Boundaries of Somatic Mutation in Rearranged Immunoglobulin Genes: 5′ Boundary Is Near the Promoter, and 3′ Boundary Is ~1 kb from V(D)J Gene

By Serge G. Lebecque and Patricia J. Gearhart

From the Department of Biochemistry, The Johns Hopkins University School of Hygiene and Public Health, Baltimore, Maryland 21205

Summary

To investigate why somatic mutations are spatially restricted to a region around the rearranged V(D)J immunoglobulin gene, we compared the distribution of mutations flanking murine V gene segments that had rearranged next to either proximal or distal J gene segments. 124 nucleotide substitutions, nine deletions, and two insertions were identified in 32,481 bp of DNA flanking the coding regions from 17 heavy and κ light chain genes. Most of the mutations occurred within a 2-kb region centered around the V(D)J gene, regardless of which J gene segment was used, suggesting that the structural information for mutation is located in sequences around and within the V(D)J gene, and not in sequences downstream of the J gene segments. The majority of mutations were found within 300 bp of DNA flanking the 5′ side of the V(D)J gene and 850 bp flanking the 3′ side at a frequency of 0.8%, which was similar to the frequency in the coding region. The frequency of flanking mutations decreased as a function of distance from the gene. There was no evidence for hot spots in that every mutation was unique and occurred at a different position. No mutations were found upstream of the promoter region, suggesting that the promoter delimits a 5′ boundary, which provides strong evidence that transcription is necessary to generate mutation. The 3′ boundary was ~1 kb from the V(D)J gene and was not associated with a DNA sequence motif. Occasional mutations were located in the nuclear matrix association and enhancer regions. The pattern of substitutions suggests that there is discrimination between the two DNA strands during mutation, in that the four bases were mutated with different frequencies on each strand. The high frequency of mutations in the 3′ flanking region and the uniqueness of each mutation argues against templated gene conversion as a mechanism for generating somatic diversity in murine V(D)J genes. Rather, the data support a model for random point mutations where the mechanism is linked to the transcriptional state of the gene.

Somatic mutation in rearranged V, D, and J gene segments encoding Igs can be described by temporal and spatial events. Concerning time, the enzymes that produce mutation are most active in recently stimulated B cells (1–3), and are less active in more differentiated cells (4–7). During the first week after immunization, mutation occurs at an extraordinary rate, estimated to be 10^{-3} mutations per base pair per generation (2), and operates during several cell divisions (8). Substitutions occur randomly throughout DNA containing rearranged V, D, and J gene segments for H and L chains, collectively termed V(D)J, and, with time, mutations are preferentially selected that change the amino acid codon in hypervariable regions to encode antibodies with higher affinity for the cognate antigen (1, 2, 6, 9–11). In fact, recent evidence suggests that mutation occurs in a subset of primary B cells that are destined to become high affinity memory cells (12). The transient generation of mutations will make it difficult to study the enzymatic activities that cause mutation.

Concerning space, the most striking enigma of mutation is its presence in a small region surrounding and including the rearranged V(D)J gene, and its absence in the region surrounding and including the C gene (13, 14). Since both V(D)J and C genes are transcribed by RNA polymerase as part of one transcript, both regions are presumably equally accessible to mutation enzymes. Therefore, there must be a structural basis, such as *cis* DNA sequences, that targets mutation to the V(D)J gene. Several theories could be proposed to explain the asymmetrical distribution of mutations. (1) Mutation may be initiated by DNA sequences in the intervening sequence between the J gene segments and the C gene, predicting that V gene segments rearranged to the first J gene segment will have mutations extending for a longer distance

 J. Exp. Med. © The Rockefeller University Press • 0022-1007/90/12/1717/11 $2.00
Volume 172 December 1990 1717–1727

that V gene segments rearranged to the last J gene segment. (2) Mutation may be initiated by the unique structure of a rearranged V(D)J gene, predicting that mutation would be centered around the V(D)J gene and extend the same distance regardless of which J gene segment was rearranged (3). The range of mutation around the V(D)J gene may be physically delimited by the promoter and enhancer regions, which interact with transcription proteins (reviewed in reference 15) to prevent mutation outside of this region.

These theories concerning the localization of mutations were addressed by sequencing 25,914 bp of 5′ and 3′ flanking sequences from 13 V_H and V_κ genes rearranged to different J gene segments and identifying somatic mutations. A comparison of the frequency and distance of the flanking mutations, in addition to flanking mutations found in 6,567 bp from genes in the literature (14, 16–18), indicated that most were located within a 2-kb region centered around the rearranged V(D)J gene, regardless of which J gene segment was used. The 5′ mutations did not extend beyond the promoter region, suggesting that transcription is necessary for mutation. Furthermore, the analysis of 135 unselected flanking mutations confirmed a previous analysis that showed that a different spectrum of mutations arise during mitosis compared with meiosis (19).

Materials and Methods

Genomic Libraries of H Chain Genes. The DNA in this study was obtained from the BALB/c strain of mice. The germline sequence of the unrearranged J_H locus was derived from a plasmid containing a 3-kb partial BamHI-EcoRI fragment from sperm DNA (provided by L. Hood, California Institute of Technology, Pasadena, CA). Genomic DNA encoding mutated antibodies was obtained from two groups of B cell lines. The first group secreted antiphosphorylcholine antibodies (20) and rearranged V gene segments from the V_HS107 family (13) next to J_H1: myelomas McPC603 and MOPC167 and hybridoma HPCG13 rearranged the V_H1 gene segment next to DFL16.1 and J_H1 to generate a 7.5-kb EcoRI fragment, and hybridoma HPCG15 rearranged to the V_H11 gene segment next to DFL16.1 and J_H1 to generate a 4.8-kb EcoRI fragment. DNA from MOPC167, HPCG13, and HPCG15 was digested with EcoRI, subjected to electrophoresis on agarose gels, and DNA in the appropriate size range was electroeluted. A viral clone containing the rearranged VDJ gene from myeloma McPC603 was obtained from L. Hood. The second group of hybridomas (provided by W. Gerhard, Wistar Institute, Philadelphia, PA.; reference 21) secreted antiinfluenza antibodies and rearranged V gene segments from the V_H7183 family next to J_H4: hybridomas H37-311, H37-45, and H37-80 rearranged the V_H76 gene segment (22) next to DFL16.2 and J_H4 to produce a 2.3-kb EcoRI fragment, and hybridoma H37-62 rearranged the V_HV2 gene segment (2, 21) next to DFL16.1 and J_H4 to produce a 2.3-kb EcoRI fragment. Genomic DNA was digested with EcoRI and size selected in this range.

Partial libraries of the EcoRI fragments were constructed in λ Zap II (Stratagene, La Jolla, CA) and plated on the *Escherichia coli* NM522 strain to avoid restriction of eucaryotic DNA by Eco K. Viral plaques were screened with a radiolabeled 2-kb BamHI-EcoRI probe containing J_H3 and J_H4. DNA from positive plaques was used to infect XL1-Blue bacteria, and the Bluescript plasmid was excised from λ Zap according to the manufacturer's directions.

