picked	No. of coding joints[a]	No. of untrimmed junctions[b]	% Untrimmed vs total junctions[c]	% P insert-containing junctions vs	
						Total[d]	Untrimmed[e]
Construct							
pSal-Bam	5	141	69	Sal 16 (16)	23	16	73
				Bam 36 (33)	52	8	15
pSpe-Bam	9	262	128	Spe 3 (3)	2	(0.8)	(33)
				Bam 34 (34)	27	9	32
pSal-Xho	9	226	111	Sal 25 (19)	23	15	65
				Xho 42 (34)	38	13	35
pSpe-Xho	9	190	93	Spe 8 (8)	4	3	75
				Xho 52 (37)	27	14	53
End							
Sal			180	41 (35)	23		69
Bam			197	70 (67)	35		24
Spe			221	11 (11)	5		64
Xho			204	94 (71)	46		44

[a] Number of standard recombinants analyzed (estimated as described in Materials and Methods).
[b] Total number of untrimmed coding joints for each of the two ends within a given construct (direct determination). The number of untrimmed ends subjected to DNA sequence analysis is in parentheses.
[c] Truncation index (percentage of untrimmed end among all coding joints).
[d] Percentage of P insert-containing junctions among all coding joints (calculated on the basis of estimate in *a*).
[e] Percentage of P insert-containing junctions among untrimmed junctions (as determined directly from the DNA sequence analysis). Parentheses indicate that a sole insert-containing junction was observed.

(Table 1), as opposed to hybrid deletions and cryptic site 6130 deletions, was determined in each case according to the *Hgi*AI digestions. Fully one third of the total number of recombinants isolated in these experiments were analyzed in this fashion.

(v) pDSJ products. The presence of precise junctions in pDSJ was indicated by digestion with *Apa*LI (or the isoschizomer *Sno*I). However, because imprecise junctions containing two or more P insert residues are likewise *Apa*LI sensitive, unambiguous identification of each of the following recombinant classes required DNA sequence analysis. A digestion pattern of six bands indicated the possibility of two precise junctions. The DNA sequences of all such candidates were determined. A pattern with five bands indicated that one of the two junctions was imprecise (with base added and/or subtracted). Sixteen of 23 such isolates were analyzed at the DNA sequence level. A pattern with four bands indicated that neither junction consisted of two fused, full-length signals without inserts. The DNA sequences of all six examples of this class were determined.

(vi) pMut-2 and pMut-27 recombinants. All pMut-27 recombinants were identified on the basis of Cam^r and diagnostic *Hgi*AI digests. The isolation of the pMut-2 recombinants was carried out as described for p12X23 (26). All chloramphenicol-resistant colonies obtained after transfection and transformation of pMut-2 were picked onto grid arrays in triplicate. Filter lifts were probed with oligonucleotides 1-, 2-, and 23-SIG as described previously (26). DNA was then prepared from candidate recombinants (positive for the first two and negative with 23-SIG) and digested with *Hgi*AI. All isolates with the proper restriction pattern were then sequenced.

Statistics. The probability of obtaining the observed number of P nucleotide inserts (n) or a larger number, on the basis of randomness, was calculated by summing the terms of the binomial distribution from $x = n$ to N, where N is the total number of insert-containing samples examined:

$$\sum_{x=n}^{N} \binom{N}{x} p^x q^{N-x}$$

The expected frequency for a particular nucleotide, p (and thus for the other three, $q = 1 - p$), was first determined by totaling the number of A, C, G, and T residues found as inserts by using the extrachromosomal assay system described here and previously (26, 30). All such junctions (coding, hybrid, or open-and-shut) were oriented so that to the left and right were the coding ends originally associated with the 12- and 23-bp spacer signals, respectively. The inserts within signal joints were tallied with the 23 signal on the right. Accordingly (out of a total of 484 inserted residues scored) the frequencies of A, C, G, and T were 0.16, 0.31, 0.38, and 0.15, respectively. The expected P nucleotide probabilities were then calculated by using these values for p, the value o (observed) shown in Table 2 for n, and likewise N for N. For example, in pSal-Bam, a 2-bp P insert adjacent to the Sal end would have the sequence GT (Table 2, Sal, column 2). On the basis of randomness and independence, the expected frequency for GT is 0.38 × 0.15, or 0.057. As shown in Table 2, there were 10 junctions that had an insert of two or more residues (so that $N = 10$); among these, 6 began with GT (i.e., $n = 6$). Summing the binomial distribution from $n = 6$ to $N = 10$ gives the P value shown in the table.

To specifically evaluate the probability of observing P inserts of a length ≥2 bp as observed in the Sal-end-containing coding joints, the calculation was as follows. The number of P inserts of a length of ≥3 bp (six) was substituted for n and the number of inserts of ≥3 bp that already contained at least two P nucleotides was substituted for N (nine). The value 0.38 (the probability of observing a G at any position within an insert, as described above) was substituted for p and the expansion was solved to give $P \leq 0.03$.

TABLE 2. Statistical analysis of P insert frequencies

Untrimmed end[a]	P insert frequency [o, N (P)][b] at insert length of ≥				
	1 bp	2 bp	3 bp	4 bp	5 bp
Sal to:	G	GT	GTC	GTCG	GTCGA
Bam (pSal-Bam)	11, 12 (<0.0002)	6, 10 ($<10^{-5}$)	4, 9 ($<10^{-4}$)	1, 2 (<0.01)	0, 2
Xho (pSal-Xho)	11, 13 (<0.0008)	4, 7 (<0.0004)	2, 7 (<0.007)	0, 5	0, 4
23 signal (pSal-Bam and pSal-Xho)	5, 7 (<0.08)	2, 4 (<0.02)	0, 1		
Bam to:	C	CC	TCC	ATCC	
Sal (pSal-Bam)	5, 13 (<0.4)	3, 5 (<0.008)	0, 2		
Spe (pSpe-Bam)	12, 18 (<0.002)	5, 10 (<0.002)	2, 6 (<0.003)	0, 3	
Spe to:	A	AC	ACT		
Bam (pSpe-Bam)	1, 1 (<0.2)	0, 1			
Xho (pSpe-Xho)	6, 8 (<0.0004)	2, 4 (<0.02)	0, 3		
12 signal (pSpe-Bam and pSpe-Xho)	3, 4 (<0.02)	1, 1 (<0.05)			
Xho to:	G	AG	GAG		
Sal (pSal-Xho)	12, 20 (<0.04)	2, 8 (<0.08)	0, 4		
Spe (pSpe-Xho)	19, 23 ($<10^{-4}$)	3, 10 (<0.02)	0, 5		
23 signal to:	G	TG			
Sal (pSal-Bam and pSal-Xho)	2, 11 (<0.96)	0, 4			
Spe (pSpe-Bam and pSal-Xho)	0, 7	0, 3			
12 signal (all constructs)	5, 12 (<0.50)	0, 11			
12 signal to 23 signal (all constructs)	C	CA			
	2, 9 (<0.82)	0, 8			
Pooled data for:	G	GT	GTC	GTCG	GTCGA
Sal	27, 32 ($<10^{-7}$)	12, 21 ($<10^{-9}$)	6, 16 ($<10^{-6}$)	1, 7 (<0.05)	0, 2
Bam	C	CC	TCC	ATCC	
	17, 31 (<0.005)	8, 15 ($<10^{-4}$)	2, 8 (0.006)	0, 3	
Spe	A	AC	ACT		
	10, 13 ($<10^{-8}$)	3, 6 (<0.003)	0, 3		
Xho	G	AG	GAG		
	31, 43 ($<10^{-5}$)	5, 18 (<0.004)	0, 9		
23 signal	G	TG			
	7, 30 (<0.97)	0, 18			
12 signal	C	CA			
	2, 9 (<0.82)	0, 8			

[a] Statistical analysis of P insert frequencies for each end is shown by junction. For example, as shown for Sal (top), the untrimmed Sal end was assayed in three different junctions: with the Bam end in pSal-Bam, with the Xho end in pSal-Xho, and with the 23-bp-spacer signal in hybrid joints. The pooled data are pooled results for each of the junctions presented above.

[b] Abbreviations: o, observed number of untrimmed junctions with P inserts of the specified length. The sequence of such inserts is given in each subheading. The absence of an entry indicates there were no junctions containing inserts (P or otherwise) of the specified length. N, number of junctions with inserts (P or otherwise) of the specified length; P, probability associated with each observed value, calculated as described in Materials and Methods.

Analysis of endogenous junctions. The following set of rules were developed so that junctional inserts, and P residues in particular, could be scored without allowing any arbitrary assignments. Assembled collections (1, 2, 4, 9–12, 14, 41) were first checked for data reported in more than one publication. Also, any junction isolated more than once from either a PCR reaction or an individual animal was counted only once. Each junction was examined for full-length ends, for the presence of inserts, and, finally, for the presence of P nucleotides. No end appeared in the final calculations unless the germ line sequence up to and including the joining signal was known. (We note that in some studies, the term "germ line" does not refer to sequences derived from germ line DNA but instead has been used incorrectly to refer to consensus sequences deduced from cDNA.) The only exception was that for the IgH analysis, we accepted two consensus D_h sequences described by Feeney (10) as being highly likely to represent germ line elements.

