Downloaded from genesdev.cshlp.org on August 22, 2022 - Published by Cold Spring Harbor Laboratory Press

V(D)J recombination: a functional definition of the joining signals

Joanne E. Hesse, Michael R. Lieber, Kiyoshi Mizuuchi, and Martin Gellert

Laboratory of Molecular Biology; National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892 USA

Two conserved DNA sequences serve as joining signals in the assembly of immunoglobulins and T-cell receptors from *V*-, *(D)*-, and *J*-coding segments during lymphoid differentiation. We have examined *V(D)J* recombination as a function of joining signal sequence. Plasmid substrates with mutations in one or both of the heptamer-spacer-nonamer sequences were tested for recombination in a pre-B-cell line active in *V(D)J* recombination. No signal variant recombines more efficiently than the consensus forms of the joining signals. We find the heptamer sequence to be the most important; specifically, the three bases closest to the recombination crossover site are critical. The nonamer is not as rigidly defined, and it is not important to maintain the five consecutive As that distinguish the consensus nonamer sequence. Both types of signals display very similar sequence requirements and have in common an intolerance for changes in spacer length >1 bp. Although the two signal types share sequence motifs, we find no evidence of a role in recombination for homology between the signals, suggesting that they serve primarily as protein recognition and binding sites.

[*Key Words*: *V(D)J* recombination; site-specific recombination; gene rearrangement; antigen receptor]

Received March 21, 1989; revised version accepted May 9, 1989.

Mature genes encoding the component chains of immunoglobulin (Ig) and T-cell receptor (TCR) proteins are assembled early in lymphoid development from germ-line arrays of variable (*V*), diversity (*D*), and joining (*J*) gene segments (for recent reviews, see Blackwell and Alt 1988; Davis 1988). By allowing for a large variety of coding sequences, this combinatorial process is the key to diversity of antigen-binding proteins. Seven loci subject to rearrangement by *V(D)J* recombination have now been identified, and a great deal is known about the highly complex and distinctive organization of the endogenous genes. Gene assembly occurs within a restricted period of early B- and T-cell development and in a preferred temporal order within each lineage.

DNA sequence motifs, found adjacent to all *V*-, *D*-, and *J*-coding elements at every locus known to rearrange, were implicated in recombination because of their strategic locations (Tonegawa 1983). Their role as joining signals was confirmed when recombinant junctions involving these sites were identified (Steinmetz et al. 1980; Höchtl et al. 1982; Selsing et al. 1984). Initial comparison of joining signals revealed a highly conserved consensus among signal sequences (Tonegawa 1983). Immediately flanking a coding element is a palindromic heptamer whose consensus form is CACAGTG. It is separated by a short spacer sequence from an AT-rich nonamer whose consensus form is ACAAAAACC. Although there is no apparent consensus among spacer sequences, discrete spacer lengths of 12 and 23 bp are highly conserved and define two signal types. Recombination of endogenous genes occurs only when one coding element is flanked by a 12-spacer signal and the other by a 23-spacer signal (the '12/23 joining rule').

Recent studies have shown that a pair of signals, one of each type, contains sufficient sequence information to direct recombination; no other specific DNA, including coding element DNA, is required for the reaction (Akira et al. 1987; Hesse et al. 1987). Thus, signals can direct recombination even when removed altogether from their Ig or TCR gene context. Standard *V(D)J* recombination generates two recombinant junctions: Coding elements are fused in a 'coding joint,' generating a region of continuous coding sequence, and joining signals are fused at their heptamers in a 'signal joint' (Lewis et al. 1984, 1985). *V(D)J* recombination can result in inversion or deletion of the intervening DNA, depending on whether the signals are in direct or opposed orientation (Lewis et al. 1984, 1985). Both deletion and inversion occur during assembly of mature Ig and TCR genes (Malissen et al. 1986; Lorenz et al. 1987; Korman et al. 1989).

It has been suggested that the signals are recognition and binding sites for recombination proteins (Tonegawa 1983). Alternatively, the common heptamer and nonamer motifs of the signals might induce transient base-pairing during recombination. The palindromic nature of the heptamer would suggest the possibility of either direct or inverted signal alignment in an intermediate structure.

Although comparison of naturally occurring signals reveals conservation of structure and a consensus among joining signal sequences, it also reveals that rearrange-

Downloaded from genesdev.cshlp.org on August 22, 2022 - Published by Cold Spring Harbor Laboratory Press

ments are possible despite a degree of variability in joining signal sequence (Kabat et al. 1987), suggesting that some features of the signals are less critical in recombination than others. Although conservation of a nucleotide at a given position may provide a clue as to its functional importance, it has been difficult to assess the effect of a given variation in signal structure on recombination. Examination of endogenous signals, while cataloging those signal variants that occur in nature, does not show how well they work. Except when a coding element has been identified in a coding joint, in which case it can be inferred that its joining signal has been used, there has been no way of determining which signal variants are functional. Among those variants identified as functional, no correlation between signal structure and recombination efficiency has been possible.

In this paper we examine recombination as a function of joining signal sequence, in a recombination assay employing plasmid substrates introduced transiently into a recombinationally active cell line. This method makes it convenient to test a variety of different substrates with altered signals and to generate large collections of products for examination. The assay allows a recombination frequency to be measured; even a frequency reduced 100-fold from that obtained with consensus signals is detected easily. Our results define further the substrate for *V(D)J* recombination, and provide insight into the role of the joining signals in this reaction. Although these results were obtained with constructed substrates and a single active cell line, they are likely to have more general validity, as judged from a comparison with natural signal sequences that is included below.

Results

The plasmid substrate pJH299, diagramed at the top of Figure 1, contains one of each type of *V(D)J* recombination signal, one with a 12-bp spacer (hereafter referred to as a '12-signal') and one with a 23-bp spacer (a 23-signal). Both use the consensus forms of heptamers and nonamers. For this study we prepared a collection of substrates identical to pJH299, except for specific alterations in one or both joining signals, and tested them for *V(D)J* recombination in the pre-B-cell line 1-8.