Genomic Libraries of κ L Chain Genes. The germline sequence of the unrearranged J_κ locus was derived from a 6.5-kb EcoRI-BamHI fragment from embryo DNA (provided by P. Leder, Harvard University, Boston, MA). Genomic DNA encoding mutated antibodies was made from B cell lines secreting antiphosphorylcholine antibodies, and from the nonproductive allele of the SP2/0 hybridoma (23). Hybridoma HPCG15 rearranged an unknown V gene segment from the $V_\kappa8$ family next to $J_\kappa2$ to generate a 4.8-kb HindIII fragment, and hybridoma SP2/0 rearranged the $V_\kappa21E1.5$ gene segment (24) from the $V_\kappa21$ family next to $J_\kappa2$ to produce a 6.2-kb HindIII fragment. DNA in these respective size ranges was prepared and cloned into λ Charon 28; libraries were screened with a radiolabeled 1-kb XbaI-HindIII fragment from the intervening sequence downstream of $J_\kappa5$. Inserts from positive clones were subcloned into pTZ plasmid (Pharmacia LKB Biotechnology, Inc., Piscataway, NJ). Myeloma MOPC167 and hybridomas HPCG10 and HPCG13 rearranged the Vκ167 gene segment from the $V_\kappa24$ family next to $J_\kappa5$; M13 clones were obtained from a previous study (14).

DNA Sequencing and Assignment of Mutations. Single-stranded DNA was prepared from pBluescript, pTZ, and M13 clones, and sequenced by the dideoxy chain termination method. Sequencing was done with 20-nucleotide primers corresponding to sequences located at 250-nucleotide intervals along 3′-J_H and J_κ intervening sequences and 5′-V_H1 and V_H76 flanking sequences. To identify mutations, sequences were compared with germline sequences or consensus sequences of DNA derived from three independent cell lines. The convention for assigning mutations around V(D)J joining sites in the third hypervariable site of the coding region was described previously (2).

Results

Identification of Flanking Mutations. The strategy was to determine how far mutation extended when V gene segments rearranged to proximal vs. distal J gene segments. Both H chain and κ L chain genes were examined. To detect mutations in the 3′ flanking regions of the J_H and J_κ loci, the germline sequences were first confirmed. DNA clones from sperm and embryo were simultaneously sequenced with the corresponding regions from hybridomas to avoid ambiguity in determining the germline sequence and identifying mutations. Approximately 2.6 kb of DNA were sequenced for each of the J_H and J_κ loci; the germline sequences are presented in Fig. 1. Several differences were noted compared with previously published sequences (25–28), which emphasizes the importance of comparing sequences from at least three DNA samples in order to establish a germline or consensus sequence.

For H chain genes, two groups of antibodies were sequenced. The first group contained antiphosphorylcholine antibodies from four B cell lines that rearranged V gene segments from the V_HS107 family to J_H1: McPC603, MOPC167, HPCG13, and HPCG15. The second group included antiinfluenza antibodies from four hybridomas that rearranged V gene segments from the V_H7183 family to J_H4: H37-311, H37-45, H37-80, and H37-62. All of the V to D to J rearrangements were in the correct reading frame. To identify mutations on the 5′ side of the V_H1 gene from the V_HS107 family, the germline sequences reported by Clarke et al. (18) and Siu et al. (29) were used. A consensus sequence for the

