MOLECULAR AND CELLULAR BIOLOGY, Oct. 1990, p. 5187–5196
0270-7306/90/105187-10$02.00/0

Vol. 10, No. 10

Distribution of Mutations around Rearranged Heavy-Chain Antibody Variable-Region Genes

GERALD W. BOTH,[1] LINDA TAYLOR,[1] JEFFREY W. POLLARD,[2] AND EDWARD J. STEELE[3]*

Laboratory of Molecular Biology, Division of Biomolecular Engineering, Commonwealth Scientific and Industrial Research Organisation, North Ryde, NSW 2113, Australia[1]; Department of Developmental Biology and Cancer, Albert Einstein College of Medicine of Yeshiva University, Bronx, New York[2]; Department of Biology, University of Wollongong, NSW 2500, and Division of Cell Biology, John Curtin School of Medical Research, Canberra, ACT 2600, Australia[3]

Received 24 May 1990/Accepted 20 July 1990

The mechanism of somatic hypermutation in the variable region of immunoglobulin genes expressed in mammalian B cells is a major unexplained phenomenon in the generation of diversity in the immune system. To evaluate possible mechanisms, the distribution of somatic mutations was examined for a group of five cloned, rearranged, somatically mutated V_H genes generated in C57BL/6j mice. These mutated V_H genes were sequenced and compared with their germ line counterparts from a point approximately 550 base pairs upstream of the transcription start site to an *Eco*RI site some 1,200 base pairs downstream of J_{H-4}. The location of the transcription start (cap) sites was also precisely determined. Most (≥94%) of the 118 mutations scored occurred between the transcription start site and the distal end of J_{H-4}. However, seven mutations occurred upstream of the transcribed region, and at least four were found downstream of J_{H-4}. The target region for the mutator mechanism therefore clearly extends into the 3′ nontranslated and 5′ nontranscribed regions. Thus, models which propose the transcribed region of the DNA as the sole substrate for the mutation process are not ruled out but are inadequate to explain the upstream distribution of somatic mutations.

The DNA sequences both within and flanking rearranged antibody variable-region genes (V_LJ_L and V_HDJ_H) expressed in B lymphocytes can mutate at a very high rate (14; for recent reviews, see reference 28). This type of genetic variation is termed somatic hypermutation to distinguish it from the other more familiar processes of immunoglobulin diversification, such as junctional diversity, combinatorial recombination, and the association of fully assembled heavy and light polypeptide chains.

The mechanism of hypermutation in mammalian B cells is unknown, but several models have been proposed. These can be segregated into those models in which the DNA is the direct substrate for mutation (5, 8, 14, 17), including gene conversion (26), or those in which the DNA becomes hypermutable as a consequence of gene expression (i.e., transcription-based models [41]). In the first type of model, the DNA is mutated directly either via specific nicking and error-prone repair enzymes (8, 14), through errors introduced by misalignment of replicating DNA templates directed by inverted or direct repeat sequences (17), or via localized amplification of DNA replication which generates errors at the normal rate but in greater number in a localized region (5). Alternatively, the model invoking gene conversion envisages that sequence information donated by other related genes is incorporated by recombination into the rearranged V gene (26). However, these DNA-based models do not state how the enzymes display the specificity which limits the mutational process to the DNA within and around rearranged V genes (41). Known DNA polymerases also display a very high copying fidelity in vivo (24, 31, 32). Thus, the DNA-based models also require the presence of new enzymes with V-region specificity.

A second type of model based on gene expression accounts for many of the features of the somatic hypermutation process (41; E. J. Steele, J. W. Pollard, L. Taylor, and G. W. Both, *in* E. J. Steele, ed., *Somatic Hypermutation in V-regions*, in press). In brief, RNA polymerase and reverse transcriptase create errors during the transfer of genetic information via a DNA → RNA → DNA copying loop. The mutated cDNA is then either integrated directly into the normal chromosomal site for that allele or first subjected to further cycles of transcription and reverse transcription on the episomal cDNA element. Whatever the mutator mechanism, long-lived memory B cells may be exposed multiple times to the mutator process in the microenvironment of a germinal center (25; A. L. M. Bothwell, W. Tao, and P. R. Blier, *in* E. J. Steele, ed., *Somatic Hypermutation in V-Regions*, in press; D. Gray, H. Skarvall, Y.-J. Lui, I. C. M. MacLennan, and T. Leanderson, *in* E. J. Steele, *Somatic Hypermutation in V-Regions*, in press).

Existing data indicate that somatic mutations appear localized to the rearranged V gene and its immediate transcribed, noncoding regions (15, 18, 20, 22, 30, 39). They are not found in constant region sequences (15) and are found rarely in unrearranged germ line V genes (44). It is important now to precisely define the distribution of somatic mutations in and around rearranged V genes, particularly in the 5′ upstream region, because this will help to identify the target region of the DNA and therefore influence ideas on the type of mutational mechanism involved. In this work, we contribute to a better definition of this target region by comparing DNA sequences of related sets of hypermutated V_H genes with their germ line counterparts.

MATERIALS AND METHODS

Cloning and sequencing rearranged somatically mutated V_H genes. Genes derived from the V_H186.2 germ line gene

* Corresponding author.

family involved in the anti-(4-hydroxy-3-nitrophenyl)acetyl (NP) response in C57BL/6j mice (6) were cloned from DNA isolated from hybridomas expressing their products (33, 34). The coding region of these genes was partially sequenced from the mRNA (10). The V_H genes from A6/24 and A20/44 are involved in the anti-idiotypic response (in C57BL/6j mice) to V_H186.2-containing anti-NP antibodies. These genes are members of the V_H205.12 germ line gene family and were also cloned from the DNA of hybridomas expressing these genes and sequenced through their coding regions (34, 35). Maps of the V_H186.2 and V_H205.12 genes are shown in Fig. 1. V_H186.2 DNA clones 3B44, 3B62, and 40.3 were provided as 4.2-, 4.2-, and 4.6-kilobase (kb) *Eco*RI fragments in pUC19 (33, 40). The V_H205.12 clones A20/44 and A6/24 were similarly obtained as 7.2-kb *Eco*RI fragments in pBR328 and λgt WES:λB vectors, respectively (33). The V_H186.2 inserts were excised with *Eco*RI and subcloned into the plasmid vector Bluescribe M13(+) (Stratagene, San Diego, Calif.) in the T3 orientation. Single-stranded DNA corresponding to the complete plus strand of each clone was rescued by using helper phage M13K07 (IBI, New Haven, Conn.). Similarly, the 2.6-kb *Bam*HI-*Eco*RI fragments encoding the entire VDJ regions from the V_H205.12 clones A20/44 and A6/24 were subcloned into Bluescribe M13(+), and single-stranded DNA was prepared. Sequence data further upstream of the *Bam*HI site for these clones were obtained by subcloning an overlapping 2.1-kb *Bgl*II-*Pst*I fragment into Bluescribe M13(−). Single-stranded DNA was again rescued. All sequencing was carried out by using the dideoxy method (38) and Sequenase (U.S. Biochemicals, Cleveland, Ohio) or T7 DNA polymerase (Pharmacia, Uppsala, Sweden). Sequencing primers corresponding to newly determined sequences were made by using an Applied Biosystems model 380A synthesizer. Thus, the sequencing strategy involved walking along the clones in segments of approximately 300 to 350 bases (Fig. 1).