Each VDJ junction was scored by its parts: 3′V, 5′D, 3′D, and 5′J. No putative D segments shorter than three residues were scored as such. No D-D junction involving a D segment shorter than four residues was scored as such. Junctions were compared with germ line elements and the maximum number of bases was assigned to each end. (Where residues might have been contributed by either of two segments, bases infrequently were assigned to both ends as necessary in order to count the maximum number of full-length ends in each case.) Where three or more D segment residues were

present but could represent either an internal or terminal D segment sequence, they were scored every time as a terminal fragment. We accepted single-base-pair mismatches no closer than 2 bp from a putative segment terminus. If the mismatch occurred at the penultimate base, both the mismatched base and the one following were scored as junctional inserts.

DNA sequence analysis. DNA sequence analysis was carried out with a Sequenase version 2 kit (U.S. Biochemical). The Lac-1 oligonucleotide (25) was used as a primer for sequencing the coding joints in pSal-Bam, in its derivatives and in pMut-2 and pMut-27. Hybrid joints were sequenced with Lac-1 or JH33 (26). The oligonucleotides JH33 or Ter-1 (26) were used to sequence signal joints. For pDSJ, right junctions were analyzed with the Ter-1 primer; the left junctions were sequenced with the Lac-1 oligonucleotide.

RESULTS

P nucleotides are added to coding ends. To investigate P insertion at coding ends, we used the extrachromosomal plasmid assay developed by Hesse et al. (16, 29). Briefly, the assay entails transfection of recombination substrates into a pre-B-like cell line, 1-8, that is active for V(D)J joining. During residence in the 1-8 cells, some of the transfected molecules become rearranged in a site-specific manner (Fig. 2A). This is detected by reisolating the plasmid DNA from transfectants about 48 h later and using it to transform *Escherichia coli* cells. Recombined molecules confer chloramphenicol resistance to *E. coli* due to activation of CAT expression (Fig. 2A).

This approach has several important features. By making use of plasmid substrates that contain minimal recombination recognition sites, and no locus-specific sequences (16), we can look at junction products in the absence of locus-specific influences. Furthermore, all recombinants are generated in a single clonal cell line, providing a standardized cellular context with which to compare results between experiments. Finally, because recombined molecules are not isolated until after they are introduced into *E. coli* and because the recombinant junctions (coding and signal joints) are extraneous to the coding sequences of the selectable marker (see Fig. 2A), we can expect to analyze the full diversity of junctions formed by the V(D)J joining machinery without the selective bias that occurs within an intact immune system.

By this approach, we tested whether the alteration of the sequence of a coding end would alter the identity of the associated junctional inserts in a pattern predicted by the P nucleotide theory. To a first approximation, if P addition is an integral part of the joining process, then P inserts would be expected to arise at statistically significant frequencies regardless of the sequence of a given coding end.

We generated a series of four closely related recombination substrates that differed only at their coding ends (as defined in Materials and Methods). In the parental construct pJH288 (30) and in our three variants (pSpe-Bam, pSal-Xho, and pSpe-Xho), restriction enzyme recognition sequences were located next to the 12- and 23-bp spacer joining signals, in positions equivalent to native V, D, or J coding elements. This design facilitated the later isolation of nontruncated junctions. As their names indicate, pSal-Bam and the three derivatives had either *Sal*I or *Spe*I recognition sites abutting their 12 signals and either *Bam*HI or *Xho*I sites adjoining their 23 signals.

After transfection, DNA samples were prepared from over 800 chloramphenicol-resistant recombinants and screened for the presence of full-length coding ends by digestion with the appropriate restriction enzymes. Representative numbers of untrimmed junctions (those containing at least one nontruncated coding end) were then subjected to DNA sequence analysis. A general summary of the experiment and our results is given in Table 1; details of the method of analysis are provided in Materials and Methods.

Upon recombination, each of the four coding ends was found to have acquired P nucleotide additions in at least some instances. DNA sequences are shown in Fig. 3. For each coding joint, residues that can be attributed to either of the input ends were so assigned (to the left or right), and those residues that do not correspond to either end, as shown in the middle, were scored as inserts (for details, see legend to Fig. 3). P nucleotides were noted as indicated (in boldface type and underlined).

We wished to rule out the possibility that some or all putative P inserts might in fact simply be N regions with a fortuitous inverse-complementary match to the coding end. To do so, we applied a statistical test. The probabilities of obtaining the observed frequencies of P nucleotide inserts through random N insertion were calculated according to the binomial distribution (Table 2, Sal, Bam, Spe, Xho, and 23 signal; details provided in Materials and Methods). Each end was evaluated, and in each case, P inserts of various lengths were considered.

As summarized in Table 2 (pooled data), all four coding ends gave rise to junctions containing P inserts. In each instance, the match between observed inserts and P nucleotide sequences, as specified by the coding end, was highly significant: $P < 0.01$. Thus, the observed P inserts were clearly distinct from N addition.

Several other properties of P insertion were established by the statistical analysis. The first was that P nucleotides were associated with some coding ends more often than with others (Figure 3; Table 2). For example, overall, the P value for P inserts at Sal coding ends was 10^{-7}, while that for the Bam coding end was 0.005 (Table 2, pooled data). Differences were evident when various endwise combinations were tested as well as for P inserts of various lengths (Table 2, Sal and Bam). The second observation was that P nucleotide inserts were short or long depending upon the particular coding end involved. For example, at the Sal end, 6 of 27 P inserts were three or more residues in length, whereas this was true of none of the P inserts (31 total) characterized at the Xho end (Table 2, pooled data, compare the numbers for P inserts with lengths of ≥ 3 bp with those of P inserts with lengths of ≥ 1 bp in each case). A third, related observation is that for the Sal end in particular, a significant number of the inserts were >2 bp in length. We considered the possibility that 3-bp P inserts were actually 2-bp P inserts that happened by chance to appear longer because of the addition of N region sequence. The probability of obtaining a third P residue at the observed frequency through random addition ($P \leq 0.03$; calculated as described in Materials and Methods) indicated that P inserts longer than two residues were not created by random N addition.

P nucleotides are not detected at signal ends. Lafaille et al. (21), in accordance with earlier observations by McCormack et al. (31), noted that P inserts appeared adjacent to coding ends but not signal ends and concluded that the P addition mechanism operated only upon the former. In the present study, a large number of junctions that incorporate signal ends were assayed for P inserts. These were of two kinds: signal joints and hybrid joints. Signal joints are reciprocally

A)	Sal		Bam
pSal-Bam	TGCAGGTCGAC		GGATCCTCTCA
4.1.20	TGCAGGTCGAC		CCTCTCA
5.1.19	TGCAGGTCGAC		CCTCTCA
5.1.17	TGCAGGTCGAC	**GTCG**G	CCTCTCA
8.7.1	TGCAGGTCGAC	**G**	TCCTCTCA
10.1.11	TGCAGGTCGAC	**GT**G	TCCTCTCA
8.3.5	TGCAGGTCGAC	**GTC**	ATCCTCTCA
8.5.10	TGCAGGTCGAC	**GTC**CC	ATCCTCTCA
11.1.1	TGCAGGTCGAC		GATCCTCTCA
10.1.9	TGCAGGTCGAC		GATCCTCTCA
8.6.4	TGCAGGTCGAC	**GT**	GATCCTCTCA
8.1.1	TGCAGGTCGAC	**G**CT	GATCCTCTCA
8.2.9	TGCAGGTCGAC	**GTC**	GATCCTCTCA
10.1.6	TGCAGGTCGAC	**G**GA	GATCCTCTCA
8.7.11	TGCAGGTCGAC	**G**	GGATCCTCTCA
8.7.12	TGCAGGTCGAC	**CC**	GGATCCTCTCA
5.1.33	TGCAGGTCGA		GGATCCTCTCA
8.1.13	TGCAGGTCG		GGATCCTCTCA
4.1.24	TGCAGGTC		GGATCCTCTCA
5.1.28	TGCAGGTC		GGATCCTCTCA
8.1.5	TGCAGGTC		GGATCCTCTCA
10.1.3	TGCAGGTC		GGATCCTCTCA
5.1.32	TGCAGGT		GGATCCTCTCA
8.6.2	TGCAGGT		GGATCCTCTCA
8.6.8	TGCAGGT	AA	GGATCCTCTCA
8.1.4	TGCAGG		GGATCCTCTCA
4.1.60	TGCAGG		GGATCCTCTCA
8.1.11	TGCAGG	G	GGATCCTCTCA
8.7.9	TGCAGG	**CC**	GGATCCTCTCA
4.1.75	TGCAGG	GGG	GGATCCTCTCA
8.1.10	TGCAG		GGATCCTCTCA
5.1.36	TGCAG	T	GGATCCTCTCA
5.1.18	TGCA		GGATCCTCTCA
8.8.12	TGC	**C**	GGATCCTCTCA
10.1.4	TGC	**C**	GGATCCTCTCA
10.1.8	TGC	**CC**	GGATCCTCTCA