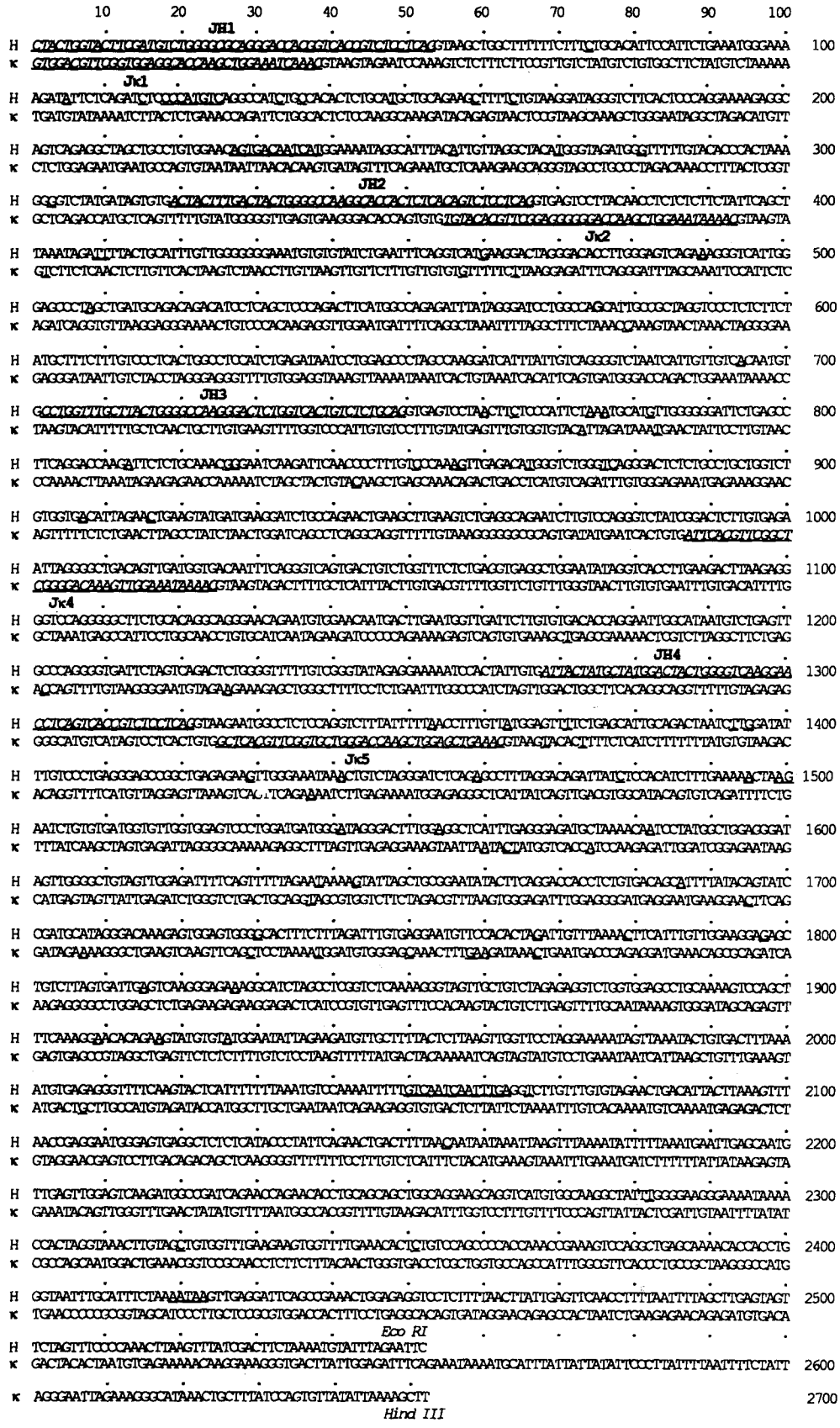

Figure 1. Germline sequences of the J_H and J_κ loci of the BALB/c mouse. H chain nucleotides are shown on the top line, and κ L chain sequences are listed on the bottom line. Numbering starts with the first nucleotide of J_H1 and $J\kappa1$. Exons are italicized, underlined, and noted above the top line for J_H gene segments and below the bottom line for $J\kappa$ gene segments. The location of mutations listed in Table 1 is underlined. EcoRI and HindIII sites at the end of the sequences are noted for H and κ introns, respectively. These sequence data are available from EMBL GenBank Data Libraries under accession number X53774 for the J_H locus and X53775 for the J_κ locus.

```
-100      -90       -80       -70       -60       -50       -40       -30       -20       -10
.         .         .         .         .         .         .         .         .         .
                                                            TGAGGAACATTCTATCACAAATAAGTAAATTCAGAAAATT  -701
CTGAGTGCTCCCATCACGGAGATGAATCTGCTATGAACAGCTCATAGGTGTGAGCCTCTACAAAAGCCATATTATTGAAAAGCCACATTGTGCCCAGACT  -601
TTGGAAAGACTGAGCTCATATCCTGAAATACAGTTATGTGTGGTTCTATTTAATTACACATTTACACTAAGGAAACATGGCAGTATGGGAATGAAGCTTG  -501
TTCTGTACACATTAACAGAGGGAAACTAAACAAAGTATGGTGAATCCCTAACCAAAAGTTAAAAAAAAAAAAAGAAAAAGAAAGAAAAAAGTGAACTACAA  -401
TATGTTTCAATGCTGTACTGAATCTGGTTTTGATGCCTTATATCTGGTATCATCAGTGACTTCAGATTGAGTCCAACTCCAGAGCATGGTATAGCAGGAA  -301
GACATGCAAATGAGTCTTCTCTCTGCCCATGAAAAACACCTTGGCCCTGACCCTGCAGCTCTGACAGAGGAGGCCTGTCCTGGATTCGATTCCCAGTTCC  -201
TCACATTCAGTCAGCACTGAACACGGACCCCTCACCATGAACTTCGGGCTCAGCTTGATTTTCCTTGTCCTTGTTTTAAAAGGTAATTTATTGAGAAGTG  -101
ATGACATCTGTTGTATGCACATGAGACAGAGAAAAATTGTTGTTTGTTTTGTTAGTGAGAGTTTTCCAACCAGTATTCTCTGTTTGCAGGTGTCCAGTGT    -1
```

Figure 2. 5′ consensus sequence of the V_H76 gene segment from the V_H7183 family. Negative numbering goes from 3′ to 5′, and position −1 corresponds to the first nucleotide 5′ of the first amino acid. The leader sequence is italicized and underlined. TATA (position −266), octamer (position −290), and heptamer (position −318) sequences are in bold. Positions where nucleotides have been mutated are underlined. These sequence data are available from EMBL GenBank Data Libraries under accession number X53776.

5′ region of the V_H76 gene from the V_H7183 family was derived from three rearranged genes and is shown in Fig. 2. The published sequence of MC101 (17), which rearranged V_HOx2 from the V_HQ52 family to J_H3, was also included in the analysis. The types of mutations are summarized in Table 1, their location is shown schematically in Fig. 3, and positions of the 5′ and 3′ mutations are underlined in Figs. 1 and 2. Each flanking mutation occurred at a different position. For the MOPC167 rearrangement, 14 mutations reported in reference 13 were confirmed, and 25 differences were found. In particular, nine substitutions, eight deletions, and two insertions reported in reference 13 were not found, and two new substitutions, three deletions, and one insertion were identified in this study. The variability may reflect errors in

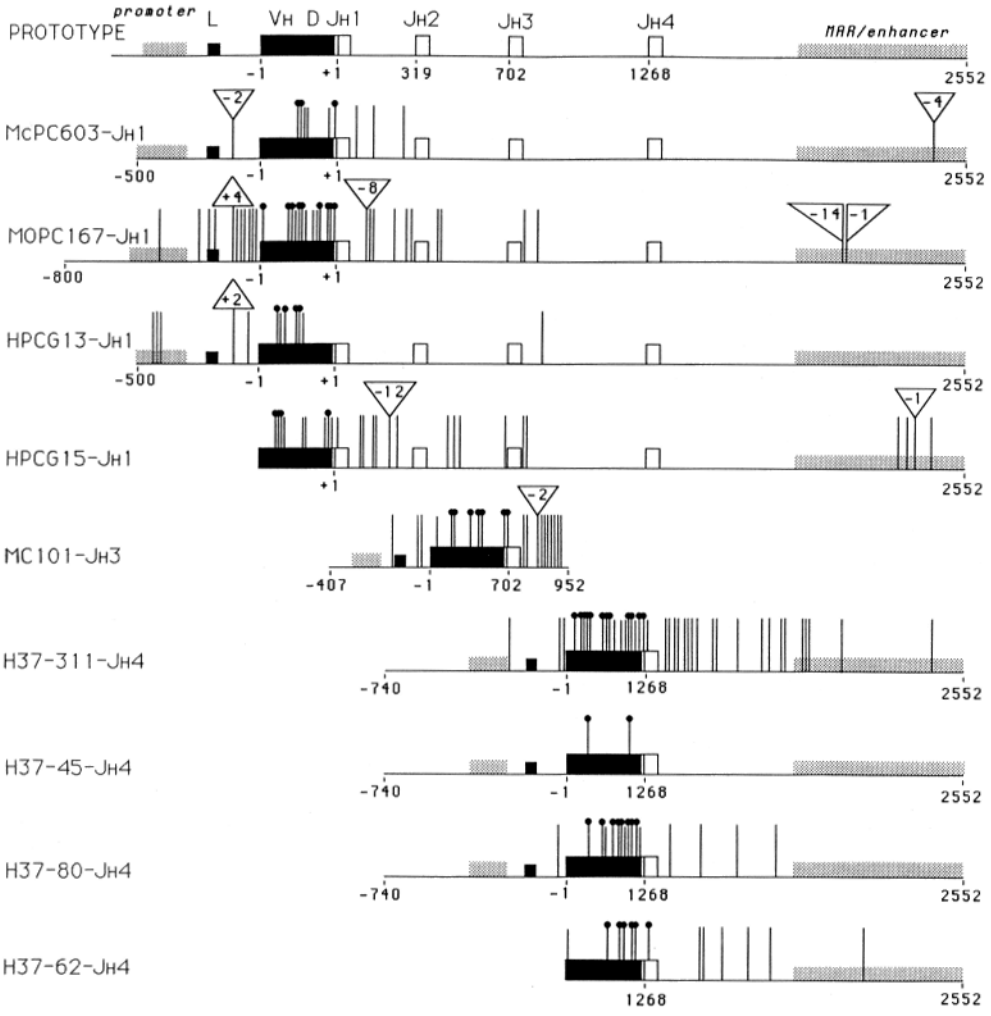

Figure 3. Location of mutations in rearranged V_H genes. The first line represents a prototype V gene segment rearranged to J_H1. Promoter, MAR, and enhancer regions are stippled. Leader (L) and V exons are denoted by black boxes; D and J exons are open boxes. Numbering corresponds to the convention in Figs. 1 and 2; the coding region is not numbered. Horizontal line for each rearranged gene represents the distance sequenced. Vertical lines show position of flanking mutations in Table 1 and coding region mutations; filled circles denote mutations causing an amino acid replacement in the coding region. Regular triangles with positive numbers represent insertions, and inverted triangles with negative numbers signify deletions.