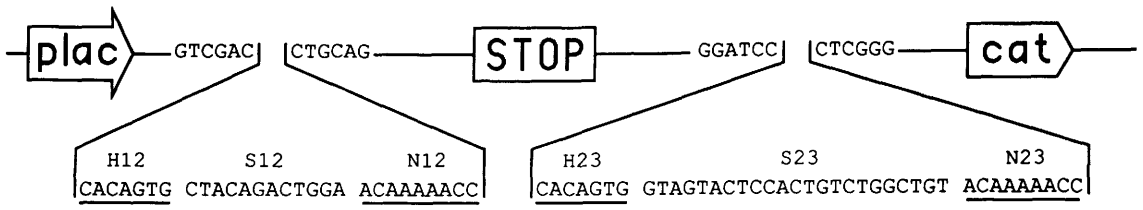
A detailed analysis of the structure and frequency of the recombination products generated from pJH299 in the 1-8 cell line in this assay has been presented (Lewis et al. 1988). Briefly, the predominant product has undergone inversion and contains both the recombinant joints generated in a standard event, a signal joint and a joint analogous to a coding joint. (The plasmid substrates contain no *V*-, *D*-, or *J*-coding element DNA.) Another *V(D)J* recombination product, a hybrid joint, which can represent up to 30% of the total, has undergone deletion and fuses one signal to the sequences flanking the other. Over 99% of the recombinant products conferring chloramphenicol resistance (Cam^r) on bacterial hosts result from *V(D)J*-mediated rearrangement using the provided joining signals. The remaining Cam^r products, representing 0.8% of the total at most, appear to result from *V(D)J*-mediated recombination using one of the intended signals and other fortuitous signal-like sequences in the plasmid backbone. We determined the background frequency of Cam^r transformants generated by direct transformation of *Escherichia coli* with substrate DNA to be 10^{-6}, corresponding to 0.0025% of the total Cam^r products and thus too low to be significant. In this study we examined the details of recombinant structures in only selected cases. A more complete study of the effects of joining signal sequences on recombinant structure will be reported separately.

Signals with altered heptamers

When the 12-signal was altered by single-base substitutions in its heptamer and tested in combination with a consensus 23-signal, it became clear that the identities of the first three nucleotides, those farthest away from the nonamer, are critical for recombination (Fig. 1A). Any of the three possible substitutions at each of these positions reduced recombination dramatically, showing the consensus sequence CAC to be required strictly. These results are consistent with those of Akira et al. (1987), who tested two pairs of signals altered within these positions and found recombination to be reduced greatly.

Figure 1. Recombination of plasmids with mutated signal sequences. (*Top*) A portion of the plasmid substrate used throughout this study. A prokaryotic transcription terminator (STOP), flanked by one of each type of *V(D)J* recombination signal, separates a prokaryotic promoter (*plac*) from the structural gene encoding chloramphenicol acetyltransferase (*cat*). *V(D)J*-mediated rearrangement in transfected pre-B cells results in inactivation of the terminator (by inversion, forming standard products, or by deletion, forming hybrid junctions), allowing subsequent *cat* gene expression in *E. coli*. The signal sequences shown are those found in pJH299, which bears heptamers and nonamers matching the consensus sequence. Also shown are the bases immediately flanking each signal. The sequences flanking the heptamer of each signal are positionally analogous to *V*-, *D*-, and *J*-coding element DNA in the endogenous Ig and TCR genes. Recombination results are shown for plasmid substrates identical to pJH299, except for the indicated alterations in heptamers (*A*), spacer lengths (*B*), nonamers (*C*), or both heptamers and nonamers (*D*). At the top of each section (except *D*) the pertinent sequences found in pJH299 are reiterated. A dot denotes identity with the pJH299 sequence. Base substitutions and insertions are indicated; a base deletion is represented by 'X.' We have assigned numbers to the positions of nucleotides in the nonamer which are shown (in *C*) just below the pJH299 sequence. Recombination frequencies have been normalized to the frequency obtained using pJH299 ($R_n = R_{test}/R_{pJH299} \times 100$). Values of $R_n < 1.0$ could include a significant contribution from the background (see text); these are given as one significant figure, whereas values >1.0 are rounded off to the nearest percentage point.

Downloaded from genesdev.cshlp.org on August 22, 2022 - Published by Cold Spring Harbor Laboratory Press



A

Plasmid	H12	H23	Rn
pJH299	CACAGTG	CACAGTG	100
pJH306	T.....	.....	1
pJH413	A.....	.....	2
pJH446	G.....	.....	2
pJH415	.C.....	.....	2
pJH414	.T.....	.....	2
pJH328	.G.....	.....	1
pJH334	..T....	.....	3
pJH420	..A....	.....	1
pJH403	..G....	.....	2
pJH421	...C...	.....	6
pJH342	...T...	.....	33
pJH417	...G...	.....	77
pJH336	....A..	.....	85
pJH338	.....C.	.....	87
pJH340	......A	.....	74
pJH419	....ACA	.....	26
pJH449	TGGCGAT	.....	0.7
pJH314	.....	T.....	1
pJH316	.....	.G.....	0.7
pJH362	.....	..T....	1
pJH364	.....	...T...	25
pJH366	.....	....A..	73
pJH368	.....	.....C.	46
pJH384	.....	......A	55
pJH322	T.....	T.....	0.1
pJH395	.G.....	.G.....	<0.4
pJH372	..T....	..T....	0.7
pJH376	...T...	...T...	0.5
pJH378	....A..	....A..	14
pJH398	.....C.	.....C.	31
pJH396	......A	......A	29
pJH392	T.....	......A	0.5
pJH374	..T....	....A..	0.4
pJH380	....A..	..T....	1
pJH400	.....C.	.G.....	0.9
pJH382	......A	T.....	<0.5