Amplification and cloning of germ line immunoglobulin sequences from total liver DNA. C57BL/6j DNA was prepared from the livers of several mice. Sequences that were likely to be specific for germ line genes of the V_H205.12 family were identified from the consensus sequence determined by us for the A20/44 and A6/24 clones. For polymerase chain reaction (PCR) amplification of germ line DNA, the 5′ primer was located 525 to 545 bases upstream and the 3′ primer lay 445 to 467 bases downstream of the cap site. To facilitate cloning of amplified DNA, *Sal*I and *Eco*RI sites were also included at the 5′ end of the upstream and downstream primers, respectively. The primer sequences were (upstream) 5′ GCGGTCGACTCAGTATGTGACAGT GACTAG 3′ and (downstream) 5′ GGAATTCAGAGTCCT CAGATGGTCAGGCT 3′.

PCR amplification (37) was performed in a 25-μl reaction mixture containing 100 ng of C57BL/6j DNA, 100 pmol of each primer, 0.2 mM of each deoxynucleoside triphosphate in 100 mM Tris hydrochloride (pH 8.3) at 25°C, 50 mM KCl, 0.01% gelatin, and 2 mM $MgCl_2$. Reaction mixtures were overlaid with paraffin oil and subjected to an 8-min denaturation at 95°C, followed by slow cooling to 25°C. Taq polymerase (Cetus Corp.) (2 U) was added, and amplification was performed in an IHB2024 (Cherlyn Electronics Ltd., Cambridge, United Kingdom) thermal cycler by five cycles of denaturation at 95°C for 30 s, annealing at 50°C for 30 s, and elongation at 70°C for 1 min. This was followed by 35 cycles of denaturation at 95°C for 30 s, annealing at 55°C for 30 s, and elongation at 70°C for 1 to 2 min. A further 2 U of Taq polymerase was added after 20 cycles. On completion, the paraffin oil was removed by aspiration, and the samples were extracted with ether before analysis of the amplified product by electrophoresis on 1.5% agarose gels.

After electrophoresis, a gel slice containing the PCR-amplified fragment was removed and the DNA was recovered by electroelution. The fragment was purified by extraction with phenol-chloroform (1:1), chloroform, and ether and recovered by precipitation with isopropanol. The amplified DNA was then digested with the appropriate restriction enzymes, repurified as described above, and ligated with bacteriophage M13 DNA at an insert-vector molar ratio of 2:1. Ligation mixtures were transformed into *Escherichia coli* JM101, and clear recombinant plaques were identified by selection on 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside–isopropyl-β-D-thiogalactopyranoside minimal plates. DNA from putative recombinants was analyzed by restriction enzyme digestion to confirm that a fragment of the appropriate size had been cloned.

Single-stranded templates prepared from recombinant plaques were sequenced by the dideoxy method (38). Upstream sequences were obtained by using a DNA primer which spanned the initiation codon of the V_H205.12 genes. Sequences in the coding region were determined by using the downstream PCR primer.

Preparation of hybridoma RNA. Hybridoma cells were grown in RPMI medium supplemented with 10% fetal calf serum, 2 mM glutamine, and 5×10^{-5} M β-mercaptoethanol. A total of 2×10^8 to 4×10^8 cells were harvested by centrifugation, washed once with 5 ml of phosphate-buffered saline, and the RNA was extracted by a modification of the published method (9). Briefly, cells were homogenized in 7.5 ml of a solution consisting of 4 M guanidine isothiocynate, 100 mM sodium acetate, and 5 mM EDTA (pH 5.0). The homogenate was layered over a 4-ml cushion of a solution consisting of 5.0 M CsCl, 100 mM sodium acetate, and 5 mM EDTA (pH 5.0) and subjected to centrifugation at 110,000 × g_{avg} for 18 h at 18°C. The RNA pellet was washed in 70% ethanol, air dried, and dissolved in 0.35 ml of sterile H_2O. RNA concentrations were determined by measurement at 260 nM.

To ensure that the RNA was recovered intact, 10-μg aliquots were subjected to 1% agarose gel electrophoresis in the presence of formaldehyde as previously described (42). The RNA was stained with ethidium bromide and visualized under short-wave UV.

Location of cap sites by primer extension. The transcription start sites for mRNAs encoding V_H186.2 and V_H205.12 heavy chains were identified by determining the length of cDNA transcripts synthesized, using the primers indicated in the legends to Fig. 2 and Fig. 3. Total hybridoma RNA isolated as described above was copied by avian myeloblastosis virus (AMV) reverse transcriptase primed by oligonucleotides complementary to known regions of the V_H186.2 and V_H205.12 mRNAs (7). The length of the cDNA products was determined by analyzing the products on a gel in parallel with DNA fragments of known size or a known nucleotide sequence. The location of the start sites are considered accurate to within one nucleotide.

RESULTS

Nucleotide sequences of related V_H genes and their germ line counterparts. The goal of this work was to define the distribution of mutations within rearranged, hypermutated V_H genes by comparing the sequences determined for related sets of genomic V_H gene clones with that of DNA

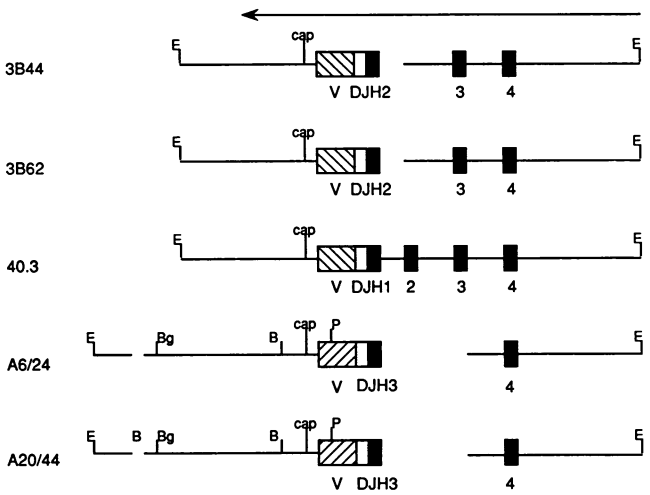


FIG. 1. Arrangement of V_H genes in genomic clones derived from hybridoma DNAs. 3B44, 3B62, and 40.3 are derived from the V_H186.2 germ line gene. A6/24 and A20/44 are derived from the V_H205.12 germ line gene. The clones were sequenced from the right hand *Eco*RI (E) site to a point approximately 600 nucleotides upstream of the transcription start (cap) site (see arrow). The *Bam*HI (B), *Bgl*II (Bg), and *Pst*I (P) sites used for subcloning are indicated.

derived from liver. The genomic clones were isolated previously from hybridomas expressing hypermutated genes derived from the V_H186.2 (three clones 3B44, 3B62, and 40.3) (10, 40) or the V_H205.12 germ line genes (two clones A20/44 and A6/24 (34, 35). The origin of the clones sequenced is described in Materials and Methods and the arrangement of the genes they contain is shown in Fig. 1. For the coding and upstream regions of the V_H186.2 set, the germ line sequence was previously determined (6, 23). For the V_H205.12 set, the upstream region in the germ line sequence was amplified from total liver DNA by PCR and sequenced. Similarly, a 1,900-base-pair portion of the downstream, germ line region (downstream of J_{H-1}, bases 1,121 to 2922) was amplified by PCR and sequenced.