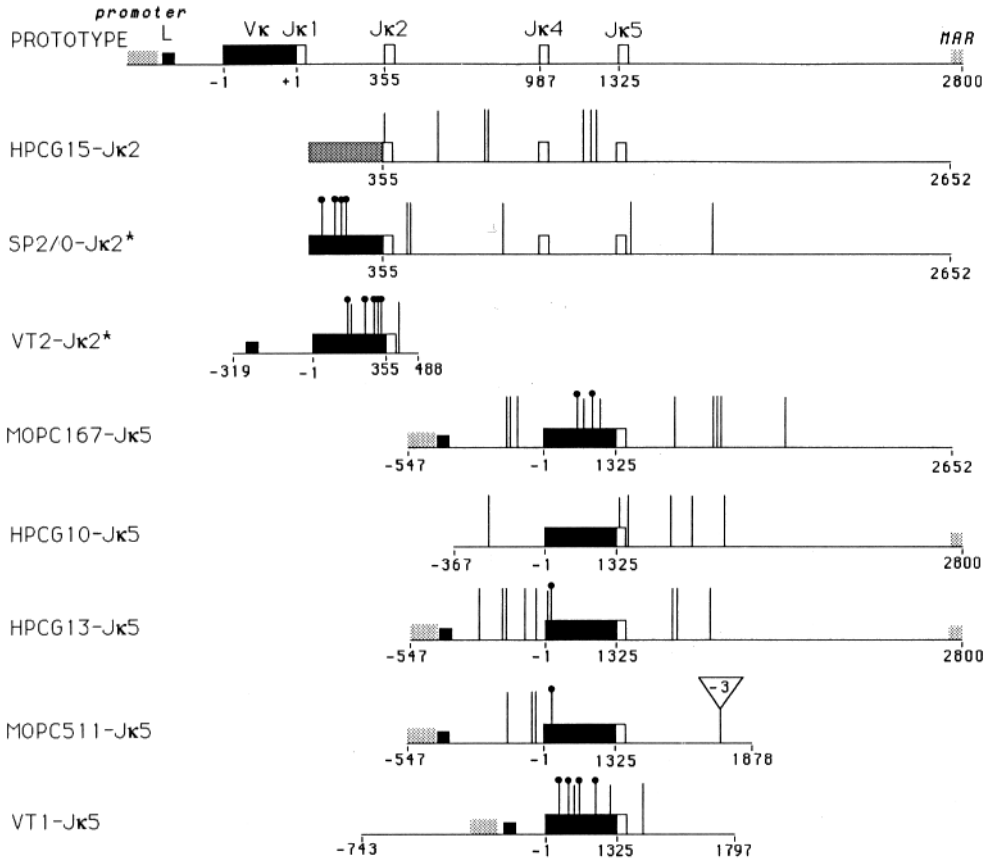

Figure 4. Location of mutations in rearranged V_κ genes. Conventions are the same as in Fig. 3. Stippled V gene segment in HPCG15 means the germline counterpart has not been sequenced. (*) Nonproductive rearrangement.

sequencing or indicate that the tumor continued to mutate in vivo, since different tumor isolates were used to clone the gene in both studies. Mutational drift has previously been observed in human lymphomas grown in vivo (30). Mutations in the coding region, also shown in Fig. 3, were identified by sequencing and compared with published data (13, 21).

For κ L chain genes, two hybridomas, HPCG15 and SP2/0, were studied that rearranged V gene segments next to $J_\kappa 2$, and three lines, MOPC167, HPCG10, and HPCG13, were analyzed that rearranged the $V_\kappa 167$ gene segment from the $V_\kappa 24$ family to $J_\kappa 5$. Sequences from the latter three clones were extended by 780 bp in the 3′ direction to search for mutation, in addition to previously reported substitutions (14). Also included in the analysis are published sequences for the VT1 gene segment from the $V_\kappa 9$ family rearranged to $J_\kappa 2$ (16), the VT2 gene segment from the $V_\kappa 23$ family rearranged to $J_\kappa 5$ (16), and the MOPC511 myeloma, which rearranged $V_\kappa 167$ next to $J_\kappa 5$ (14). Most of the V to J rearrangements were in the correct reading frame except for SP2/0 and VT2, which had two extra bases at the joining site. The location and types of mutations are shown in Fig. 4 and Table 1, and positions of the 3′ mutations are underlined in Fig. 1. Each flanking mutation is located at a unique position in the sequence, analogous to the H chain mutations. The coding region mutations in SP2/0 were determined by sequencing; other exon substitutions shown in Fig. 4 were taken from published data (14, 16).

Distance of Mutation. On the 5′ side, mutation was detected as far as 428 bp upstream of the first V region codon; most of the mutations, 32 of 36, occurred between the V(D)J gene and the promoter region, which is defined by the TATA, octamer, and heptamer sequences, and adjacent regions that bind transcription factors (15). On the 3′ side, mutation was detected as far away as 2,366 bp from the last codon of the rearranged V(D)J gene, but the majority of mutations, 90 of 99, occurred within 1 kb of the rearranged gene. In the H chain locus for McPC603, MOPC167, and HPCG15, three substitutions and four deletions were found >2 kb from the rearranged gene in the enhancer region and the matrix association region (MAR),[1] which anchors DNA to the nuclear matrix (31). It is not clear if the distal mutations are caused by the same mechanism that produces mutations proximal to the V(D)J gene, or are due to the reported instability of the region after cloning (32). In the κ chain locus, the frequency of mutations abated before the MAR (33) and enhancer.

Frequency of Mutation. The global frequency of flanking mutations calculated for all the nucleotides analyzed is 0.4% (135 mutations of 32,481 nucleotides sequenced). However, without knowledge of the actual borders of the region undergoing mutation, the meaning of this frequency is limited

[1] *Abbreviation used in this paper:* MAR, matrix association region.

Table 1. *Mutations Flanking Rearranged V(D)J Genes*

H chain	Nucleotide*	Mutation	H chain	Nucleotide	Mutation	L chain	Nucleotide	Mutation
M603	−132–133	Deletion	MC101	−149	A to G	HPCG15	579	C to T
	74	C to T		−52	C to A		773	A to G
	148	T to C		−35	A to T		783	T to A
	270	T to A		774	A to G		1171	T to A
	2420–2423	Deletion		782	G to C		1202	C to G
				826–827	Deletion		1226	A to G
M167	−406	T to A		851	C to A			
	−248	T to C		857	G to A	SP2/O	457	G to A
	−207	T to A		866	T to G		464	T to C
	−183	T to C		876	T to C		843	C to G
	−121	Insert 4 bp		877	C to T		1368	T to A
	−96	T to A		907	A to G		1695	C to A
	−74	T to G		916	C to A			
	−63	C to T				VT2	402	T to C
	−46	T to C	H37-311	−242	C to A			
	−19	T to C		−28	G to A	M167	−130	A to G
	−17	T to C		−9	G to A		−127	A to G
	118–125	Deletion		1353	A to C		−107	A to G
	133	C to T		1363	A to G		1563	C to T
	136	C to T		1393	T to G		1729	C to T
	234	A to T		1395	G to A		1738	T to G
	281	G to A		1429	G to A		1750	C to T
	303	G to A		1441	A to C		2007	G to A
	409	T to C		1459	A to T			
	410	T to C		1478	C to G	HPCG10	−228	C to G
	764	C to T		1541	A to G		1373	T to C
	813	A to T		1554	A to G		1564	T to A
	2050–2063	Deletion		1638	T to G		1637	T to G
	2066	Deletion		1730	G to C		1767	C to T
				1767	G to A			