B

Plasmid	S12	S23	Rn
pJH299	CTACAGACTGGA	GTAGTACTCCACTGTCTGGCTGT	100
pJH348	.....T.....	.....	37
pJH388	.....	.....G.....	38
pJH391	.....T.....	.....G.....	6
pJH447	.....	.....GT.....	3
pJH346	.....X.....	.....	15
pJH386	.....	.....X.....	41
pJH402	.....X.....	.....X.....	4
pJH309	.....XX.....	.....	4
pJH318	.....	.....XX.....	2
pJH324	.....XX.....	.....XX.....	0.7
pJH311	....XXX.....	.....	4
pJH320	.....	.....XXX.....	1
pJH326	....XXX.....	.....XXX.....	0.9
pJH407	.....T.....	.....X.....	6
pJH438	.....X.....	.....G.....	3

C

Plasmid	N12	N23	Rn
pJH299	ACAAAAACC 123456789	ACAAAAACC 123456789	100
pJH344	....T....	.....	87
pJH350	....G....	.....	80
pJH390	...CG....	.....	34
pJH405	.....G...	.....	3
pJH439	......G..	.....	14
pJH422	CA....C..	.....	17
pJH429	.....	....G....	22
pJH433	.....	...CG....	22
pJH431	.....	.....G...	4
pJH441	....G....	....G....	34
pJH445	...CG....	...CG....	4
pJH443	.....G...	.....G...	<0.2
pJH425	TGTCTCTGA	.....	5
pJH434	.....	TGTCTCTGA	1
pJH448	.....	TGTCT..GA	1

D

Plasmid	H12	S12	N12	H23	S23	N23	Rn
pJH423	....A..	.....	....T....	.....	.....	.....	33
pJH427	....A..	.....	...CG....	.....	.....	.....	8

Figure 1. (*See facing page for legend.*)

Downloaded from genesdev.cshlp.org on August 22, 2022 - Published by Cold Spring Harbor Laboratory Press

The next adjacent base, the middle A of the heptamer, was less stringently specified; the level of recombination varied between 6% and 77%, depending on the base substituted at this position. Replacing A with T here was of particular interest; because of the palindromic consensus sequence, this change could be viewed as reversing the orientation of the heptamer, thus possibly allowing the end of the heptamer internal to the signal to be active in joining. In fact, the vast majority of recombinants generated with this signal sequence still reflected a heptamer active at the nonamer–distal end rather than the internal end (data not shown). Recombinants ligated at the internal end do occur but not any more frequently with this signal than with others in which the middle base of the heptamer is unchanged from the consensus.

At each of the remaining positions, the three nonamer–proximal positions in the heptamer, we tested one transition substitution; these changes had little effect on recombination. Even when tested in concert (pJH419), these alterations supported a level of recombination reduced only fourfold from that obtained with a consensus 12-signal.

Some of these substitutions were also tested in the 23-signal, each being paired for recombination with a consensus 12-signal. The results, shown in Figure 1A, reveal that the base identity requirements for the heptamer of the 23-signal parallel those of the 12-signal; the nonamer–distal CAC sequence is critical and the middle A less so, whereas transition mutations at the nonamer–proximal bases show only minor effects.

Could activity of a defective heptamer be restored by also mutating the heptamer of the partner signal so as to restore the sequence homology between them? We constructed two series of substrates to test this possibility (Fig. 1A). In one series, each plasmid has identical alterations in both its 12- and 23-signal heptamers. The other is a reverse series of paired alterations that takes into account the palindromic nature of the heptamer (so that homology is maintained between the heptamers aligned in inverted orientation). All possible combinations in each series, except for one of the reverse combinations, were tested. As shown in Figure 1A, we found no case in which the presence of a corresponding substitution on the partner signal restored recombination. In fact, for those examples within the detection limits of our assay, we found all doubly altered substrates to recombine at a level lower than either of the alterations tested singly.

No recombination was seen with a signal totally lacking a recognizable heptamer. We prepared a 12-signal with our standard spacer sequence and a consensus nonamer, but with a heptamer substituted at every residue, and tested it for recombination with a consensus 23-signal (pJH449). Although some Cam^r transformants were isolated, none had a structure consistent with *V(D)J*-like rearrangement. In contrast, a 'nonamer-less' signal was active. Recombination was 5% and 1% with a nonamer-less 12-signal and 23-signal, respectively, each tested with a consensus signal of the opposite type (pJH425, pJH434: Fig. 1C). Structural analysis confirmed that the recombinants generated with each of these substrates had indeed undergone *V(D)J*-mediated rearrangement events.

Signals with altered spacers

We prepared signals that had the consensus heptamer and nonamer sequences but with an altered spacer length between them (Fig. 1B). Changes from 12 and 23 bp were made by inserting or deleting bases close to the midpoints of the existing sequences. Otherwise, the spacer sequences were unchanged from the standard sequences we chose to use in all the rest of our substrates. Recombination was reduced only modestly when a normal signal was paired with a partner having a change of +1 or −1 in spacer length; partners of spacer lengths 13 × 23, 12 × 24, 11 × 23, and 12 × 22 recombined at levels between 15% and 41% of that observed using the 12 × 23 pair. When both partners had spacer lengths altered by 1 base, recombination fell more drastically, to 6% or less. This was true whether the pair had like changes (13 × 24 and 11 × 22) or opposed changes (13 × 22 and 11 × 24), in which the overall combined spacer lengths of both signals would be conserved. Recombination was very low also, 4% or less, for spacer length changes >1. This included the 12 × 25 pair and the pairs 10 × 23, 12 × 21, 9 × 23, and 12 × 20, in which one normal partner was paired with another missing 2 or 3 bases from its spacer and the 12 × 25 pair. Again, potentially compensatory alterations of −2 or −3 in the other signal (10 × 21 and 9 × 20) did not restore activity.