B)	Spe		Bam
pSpe-Bam	GGTCGACTAGT		GGATCCTCTCA
34.2.6	GGTCGACTAGT		CCTCTCA
35.4.6	GGTCGACTAGT		ATCCTCTCA
32.3.6	GGTCGACTAGT	**A**A	GATCCTCTCA
42.1.2	GGTCGACTAG	G	GGATCCTCTCA
42.2.12	GGTCGACTAG	**C**	GGATCCTCTCA
32.3.7	GGTCGACTA		GGATCCTCTCA
34.2.3	GGTCGACTA		GGATCCTCTCA
35.4.3	GGTCGACTA		GGATCCTCTCA
45.4.3	GGTCGACTA		GGATCCTCTCA
33.3.11	GGTCGACTA	**C**	GGATCCTCTCA
43.1.2	GGTCGACTA	CG	GGATCCTCTCA
34.3.12	GGTCGACT		GGATCCTCTCA
45.2.7	GGTCGACT	CC**CC**	GGATCCTCTCA
32.3.5	GGTCGAC	CA	GGATCCTCTCA
18.1.1	GGTCGAC	GTA	GGATCCTCTCA
33.3.10	GGTCGAC		GGATCCTCTCA
44.5.1	GGTCGAC		GGATCCTCTCA
45.2.1	GGTCGAC		GGATCCTCTCA
33.3.5	GGTCGAC	**C**	GGATCCTCTCA
43.2.6	GGTCGAC	**C**	GGATCCTCTCA
44.2.12	GGTCGAC	CAA	GGATCCTCTCA
42.5.1	GGTCGA	**TCC**	GGATCCTCTCA
33.2.11	GGTCGA	T**TCC**	GGATCCTCTCA
45.2.9	GGTGG		GGATCCTCTCA
32.2.4	GGTCG	T**C**	GGATCCTCTCA
32.2.5	GGTCG	**CC**	GGATCCTCTCA
33.2.1	GGTCG	CC**CC**	GGATCCTCTCA
33.3.6	GGTC		GGATCCTCTCA
44.1.4	GGTC		GGATCCTCTCA
33.2.6	GG		GGATCCTCTCA
44.5.4	-11(A)		GGATCCTCTCA
42.5.5	-11(A)	**C**	GGATCCTCTCA
35.3.4	-16(G)	T	GGATCCTCTCA
27.1.10	-17(G)		GGATCCTCTCA

C)	Sal		Xho
pSal-Xho	TGCAGGTCGAC		CTCGAGGATCC
28.3.11	TGCAGGTCGAC		GGATCC
38.4.8	TGCAGGTCGAC		GGATCC
73.1.5	TGCAGGTCGAC	**GT**AA	GAGGATCC
28.2.17	TGCAGGTCGAC		GAGGATCC
29.4.5	TGCAGGTCGAC		GAGGATCC
72.1.7	TGCAGGTCGAC	**G**	CGAGGATCC
73.1.7	TGCAGGTCGAC	**G**GG	CGAGGATCC
64.1.1	TGCAGGTCGAC	**GTC**TC	CGAGGATCC
65.1.6	TGCAGGTCGAC	**G**	TCGAGGATCC
28.2.12	TGCAGGTCGAC	**G**	TCGAGGATCC
73.1.1	TGCAGGTCGAC		CTCGAGGATCC
28.2.8	TGCAGGTCGAC	**G**	CTCGAGGATCC
72.2.6	TGCAGGTCGAC	**G**	CTCGAGGATCC
73.2.5	TGCAGGTCGAC	**GTC/AG**	CTCGAGGATCC
72.1.3	TGCAGGTCGAC	**GT**GAG**G**	CTCGAGGATCC
72.1.1	TGCAGGTCGAC	CC**G**	CTCGAGGATCC
73.2.7	TGCAGGTCGAC	TGAGA**AG**	CTCGAGGATCC
22.1.4	TGCAGGTCGA	**G**	CTCGAGGATCC
28.2.15	TGCAGGTCG		CTCGAGGATCC
29.4.11	TGCAGGTCG		CTCGAGGATCC
29.2.10	TGCAGGTCG	**G**	CTCGAGGATCC
64.1.4	TGCAGGTCG	CT	CTCGAGGATCC
23.2.8	TGCAGGTC	CC	CTCGAGGATCC
29.3.8	TGCAGGT	A	CTCGAGGATCC
28.4.9	TGCAGGT	**G**	CTCGAGGATCC
38.2.12	TGCAGGT		CTCGAGGATCC
38.3.2	TGCAGG	**G**	CTCGAGGATCC
38.3.5	TGCAGG		CTCGAGGATCC
28.3.1	TGCAG		CTCGAGGATCC
29.2.17	TGCAG		CTCGAGGATCC
38.3.7	TGCAG		CTCGAGGATCC
64.1.6	TGCAG		CTCGAGGATCC
28.4.6	TGCAG	C	CTCGAGGATCC
23.2.2	TGCA	C	CTCGAGGATCC
23.2.9	TGCA	A	CTCGAGGATCC
28.2.19	TG		CTCGAGGATCC

D)	Spe		Xho
pSpe-Xho	GGTCGACTAGT		CTCGAGGATCC
62.2.2	GGTCGACTAGT		AGGATCC
47.1.4	GGTCGACTAGT	**AC**	AGGATCC
49.3.1	GGTCGACTAGT	**A**	CGAGGATCC
63.5.1	GGTCGACTAGT	**A**	CGAGGATCC
49.1.5	GGTCGACTAGT	**AC**C	CGAGGATCC
63.2.9	GGTCGACTAGT	**A**	CTCGAGGATCC
46.1.2	GGTCGACTAGT	CG**G**	CTCGAGGATCC
47.1.3	GGTCGACTAG	**AG**	CTCGAGGATCC
63.2.10	GGTCGACTAG	**AG**	CTCGAGGATCC
49.1.2	GGTCGACTAG	AATC	CTCGAGGATCC
62.5.9	GGTCGACTAG		CTCGAGGATCC
63.2.12	GGTCGACTAG		CTCGAGGATCC
53.1.7	GGTCGACT	**G**	CTCGAGGATCC
62.2.1	GGTCGACT	**G**	CTCGAGGATCC
63.2.5	GGTCGACT	**G**	CTCGAGGATCC
63.3.4	GGTCGACT	C**G**	CTCGAGGATCC
49.3.6	GGTCGAC	CC**G**	CTCGAGGATCC
62.2.12	GGTCGA	**G**	CTCGAGGATCC
63.2.4	GGTCGA	**G**	CTCGAGGATCC
47.1.5	GGTCGA		CTCGAGGATCC
48.1.1	GGTCGA		CTCGAGGATCC
62.2.6	GGTCGA		CTCGAGGATCC
63.4.4	GGTCGA		CTCGAGGATCC
48.1.4	GGTCG	CCCAT	CTCGAGGATCC
62.4.11	GGTCG	**G**	CTCGAGGATCC
63.4.2	GG	AGA	CTCGAGGATCC
63.4.8	GG	**AG**	CTCGAGGATCC
62.4.5	G		CTCGAGGATCC

FIG. 3. P Inserts within untrimmed coding joints. (A through D) Products isolated with pSal-Bam, pSpe-Bam, pSal-Xho, and pSpe-Xho, as indicated. The first line in each panel shows the sequences of the two relevant, nontruncated coding ends. The recombinants containing untrimmed Sal or Spe coding ends are shown at the top portion of each panel, those in which both coding ends are intact are grouped in the middle, and those with intact Bam or Xho coding ends are shown at the bottom. For consistency, where a residue might be assigned to either of the two coding ends it was assigned to the untrimmed coding end. Inserted residues are shown in the center; for each junction, P inserts are underlined and in boldface and N residues are in regular typeface. All recombinants listed are independent isolates.

related to coding joints (23) and can be isolated (as in the present study) with substrates in which recombination sites are oriented so as to promote inversional rearrangement (e.g., Fig. 2). In a signal joint, a 12-bp spacer signal and a 23-bp spacer signal are fused, but the ends are almost always connected without truncation and/or base addition. Signal joint formation thus creates an *Hgi*AI (*Sno*I or *Apa*LI) recognition site (30). This feature facilitates identification of the rare signal joint with either an insert and/or loss of residues. The other signal-containing junction mentioned above, a hybrid joint, is an alternative V(D)J joining product in which coding-to-signal end fusions occur (22 and references cited therein; 25). With the present constructs, one of two possible hybrid joint conformations is recoverable; that in which a coding end (represented by the Sal or Spe end, as the case may be) has become connected to a 23 signal (Fig. 2A). The restriction pattern of these recombinants obtained by *Hgi*AI digestion is also distinctive.