continued

because mutations decrease with distance. To define the region that is most susceptible to mutation, the cumulative data from 17 H and L chain rearrangements reported in Figs. 3 and 4 were graphed as frequency of mutation vs. distance from the V(D)J coding region. As shown in Fig. 5, the frequency is relatively constant for 300 bp upstream and 850 bp downstream of V(D)J genes, such that mutations occurred at equal frequencies of 0.8% in both 5′ and 3′ regions. Thus, the number of mutations in these regions were used to calculate the following frequencies. Comparing flanking region mutations to coding region mutations, the frequencies differ by about twofold, with 0.8% (121 mutations/15,680 bp sequenced) occurring in the introns and 1.9% (105 mutations/5,600 bp) occurring in the exons of 350-bp length. The latter frequency is probably higher due to selection for replacement substitutions that change the protein sequence. If corrected for selection, the frequencies would likely be similar, implying that the hypermutation process operates at a high rate in the absence of selection. Comparing flanking region mutations in H chains vs. κ L chains, the frequency is slightly higher for H chain rearrangements (0.9%; 85 mutations/9,307 bp), than for κ L chain rearrangements (0.5%; 36 mutations/7,093 bp). Comparing flanking region mutations in κ L chains for productively vs. nonproductively rearranged alleles, the frequencies are approximately equivalent, with 0.5% (32 mutations/5,849 bp) for productive alleles and 0.3% (4 mutations/1,244 bp) for nonproductive alleles. The latter result confirms the conclusion of Roes et al. (34) that mutation occurs at similar frequencies on both productive and nonproductive alleles.

Types of Mutations. As listed in Table 1, 124 substitutions, nine deletions spanning 1–14 nucleotides, and two insertions

Table 1. *(continued)*

H chain	Nucleotide*	Mutation	H chain	Nucleotide	Mutation	L chain	Nucleotide	Mutation
HPCG13	−428	T to C		1815	A to C	HPCG13	−265	T to G
	−410	T to G		1827	A to T		−172	A to T
	−404	T to C		1909	A to C		−157	T to C
	−111	Insert 2 bp		1917	A to G		−81	A to G
	−45	A to G		1926	A to T		−33	C to A
	856	A to G		2058	A to T		1560	A to T
				2423	A to G		1574	A to C
HPCG15	105	A to C					1707	A to G
	115	C to A	H37-80	−31	C to T			
	159	C to G		1371	T to G	M511	−139	T to C
	164	C to A		1495	A to T		−26	T to A
	227–238	Deletion		1643	G to A		−23	C to T
	256	A to C		1797	G to C		1758–1760	Deletion
	460	G to A						
	489	A to C	H37-62	1499	A to G	VT1	1437	A to T
	508	A to G		1500	G to T			
	694	A to G		1582	A to C			
	760	A to G		1686	A to G			
	776	A to G		1779	C to T			
	2282	T to C		2155	C to T			
	2320	C to T						
	2351	Deletion						
	2419	A to T						

* Negative numbers correspond to position in 5′ flanking sequences so that −1 is the first nucleotide 5′ of the first amino acid (see Fig. 2 for example). Positive numbers correspond to numbering in Fig. 1 for 3′ flanking sequences.

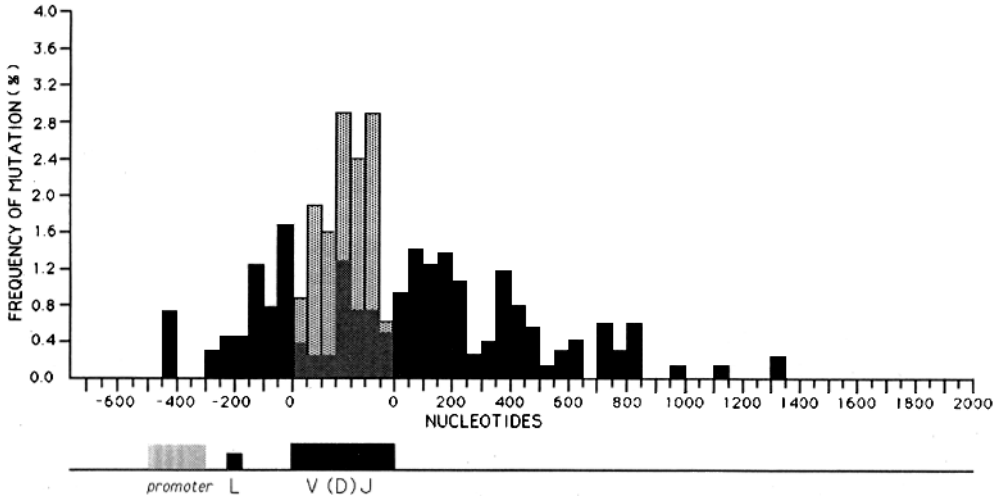

Figure 5. Frequency of mutation vs. distance from rearranged V(D)J gene. The data from 17 H and κ V genes rearranged to different J gene segments are compiled in this figure. Bottom drawing shows position of promoter, leader, and V(D)J gene. Abscissa of graph is numbered as follows: negative numbers correspond to 5′ flanking nucleotides; coding region is not numbered but represents 350 bp; and positive numbers correspond to 3′ flanking nucleotides. Ordinate designates the frequency of mutation per 50 nucleotide increments, calculated as the number of mutations in the increment, divided by 50, divided by the number of times the increment was sequenced, multiplied by 100. Black bars represent frequency of flanking mutations. The frequency in coding region is differentially shaded to indicate replacement and silent mutations: light stippled bars are replacement mutations and dark stippled bars are silent substitutions.

Table 2. *Pattern of Substitutions in Flanking Sequences*

From	To A	To T	To C	To G
A	-	9.2 ± 1.4	6.8 ± 1.1	16.2 ± 1.7
T	5.4 ± 1.1	-	11.1 ± 1.8	6.1 ± 0.7
C	7.9 ± 1.2	17.7 ± 2.3	-	5.8 ± 1.1
G	10.6 ± 1.1	0.7 ± 0.4	2.5 ± 0.5	-

The data are corrected for base composition, so that the numbers represent percent substitutions expected from a sequence with equal numbers of A, T, C, and G. The SD is calculated within 95% confidence limits.

of two and four nucleotides were observed. The pattern of substitutions is shown in Table 2, and the data were analyzed as follows. Mutations were scored from the sense strand of DNA so that an A to T substitution means that an A:T pair was replaced by a T:A pair. Substitutions were corrected for the base composition of each sequence according to Gojobori et al. (35), so that all four bases would be equally frequent. This was necessary because C is underrepresented in Ig flanking sequences. Mean percentages were weighted by the number of mutations per gene, and the SD was calculated within 95% confidence limits in a sample size of 16 sequences. Transitions, 56%, occurred slightly more frequently that transversions, 44%. None of the C to T changes occurred at a CG dinucleotide in the germline sequence, which is a site where C is preferentially methylated (36) and may undergo subsequent chemical deamination to thymidine (37). Asymmetry of mutation, indicative of strand bias, was observed in nearly every category; e.g., A to T occurred more frequently on one strand than it did on the opposite strand, as measured by a T to A change. Strand discrimination was particularly evident in C to A mutations, which occurred 10 times more frequently than corresponding G to T substitutions. Conversely, if scored from the antisense strand, G mutated to T significantly more frequently than the corresponding C to A replacement.