We tested substrates bearing pairs of identical signals (12 × 12 and 23 × 23) and found—as we had when testing a nonamer-less signal—that recombination was indeed possible but inefficient. Plasmids conferring Cam^r were isolated with the structure expected to result from *V(D)J*-mediated inversion, but at a frequency of only 2% of that obtained using a 12 × 23 pair of signals (data not shown). In a related experiment using substrates designed to integrate into the host-cell genome, Akira et al. (1987) had failed to detect recombination between a 23 × 23 pair of signals.

Because the center-to-center distance from heptamer to nonamer is 20 bp for a 12-signal and 31 bp for a 23-signal, roughly consistent with integral numbers of helical turns, one can consider the possibility of specific contacts between recombination proteins and DNA along one face of the helix. Such contacts might be retained by mutations that change spacer length by integral numbers of turns. The 12 × 12 and 23 × 23 spacer combinations are a test of this idea, in one sense, and are found to work poorly. It would be interesting to test unequal combinations such as 12 × 34 or 23 × 34.

Signals with altered nonamers

Signals altered from the consensus at various residues within their nonamers were tested for recombination with a partner signal of consensus sequence or with each other. The results are shown in Figure 1C. For conve-

Downloaded from genesdev.cshlp.org on August 22, 2022 - Published by Cold Spring Harbor Laboratory Press

nience, we have assigned numbers to the nonamer base positions, designating the position adjacent to the spacer as 1 and the most distal position as 9. We focussed our efforts on testing mutations at positions 5, 6, and 7; among endogenous signals these are the most conserved positions (Table 1). Substituting G or T for the A in position 5 in the nonamer of the 12-signal had little effect on recombination. A double substitution of CG replacing AA at the fourth and fifth positions had only a modest effect, reducing recombination about threefold. A more dramatic effect was observed when we tested single-base substitutions of G for A at the sixth and seventh positions; these replacements reduced recombination to 3% and 14%, respectively, of that obtained using a consensus signal. Single-base substitutions at other positions in the nonamer were not tested. When the A in position 7 was changed (this time to C) in concert with substitutions at positions 1 and 2, activity dropped to 17%, comparable to the level observed for the single change at position 7. This implies that the changes at positions 1 and 2 have little effect and that the substitution at the seventh position is the dominant feature of this nonamer that contributes to the reduction in activity. It should be pointed out that this triple mutant in the nonamer signal sequence can be interpreted instead as a consensus nonamer and a 1-base deletion of the spacer sequence at the nonamer-proximal A. This view, combined with the observation of a comparable recombination frequency obtained with pJH346 (Fig. 1B), supports further the notion that the length, but not the sequence, is the important feature of the spacer.

A similar pattern emerged when we measured recombination using some of the same substitutions on the 23-signal, this time paired with a consensus 12-signal (Fig. 1C); substitutions at the fifth, or fourth and fifth, residues in the nonamer of the 23-signal reduced recombination a modest 4-fold, whereas replacement of the A at the sixth position resulted in a 25-fold reduction of activity. Although the A in the sixth position, and perhaps the seventh, might be necessary for efficient recombination, together they are not sufficient. A nonamer retaining AA at these positions but substituted at each of the other seven yielded only the basal level of activity observed with an altogether nonamer-less 23-signal.

When signals with identical substitutions in their nonamers were paired with each other and tested, we found, as we had with the heptamer mutants, that restoring homology between altered signals did not restore activity. In two examples, pJH443, and pJH445, the doubly altered substrates recombined much less efficiently than the substitutions tested singly; and in a third, pJH441, recombination was comparable to that of single mutants.

What would be the effects of combining changes chosen among those having little effect? We prepared two 12-signals, each with an alteration in the heptamer and in the nonamer; the substitution of A for G in one of the nonamer–proximal heptamer residues and substitution at the fifth position or the fourth and fifth positions in the nonamer. These signals were tested for recombination with a 23-signal of consensus sequence (Fig. 1D). When tested singly, the alterations in pJH423 had recombined at 85% and 87%, respectively; when combined, the frequency was reduced to 33%. The substitutions in pJH427 had yielded 85% and 34% recombination, respectively; the frequency when they were combined in the same signal dropped to 8%. Thus, as we had observed earlier with pJH419, which was substituted at all three of the nonamer–proximal positions of the 12-signal, multiple changes in the same signal did seem to exert a concerted effect; but the signals in all examples still allowed substantial recombination.

Discussion

The results reported here suggest that the consensus joining signal sequence in *V(D)J* recombination is indeed the most efficient signal; we found no signal sequence variant to recombine more efficiently than the consensus form. At the same time, we find many signal variants to be functional, exhibiting a wide range of recombination frequencies. These quantitative differences

Table 1. *Sequence conservation at the heptamer and nonamer among naturally occurring joining signals*

	C	A	C	A	G	T	G	A	C	A	A	A	A	A	C	C
G	0	<1	0	3	**82**	<1	**76**	9	3	6	8	3	0	2	4	4
A	0	**99**	0	**87**	10	3	13	**72**	5	**83**	**73**	**91**	**97**	**87**	4	5
C	**100**	0	**99**	<1	4	11	9	11	**86**	8	9	1	<1	4	**84**	**76**
T	0	0	<1	9	4	**85**	2	8	6	3	10	5	2	7	8	15

The frequency (%) of occurrence of the four possible bases at each position in the heptamer and nonamer is shown. The consensus sequences for the heptamer and nonamer are underlined. The data are summarized from a compilation of 226 published Ig and TCR joining signal sequences (Hesse 1989). The majority derive from mouse and human loci, although some examples from other species are included. Joining signals were included whether or not they are known to be functional variants, although those thought to flank nonfunctional pseudogenes (16 examples) were excluded from the calculation. The frequencies of the consensus nucleotides are shown in boldface type. The distributions of signal changes from the consensus are as follows. Only 28 signals were perfect matches to the consensus signal sequence. There were 63 signals with a single mismatch (9 with the alteration in the heptamer, and 54 in the nonamer) and 44 with 2 mismatches (3 with both in the heptamer, 9 with both in the nonamer, and 32 with one mismatch in each). Signals with 3, 4, and >4 mismatches accounted for 46, 24, and 21 examples, respectively.