A collection of imprecise signal joints as well as a large number of hybrid junctions were identified by digesting approximately 300 of the DNA samples with *Hgi*AI. All independent isolates that contained at least one untrimmed end according to DNA sequence analysis are shown in Fig. 4. The results of our statistical analysis were unambiguous. P inserts were absent from signal ends, whether such ends occurred within a signal joint or a hybrid joint. The *P* values associated with the few apparent P nucleotides that were observed in each case were 0.8 and 0.5, respectively (Table 2, 23 signal and 12 signal). P inserts were demonstrated, however, at the coding ends of hybrid joints (Fig. 4A; Table 2, Sal and Spe [$P \leq 0.02$ for all cases]). These results provide statistical support for the coding end specificity of P inserts noted earlier (21, 31).

Is evidence of P addition hidden at signal joints? On the basis of the above analysis, one might conclude that the P insertion mechanism only modifies coding ends. However, when considered in the context of proposed models for P insertion (21, 27), P nucleotides might well be added to signal ends and yet be obscured within signal joints.

Two models have been put forward to account for the presence of P nucleotides. In one model (21), a dinucleotide is removed from one strand of the cut coding end and transferred to the other strand (Fig. 1A, left). In a second model (27), the two strands of the coding terminus are first joined to one another to form a sealed hairpin, after which the hairpin is opened by nicking one strand at a position a few bases pairs in from the end (Fig. 1A, right). The proposed order of single-strand nicks and ligations differ between these two models, but both posit an intermediate structure with staggered ends. In this intermediate, one strand has been extended at the expense of the other (Fig. 1A, middle). Thereafter, according to either model, the staggered termini meet with varying fates: when ligated directly, P nucleotides appear as junctional inserts appended to full-length ends (Fig. 1A, bottom), but more often other operations intervene which remove the added bases before end-joining can occur.

The 12- and 23-bp spacer signal ends, which are palindromic, bear an inverse-complementary relationship to one another as oriented for joining. By either of the above proposals, P nucleotide addition at signal ends would generate a pair of termini with complementary single-stranded overhangs (Fig. 1B). Annealing of complementary overhangs [arguably a favored event in V(D)J joining] (14) followed by sealing of two single-strand nicks would reproducibly create a junction without insertion or truncation at

A)

	Sal		23 signal
	TGCAGGTCGAC		CACAGTGGTAG
29.6.5	TGCAGGTCGAC	**G**G	G
26.2.4	TGCAGGTCGAC		AGTGGTAG
29.6.9	TGCAGGTCGAC		CACAGTGGTAG
38.4.5	TGCAGGTCGAC		CACAGTGGTAG
28.2.10	TGCAGGTCGAC	**G**	CACAGTGGTAG
29.5.6	TGCAGGTCGAC	A	CACAGTGGTAG
29.6.2	TGCAGGTCGAC	C	CACAGTGGTAG
28.5.12	TGCAGGTCGAC	**G**A	CACAGTGGTAG
29.4.12	TGCAGGTCGAC	**GT**	CACAGTGGTAG
72.1.5	TGCAGGTCGAC	**GT**GGAA	CACAGTGGTAG
28.5.5	TGCAGGTCGA		CACAGTGGTAG
29.4.3	TGCAGGTCGA		CACAGTGGTAG
29.2.5	TGCAGGTCG		CACAGTGGTAG
64.1.2	TGCAGGTCG		CACAGTGGTAG
28.5.1	TGCAGGTCG	T	CACAGTGGTAG
8.1.6	TGCAGGTCG	GG**G**	CACAGTGGTAG
8.2.1	TGCAGGTCG	CA	CACAGTGGTAG
4.1.18	TGCAGGT	AGGA	CACAGTGGTAG
4.1.41	TGCAGGT		CACAGTGGTAG
8.6.5	TGCAG	TC	CACAGTGGTAG

	Spe		23 signal
	GGTCGACTAGT		CACAGTGGTAG
63.4.11	GGTCGACTAGT	C	CACAGTGGTAG
63.3.7	GGTCGACTAGT		CACAGTGGTAG
46.1.4	GGTCGACTAGT	**AC**	CACAGTGGTAG
49.3.2	GGTCGACTAGT	**A**	CACAGTGGTAG
53.1.12	GGTCGACTAG	G**G**	CACAGTGGTAG
44.2.1	GGTCGACTAG		CACAGTGGTAG
44.4.6	GGTCGACT	CCC	CACAGTGGTAG
45.3.12	GGTCGACT		CACAGTGGTAG
62.4.7	GGTCGACT	A	CACAGTGGTAG
62.2.8	GGTCGACT		CACAGTGGTAG
48.3.6	GGTCGAC	GGA	CACAGTGGTAG
62.2.3	GGTCGAC		CACAGTGGTAG
27.2.1	GGTCGAC		CACAGTGGTAG
27.1.4	GGTCGA		CACAGTGGTAG
32.2.6	GGTCGA		CACAGTGGTAG
62.4.1	GGTCGA		CACAGTGGTAG
32.2.10	GGTCG		CACAGTGGTAG
64.1.2	GGTCG		CACAGTGGTAG
32.3.9	-13(G)	TA	CACAGTGGTAG

B)

	12 signal		23 signal
	GTAGCACTGTG		CACAGTGGTAG
43.2.9	GTAGCACTGTG	GGTC	CACAGTGGTAG
5.1.2	GTAGCACTGTG	GTT	CACAGTGGTAG
8.1.3	GTAGCACTGTG	TGGA	CACAGTGGTAG
8.1.8	GTAGCACTGTG	T	CACAGTGGTAG
8.5.8	GTAGCACTGTG	AGG**G**	CACAGTGGTAG
8.5.9	GTAGCACTGTG	GG**G**	CACAGTGGTAG
8.8.7	GTAGCACTGTG	**C**C	CACAGTGGTAG
29.4.9	GTAGCACTGTG	**C**C	CACAGTGGTAG
27.1.5	GTAGCACTGT		CACAGTGGTAG
43.4.3	GTAGCACTG		CACAGTGGTAG
5.1.7	GTAGCACTG	C**G**	CACAGTGGTAG
63.3.5	GTAGC		CACAGTGGTAG
28.5.3	-59(G)		CACAGTGGTAG

FIG. 4. P inserts within hybrid and signal joints. (A) Top portion shows hybrids in which the Sal coding end is connected to the 23-bp spacer signal. Products were isolated from pSal-Bam and pSal-Xho. Bottom portion shows hybrids in which the Spe end is joined to the 23 signal. Products were isolated from pSpe-Bam and pSpe-Xho. (B) Signal joints exhibiting base loss or addition. All details are as described in the legend to Fig. 3.

the crossover site. The precise stereotyped structure typical of signal joints therefore might in fact arise as a consequence of P nucleotide addition (Fig. 1A and B), but P inserts would be consistently absent from these junctions.

We tested for covert P addition by manipulating the degree of complementarity between joined ends. In one experiment, we decreased the complementarity between signal ends; in a second, we increased it at coding ends. Reduction of the base-pairing potential between proposed P-overhang intermediates (Fig. 1C) should radically alter the characteristic precision of signal joints, if they are usually formed by the pathway shown in Fig. 1B. Likewise, by creating the opportunity for annealing between the overhang intermediates of coding ends (by substituting signal sequences for coding ends), we might detect a corresponding increase in precision in the coding joint connections, as shown in Fig. 1B.

Substrates (pMut-2 and pMut-27, Fig. 5A) in which base-pairing of putative intermediate structures is disrupted were constructed in the course of a separate study (25a). The pertinent difference between these plasmids and pSal-Bam was the alteration of the second base of the 12-bp-spacer signal heptamer from CACTGTG to CACTGAG (Fig. 1C). The 1-bp change eliminates two out of four possible base-pairing interactions upon attempted annealing of the postulated overhang intermediates (Fig. 1C, middle).

The DNA sequences of all verifiably independent isolates recovered upon transfection of pMut-2 and pMut-27 are shown in Fig. 5A. To test our hypothesis, it was necessary to introduce the base pair change within 2 bp of the heptamer border, placing the alteration within a region of the heptamer that is critical for joining signal function (17). As a result, only one or two (sometimes no) recombinants were recovered per transfection. Despite the severe depression in recombination frequency and the loss of potential base-pairing between postulated P nucleotide-modified intermediates, there were no examples in which the signal ends were truncated. The mutant joining signal eliminated efficient recombination without causing the signal joints to acquire coding joint character.