Discussion

Mutations Are Centered around the V(D)J Gene. The spatial restriction of somatic mutations implies that *cis* sequences direct the mutational mechanism to rearranged V(D)J genes. One possibility is that there are conserved DNA sequences near the J gene segments that initiate mutation, since there are only four J gene segments compared with several hundred V gene segments. Another possibility is that the V(D)J gene provides the structural information to target mutations. To test these two hypotheses, we compared the distribution of mutations in DNA adjacent to V gene segments that had rearranged to either proximal (J_H1, $J_\kappa2$) or distal (J_H4, $J_\kappa5$) J gene segments. The first theory predicts that mutations will occur next to all the J gene segments, whereas the second theory predicts that mutations will predominantly be found around the J gene segment attached to the V gene segment. As shown in Figs. 2 and 3, and summarized in Fig. 5, our analysis strongly supports the second interpretation that the V(D)J gene is the epicenter of mutation, with the frequency decreasing progressively in both 5′ and 3′ directions. This was most striking when comparing the range of mutation in V_H gene segments rearranged to J_H1 vs. J_H4 in Fig. 3. DNA located 500 bp downstream of J_H4 had no mutation when a V_H gene segment rearranged distally to J_H1 in the antiphosphorylcholine cell lines, but the same region had many mutations when a V_H gene segment rearranged to J_H4 in the H37 hybridomas. Thus, the range of mutation depended on its distance from the V(D)J gene, suggesting that the structural information for mutation is located in sequences surrounding and including the V(D)J gene, rather than in sequences downstream of the J gene segments.

Promoter May Delimit the 5′ Boundary. The distribution of flanking mutations in Fig. 5 implied the existence of a 5′ boundary, since the 5′ mutations extended for a shorter distance than the 3′ mutations. Most of the 5′ mutations occurred between the coding region and the promoter region; four were located within the promoter; and none were found within 350 bp upstream of the promoter from five genes that were extensively sequenced. The promoter boundary thus provides strong evidence that mutation is related to the transcriptional state of the V(D)J gene. Since a plethora of proteins bind to sequences within the promoter, it is possible that transcription proteins are physical obstacles that interrupt the mutation process. Alternatively, transcription itself may be required for mutation; for example, as RNA polymerase unwinds and exposes the two DNA strands as single-stranded regions during transcription, they may become accessible to enzymes generating mutations. The presence of four mutations 5′ of the transcription start site in V_H1 H chain genes from the MOPC167 and HPCG13 does not support a recent hypothesis that mutations are introduced into DNA via RNA transcripts (38).

3′ Boundary is ~1 kb from the V(D)J Gene. Because the 5′ boundary is near the promoter region, it was conceivable that the 3′ boundary may be near the enhancer region, which also binds transcription factors. MAR may also act as a physical boundary since it anchors DNA to the nuclear matrix and divides the V(D)J and C genes into separate domains. However, unlike the 5′ flanking region where mutations abruptly terminated in the promoter, there was no DNA sequence motif in the 3′ flanking region that denoted a boundary. Instead, there was a gradient of mutation, so that mutations decreased with increasing distance from the V(D)J gene over a distance of ~1 kb. In the H chain locus, mutations appeared in the enhancer and MAR, which are located 500 bp from J_H gene segments and are within the range of mutation. In the κ chain locus, MAR and the enhancer are situated >1,000 bp from J_κ gene segments, and it is unlikely that they will contain mutations because they are so far away. In fact, a recent analysis of κ transgenes showed no mutations in DNA containing the MAR, enhancer, and C_κ gene (28). Thus, although there were a few mutations that occurred

at a greater distance, the majority of 3′ flanking mutations were located within 1 kb of the V(D)J gene, which implies that the mutational machinery operates over a limited distance and then stops.

High Frequency of 3′ Flanking Mutations Excludes Gene Conversion as a Genetic Mechanism. The pattern of mutation around murine V(D)J genes is markedly different from chicken V(D)J genes, which undergo gene conversion in the coding region (39). In murine genes, 50% of the mutations occurred in the coding region and 50% were found in the 3′ flanking region (Fig. 5), whereas in chicken genes, 95% of the mutations were in the coding region, and 5% were in the 3′ flanking region (39, 40). Therefore, the strongest argument against gene conversion generating somatic mutations in mice is the high frequency of mutations in the 3′ flanking region, which is present as a single copy gene and has no homologous donor for conversion. By the same reasoning, it has been proposed that the 3′ flanking mutations in chicken Ig genes are not due to gene conversion, but are the result of point mutations (40). Furthermore, every mutation in both 5′ and 3′ murine flanking sequences was located at a unique position; there were no hot spots and no repeated mutations, which are characteristic of template-induced gene conversion. Although there is a possibility that gene conversion could produce somatic mutations at a low frequency in the coding region, the equivalent high frequencies of mutation in coding and flanking regions of murine genes implies that one mechanism is acting on both regions, and it is not gene conversion.

Mutations May Occur Preferentially on One DNA Strand. We have previously reported that the types of mutations that occur somatically in Ig genes represent a distinctly different pattern compared with mutations that arise during meiosis (19). The pattern of 124 substitutions in the flanking regions, summarized in Table 2, confirms and extends the analysis with the following observations. First, somatic mutations do not predominantly reflect the chemical deamination of 5-methylcytosine to thymidine, which comprise almost half of the mutations that arise during meiosis (41). Thus, somatic mutations are not predominantly due to a chemical process, but more likely represent errors introduced by DNA polymerases during repair or replication. The high error rate may result from a known polymerase working under suboptimal conditions, or result from many rounds of repair. Second, the somatic mutational process appears to affect the two DNA strands asymmetrically in that the four bases are not mutated equally on each strand (2), whereas meiotic mutation affects the four bases with the same frequency on both strands (41). There are conflicting reports as to whether strand discrimination is found among mutations in introns (42, 43). The bias in flanking region mutations could be caused by local nucleotide sequences, which are different on each strand, or could be due to preferential mutation or repair on the sense or antisense strand during transcription (44), DNA replication (45), or repair by methylation (46). Strand discrimination is consistent with the apparent requirement for transcription in generating mutations, since both strands are handled differently during transcription.