Downloaded from genesdev.cshlp.org on August 22, 2022 - Published by Cold Spring Harbor Laboratory Press

may be relevant to certain temporal and developmental biases in endogenous gene rearrangements (Yancopoulos et al. 1984; Perlmutter et al. 1985; Reth et al. 1986a; Yancopoulos and Alt 1986).

Our results show also that many signal sequence changes are tolerated without seriously affecting the recombination frequency, suggesting that many of the coding elements flanked by such a nonconsensus signal could, in fact, be used readily. Figure 2 gives a general summary of the mutational results. This tolerance for signal sequence changes emphasizes that the nucleotide sequence of the signal cannot be solely responsible for determining the recombination sites of *V(D)J* rearrangement in lymphoid cells, because a large number of usable signal-like sequences are expected to be distributed along the entire chromosome. Indeed, the aberrant use of signal-like sequences, apparently consequent to the escape of recombination from its normal regulation, has been reported; signal-like sequences identified at sites of chromosomal translocations and inversions in lymphomas and leukemias are implicated in recombination in which the flanking DNA, unrelated to coding-element DNA but positionally equivalent, becomes joined to a coding element from an Ig or TCR locus (for review, see Showe and Croce 1987). One frequently discussed possible contribution to normal site determination is the inaccessibility to the recombination enzyme(s) of portions of the chromosome not expected to be rearranged during lymphoid differentiation (Yancopoulos and Alt 1985).

Despite its apparent symmetry, the heptamer has a distinct functional sidedness in the *V(D)J* recombination reaction; the end distal to the nonamer is used for the recombination crossover, whereas the other end remains unaltered within the joining signal. Underscoring this functional asymmetry, our results demonstrate that the nonamer–distal bases of the heptamer are critical for normal levels of activity, whereas the nonamer–proximal bases are not. Further study should reveal whether mutations in these distal positions affect the type of product as well as the recombination frequency.

Because the consensus sequence for the heptamer is symmetrical, a likely determinant of its functional sidedness is its positional relationship to the nonamer. The middle AT of the heptamer also might be expected to affect sidedness because it is the heptamer's only departure in sequence from a perfect palindrome. We find that changing the middle A to a T does not alter which side of the heptamer is recombined, as long as the rest of the signal retains its consensus sequence. However, in the absence of a recognizable nonamer, the identity of the middle base (A or T) now determines which end of the heptamer is primarily active in the reaction (data not shown). Thus, the nonamer seems to have the major influence on the functional polarity of the signal, with the central base of the heptamer playing a secondary part. The symmetry of the heptamer, except that it allows the use of either end, is not important for recombination; a heptamer with the three nonamer–proximal bases changed, completely disrupting its rotational symmetry, still recombined quite efficiently.

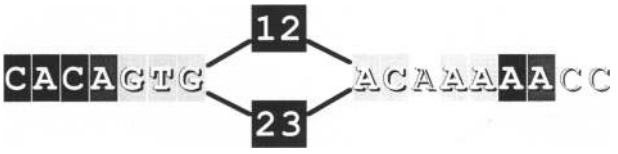


Figure 2. Contribution of individual nucleotides to *V(D)J* signal use. The heptamer and nonamer consensus sequences are shown with the most essential nucleotides in dark stippling, less essential nucleotides in medium stippling, and nucleotides whose alteration affects recombination only slightly, in light stippling. Nucleotides that were not altered in the present work are shown without stippling. Dark stippling of the boxes enclosing spacer length indicates that changes in these lengths were accommodated poorly.

Our observation that the nonamer–distal sequence of the heptamer is functionally critical fits well with its marked conservation among endogenous joining signals. Hundreds of naturally occurring joining signal sequences are now known. Although each consensus nucleotide is conserved significantly [none <72% in the compilation summarized in Table 1 (or 63% in the compilation of Akira et al. 1987)], preservation of the heptamer sequence CAC immediately flanking the coding element is especially striking. Each of these nucleotides is found here in at least 99% of known signals when those nucleotides flanking nonfunctional pseudogenes are omitted from consideration. The middle base of the heptamer is slightly less conserved; A is found here in 87% of naturally occurring signals (Table 1). This also is consistent with the somewhat less stringent requirement we measured for this nucleotide.

Considering that each of the three nonamer–proximal nucleotides of the heptamer also is well-conserved among endogenous signals (76–85%: Table 1; see also Akira et al. 1987), the results of substitutions at these positions were somewhat surprising. We find that the bases at these positions need not be specified stringently for efficient recombination, so that the heptamer could be redefined functionally as a tetramer, with the three nonamer–proximal positions forming part of the spacer. Conservation of this GTG sequence (and the rotational symmetry that results) may serve some purpose not related directly to the recombination mechanism. In view of this possibility, it is noteworthy that as a group, joining signals flanking TCR gene elements appear to vary from the consensus sequence in the nonamer–proximal region more often than those flanking Ig gene elements (Hesse 1989).

Endogenous *V(D)J* recombination has been seen to result in joining directed by a joining signal (or a sequence highly resembling a joining signal) and what appear to be isolated heptamers. Although some of these examples seem to be rare aberrant events (Seidman and Leder 1980; Höchtl et al. 1983; Kelley et al. 1985), others seem to be conserved events with possible physiological roles (Durdik et al. 1984; Moore et al. 1985; Siminovitch et al.

Downloaded from genesdev.cshlp.org on August 22, 2022 - Published by Cold Spring Harbor Laboratory Press

1985; Stavnezer et al. 1985; Kleinfeld et al. 1986; Reth et al. 1986b). In our assay as well, a significant level of basal recombination activity is observed when one signal totally lacks a recognizable nonamer. This result emphasizes further the existence of other contributions to the determination of site usage in *V(D)J* recombination, because sequences resembling just the heptamer, without an associated nonamer, would be expected to occur even more frequently at random along the chromosome than fully signal-like sequences.