In a reciprocal experiment, a plasmid, pDSJ, was constructed in which the two coding ends were replaced by signal ends. In this four-signal plasmid (Fig. 2B), both junctions arising from V(D)J joining (not just the signal joint) are created from ends that possess inverse complementarity. We wished to determine whether the coding joints resulting from pDSJ rearrangement would assume a signal joint-like structure lacking N regions, overt signs of P addition in the form of P inserts, and evidence of end truncation. pDSJ was transfected into 1-8 cells, and analyzed by a diagnostic *Apa*LI digestion as detailed in Materials and Methods. The DNA sequences of 26 of a total of 32 isolates were determined (Fig. 5B).

Because of the symmetry of the pDSJ substrate, we could not anticipate whether the right junction or the left junction would become the coding joint or whether, in fact, two types of junction would be distinguishable. What we found was that in almost all isolates, one junction had been processed as a coding joint while the other had the characteristics of a signal joint. To be specific, the fine structure of all of the left junctions shown above the dashed line in Fig. 5B was consistent with that of signal joints; all contained two untrimmed signal ends (with occasional base inserts). The corresponding right junctions all exhibited base loss and/or addition, as is typical of coding joints. Below the line in Fig. 5B are isolates in which these identities were reversed, with the left junctions corresponding to coding joints and the right junctions appearing to be the signal joints. In one recombinant (D57; below the solid line, bottom), neither junction contained two untrimmed ends.

Thus, in this experiment, the joining reaction displayed a fundamental asymmetry despite the symmetry of the input substrate. There were no examples of a pDSJ recombinant in which both junctions were precisely joined (i.e., with neither base loss nor addition). The two isolates, D26 and D70, in which truncation had not occurred had, in each case, acquired a P insert within one of their two product junctions. Thus, the D26 and D70 junctions could not have formed via the pathway in Fig. 1B because, had this been the case, P residues, as inserts, should not have been observed.

Taken together, these experiments are consistent with a P addition process that targets coding rather than signal ends during V(D)J joining.

P nucleotides at endogenous loci. To complement our studies with artificial substrates, we wished to determine whether any consistent variation in the frequency of P inserts, e.g., between loci or during development, would come to light upon examination of endogenously generated V(D)J joining products. In the case of N region insertion, an absence early in ontogeny gives way to a steady increases during development, suggestive of developmental regulation of Tdt activity (1, 2, 10, 11, 14, 32). Possibly, the incidence of P inserts might also fluctuate in some informative pattern. Furthermore, it was of interest to establish the frequency with which P inserts longer than two residues are observed in the physiological context.

A preliminary analysis focused on murine light-chain genes because rearranged kappa and lambda genes usually lack junctional inserts of any kind, N or P. In the collection of Kabat et al. (18), we counted 77 murine light-chain junctions (disregarding redundant listings of somatic variants) that are derived from known germ line V and J segments and that contain an untrimmed end. Not one contained a P insert. However, this absence of P insertion was not consistent with the fine structures of kappa gene recombinants present upon the circular DNA molecules excised by repeated rounds of $V_\kappa J_\kappa$ joining. While the Kabat collection is largely limited to expressed, functional, light-chain genes, excision products contain coding joints that have presumably been passed over by positive selection. Among 16 excision products analyzed by Harada and Yamagishi (15), one of two untrimmed V segments was appended by a 3-bp P insert, and one of six untrimmed J segments had a 2-bp P insert. One- and 2-bp P inserts have also been found in human kappa light-chain gene junctions (33). These data strongly suggest that $V_\kappa J_\kappa$ junctions containing P inserts are generated with some regularity at the light-chain locus (as reflected by excision products) but that inserts are subsequently counterselected in some unknown fashion. Selection against N regions may be a general problem in the analysis of endogenous data, as has been suggested to occur at other loci as well (2, 5).

Without, as was hoped, finding a situation in which P inserts are clearly absent, we turned to an in-depth analysis of two loci, the TCRβ locus and the IgH locus. In both cases, a large number of precursor germ line sequences had been fully defined, and the junction data were extensive enough to permit comparisons between populations grouped according to several different parameters. We surveyed eight large-scale collections of IgH and TCRβ junctions. In two cases (1, 4), unpublished and/or formatted sequences were generously provided by the authors.

P nucleotide addition was of peripheral concern in some of

A)

	CODING JOINT			SIGNAL JOINT		
	CTGCAGGTCGA		GATCCTCTCAT	GTACCACTGAG		CACAGTGATCC
241	CTGCAGGTCGA		TCCTCTCAT	GTACCACTGAG		CACAGTGATCC
027-6	CTGCAGGTCGA		TCCTCTCAT	GTACCACTGAG	**G**	CACAGTGATCC
027-10	CTGCAGGTCGA	AA	CCTCTCAT	GTACCACTGAG	T	CACAGTGATCC
LC20-3	CTGCAGGTCGA		TCCTCTCAT	GTACCACTGAG		CACAGTGATCC

B)

	LEFT JUNCTION			RIGHT JUNCTION		
	GGAGCACTGTG		CACAGTGGTAG	ACACCAGTGTG		CACAGTGCTAC
D1	GGAGCACTGTG		CACAGTGGTAG	ACACCAGT	A	AGTGCTAC
D4	GGAGCACTGTG		CACAGTGGTAG	ACACCAGTGT	CGG	TAC
D5	GGAGCACTGTG		CACAGTGGTAG	ACACCAGTGTG		CTAC
D7	GGAGCACTGTG		CACAGTGGTAG	ACAC		CAGTGCTAC
D22	GGAGCACTGTG		CACAGTGGTAG	ACACCAGTGTG	G	GTGCTAC
D26	GGAGCACTGTG		CACAGTGGTAG	ACACCAGTGTG	**CA**	CACAGTGCTAC
D60	GGAGCACTGTG		CACAGTGGTAG	ACACCAGT		GTGCTAC
D61	GGAGCACTGTG		CACAGTGGTAG	ACACCA	GAACA	CACAGTGCTAC
D66	GGAGCACTGTG		CACAGTGGTAG	ACAC	G	CAGTGCTAC
D68	GGAGCACTGTG	**G**	CACAGTGGTAG	ACACCAG	A	GCTAC
D67	GGAGCACTGTG	ACTT	CACAGTGGTAG	ACAC		CAGTGCTAC
D70	GGAGCACTGTG		CACAGTGGTAG	ACACCAGTGTG	**C**CT	CACAGTGCTAC
D72	GGAGCACTGTG		CACAGTGGTAG	ACACCAGTGT		-11(A)
D78	GGAGCACTGTG		CACAGTGGTAG	ACACC	C	GCTAC
D80	GGAGCACTGTG		CACAGTGGTAG	ACACCAGTGTG	**CAC**TG	C
D82	GGAGCACTGTG	G**G**	CACAGTGGTAG	ACACCAG	A	-11(A)
D84	GGAGCACTGTG		CACAGTGGTAG	ACACCAGTGTG		CAGTGCTAC
D86	GGAGCACTGTG	AGGGC	CACAGTGGTAG	ACAC		CAGTGCTAC
D88	GGAGCACTGTG		CACAGTGGTAG	ACAC		CAGTGCTAC
D65	GGAGC	**TGTG**	CACAGTGGTAG	ACACCAGTGTG	**C**C	CACAGTGCTAC
D89	GGAGC	**G**	CACAGTGGTAG	ACACCAGTGTG	**G**	CACAGTGCTAC
D57	-12(C)		AG	ACAC	T	GTGCTAC

FIG. 5. Analysis of P inserts: complementary versus noncomplementary signal ends. (A) Sequences of pMut-2 and pMut-27 recombinants. (B) Sequences of pDSJ recombinants. The first line in each panel gives the sequence of the untrimmed end comprising the junctions in each construct. Ambiguous residues were assigned as in Fig. 3. Inserts and P nucleotides are designated as described in the legend to Fig. 3. Recombinants with the same sequence were listed only if derived from independent transfections.