Nucleotide changes in the 5′-flanking regions. Flanking (5′) and coding sequences for the V_H186.2 and V_H205.12 germ line and hybridoma DNAs are shown in Fig. 2 and 3, respectively. There are many nucleotide changes throughout the coding regions which have been previously documented (10, 34, 35, 40). In addition, five changes in 3B62 fell ≥375 base pairs upstream of the cap site (Fig. 2), and two changes in A6/24 fell 21 and 73 base pairs upstream of the cap site (Fig. 3).

The germ line counterpart of the functional V_H genes in hybridomas A6/24 and A20/44 was obtained by PCR amplification of the V_H205.12 family in C57BL/6j liver DNA. A total of 24 clones were partially or completely sequenced. Of these, nine were considered to be representatives of the germ line V_H205.12 gene utilized in the hybridomas (see below), but at least eight other V_H genes were obtained, some of them more than once. Clones 1 through 7 (Fig. 4) are representatives of the 5′ regions of these V_H205.12-related genes. All these genes were clearly different in their coding regions (data not shown) and markedly different in their 5′ noncoding regions. However, none of these germ line V_H genes carried the 5′ upstream changes seen at bases 507 and 559 in A6/24. It is very likely therefore, that these represent genuine somatic mutations.

In some cases, PCR-amplified clones differed from each other by only a single nucleotide. These changes were scattered and were predominantly A/T-to-G/C transitions. None of these occurred at base 507 or 559. We have interpreted these as Taq polymerase-induced mutations (21, 37, 43). The V_H205.12 germ line sequence (Fig. 3) is the consensus from nine clones.

Nucleotide changes in the 3′-flanking regions. The downstream sequences determined for the V_H186.2 and V_H205.12 genes are shown in Fig. 5. The germ line sequence was not determined beyond base 2869; however, the sequences of all five clones were identical downstream of base 2,578 (Fig. 5).

In 11 clones, obtained by three independent PCR amplifications of total C57BL/6j liver DNA, two distinct sequences in the germ line were obtained at bases 2611 and 2613 (Fig. 5A). Six of the clones had the sequence AGA, shown as the germ line sequence (Fig. 5A), and five had the sequence GGG. We also observed these nucleotide differences at similar frequency in clones from DNA separately amplified by PCR from hybridoma 3D61 DNA (10) by using a primer within J_{H-4} (5′) and a primer spanning the *Xba*I site (3′) (position 2908, Fig. 5B). The sequence GGG is also found at the same location in BALB/c DNA, although in the latter other nucleotide differences from C57BL/6j are also apparent (16, 19).

Determination of transcription start sites. Transcription start sites were located by primer extension using reverse transcriptase to copy mRNA isolated from the relevant hybridoma cell lines. The size of the primer-extended cDNA was estimated by comparing it with DNA fragments of known size and sequence (Fig. 6A). To obtain an estimate accurate to within one nucleotide, the cDNAs were also run on a polyacrylamide sequencing gel adjacent to a known nucleotide sequence (Fig. 6B). The cap sites are located 48 and 53 nucleotides upstream of the ATG start codon for the V_H186.2 and V_H205.12 groups, respectively (Fig. 2 and 3).

Distribution of nucleotide changes. The mutational events scored from the comparison of the hybridoma DNA sequences with the appropriate germ line sequences for these two gene families totalled 118. Mutational events were defined as either single-base substitutions, single-base deletions, or complete codon deletions not attributable to N-region additions or junctional errors (nucleotide changes included in the latter category were defined as those occurring within 10 nucleotides on either side of the V-D and D-J junctions). The incidence of mutation, expressed as a percentage, is plotted against the position in the gene (Fig. 7). The cap sites represent a logical reference point about which to align the sequences to compare the incidence of mutation in transcribed and nontranscribed regions of the DNA. Clearly, there is a sharp increase (>10-fold) in mutation frequency on the 3′ side of the cap site compared with that on the immediate 5′ side. This elevated level of mutation ranged from 1 to 7% of bases sequenced and terminated approximately 210 bases beyond J_{H-4} (Fig. 7).

Nature of the nucleotide changes in nontranslated flanking regions. The types of nucleotide changes in nontranslated introns and 5′- and 3′-flanking regions are summarized in Table 1. Of 75 mutations scored within these regions, 50.7% were transitions, 44% were transversions, and 5.3% were single-base deletions. No base additions were observed.

DISCUSSION

In this work, we have systematically sequenced two groups of somatically mutated V_H genes and compared them with their germ line sequences. The sequences in this work