Model for Somatic Mutation in V(D)J Genes. Mutation in murine antibodies appears to be due to point mutations that arise during DNA repair or replication (14, 47, 48). However, molecular details of the mutational process remain unknown and must take into account the following observations. (*a*) The majority of mutations on the 5′ side are located downstream of the promoter, suggesting that transcriptional machinery is involved in generating mutations. (*b*) Mutations are centered around the V(D)J gene irrespective of which J gene segment is used, implying that the structural information for mutation is located in sequences surrounding and including the rearranged gene rather than in sequences downstream of the J gene segments. (*c*) Mutations occur at approximately the same frequency in the 5′ flanking region, the coding region, and the 3′ flanking region. (*d*) Every mutation is unique and occurs at a different position.

We can propose a model that accounts for most of the observations presented in this paper. As transcription starts at the promoter site, RNA polymerase unwinds the DNA helix and exposes short single-stranded regions that become accessible to enzymes causing mutation. Unwinding may also cause torsional stress in the V(D)J domain, which is anchored to the nuclear matrix by the MAR. Torsional stress can induce local regions to adopt cruciform configurations (49); we and others have previously proposed that transient palindromes, which are in greater abundance around the V(D)J gene than in random sequences, could be sites for mismatch repair (19, 50). Alternatively, inverted repeat sequences could act as pause sites, causing DNA polymerase to stall and insert wrong nucleotides. Mutations occur at a greater frequency near the V(D)J gene because the V gene segment or a combination of the V, D, and J gene segments encodes the information, i.e., secondary structures, to cause mutation. Unrearranged V gene segments may contain the structural information but do not mutate because they do not undergo a high rate of transcription, which is achieved only after rearrangement brings the V gene in proximity to the enhancer. Unrearranged J gene segments do not contain sufficient information because they do not mutate at a high frequency even when they are transcribed, unless a V gene segment is nearby. For example, in the antiphosphorylcholine cell lines shown in Fig. 3, unrearranged J_H4 gene segments distal to highly mutated VDJ genes had no mutation, although they were transcribed at the same rate as mutated VDJ genes. Transcription may cause preferential mutation on one of the DNA strands by making it more accessible to mutation enzymes. However, although transcription may be necessary for mutation, it is not sufficient, since V(D)J genes do not mutate at a high frequency in tissue culture cell lines where they are transcribed; other unknown factors and enzymes must be involved. Unique mutations are introduced by DNA polymerase during repair or replication; eucaryotic polymerases are known to produce predominantly substitutions and deletions, in accord with the mutational spectrum for somatic mutations. Experimental proof for these theories will require systems where genetically altered V(D)J genes can undergo mutation and the enzymes can be studied.

We thank D'Ann Caron and Jia-Jia Mai for technical assistance, Walter Gerhard for antiinfluenza hybridomas, and Nina Levy, Peter Schweitzer, and Asad Umar for insightful comments on the manuscript.

This work was supported by American Cancer Society grant IM491 and National Institutes of Health grant GM-42975. S. Lebecque was a visiting Fulbright Scholar from Belgium.

Address correspondence to Patricia J. Gearhart, Department of Biochemistry, The Johns Hopkins University School of Hygiene and Public Health, 615 N. Wolfe Street, Baltimore, MD 21205.

Received for publication 9 July 1990 and in revised form 24 August 1990.

References

1. Griffiths, G.M., C. Berek, M. Kaartinen, and C. Milstein. 1984. Somatic mutation and the maturation of immune response to 2-phenyl oxazolone. *Nature (Lond.).* 312:271.
2. Levy, N.S., U.V. Malipiero, S.G. Lebecque, and P.J. Gearhart. 1989. Early onset of somatic mutation in immunoglobulin V_H genes during the primary immune response. *J. Exp. Med.* 169:2007.
3. Manser, T. 1989. Evolution of antibody structure during the immune response. The differentiative potential of a single B lymphocyte. *J. Exp. Med.* 170:1211.
4. French, D.L., R. Laskov, and M.D. Scharff. 1989. The role of somatic hypermutation in the generation of antibody diversity. *Science (Wash. DC).* 244:1152.
5. Siekevitz, M., C. Kocks, K. Rajewsky, and R. Dildrop. 1987. Analysis of somatic mutation and class switching in naive and memory B cells generating adoptive primary and secondary responses. *Cell.* 48:757.
6. Clafin, J.L., J. Berry, D. Flaherty, and W. Dunnick. 1987. Somatic evolution of diversity among anti-phosphocholine antibodies induced with *Proteus morganii. J. Immunol.* 138:3060.
7. Schlomchik, M.J., A. Marshak-Rothstein, C.B. Wolfowicz, T.L. Rothstein, and M.G. Weigert. 1987. The role of clonal selection and somatic mutation in autoimmunity. *Nature (Lond.).* 328:805.
8. McKean, D., K. Huppi, M. Bell, L. Staudt, W. Gerhard, and M. Weigert. 1984. Generation of antibody diversity in the immune response of BALB/c mice to influenza virus hemagglutinin. *Proc. Natl. Acad. Sci. USA.* 81:3180.
9. Rodwell, J.D., P.J. Gearhart, and F. Karush. 1983. Restriction in IgM expression. IV. Affinity analysis of monoclonal anti-phosphorylcholine antibodies. *J. Immunol.* 130:313.
10. Wysocki, L., T. Manser, and M.L. Gefter. 1986. Somatic evolution of variable region structures during an immune response. *Proc. Natl. Acad. Sci. USA.* 83:1847.
11. Kocks, C., and K. Rajewsky. 1988. Stepwise intraclonal maturation of antibody affinity through somatic hypermutation. *Proc. Natl. Acad. Sci. USA.* 85:8206.
12. Linton, P.-J., D.J. Decker, and N.R. Klinman. 1989. Primary antibody-forming cells and secondary B cells are generated from separate precursor cell subpopulations. *Cell.* 59:1049.
13. Kim, S., M. Davis, E. Sinn, P. Patten, and L. Hood. 1981. Antibody diversity: somatic hypermutation of rearranged V_H genes. *Cell.* 27:573.
14. Gearhart, P.J., and D.F. Bogenhagen. 1983. Clusters of point mutations are found exclusively around rearranged antibody variable genes. *Proc. Natl. Acad. Sci. USA.* 80:3439.
15. Calame, K., and S. Eaton. 1988. Transcriptional controlling elements in the immunoglobulin and T cell receptor loci. *Adv. Immunol.* 43:235.
16. Pech, M., J. Hochtl, H. Schnell, and H.G. Zachau. 1981. Differences between germ-line and rearranged immunoglobulin V_κ coding sequences suggest a localized mutation mechanism. *Nature (Lond.).* 291:668.
17. Kataoka, T., T. Nikaido, T. Miyata, K. Moriwaki, and T. Honjo. 1982. The nucleotide sequences of rearranged and germline immunoglobulin V_H genes of a mouse myeloma MC101 and evolution of V_H genes in mouse. *J. Biol. Chem.* 257:277.
18. Clarke, C., J. Berenson, J. Goverman, P.D. Boyer, S. Crews, G. Siu, and K. Calame. 1982. An immunoglobulin promoter region is unaltered by DNA rearrangement and somatic mutation during B-cell development. *Nucleic Acids Res.* 10:7731.
19. Golding, G.B., P.J. Gearhart, and B.W. Glickman. 1987. Patterns of somatic mutations in immunoglobulin variable genes. *Genetics.* 115:169.
20. Gearhart, P.J., N.D. Johnson, R. Douglas, and L. Hood. 1981. IgG antibodies to phosphorylcholine exhibit more diversity than their IgM counterparts. *Nature (Lond.).* 291:29.
21. Caton, A.J., G.G. Brownlee, L.M. Staudt, and W. Gerhard. 1986. Structural and functional implications of a restricted antibody response to a defined antigenic region on the influenza virus hemagglutinin. *EMBO (Eur. Mol. Biol. Organ.) J.* 5:1577.
22. Riley, S.C., S.J. Connors, N.R. Klinman, and R.T. Ogata. 1986. Preferential expression of variable region heavy chain gene segments by predominant 2,4-dinitrophenyl-specific BALB/c neonatal antibody clonotypes. *Proc. Natl. Acad. Sci. USA.* 83:2589.
23. Shulman, M., C.D. Wilde, and G. Kohler. 1978. A better cell line for making hybridomas secreting specific antibodies. *Nature (Lond.).* 276:269.
24. Heinrich, G., A. Traunecker, and S. Tonegawa. 1984. Somatic mutation creates diversity in the major group of mouse immunoglobulin κ light chains. *J. Exp. Med.* 159:417.
25. Sakano, H., R. Maki, Y. Kurosawa, W. Roeder, and S. Tonegawa. 1980. Two types of somatic recombination are necessary for the generation of complete immunoglobulin heavy-chain genes. *Nature (Lond.).* 286:676.
26. Gough, N.M., and O. Bernard. 1981. Sequences of the joining region genes for immunoglobulin heavy chains and their role in generation of antibody diversity. *Proc. Natl. Acad. Sci. USA.* 78:509.
27. Max, E.E., J.V. Maizel, and P. Leder. 1981. The nucleotide sequence of a 5.5 kilobase DNA segment containing the mouse κ immunoglobulin J and C region genes. *J. Biol. Chem.* 256:5116.