TABLE 3. P Nucleotides within endogenously generated junctions

Locus[a]	No. of ends analyzed	No. of untrimmed ends[b]	No. of untrimmed junctions with inserts	No. of full-length ends with P nucleotides	No. of full-length ends with P nucleotides according to P insert length of:				% of untrimmed total ends[c]	% of P inserts among:	
					1 bp	2 bp	3 bp	4 bp		Total ends[d]	Untrimmed junctions only[e]
IgH											
Adult											
3'V	103	26	25	10	3	6	1		25	10	38
5'D	474	43	38	22	10	10	1	1	9	5	51
3'D	584	214 (165)	140	83	62	14	7		37	14	39 (50)
5'J	562	119 (85)	79	44	24	19	1		21	8	37 (52)
Fetal and neonatal											
3'V	139	20	5	3	1	2			14	2	12
5'D	214	11	6	5	5				5	2	45
3'D	225	135 (40)	15	15	10	5			60	7	11 (37)
5'J	308	81 (33)	23	16	8	8			26	5	20 (48)
TCR											
Adult											
3'V	687	91	80	70	49	17	4		13	10	78
5'D	655	186	144	100	41	45	16		28	15	54
3'D	655	74 (73)	51	28	18	10			11	4	38 (38)
5'J	700	174 (173)	132	78	34	36	8		18	11	45 (45)
Fetal and neonatal											
3'V	387	108	54	49	35	13	1		28	13	45
5'D	365	86	48	40	23	15	2		24	11	47
3'D	330	75 (48)	26	17	12	5			23	5	13 (35)
5'J	415	90 (78)	34	25	9	15	1		22	6	28 (67)

[a] For more detail on analysis of IgH junctions, see references 1, 9, 10, 12, 14, and 40a. For more detail on analysis of TCRβ data, see references 2, 4, and 11. Scoring is described in Materials and Methods.
[b] Values in parentheses discount ambiguous junctions (see text).
[c] Calculated without the correction (in parentheses) shown in *b*.
[d] Values are for P inserts among all junctions.
[e] Calculations that discount ambiguous junctions (*b*) are shown in parentheses.

the studies, and the format in which junction sequences were displayed was very different from one publication to the next. In order to provide a consistent basis for comparison, we found it necessary (although tedious) to examine the data junction by junction. Full-length ends, inserts, and P residues were scored according to a set of rules based upon consistent assignments of residues to each of these categories (complete details are given in Materials and Methods). There was variable agreement between our determinations and those of different authors; this affirmed the importance of reevaluating all sequences according to standardized criteria prior to any comparison between studies. The data in Table 3 summarize the results of our analysis.

Between 2 and 15% of all ends incorporated into endogenous V(D)J junctions were associated with P inserts. The proportion of untrimmed ends associated with P inserts ranged from 11 to 78%. Although 95% of these P nucleotide inserts were one to two residues in length, 5% were 3 bp long, and there was one example of a 4-bp P insert (Table 3, IgH locus). For both the TCRβ and IgH loci, fetal and neonatal samples exhibited a lower level of P insertion than the equivalent adult sample. The largest ontogenetic differences were associated with 3' D ends in each case.

Before concluding that P insertion varies during ontogeny, other explanations for the observed differences must be considered. First, the number of full-length D ends within fetal and neonatal samples might have been inflated (leading to lowered P insert ratios) because fetal junctions tend to lack N regions and crossover sites are often located within regions with little homology between component segments (1, 10, 14). In a large number of the fetal (but not adult) junctions, therefore, there is uncertainty in the assignment of residues within the junctional region. As with the introduced substrate data, in order to conform to a consistent set of rules, residues were assigned so that, where possible, a full-length end would be scored. As a consequence, the apparent P insert/untrimmed end ratio for the fetal and neonatal samples may have been depressed relative to its actual value. We suspect that this was in fact the case, as illustrated by the alternative set of calculations shown in Table 3 (first and last column, values in parentheses). When all ambiguous 3' D-5' J junctions were excluded from the analysis, the discrepancy between fetal-neonatal and adult was much reduced for the IgH locus and disappeared completely at the TCRβ locus (Table 3).

Another pronounced difference between the fetal-neonatal sample and that of the adult was the P insert frequency associated with the 3'Vβ coding end (78% in the case of the adult compared with 45% for the fetal-neonatal sample). The 3'Vβ coding end data were dominated by a single Vβ segment, Vβ17a. This segment accounted for the majority of the adult 3'Vβ ends, and about half of those in the fetal-neonatal collection (2, 4). There were no homologies at the Vβ17a-Dβ junctions in either the adult or fetal collections that could account for these differences. However, because the Vβ17a data were derived predominantly from TCR

surface receptor-positive cells, either positive or negative immune selection could have biased the samples. In order to compare samples that, except for the age of the mouse, are as similar as possible, a comparison was limited to thymocyte-derived Vβ17a-containing junctions isolated from only one strain of mice (2). In this case, the sample sizes are small but suggest that significant differences may not exist (8 of 14 untrimmed neonatal junctions had P inserts versus 11 of 15 untrimmed adult junctions).

We conclude that any actual ontogenic fluctuation in P insert frequency is too subtle to be revealed even with the extensive junctional data included in this survey. By way of contrast, a very striking variation in N insertion occurs through ontogeny (1, 2, 10, 11, 14, 32). Thus, P insertion is not as tightly regulated as N addition, if it is regulated at all.

DISCUSSION

In other site-specific recombination systems, cutting and joining are energetically coupled operations carried out by a single site-specific enzyme (8). By contrast, V(D)J recombination appears to accomplish these operations through an ill-defined collaboration between several activities (reviewed in reference 22). Such complexity may well account for the lack of success in recreating the V(D)J joining reaction in a test tube, despite nearly a decade of effort in a number of laboratories. The only known enzyme to be implicated (through molecular genetic analyses) in the reaction is terminal deoxynucleotidyl transferase (19 and references cited therein). Although likely critical for N insertion, terminal deoxynucleotidyl transferase activity is not essential for V(D)J joining. By all appearances, V(D)J rearrangement is not the culmination of an orderly stringing together of nucleolytic, polymerization, and ligation operations, nor (judging from the phenotype of mice with severe combined immune deficiency [SCID]) should we expect that all functions playing a role in V(D)J joining are specifically dedicated to the process. Instead, some activities that participate in V(D)J joining may also feature in generalized recombination and/or repair mechanisms for the eucaryotic cell. There being no a priori guidelines in the interpretation of the fine structure of V(D)J junctions, some aspects may be indicative of an intrinsic, necessary property of the joining mechanism and some may not. We interpret our present findings in this context.

The specificity of P nucleotide addition. We have demonstrated a statistically highly significant correlation between the sequences at the tip of the coding ends and the predicted P nucleotide pattern within inserts (Table 2, pooled data). Thus, our data prove the hypothesis that P inserts exist (21). An unanticipated result was that the P nucleotide frequency within coding joints was variable. This variation was end-specific. For example, overall, P inserts were more frequently seen at the Sal end than at the Bam end (Fig. 3; Table 1; Table 2, pooled data). End-associated differences in P insert frequency were apparent even between the two coding ends of the same construct (e.g., pSal-Bam [Fig. 3A]; see values in Tables 1 and 2). Interestingly, a study reported by Kallenbach et al. (19) employed a plasmid substrate in which the 12- and 23-bp spacer signals associated with Bam and Sal coding ends were opposite to the arrangement in pSal-Bam. The *P* values we calculate from their data were surprisingly close to what was found here (0.83 for the Bam end, and 0.0003 for the Sal end versus 0.4 and 0.0002, respectively; Table 2, Sal and Spe). This rules out the possibility that observed differences in P insert frequency were a function of the signal arrangement. Furthermore, in our experiments, care was taken so that all recognized variables of the system (cell line, substrate structure, etc.) were held constant; we conclude that the DNA sequence of the coding end itself is the primary determinant of the observed P insert variation.

We have avoided the assertion that P nucleotide addition varies among the four substrates tested, because it is possible that the different frequencies with which we observe P nucleotides in the final products instead reflect differential eradication of preexisting P residues. This distinction is important: if P nucleotide addition is irregular, then the presence of P residues must not be central to the joining reaction. If, instead, P nucleotides are always added to ends (after which they are subject to sequence-influenced removal), P residues may indeed play an important role in joining.

Of the two possibilities, we first consider that P addition is consistent, but the added nucleotides are only variably preserved in the final junction products. Our data in fact suggest that different coding ends are subject to conspicuous and reproducible differences in the degree of truncation they exhibit upon incorporation into coding joints. As a consequence, one might imagine that the frequency of P insertion would also vary according to end identity. The two extreme cases were the Spe and Xho ends. Only 5% of the coding joints exhibited nontruncated Spe ends, whereas 46% of the coding joints had nontruncated Xho ends (Table 1, summary according to end). Not surprisingly, the percentage of all coding joints that contained P inserts was correspondingly lower for the Spe end (2%) than for the Xho end (15%).

However, the observed sequence-influenced truncation was an unlikely explanation for the sequence-dependent frequency of P inserts at untrimmed ends (Table 2). As shown in Table 1 (summary according to end), neither a positive nor a negative correlation existed between the frequency with which an end remained full-length and the likelihood of discovering P inserts among those examples that escaped truncation. Untrimmed Bam ends, for example, have fewer P inserts; if this is caused by P removal, then there must be a second distinct truncation activity with unusual discriminatory properties that allow it to subtract only P nucleotides from a recombination intermediate. The single-stranded region of proposed recombination intermediates (Fig. 1) includes non-P bases, but the agent that eradicates P nucleotides must somehow be able to distinguish P from non-P in order to account for a fluctuation in the ratio of P inserts to untrimmed ends. This type of explanation may be correct, but at present it requires overly intricate and unprecedented activities.