Although a nonamer-less signal allowed recombination at levels detected easily by this assay, a heptamer-less signal did not. This suggests that the heptamer is the more essential element of the two. In addition to not being altogether essential in the reaction, the nonamer is not as rigidly defined a sequence motif as is the heptamer. Among naturally occurring signals, although conservation of consensus nucleotides is >90% at two positions in the nonamer, it ranges from 72% to 87% at the remaining seven (Table 1), showing that, overall, some degree of sequence variation in this motif is functionally acceptable. In our experiments as well, a range of sequence variants can direct efficient recombination. Furthermore, there appears little need to maintain the five consecutive As that are such a prominent sequence feature of the consensus nonamer; a change of the middle A is almost inconsequential, and change of the two As in positions 4 and 5 reduces activity only threefold. This suggests that neither melting nor bending, potentially facilitated by an A tract, plays a significant role here.

Those nucleotides of the nonamer identified in our assay as functionally important also are highly conserved (e.g., the As at positions 6 and 7 are found in 97% and 87% of known signals, respectively). However, as was observed with the heptamer, not all well-conserved nonamer nucleotides are found to be critical in our assay (e.g., the As at positions 4 and 5 are found in 73% and 91% of known signals, respectively).

The relatively less stringent sequence requirements for the nonamer prompted us to reevaluate the results of changes in spacer length. Could these be reinterpreted as revealing instead the effects of signals with correct spacers but unusual nonamers? When we highlighted the 'new' nonamers in these signals, some plausible correlations emerged, but not enough to show that spacer length is indeed a fixed parameter. We have no unequivocal evidence that alternative spacer lengths are acceptable, whereas we have strong evidence that some alternative nonamers are acceptable. A critical test would demand the construction of many more nonamer mutants.

One model that can be excluded is that it is the heptamer whose position is defined relative to the nonamer. Mutants with altered spacers would have new heptamers with nonamer–distal bases substituted in ways that, when tested in signals of standard spacer length, allow only minimal recombination. In fact, some spacer length mutants allow recombination levels up to 41%. In addition, recombinants derived from signals with altered spacer length contain precise signal joints (data not shown) that used the end of the consensus heptamer rather than the end of a new heptamer.

Recombination models that invoke direct DNA–DNA interaction via sequence homology between the signals are inconsistent with our results. The stem-and-loop model, still popularly used to depict this reaction, proposes a recombination intermediate in which each strand of a 12-signal reanneals with that of a 23-signal via base-pairing between their respective heptamers and nonamers (Sakano et al. 1979; Early et al. 1980; Watson et al. 1987). Alternatively, a four-strand intermediate structure could be stabilized by base-pairing between the respective heptamers and nonamers of directly aligned signals or between the heptamers of inversely aligned signals. These models predict that recombination might be less efficient between partners having disrupted homology and that activity might be restored by pairing signals correspondingly altered so as to restore homology between them. These predictions were not borne out when tested. First, although recombination was, in fact, less efficient in some examples in which homology between signals was reduced, interruption of homology had little effect on activity in other examples. Second, we found no example in which activity was restored by pairing correspondingly altered signals when we tested double mutants that restored homology between heptamers (aligned in either direct or inverse orientation) or between nonamers. Thus, the heptamer and nonamer motifs do not appear to be recognized by DNA—DNA interaction. The lack of a role for base-pairing in the joining reaction also has been inferred from other data indicating a lack of directionality in the strand exchanges (Lewis et al. 1988). Our results indicate that recognition of spacer length is also intrinsic to each signal rather than a reflection of inter-signal spatial arrangement.

The lack of evidence for direct contact between signals leaves the attractive alternative that the requirement for the characteristic signal structures is manifested at the level of DNA–protein interaction instead. The signals may define recognition and binding sites for the recombination machinery. Those sequence alterations we tested that reduce activity may identify essential contact points.

Furthermore, 12- and 23-signals exhibit the same potential contact points. In our experiments the two types of signals display very similar profiles of sequence and spacing requirements. In all examples tested, the effect of a given alteration was roughly the same whether the change resided on the 12-signal (paired for reaction with a consensus 23-signal) or on the 23-signal (paired with a consensus 12-signal). We noted a slight difference only with respect to the severity of the effect; altered 23-signals were generally more defective than their 12-signal counterparts. Thus, the shared heptamer and nonamer motifs, in lieu of participating in base-pairing, may instead reflect protein contact points in common, possibly indicating that the signals interact with the same protein species.

Downloaded from genesdev.cshlp.org on August 22, 2022 - Published by Cold Spring Harbor Laboratory Press

Methods

Cell culture and transfection

Cell culture and DEAE–dextran transfection were carried out as described previously (Hesse et al. 1987; Lieber et al. 1987). The Abelson murine leukemia virus-transformed lymphoid cell line 204-1-8 (McKearn and Rosenberg 1985), referred to here as 1-8, was used throughout this study. This cell line has been shown previously to be active in recombination with the plasmid assay (Lieber et al. 1987).

Recovery of plasmid DNA from transfected cell lines and DNA transformation of E. coli

Procedures used here were as described (Hesse et al. 1987), with the exceptions that competent cells (*E. coli* strain DH5α) were purchased from Bethesda Research Laboratories in some experiments, and the incubation time at 37°C of transformed cells prior to spreading on agar plates was increased to 120 min in all experiments. The longer recovery period improves the plating efficiency on chloramphenicol plates without allowing for significant cell growth (J. Hesse, unpubl.).