28. Hackett, J., B.J. Rogerson, R.L. O'Brien, and U. Storb. 1990. Analysis of somatic mutations in κ transgenes. *J. Exp. Med.* 172:131.
29. Siu, G., E.A. Springer, H.V. Huang, L.E. Hood, and S.T. Crews. 1987. Structure of the T15 V_H gene subfamily: identification of immunoglobulin gene promoter homologies. *J. Immunol.* 138:4466.
30. Cleary, M.L., T.C. Meeker, S. Levy, E. Lee, M. Trela, J. Sklar, and R. Levy. 1986. Clustering of extensive somatic mutations in the variable region of an immunoglobulin heavy chain gene from a human B cell lymphoma. *Cell.* 44:97.
31. Cockerill, P.N., M. Yuen, and W.T. Garrard. 1987. The enhancer of the immunoglobulin heavy chain locus is flanked by presumptive chromosomal loop anchorage elements. *J. Biol. Chem.* 262:5394.
32. Marcu, K.B., J. Banerji, N.A. Penncavage, R. Lang, and N. Arnheim. 1980. 5' Flanking region of immunoglobulin heavy chain constant region genes displays length heterogeneity in germlines of inbred mouse strains. *Cell.* 22:187.
33. Cockerill, P.N., and W. Garrard. 1986. Chromosomal loop anchorage of the kappa immunoglobulin gene occurs next to the enhancer in a region containing topoisomerase II sites. *Cell.* 44:273.
34. Roes, J., K. Huppi, K. Rajewsky, and F. Sablitzky. 1989. V gene rearrangement is required to fully activate the hypermutation mechanism in B cells. *J. Immunol.* 142:1022.
35. Gojobori, T., W. Li, and D. Graur. 1982. Patterns of nucleotide substitution in pseudogenes and functional genes. *J. Mol. Evol.* 18:360.
36. Razin, A., and A.D. Riggs. 1980. DNA methylation and gene function. *Science (Wash. DC).* 210:604.
37. Coulondre, C., J.H. Miller, P.J. Farabaugh, and W. Gilbert. 1978. Molecular basis of base substitution hotspots in *Escherichia coli*. *Nature (Lond.).* 274:775.
38. Steele, E.J., and J.W. Pollard. 1987. Hypothesis: somatic hypermutation by gene conversion via the error prone DNA → RNA → DNA information loop. *Mol. Immunol.* 24:667.
39. Reynaud, C.-A., V. Anquez, H. Grimal, and J.-C. Weill. 1987. A hyperconversion mechanism generates the chicken light chain preimmune repertoire. *Cell.* 48:379.
40. Parvari, R., E. Ziv, F. Lantner, D. Heller, and I. Schechter. 1990. Somatic diversification of chicken immunoglobulin light chains by point mutations. *Proc. Natl. Acad. Sci. USA.* 87:3072.
41. Li, W.-H., C.-I. Wu, and C.-C. Luo. 1984. Nonrandomness of point mutation as reflected in nucleotide substitutions in pseudogenes and its evolutionary implications. *J. Mol. Evol.* 21:58.
42. T. Manser. 1990. Regulation, timing and mechanism of antibody V gene somatic hypermutation: lessons from the arsonate system. *In* Somatic Hypermutation in V-Regions. E.J. Steele, editor. CRC Press, Inc., Boca Raton, FL. In press.
43. Kaartinen, M., S. Kulp, and O. Makela. 1990. Characteristics of selection-free mutations and effects of subsequent selection. *In* Somatic Hypermutation in V-Regions. E.J. Steele, editor. CRC Press, Inc., Boca Raton, FL. In press.
44. Mellon, I., G. Spivak, and P.C. Hanawalt. 1987. Selective removal of transcription-blocking DNA damage from the transcribed strand of the mammalian DHFR gene. *Cell.* 51:241.
45. Wu, C., and N. Maeda. 1987. Inequality in mutation rates of the two strands of DNA. *Nature (Lond.).* 327:169.
46. Hare, J.T., and J.H. Taylor. 1985. One role for DNA methylation in vertebrate cells is strand discrimination in mismatch repair. *Proc. Natl. Acad. Sci. USA.* 82:7350.
47. Chien, N.C., R.R. Pollock, C. Desaymard, and M.D. Scharff. 1988. Point mutations cause the somatic diversification of IgM and IgG2a antiphosphorylcholine antibodies. *J. Exp. Med.* 167:954.
48. Wysocki, L.J., M.L. Gefter, and M.N. Margolies. 1990. Parallel evolution of antibody variable regions by somatic processes: consecutive shared somatic alterations in V_H genes expressed by independently generated hybridomas apparently acquired by point mutation and selection rather than by gene conversion. *J. Exp. Med.* 172:315.
49. Panayotatos, N., and R.D. Wells. 1981. Cruciform structures in supercoiled DNA. *Nature (Lond.).* 289:466.
50. Kolchanov, N.A., V.V. Solovyov, and I.B. Rogozin. 1987. Pecularities of immunoglobulin gene structures as a basis for somatic mutation emergence. *FEBS (Fed. Eur. Biochem. Soc.) Lett.* 214:87.