Another way in which P additions might be underrepresented in certain junctions is if the unpaired P nucleotides at an end were to anneal to the opposite strand of the partner end and create favored alignments for joining (27). The resulting junctions would lack evidence of any insert. We can evaluate this possibility by looking for a deficit of P inserts within junctions whose endpoints are consistent with potential P overlap alignment. According to the data shown in Fig. 3, while 47% of all junctions have P inserts, 31% of the junctions with P overlap alignments likewise have P inserts. Although this deficit suggests that alignments by P extension may occur, the effect is not particularly striking. Moreover, the results with pDSJ (Fig. 5), a substrate that maximizes the opportunity for overlap alignment (Fig. 1C), show that this does not take place in a predictable, recurrent fashion. As can be appreciated from both the frequency and length of P inserts where they are favored (in the case of Sal

ends, Fig. 3, for example), it is unlikely that a propensity for P overlap alignments in some cases and not others is a significant factor in the variable observance of P inserts.

By far, the simplest explanation for our results is that P nucleotides are not introduced at all ends prior to joining. Rather than supposing that in certain cases P additions are either lost or hidden at higher frequencies, our data indicate that the initial acquisition of P nucleotides is an irregular, sequence-specific event.

P nucleotide addition at endogenous loci. To more fully characterize P insertion, we took advantage of the fact that a number of large-scale studies of the junctional diversity among assembled TCR and Ig genes have been published within the last 2 years. These include collections from both surface receptor-positive and -negative cells, from precursor as well as mature lymphocytes, and from cells at various stages of ontogeny in both mice and humans. A comparative analysis of P insertion upon rearrangement of physiological substrates could yield unique information. While the sequence-specific differences detected with the plasmid assay would be averaged out, other significant variation, or lack thereof, might emerge.

We computed the numbers of P nucleotides within junctions by following a set of rules outlined in Materials and Methods. While the percentages of P inserts varied, no systematic differences emerged (Table 3). As detailed in Results, the perceived age-specific differences could well have arisen after recombination, through selection, or as a secondary consequence of age-related differences in junction structure. Furthermore, there was no consistent variation in P insertion with regard to gene segment identity (e.g., whether an end is a 3′V or 5′J). Finally, in accord with introduced substrate data (this study and reference 19), coding ends originally associated with one type of signal (e.g., a 12-bp spacer signal as opposed to a 23-bp spacer signal) did not preferentially acquire inserts.

These results support the view that P insertion is a general feature of V(D)J joining: although low and variable in appearance, P inserts were detected at all ends examined in detail. The percentage of endogenous P inserts at untrimmed ends was comparable to the ratio of P inserts to untrimmed junctions we obtained using introduced substrates (compare Table 3, last column, with Table 1, summary according to construct).

P nucleotides and N regions. Both P and N inserts must initially be created by addition of nucleotides onto a free end during joining. However, it was not known whether N residues could be added onto ends that already have a P nucleotide extension. Junctions with composite inserts containing both P and N residues might arise in either of two ways: either a P-modified end acquired an N region prior to joining, or P and N nucleotides were each contributed by different ends. We therefore made note of all junctions in which both ends were untrimmed and that also contained inserts. Seven of 27 joints from the endogenous survey of the IgH locus that fell into this category had composite inserts in which an N region was sandwiched between two P inserts. One of these junctions (sequence 109 from Decker et al. [9]) had two P residues at both borders of the insert, with a 3-bp N region between. With introduced substrates, we here obtained one example, out of 10 doubly untrimmed, insert-containing junctions, in which an N region was bounded by P nucleotides on both sides of the insert (Fig. 3, 72.1.3). Among the pSal-Bam recombinants reported by Lieber et al. (30), one of two candidate junctions has a 2-base P insert on both sides of a 3-base N region. Thus, such sandwich junctions are encountered with regularity among doubly untrimmed, insert-containing recombinants, and, in particular, the existence of junctions in which the P inserts on both sides are longer than one residue strongly suggests that N regions can be added to ends that possess preexisting P residues. Thus, P residue-modified termini must persist long enough to be exposed to the action of terminal transferase or a similar type of activity. This excludes one class of models in which, for example, hairpin ends are not opened for resolution until immediately prior to joining.

The length of P inserts. Very long P nucleotide stretches (as long as 15 bp) have been documented in TCRγ and -δ junctions derived from SCID mice (6, 13, 20, 40). Although the defect in SCID mice has not yet been defined at the molecular level, it all but eliminates productive V(D)J joining. Significantly, SCID mice exhibit a general DNA-repair defect in addition to an inability to form V(D)J junctions (reviewed in reference 3).

The long P inserts within SCID junctions are anomalous and are not necessarily indicative of length variation in nonmutant cells. In normal, non-SCID cells, P inserts of >2 bp in length have been reported (20, 36, 40), and here, in a large-scale analysis, we find that 5% of all P nucleotide inserts in normal TCRβ and IgH junctions were ≥3 bp long (Table 4). This frequency is not so high that extended P inserts in non-SCID junctions could be considered a typical variation in the P addition step. An alternative possibility is that the longer inserts are actually composites formed from a 2-base P plus fortuitous N region sequence.

Results from the plasmid assay were enlightening in this regard. Our data show (see Results) that excess P nucleotides at Sal coding ends were not likely to have been introduced through random N addition and should indeed be regarded as P inserts of greater than two residues. Furthermore, the appearance of long P inserts was end-specific. The Xho end, for example, was never found associated with a P insert of greater than two residues. We draw two conclusions from these results: (i) P inserts of >2 bp arise in non-SCID cells at a statistically significant frequency, and (ii) the length of P inserts may be directly influenced by the sequence to which they are appended.

Asymmetry of V(D)J joining. The asymmetry of V(D)J joining has been clearly demonstrated by the pDSJ results reported here. A symmetrical substrate, in which each of the four ends has signal identity, forms two distinctive junctions when rearranged. One junction had the properties of a signal joint, while the other had the fine-structural features of a coding joint. The coding joint exhibited features including base loss and addition as well as P insertion.

The differences observed between the grossly reciprocal structures of coding and signal joints are usually interpreted to mean that mechanistically distinct cutting and joining operations create the signal and coding joints in turn (3, 21, 27, 31, 39). This view cannot be fully correct, because the existence of coding-to-signal connections, as in a hybrid and open-and-shut junctions (25 and cited therein), is difficult to thereby rationalize. If P addition is the step at which asymmetry is introduced into V(D)J joining, the joining process is probably thereafter symmetrical. That is, if ligation of signal ends is carried out by a site-specific ligase, while coding end formation is accomplished by nonspecific cellular end-joining machinery, it is hard to conceive how the two operations can mix so seamlessly in hybrid joint formation. It is far more probable that both coding and signal joint formation are created by one and the same joining mecha-

nism, whether specific or nonspecific enzymatic machinery is responsible.

An alternative view is that the apparent asymmetry in V(D)J joining may be the result of differential occlusion of coding and signal ends by site-specific components of the V(D)J joining machinery. Although in pDSJ, all four ends have signal identity, only two of the four ends are site-specifically engaged by the recombination apparatus, and only those two ends may be protected from base pair loss and addition and P nucleotide modification.

Mechanistic implications. Our results, based on the evaluation of both introduced and endogenous joining substrates, affirm the generality of P nucleotide insertion and are consistent with the view that P nucleotide addition occurs along with other end-processing operations during coding joint formation. However, at the same time, we have found that junctions constructed from certain coding ends are more apt to contain P nucleotide inserts than others, that certain ends are prone to truncation, and that the lengths of P inserts can vary. These observations are inconsistent with a mechanism in which a dinucleotide transfer is an obligatory part of the joining operation (Fig. 1A, left) (2). Instead, the initial P nucleotide addition step may be stochastic, and the number of nucleotides involved is not fixed.

A hairpin intermediate, as suggested by Lieber (27), can accommodate the data. If, as proposed, hairpins form only at coding ends in the course of site-specific cleavage, the length of a P insert (from 0 to >2 bp) could then be dictated by the position of the single-strand nick that opens the hairpin. This nicking step might be influenced by the DNA sequence of the hairpin end or perhaps by some aspect of tertiary structure. Depending upon whether a terminus with a 3′, 5′, or no overhang is created upon opening the hairpin, the probability of incorporating P inserts into junctions upon joining ought to vary. By this model, a sequence-dependent variation in P insert frequency and the mechanism(s) that causes sequence-specific differences in truncation ought to be independent of one another, which is in fact what we observe (Table 1, summary according to construct).

There is precedent for hairpin formation in other site-directed DNA transactions. A hairpin intermediate has been proposed in the excision of plant transposable elements (7, 35) and has been demonstrated as an alternative product in λ site-specific recombination under conditions where normal strand exchange is blocked (34). We note, however, that there has been no systematic analysis of insertion at non-lymphoid junctions that would serve to rule out the possibility that P inserts (or hairpins) may occur in the context of more general, illegitimate recombination. In one such study (37), a number of inserts appear to fit the P insert pattern; however, the data are not amenable to statistical analysis, so the question remains open.