Plasmid constructions

pJH299 and pJH298, from which pJH299 was derived, are members of a family of plasmids closely related in structure, having a common plasmid backbone and differing only in the details of the joining signals included. A complete description of the construction of pJH298 has been given (Lieber et al. 1988). Briefly, the plasmid backbone is a fusion of pUC13 and the early region of the polyoma virus genome, which allows the plasmids to replicate autonomously in *E. coli* and murine cells, respectively. Various DNA fragments pertinent to this recombination assay are inserted into the multipurpose cloning site of pUC13, including a prokaryotic promoter, the structural gene encoding chloramphenicol acetyltransferase, (*cat*), and the intervening transcription terminator (see Fig. 1). The terminator is flanked on the promoter–proximal side by a unique *Sal*I site and on the *cat* gene side by a unique *Bam*HI site. DNA fragments containing *V(D)J* recombination signals are inserted into these unique sites. (These signals are the only DNA segments related to Ig genes or TCR genes included on the plasmid.) A 39-bp fragment containing a 12-signal, created by annealing of synthesized complementary oligonucleotides with protruding ends compatible with *Sal*I, is inserted into the *Sal*I site of the backbone. A 49-bp fragment containing a 23-signal, made similarly but with protruding ends compatible with *Bam*HI, is inserted into the *Bam*HI site of the backbone. (Lengths of these fragments vary slightly for those signals having altered spacer lengths.) pJH299 contains signals bearing the consensus sequence versions of heptamers and nonamers. The 12- and 23-spacer sequences chosen for use here (see Fig. 1) have been used routinely in the past in plasmid substrates for this assay and were identified originally from naturally occurring endogenous joining signals known to be functional (Hesse et al. 1987). All other substrates used in this study were derived from pJH299 using standard cloning techniques by substitution of one or both signal-bearing fragments. We confirmed the sequences of both signals in each substrate using standard DNA sequencing methods.

Recombination assay

The assay for measuring levels of *V(D)J* recombination activity in cell cultures contains refinements (Lieber et al. 1987) of that originally reported (Hesse et al. 1987). Briefly, pre-B-cell lines are transfected with plasmid substrates containing recombination signals and cultured for 48 hr. During this time, some plasmids become rearranged by *V(D)J* recombination. Plasmid DNA subsequently recovered from the transfected cells is introduced into *E. coli*, and recombinant molecules, in which *cat* gene expression has been activated by *V(D)J*-mediated rearrangement, are identified by growth of transformants on selective medium. All plasmid DNA confers ampicillin resistance (Amp^r) on bacterial hosts. Recombinant molecules also confer Cam^r. The ratio of doubly resistant (Amp^rCam^r) transformants to Amp^r transformants reflects the fraction of DNA that has undergone *V(D)J*-mediated rearrangement. A more meaningful ratio can be obtained by focusing on those molecules that have replicated in the eukaryotic cells (Lieber et al. 1987). The ratio of Amp^rCam^r to Amp^r among replicated molecules is designated *R* and represents the recombination frequency.

The mean *R* for pJH299 was 0.04 with a standard deviation between transfection experiments of 55% of the mean. To compensate for variability between experiments, transfections were done in groups (test substrates plus pJH299, in duplicate). The values obtained using test substrates (R_{test}) were normalized against the average value of the duplicate pJH299 controls to obtain R_n ($R_n = R_{test}/R_{pJH299} \times 100$). R_n values for most substrates were determined in at least two independent transfection experiments. The standard deviation for R_n between duplicates within a transfection experiment was ~40% of the mean for R_n values <3, 20% for R_n between 3 and 20, and 14% for R_n >20.

Acknowledgments

We thank Susanna Lewis for many fruitful discussions during the course of this work and for comments on the manuscript. M.R.L. is a Lucille P. Markey Scholar, and this work was supported in part by a grant from the Lucille P. Markey Charitable Trust. M.R.L. also acknowledges the support of the Council for Tobacco Research (grant 1867).

References

Akira, S., K. Okazaki, and H. Sakano. 1987. Two pairs of recombination signals are sufficient to cause immunoglobulin V-(D)-J joining. *Science* **238:** 1134–1138.

Blackwell, T.K. and F.W. Alt. 1988. Immunoglobulin genes. In *Molecular immunology* (ed. B.D. Hames and D.M. Glover), pp. 1-60. IRL Press, Washington, D.C.

Davis, M.M. 1988. T cell antigen receptor genes. In *Molecular immunology* (ed. B.D. Hames and D.M. Glover) pp. 61-79. IRL Press, Washington, D.C.

Durdik, J., M.W. Moore, and E. Selsing. 1984. Novel κ light-chain gene rearrangements in mouse λ light chain producing B lymphocytes. *Nature* **307:** 749–752.

Early, P., H. Huang, M. Davis, K. Calame, and L. Hood. 1980. An immunoglobulin heavy chain variable region gene is generated from three segments of DNA: V_H, D and J_H. *Cell* **19:** 981–992.

Hesse, J.E. 1989. "*V(D)J* recombination: Functional definition of the joining signals and studies of the mechanism." Ph.D. thesis. University of Chicago, Chicago, Illinois.

Hesse, J.E., M.R. Lieber, M. Gellert, and K. Mizuuchi. 1987. Extrachromosomal DNA substrates in pre-B cells undergo in-

Downloaded from genesdev.cshlp.org on August 22, 2022 - Published by Cold Spring Harbor Laboratory Press

version or deletion at immunoglobulin V-(D)-J joining signals. *Cell* **49:** 775–783.

Höchtl, J., C.R. Müller, and H.G. Zachau. 1982. Recombined flanks of the variable and joining segments of immunoglobulin genes. *Proc. Natl. Acad. Sci.* **79:** 1383–1387.

Höchtl, J. and H.G. Zachau. 1983. A novel type of aberrant recombination in immunoglobulin genes and its implications for V-J joining. *Nature* **302:** 260–263.

Kabat, E.A., T.T. Wu, M. Reid-Miller, H.M. Perry, and K.S. Gottesman. 1987. *Sequences of proteins of immunological interest*, fourth ed. National Institutes of Health, Bethesda.