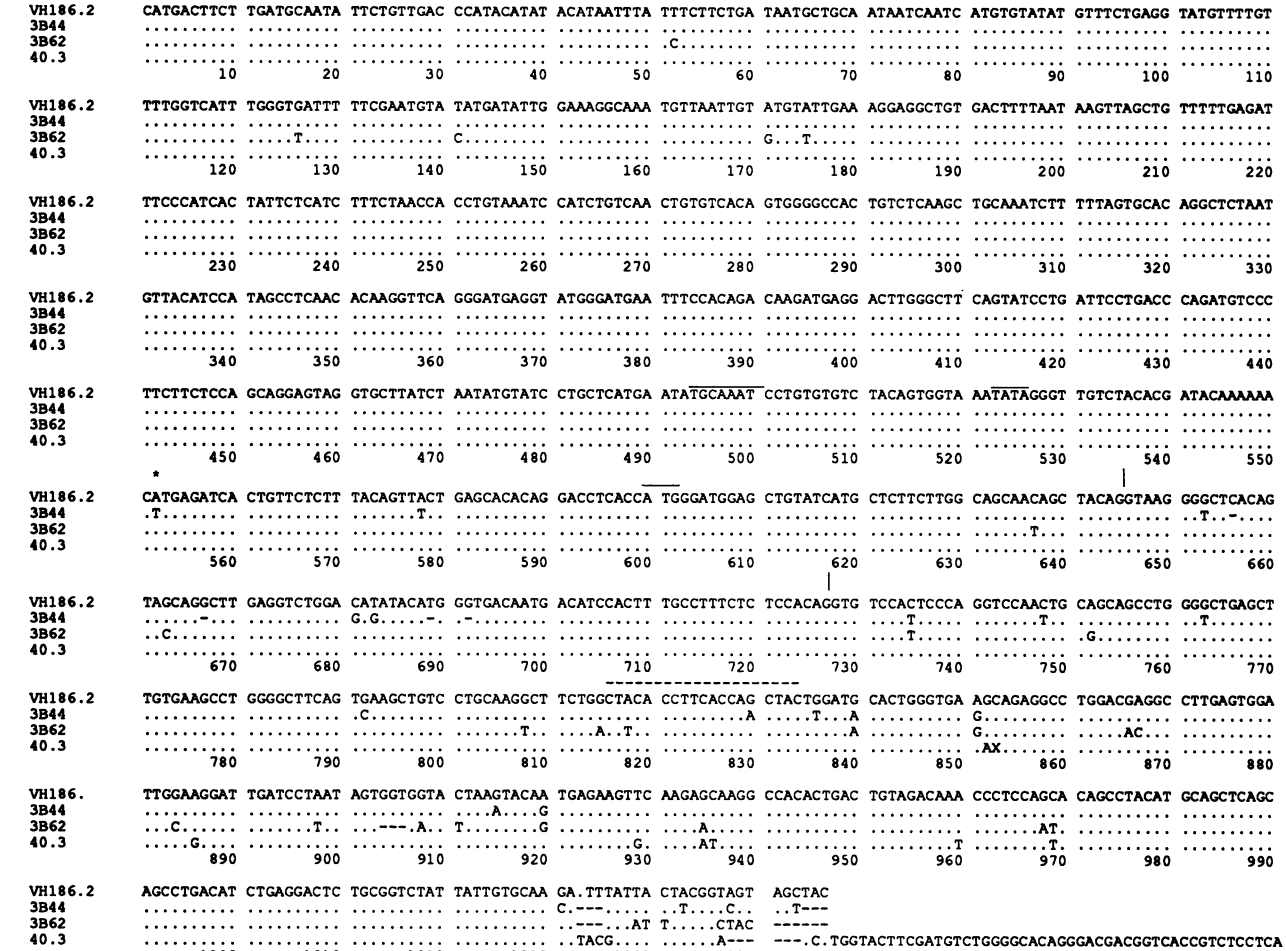


FIG. 2. Nucleotide sequences of the 5′ nontranslated regions and adjacent coding sequences of the active V_HDJ_H alleles from hybridomas 3B44, 3B62, and 40.3 compared with the V_H186.2 germ line sequence (23; also see update in the EMBL data base). The published sequence for the coding region of 3B44 (10) originally differed from our sequence at several positions. However, a recently published corrigendum (2) is in complete agreement with our data. In codon positions 43 and 49 of 3B62 and position 13 of A6/24, the sequence determined by us differs from the sequence determined by others (1, 10, 34). Symbols: ., identity with the germ line sequence; -, a gap relative to the germ line sequence; |, likely splice sites; *, transcription start or cap sites; X, an unidentified base. The start codon, ATG, plus upstream promoter elements are overlined (13, 29). The primer sequence used for cap site determinations was complementary to bases 816 to 834.

(except for the V_H186.2 germ line sequence) were determined independently and compared with the previously published sequences for the V-D-J and flanking regions (6, 10, 23, 34, 35, 40). At all but three sites, i.e., at codons 43 and 49 of 3B62 and codon 13 of A6/24 (see the legend to Fig. 2), the sequences agree with previously published work (1, 2, 10, 35). The close agreement between these two sets of data confirms the accuracy of our data and renders negligible the number of apparent mutations ascribable to sequencing errors and cloning artifacts. In addition, there were only four positions in approximately 18,000 bases sequenced at which a base could not be identified from sequencing a single strand.

An earlier literature survey of published murine and human immunoglobulin gene sequences indicated that very few mutations occurred 5′ to the region coding for the signal peptide or leader sequence (41). Of a total of 121 mutational events scored in this survey, four (or 3.3%) lay 5′ to the signal peptide region (22, 36) and one (0.8%) almost certainly lay upstream of the cap site (20), i.e., ≥96% of mutational changes occurred within the transcription unit (41). The conclusions drawn were limited by the amount of sequence data that was available for the region upstream of the leader sequence of these mutated genes (a median value of only ~120 bases). In the present work, we have greatly improved this situation by systematically sequencing at least 550 bases 5′ to the cap site for two groups of related, somatically mutated V_H genes and by locating transcription start sites. The location of the cap sites determined for the mRNAs in this study are consistent with those determined for other immunoglobulin genes (13, 29) and for the V_H186.2 gene, which was determined independently (3).

The distribution of the 118 mutational events established here (Fig. 7) tentatively identifies the target region for hypermutation which clearly extends into the 5′- and 3′-flanking regions of these V_H genes. The distribution of mutations here is similar to that described for 44 somatic mutations documented earlier for the single V_H M167 gene expressed in the phosphorylcholine-binding myeloma, MOPC167 (22). However, in this study it was not determined whether any mutations fell upstream of the transcription start site.

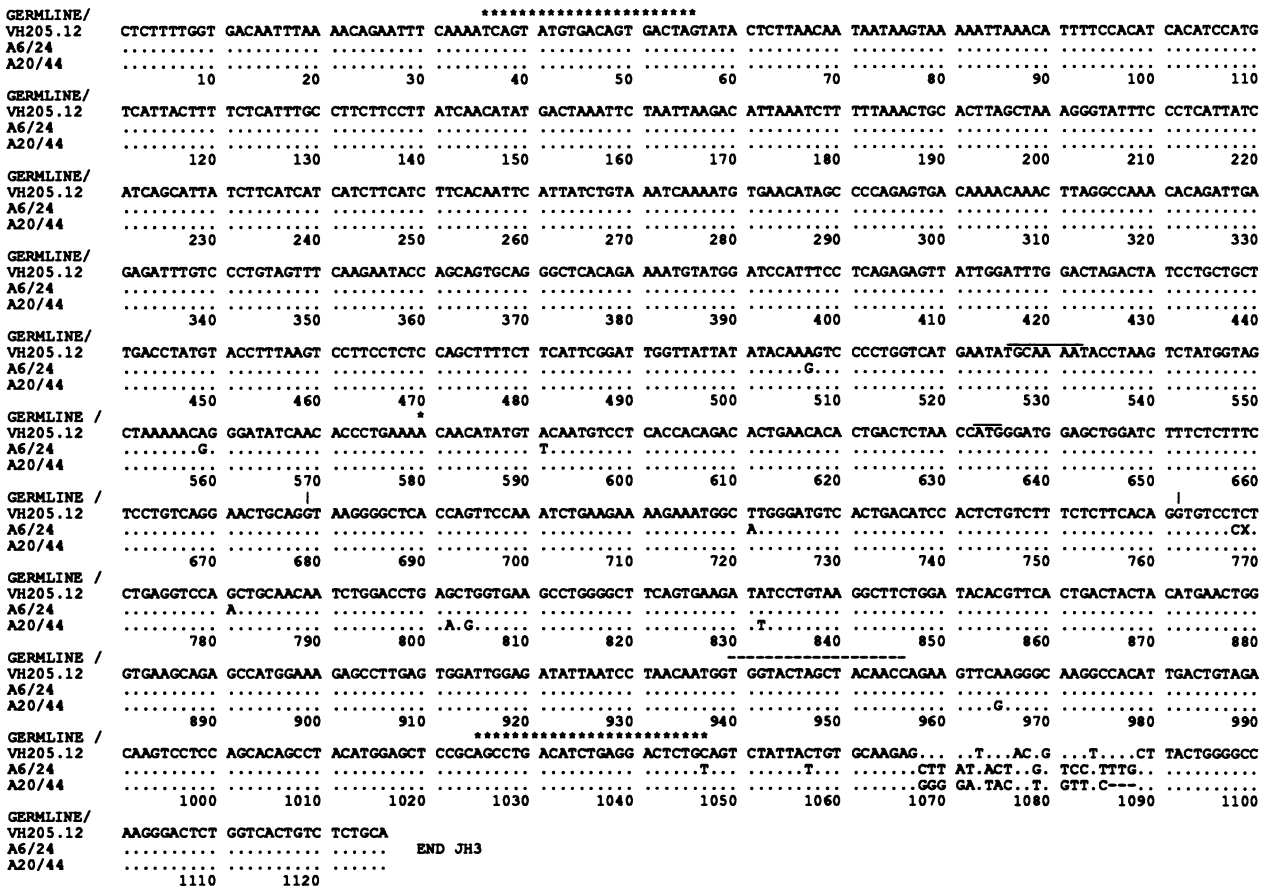


FIG. 3. Nucleotide sequence of the 5′ nontranslated regions and adjacent coding sequences of the active V_HDJ_H alleles from hybridomas A6/24 and A20/44 compared with the V_H205.12 germ line sequence. Asterisks indicate the primer sequences used for PCR amplification of the germ line genes. The primer sequence used for cap site determinations was complementary to bases 940 to 956. Other annotation is as described in the legend to Fig. 1.