We favor a more general model in which hairpin formation, and thus P nucleotide addition, play a role in DNA damage repair. Without being intrinsic to the V(D)J joining operation, P nucleotides may be observed by virtue of their introduction through an incidental enzymatic activity that intervenes during the recombination process. It could be that certain types of broken ends are sealed by hairpin formation in order to prevent exonucleolytic loss and/or to delay potentially disruptive interactions between broken ends and other chromosomal sites until the breaks can be mended. In fact, the SCID phenotype (reviewed in reference 29) suggests that a link between P insert activity, general DNA repair, and the V(D)J joining operation does exist. This link could be hairpin formation and/or processing. As examples, the SCID defect could cause excessive hairpin formation or interfere with its eventual resolution. An exciting recent development is the physical detection of hairpin coding ends in SCID thymocytes (38).

ACKNOWLEDGMENTS

We thank L. Czyzyk for superlative technical assistance; E. B. Lewis, D. Mathog, and H. Lipshitz for advice on the statistical analysis; and B. Wold, J. Kobori, L. Hood, P. Fahnestock, H. Lipshitz, and P. Bjorkman for comments on the manuscript. We thank M. Gellert for comments and for communicating his results prior to publication. We thank J. Teale for providing unpublished sequence data. J.T.M. gratefully acknowledges the support of L. Hood and NIH grant GM40867 to L. Hood.

This work was funded by research grant IM-599 from the American Cancer Society to S.M.L.

REFERENCES

1. **Bangs, L. A., I. Sanz, and J. M. Teale.** 1991. Comparison of D, J_H, and junctional diversity in the fetal, adult, and aged B cell repertoires. J. Immunol. **146:**1996–2004.
2. **Bogue, M., S. Candéias, C. Benoist, and D. Mathis.** 1991. A special repertoire of α:Lβ T cells in neonatal mice. EMBO J. **10:**3647–3654.
3. **Bosma, M. J., and A. M. Carroll.** 1991. The Scid mouse mutant: definition, characterization, and potential uses. Annu. Rev. Immunol. **9:**323–350.
4. **Candéias, S., C. Waltzinger, C. Benoist, and D. Mathis.** 1991. The Vβ17+ T cell repertoire: skewed Jβ usage after thymic selection; dissimilar CDR3s in CD4+ versus CD8+ cells. J. Exp. Med **174:**989–1000.
5. **Carlsson, L., C. Övermo, and D. Holmerg.** 1992. Selection against N-region diversity in immunoglobulin heavy chain variable regions during the development of pre-immune B cell repertoires. Int. Immunol. **4:**549–553.
6. **Carroll, A. M., and M. J. Bosma.** 1991. T-lymphocyte development in SCID mice is arrested shortly after the initiation of T-cell receptor δ gene recombination. Genes Dev. **5:**1357–1366.
7. **Coen, E. S., R. Carpenter, and C. Martin.** 1986. Transposable elements generate novel spatial patterns of gene expression in *Antirrhinum majus*. Cell **47:**285–296.
8. **Craig, N. L.** 1988. The mechanism of conservative site-specific recombination. Annu. Rev. Genet. **22:**77–105.
9. **Decker, D. J., N. E. Boyle, J. A. Koziol, and N. R. Klinman.** 1991. The expression of the Ig H chain repertoire in developing bone marrow B lineage cells. J. Immunol. **146:**350–361.
10. **Feeney, A. J.** 1990. Lack of N regions in fetal and neonatal mouse immunoglobulin V-D-J junctional sequences. J. Exp. Med **172:**1377–1390.
11. **Feeney, A. J.** 1991. Junctional sequences of fetal T cell receptor β chains have few N regions. J. Exp. Med **174:**115–124.
12. **Feeney, A. J.** 1991. Predominance of the prototypic T15 anti-phosphorylcholine junctional sequence in neonatal pre-B cells. J. Immunol. **147:**4343–4350.
13. **Ferrier, P., L. R. Covey, S. C. Li, H. Suh, B. A. Malynn, T. K. Blackwell, M. A. Morrow, and F. W. Alt.** 1990. Normal recombination substrate VH to DJH rearrangements in pre-B cell lines from scid mice. J. Exp. Med **171:**1909–1918.
14. **Gu, H., I. Förster, and K. Rajewsky.** 1990. Sequence homologies, N sequence insertion and J_H gene utilization in $V_HD_HJ_H$ joining: implications for the joining mechanism and the ontogenetic timing of Ly1 B cell and B-CLL progenitor generation. EMBO J. **9:**2133–2140.
15. **Harada, K., and H. Yamagishi.** 1991. Lack of feedback inhibition of V-kappa gene rearrangement by productively rearranged alleles. J. Exp. Med **173:**409–415.
16. **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.
17. **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.
18. **Kabat, E. A. (ed.), T. T. Wu, H. M. Perry, K. S. Gottesman, and C. Foeller.** 1991. Sequences of proteins of immunological interest, 5th ed., vol. 2. National Institutes of Health publication no. 91-3242. U.S. Department of Health and Human Services, Bethesda, Md.
19. **Kallenbach, S., N. Doyen, M. F. D'Andon, and F. Rougeon.** 1992. Three lymphoid-specific factors account for all junctional diversity characteristic of somatic assembly of T-cell receptor and immunoglobulin genes. Proc. Natl. Acad. Sci. USA **89**: 2799–2803.
20. **Kienker, L. J., W. A. Kuziel, B. A. Garni-Wagner, V. Kumar, and P.W. Tucker.** 1991. T cell receptor γ and δ gene rearrangements in SCID thymocytes: similarity to those in normal thymocytes. J. Immunol. **147**:4351–4359.
21. **Lafaille, J. J., A. DeCloux, M. Bonneville, Y. Takagaki, and S. Tonegawa.** 1989. Junctional sequences of T cell receptor γδ genes: implications for γδ T cell lineages and for a novel intermediate of V-(D)-J joining. Cell **59**:859–870.
22. **Lewis, S., and M. Gellert.** 1989. The mechanism of antigen receptor gene assembly. Cell **58**:585–588.
23. **Lewis, S., A. Gifford, and D. Baltimore.** 1984. Joining of Vk to Jk gene segments in a retroviral vector introduced into lymphoid cells. Nature (London) **308**:425–428.
24. **Lewis, S., A. Gifford, and D. Baltimore.** 1985. DNA elements are asymmetrically joined during the site-specific recombination of kappa immunoglobulin genes. Science **228**:677–685.
25. **Lewis, S., J. E. Hesse, K. Mizuuchi, and M. Gellert.** 1988. Novel strand exchanges in V(D)J recombination. Cell **55**:1099–1107.
25a. **Lewis, S. M.** Unpublished data.
26. **Lewis, S. M., and J. E. Hesse.** 1991. Cutting and closing without recombination in V(D)J joining. EMBO J. **10**:3631–3639.
27. **Lieber, M. R.** 1991. Site-specific recombination in the immune system. FASEB J. **4**:2934–2944.
28. **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.
29. **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.
30. **Lieber, M. R., J. E. 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.
31. **McCormack, W. T., L. W. Tjoelker, L. M. Carlson, B. Petryniak, C. Barth, E. Humphries, and C. B. Thompson.** 1989. Chicken IgL gene rearrangement involves deletion of a circular episome and addition of single nonrandom nucleotides to both coding segments. Cell **56**:785–791.
32. **McVay, L. D., S. R. Carding, K. Bottomly, and A. C. Hayday.** 1991. Regulated expression and structure of T cell receptor γ/δ transcripts in human thymic ontogeny. EMBO J. **10**:83–91.
33. **Milstein, C., J. Even, J. M. Jarvis, A. Gonzalez-Fernandez, and E. Gherardi.** 1992. Non-random features of the repertoire expressed by the members of one Vκ gene family and of the V-J recombination. Eur. J. Immunol. **22**:1627–1634.
34. **Nash, H. A., and C. A. Robertson.** 1989. Heteroduplex substrates for bacteriophage lambda site-specific recombination: cleavage and strand transfer products. EMBO J. **11**:3523–3533.
35. **Peacock, W. J., E. S. Dennis, W. L. Gerlach, M. M. Sachs, and D. Schwartz.** 1984. Insertion and excision of *Ds* controlling elements in maize. Cold Spring Harbor Symp. Quant. Biol. **45**:347–354.
36. **Reynaud, C., V. Anquez, and J. Weill.** 1991. The chicken D locus and its contribution to the immunoglobulin heavy chain repertoire. Eur. J. Immunol. **21**:2661–2670.
37. **Roth, D. B., X.-B. Chang, and J. Wilson.** 1989. Comparison of filler DNA at immune, nonimmune, and oncogenic rearrangements suggests multiple mechanisms of formation. Mol. Cell. Biol. **9**:3049–3057.
38. **Roth, D. B., J. P. Menetski, P. 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.
39. **Roth, D. B., P. B. Nakajima, J. P. Menetski, M. J. 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.
40. **Schuler, W., N. R. Reutsch, 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.
41. **Teale, J.** Unpublished data.