Kelley, D.E., L.M. Wiedemann, A.-C. Pittet, S. Strauss, K.J. Nelson, J. Davis, B. Van Ness, and R.P. Perry. 1985. Nonproductive kappa immunoglobulin genes: Recombinational abnormalities and other lesions affecting transcription, RNA processing, turnover, and translation. *Mol. Cell Biol.* **5:** 1660–1675.

Kleinfeld, 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:** 843–845.

Korman, A.J., J. Maruyama, and D.H. Raulet. 1989. Rearrangement by inversion of a T cell receptor δ variable region gene located 3′ of the δ constant region gene. *Proc. Natl. Acad. Sci.* **86:** 267–271.

Lewis, S., A. Gifford, and D. Baltimore. 1984. Joining of V_K to J_K gene segments in a retroviral vector introduced into lymphoid cells. *Nature* **308:** 425–428.

———. 1985. DNA elements are asymmetrically joined during the site-specific recombination of kappa immunoglobulin genes. *Science* **228:** 677–685.

Lewis, S.M., J.E. Hesse, K. Mizuuchi, and M. Gellert. 1988. Novel strand exchanges in V(D)J recombination. *Cell* **55:** 1099–1107.

Lieber, M.R., J.E. Hesse, M. Mizuuchi, and M. Gellert. 1987. Developmental stage specificity of the lymphoid V(D)J recombination activity. *Genes Dev.* **1:** 751–761.

Lieber, M.R., J.E. Hesse, S. Lewis, G.C. Bosma, N. Rosenberg, K. Mizuuchi, M.J. Bosma, and M. Gellert. 1988. The defect in murine severe combined deficiency: Joining of signal sequences but not coding segments in V(D)J recombination. *Cell* **55:** 7–16.

Lorenz, W., B. Straubinger, and H.G. Zachau. 1987. Physical map of the human immunoglobulin K locus and its implications for the mechanisms of V_K-J_K rearrangement. *Nucleic Acids Res.* **15:** 9667–9676.

Malissen, M., C. McCoy, D. Blanc, J. Trucy, C. Devaux, A. Schmitt-Verhulst, F. Fitch, L. Hood, and B. Malissen. 1986. Direct evidence for chromosomal inversion during T-cell receptor β-gene rearrangements. *Nature* **319:** 28–33.

McKearn, J.P. and N. Rosenberg. 1985. Mapping cell surface antigens on mouse pre-B cell lines. *Eur. J. Immunol.* **15:** 295–298.

Moore, M.W., J. Durdik, D.M. Persiani, and E. Selsing. 1985. Deletions of κ chain constant region genes in mouse λ chain-producing B cells involve intrachromosomal DNA recombinations similar to V-J joining. *Proc. Natl. Acad. Sci.* **82:** 6211–6215.

Perlmutter, R.M., J.F. Kearney, S.P. Chang, and L.E. Hood. 1985. Developmentally controlled expression of immunoglobulin V_H genes. *Science* **227:** 1597–1601.

Reth, M.G., S. Jackson, and F.W. Alt. 1986a. V_HDJ_H formation and DJ_H replacement during pre-B differentiation: Nonrandom usage of gene segments. *EMBO J.* **5:** 2131–2138.

Reth, M., P. Gehrmann, E. Petrac, and P. Wiese. 1986b. 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:** 840–842.

Sakano, H., K. Hüppi, G. Heinrich, and S. Tonegawa. 1979. Sequences at the somatic recombination sites of immunoglobulin light-chain genes. *Nature* **280:** 288–294.

Seidman, J.G. and P. Leder. 1980. A mutant immunoglobulin light chain is formed by aberrant DNA- and RNA-splicing events. *Nature* **286:** 779–783.

Selsing, E., J. Voss, and U. Storb. 1984. Immunoglobulin gene 'remnant' DNA—Implications for antibody gene recombination. *Nucleic Acids Res.* **12:** 4229–4245.

Showe, L.C. and C.M. Croce. 1987. The role of chromosomal translocations in B- and T-cell neoplasia. *Annu. Rev. Immunol.* **5:** 253–277.

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.

Stavnezer, J., O. Kekish, D. Batter, J. Grenier, I. Balazs, E. Henderson, and B.J.M. Zegers. 1985. Aberrant recombination events in B cell lines derived from a κ-deficient human. *Nucleic Acids Res.* **13:** 3495–3514.

Steinmetz, M., W. Altenburger, and H.G. Zachau. 1980. A rearranged DNA sequence possibly related to the translocation of immunoglobulin gene segments. *Nucleic Acids Res.* **8:** 1709–1720.

Tonegawa, S. 1983. Somatic generation of antibody diversity. *Nature* **302:** 575–581.

Watson, J.D., N.H. Hopkins, J.W. Roberts, J.A. Steitz, and A.M. Weiner. 1987. Generation of immunological specificity. In *Molecular biology of the gene* fourth ed., pp. 832–897. Benjamin/Cummings, Menlo Park, California.

Yancopoulos, G.D. and F.W. Alt. 1985. Developmentally controlled and tissue-specific expression of unrearranged V_H gene segments. *Cell* **40:** 271–281.

———. 1986. Regulation of the assembly and expression of variable-region genes. *Annu. Rev. Immunol.* **4:** 339–368.

Yancopoulos, G.D., S.V. Desiderio, M. Paskind, J.F. Kearney, D. Baltimore, and F.W. Alt. 1984. Preferential utilization of the most J_H-proximal V_H gene segments in pre-B-cell lines. *Nature* **311:** 727–733.

Downloaded from genesdev.cshlp.org on August 22, 2022 - Published by Cold Spring Harbor Laboratory Press

V(D)J recombination: a functional definition of the joining signals.

J E Hesse, M R Lieber, K Mizuuchi, et al.

Genes Dev. 1989, **3:**
Access the most recent version at doi:10.1101/gad.3.7.1053

References This article cites 32 articles, 8 of which can be accessed free at:
http://genesdev.cshlp.org/content/3/7/1053.full.html#ref-list-1

License

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here.**

Copyright © Cold Spring Harbor Laboratory Press