For the V_H205.12 family, 24 clones obtained by PCR amplification of genomic DNA were sequenced. Nine germ line and eight related clones were obtained, some of them more than once. However, the base changes in A6/24 which occurred 21 and 73 nucleotides upstream of the cap site (Fig. 3) were not found in any of these clones (Fig. 4). If a progenitor gene with 5′ sequences identical to A6/24 existed in liver DNA, we could reasonably have expected to amplify it by PCR, given that A6/24 is very similar to the V_H205.12 germ line gene and 9 of 24 clones characterized were of this

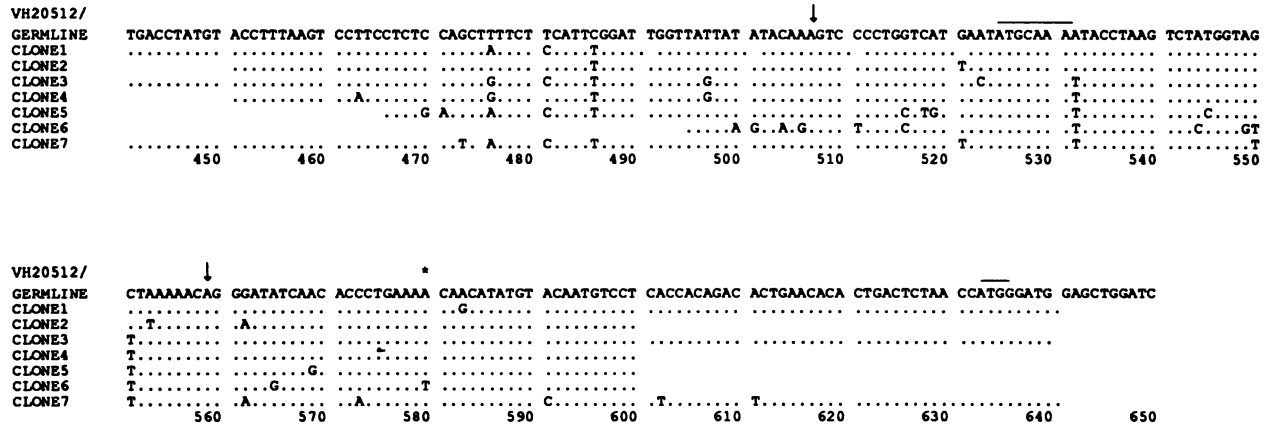
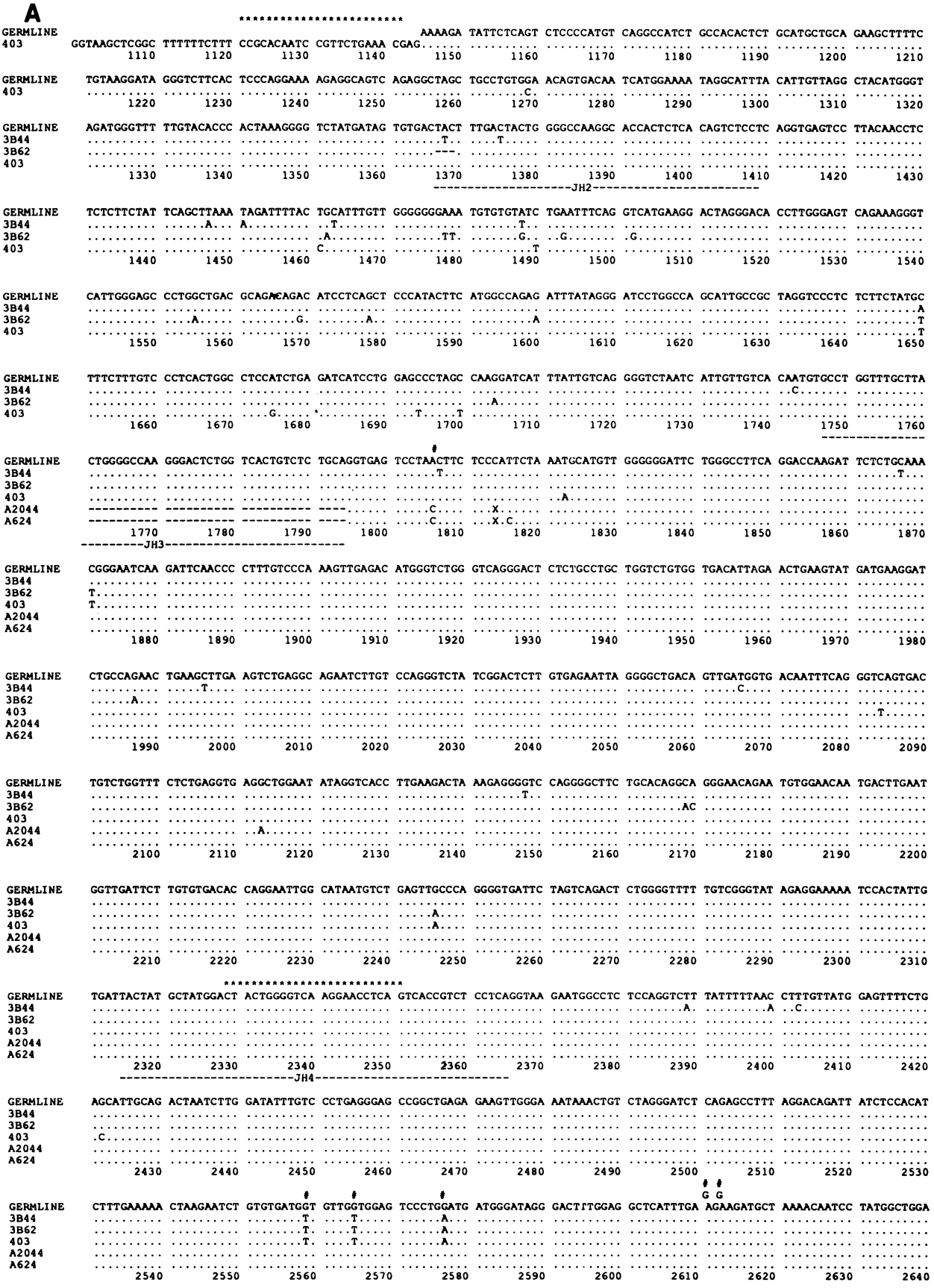


FIG. 4. Partial nucleotide sequences for the upstream regions of V_H205.12-related germ line genes amplified by PCR from total C57BL/6j liver DNA. These sequences are compared with the germ line V_H205.12 gene sequence which was the progenitor of hybridomas A6/24 and A20/44. Vertical arrows indicate the location of the two putative somatic mutations in hybridoma A6/24 (Fig. 3). Blank regions indicate where the sequence is not yet determined. Other annotations are described in the legend to Fig. 1.


A
GERMLINE
403
3B44
3B62
A2044
A624
JH2
JH3
JH4

B

```
GGGATAGTTG GGGCTGTAGT TGGAGATTTT CAGTTTTTAG AATAAAAGTA TTAGTTGTGG AATATACTTC AGGACCACCT CTGTGACAGC ATTTATACAG TATCCGATGC
      2650       2660       2670       2680       2690       2700       2710       2720       2730       2740       2750

ATAGGGACAA AGAGTGGAGT GGGGCACTTT CTTTAGATTT GTGAGGAATG TTCCGCACTA GATTGTTTAA AACTTCATTT GTTGGAAGGA GAGCTGTCTT AGTGATTGAG
      2760       2770       2780       2790       2800       2810       2820       2830       2840       2850       2860
                                              **********************

TCAAGGGAGA AAGGCATCTA GCCTCGGTCT CAAAAGGGTA GTTGCTGTCT AGAGAGGTCT GGTGGAGCCT GCAAAAGTCC AGCTTTCAAA GGAACACAGA AGTATGTGTA
      2870       2880       2890       2900       2910       2920       2930       2940       2950       2960       2970

TGGAATATTA GAAGATGTTG CTTTTACTCT TAAGTTGGTT CCTAGGAAAA ATAGTTAAAT ACTGTGACTT TAAAATGTGA GAGGGTTTTC AAGTACTCAT TTTTTTAAAT
      2980       2990       3000       3010       3020       3030       3040       3050       3060       3070       3080

GTCCAAAATT CTTGTCAATC AGTTTGAGGT CTTGTTTGTG TAGAACTGAT ATTACTTAAA GTTTAACCGA GGAATGGGAG TGAGGCTCTC TCATAACCTA TTCAGAACTG
      3090       3100       3110       3120       3130       3140       3150       3160       3170       3180       3190

ACTTTTAACA ATAATAAATT AAGTTTCAAA TATTTTTAAA TGAATTGAGC AATGTTGAGT TGGAGTCAAG ATGGCCGATC AGAACCAGAA CACCTGCAGC AGCTGGCAGG
      3200       3210       3220       3230       3240       3250       3260       3270       3280       3290       3300

AAGCAGGTCA TGTGGCAAGG CTATTTGGGG AAGGGAAAAT AAAACCACTA GGTAAACTTG TAGCTGTGGT TTGAAGAAGT GGTTTTGAAA CACTCTGTCC AGCCCCACCA
      3310       3320       3330       3340       3350       3360       3370       3380       3390       3400       3410

AACCGAAAGT CCAGGCTGAG CAAAACACCA CCTGGGTAAT TTGCATTTCT AAAATAAGTT GAGGATTCAG CCGAAACTGG AGAGGTCCTC TTTTAACTTA TTGAGTTCAA
      3420       3430       3440       3450       3460       3470       3480       3490       3500       3510       3520

CCTTTTAATT TTAGCTTGAG TAGTTCTAGT TTCCCCAAAC TTAAGTTTAT CGACTTCTAA AATGTATTTA GAATTC
      3530       3540       3550       3560       3570       3580       3590       3600
```

FIG. 5. Nucleotide sequences of the 3'-flanking regions of the active V_HDJ_H alleles from hybridomas 3B44, 3D62, 40.3, A20/44, and A6/24 compared with the germ line sequence. The downstream germ line sequence was determined from clones amplified by PCR from total liver DNA by using primers (*****) near J_{H-1} (bases 1121 to 1142) and across the *Xba*I site (base 2908, Fig. 5B). Other clones were obtained by amplification of the J_{H-4}-*Xba*I region. # signifies base differences which may be allelic, rather than somatic. The identity of base 2310 in the germ line sequence was ambiguous; it is assigned as G on the basis of the consensus sequence. (B) Germ line sequence was determined from PCR-amplified clones to base 2869. All hybridomas were identical in sequence from base 2641 to the 3' end.

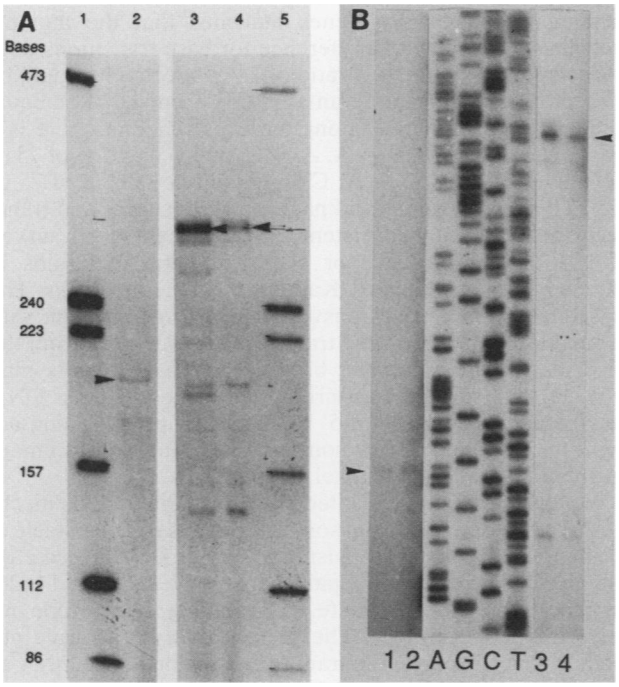


FIG. 6. Determination of the 5' termini of mRNAs from hybridomas expressing V_H186.2 and V_H205.12 genes. Total RNA was isolated from the following hybridoma cell lines: B1-8, which expresses mRNA for an unmutated V_H186.2 gene rearranged to J_{H-2} (6); A6/24, expressing mRNA representing the V_H205.12 gene, and A20/44, also expressing mRNA of V_H205.12. Primers which had been used for sequence determinations were radiolabeled with [γ-^{32}P]ATP at their 5' ends and used to initiate cDNA synthesis by avian myeloblastosis virus reverse transcriptase, using the total hybridoma RNA as the template. Primers used for V_H genes are indicated in the legends to Figs. 2 and 3. cDNAs were analyzed by electrophoresis on 3% polyacrylamide gels containing 7.0 M urea (7), using denatured DNA fragments derived from a plasmid of known sequence as markers. (A) Marker fragments of indicated sizes (tracks 1 and 5); cDNAs copied from B1-8 (track 2), A20/44, (track 3) and A6/24 (track 4) RNAs, respectively. Arrowheads indicate the full-length cDNA product in each case. (B) cDNAs synthesized in panel A were analyzed by electrophoresis on a 5% sequencing gel next to a known nucleotide sequence derived from a rotavirus VP7 gene (4). Separate photographic exposures of the same gel were aligned to form the figure. Arrowheads indicate the position of the cDNAs which were loaded twice. Tracks 1 and 2, cDNA from track 2 of panel A; tracks 3 and 4, cDNA from track 3 of panel A.

type. Therefore, these changes almost certainly represent genuine somatic mutations. For the V_H186.2 group, the 3B62 gene possesses a cluster of five base changes beginning 375 nucleotides upstream of the cap site. One of these creates a *Rsa*I site, and we are using this fact together with PCR techniques to investigate their origin. The region of DNA corresponding to bases 11 to 1011 (Fig. 2) was amplified from total liver DNA (C57BL/6j) and from hybridoma 3B44 and 3B62 DNAs. When these amplified DNAs were digested with *Rsa*I, 3B62 DNA was cut, whereas the germ line and 3B44 DNAs were not (H. Rothenfluh, L. Taylor, G. Both and E. Steele, unpublished results). These preliminary results therefore suggest that the upstream changes in 3B62 are also somatic mutations since they appear not to be present in the germ line DNA typical of the region encoding V_H186.2-related sequences. However, we cannot rule out the unlikely possibility that progenitor sequences for 3B62 (and A6/24) exist elsewhere in the genome in related sequences which were not amplified by the PCR primers used here. Thus, the 5' boundary of mutation remains to be established but mutations clearly occur upstream of the cap site in the nontranscribed flanking region. Sequencing of this region in a larger set of mutated V_H186.2 genes should define the 5' boundary of mutation. This has been initiated.

The 3' boundary for hypermutation clearly extends beyond J_{H-4}, but we are uncertain of its precise location. We have placed the boundary at a region about 210 bases downstream of J_{H-4} in Fig. 7. This takes in a cluster of nine recurrent mutations (Fig. 5A, bases 2559 to 2577), i.e., mutations to the same base in the same position in independently arising somatically mutated genes. However, it is not certain that these are genuine somatic mutations, given the

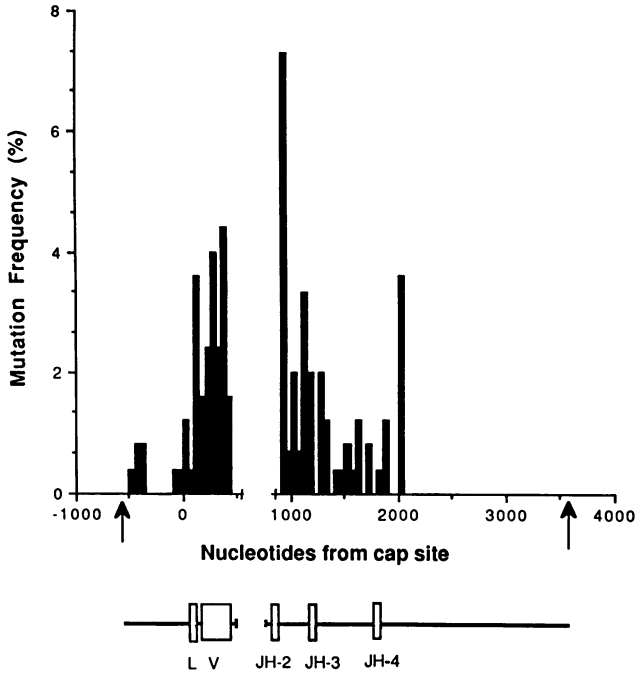


FIG. 7. Distribution of mutations within and around somatically mutated V_HDJ_H loci. Mutation frequency (%) is the number of mutational events per 100 bases sequenced. The sequences were determined in the region between the arrows. The data, grouped at 50-base-pair intervals, were derived from Fig. 2, 3, and 5, and encompass 118 mutational events. A mutational event is defined as either a single-base substitution, a single-base deletion, or a complete codon deletion. The four bases which could not be determined (labeled X in Figs. 2, 3, and 5) were not scored. The sets of sequences have been aligned first by the nucleotide specifying the initiation of RNA synthesis (cap site [*], in Fig. 2, 3, and 6). The first nine 50-base-pair intervals downstream of the cap take in most of the V region to position 1001 in V_H186.2 (Fig. 2) and base 1029 in V_H205.12 (Fig. 3) which are 29 and 31 nucleotides, respectively, from the 3′ end of the V-region sequence. The discontinuity takes in the V-D junction through to the end of J_{H-2}. Thus, any differences in the V-D junction, which might include potential N-region additions were not scored (e.g., the two mutations in A6/24 at positions 1047 and 1057 in Fig. 3 are not included in this histogram). The discontinuity is also present because only one sequence (40.3) was available through the region. Between J_{H-2} and J_{H-3}, the data are based on sequences of only three genes (3B44, 3B62, and 40.3); for all other downstream regions, the data are derived from all five genes which have been aligned at the end of the J_{H-3} region (position 1794 in Fig. 5A).

nearby sequence polymorphisms at bases 2611 and 2613. Our present interpretation is that these GGG-to-AGA changes may represent allelic differences in C57BL/6j mice. Alternatively, since these changes arose in two independent PCR amplifications of this region (in liver and hybridoma 3D61 DNAs), there might be a hot spot for Taq polymerase-induced mutations at these sites. If the recurrent base changes do not represent genuine somatic mutations, then the boundary contracts to a region about 60 base pairs 3′ to J_{H-4}. Sequencing of this region in a larger set of V_H186.2 genes will better define the 3′ boundary of mutation.

Mutations occurring in the nontranslated introns and 5′- and 3′-flanking regions of V_H genes should not be subjected to antigenic selection. Thus, the types of mutations in these regions may represent the error specificity of the enzymes involved in the somatic mutation process (M. Kaartinen, S. Kulp, and O. Makela, *in* E. J. Steele, ed., *Somatic Hypermutation in V-Regions*, in press; T. A. Kunkel, in E. J. Steele, ed., *Somatic Hypermutation in V-Regions*, in press). Analysis of the types of mutation within the nontranslated flanking regions of V_H genes indicates that the mutator mechanism may have a preference for base transitions over transversions (Table 1). A ratio of 1:2 respectively would be evident if transitions and transversions occurred at an equal frequency. In addition, among the transversions, there is a bias towards A → T, T → A, A → C, and G → T (25 of 75 or 33%) over T → G, C → A, C → G, and G → C (8 of 75 or 9%). These proportions and pattern of transitions and transversions are broadly consistent with the recent larger survey of 234 "selection-free," or silent, somatic mutations in immunoglobulin V genes (Kaartinen et al., in press). The biased pattern amongst transversions is also consistent with the ability of AMV reverse transcriptase to extend from the corresponding mismatched bases (27). These results may indicate that reverse transcriptase or a low-fidelity DNA polymerase lacking the 3′-5′ exonuclease proofreading activity may operate during somatic hypermutation in immunoglobulin V_H genes (Kunkel, in press).

TABLE 1. Types of nucleotide changes in the nontranslated regions of immunoglobulin V_H genes

Type of nucleotide change	No. of mutations (n = 75)
Transitions	
A → G	7
T → C	7
G → A	12
C → T	12
Single-base deletions	4
Transversions	
C → G	1
G → C	2
A → C	4
T → G	2
A → T	7
T → A	5
C → A	3
G → T	9

We have critically evaluated elsewhere the various mechanisms proposed to explain somatic hypermutation (Steele et al., in press). These mechanisms are either DNA-based e.g., site-directed error-prone repair, gene conversion (8, 14, 26) or gene expression-based (e.g., involving error-prone reverse transcription) (41). The present data do not rule out models which propose the transcribed region of the DNA as the primary target for the mutation process. However, this type of model is insufficient to fully account for the observed distribution of mutations in the 5′ nontranscribed region unless various ad hoc processes are proposed to explain their origin, e.g., the use of minor upstream transcription initiation sites in hypermutating B cells. Transcription from 5′ upstream start sites was proposed to explain the appearance of hygromycin-resistant cells after transfection with promoter-negative retroviral vectors carrying the hygromycin resistance gene (11, 12).

In summary, the data suggest that the transcribed region is the primary target for the mutational mechanism, but the mutations which fall outside this region require other explanations, e.g., gene conversion or heteroduplex-induced gene mutagenesis (26; Steele et al., in press). Thus, the real

mechanism of somatic hypermutation may turn out to be a combination of the various published models.

ACKNOWLEDGMENTS

We thank K. Rajewsky and his group for the clones of the various V_H genes, hybridomas, and C57BL/6j DNA. We are grateful to R. V. Blanden, H. M. Temin, C. Kocks, and K. Rajewsky for critical discussions both on the work and manuscript.

This research was supported by funds from a University of Wollongong-Commonwealth Scientific and Industrial Research Organization collaborative research grant, an NHMRC (Australia) project grant and a Medical Research Council (United Kingdom) grant to J.W.P.

LITERATURE CITED

1. **Allen, D., T. Simon, F. Sablitzky, K. Rajewsky, and A. Cumano.** 1988. Antibody engineering for the analysis of affinity maturation of an anti-hapten response. EMBO J. **7**:1995–2001.
2. **Allen, D., T. Simon, F. Sablitzky, K. Rajewsky, and A. Cumano.** 1989. Antibody engineering for the analysis of affinity maturation of an anti-hapten response (Corrigendum). EMBO J. **8**:2444.
3. **Ballard, D. W., and A. L. M. Bothwell.** 1986. Mutational analysis of the immunoglobulin heavy chain promoter region. Proc. Natl. Acad. Sci. USA **83**:9626–9630.
4. **Both, G. W., J. S. Mattick, and A. R. Bellamy.** 1983. Serotype-specific glycoprotein of simian 11 rotavirus: coding assignment and gene sequence. Proc. Natl. Acad. Sci. USA **80**: 3091–3095.
5. **Bothwell, A. L. M.** 1984. The genes encoding anti-NP antibodies in inbred strains of mice, p. 19–34. *In* M. I. Greene and A. Nisonoff (ed.), The biology of idiotypes. Plenum Publishing Corp., New York.
6. **Bothwell, A. L. M., M. Paskind, M. Reth, T. Imanishi-Kari, K. Rajewsky, and D. Baltimore.** 1981. Heavy chain variable region contribution to the NPb family of antibodies: somatic mutation evident in a 2a variable region. Cell **24**:625–637.
7. **Boyle, D. B., B. E. H. Coupar, A. J. Gibbs, L. J. Siegman, and G. W. Both.** 1987. Fowlpox thymidine kinase: nucleotide sequence and relationships to other thymidine kinases. Virology **156**:355–365.
8. **Brenner, S., and C. Milstein.** 1966. Origin of antibody variation. Nature (London) **211**:242–243.
9. **Chirgwin, J. M., A. E. Przybula, R. J. MacDonald, and W. J. Rutter.** 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry **18**: 5294–5299.
10. **Cumano, A., and K. Rajewsky.** 1986. Clonal recruitment and somatic mutation in the generation of immunological memory to the hapten NP. EMBO J. **5**:2459–2468.
11. **Dornburg, R., and H. M. Temin.** 1988. Retroviral vector system for study of cDNA gene formation. Mol. Cell. Biol. **8**: 2328–2334.
12. **Dougherty, J. P., and H. M. Temin.** 1987. A promoterless retroviral vector indicates that there are sequences in U3 required for 3′ RNA processing. Proc. Natl. Acad. Sci. USA **84**:1197–1201.
13. **Falkner, F. G., and H. G. Zachau.** 1984. Correct transcription of an immunoglobulin K gene requires an upstream fragment containing conserved sequence elements. Nature (London) **310**: 71–74.
14. **Gearhart, P. J.** 1982. Generation of immunoglobulin variable gene diversity. Immunol. Today **3**:107–112.
15. **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**:3439–3443.
16. **Gillies, S. D., S. L. Morrison, V. T. Oi, and S. Tonegawa.** 1983. A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. Cell **33**:717–728.
17. **Golding, G. B., P. J. Gearhart, and B. W. Glickman.** 1987. Patterns of somatic mutations in immunoglobulin variable genes. Genetics **115**:169–176.
18. **Gorski, J., P. Rollini, and B. Mach.** 1983. Somatic mutations of immunoglobulin variable genes are restricted to the rearranged V gene. Science **220**:1179–1180.
19. **Gough, N. M., and O. Bernard.** 1981. Sequences of the joining region genes for immunoglobulin heavy chains and their role in the generation of antibody diversity. Proc. Natl. Acad. Sci. USA **78**:509–513.
20. **Kataoka, T., T. Nikaido, T. Miyata, and T. Honjo.** 1982. The nucleotide sequences of rearranged and germline immunoglobulin V_H genes in the mouse. J. Biol. Chem. **257**:277–285.
21. **Keohavong, P., and W. G. Thilly.** 1989. Fidelity of DNA polymerases in DNA amplification. Proc. Natl. Acad. Sci. USA **86**:9253–9257.
22. **Kim, S., M. Davis, E. Sinn, P. Patten, and L. Hood.** 1981. Antibody diversity: somatic hypermutation of rearranged V_H genes. Cell **27**:573–581.
23. **Krawinkel, U., G. Zoebelein, and A. L. M. Bothwell.** 1986. Palindromic sequences are associated with sites of DNA breakage during gene conversion. Nucleic Acids Res. **14**:3871–3882.
24. **Kunkel, T. A.** 1988. Exonucleolytic proofreading. Cell **53**: 837–840.
25. **Lui, Y.-J., D. E. Joshua, G. T. Williams, C. A. Smith, J. Gordon, and I. C. M. MacLennan.** 1989. Mechanism of antigen-driven selection in germinal centers. Nature (London) **342**: 929–931.
26. **Maizels, N.** 1989. Might gene conversion be the mechanism of somatic hypermutation of mammalian immunoglobulin genes? Trends Genet. **5**:4–8.
27. **Mendelman, L. V., J. Petruska, and M. F. Goodman.** 1990. Base misrepair extension kinetics. J. Biol. Chem. **265**:2338–2346.
28. **Moller, G. (ed.).** 1987. Role of somatic mutation in generation of lymphocyte diversity. Immunological Reviews, vol. 96.
29. **Parslow, T. G., D. L. Blair, W. J. Murphy, and D. K. Granner.** 1984. Structure of the 5′ ends of immunoglobulin genes: a novel conserved sequence. Proc. Natl. Acad. Sci. USA **81**:2650–2654.
30. **Pech, M., J. Hochtl, H. Schnell, and H. G. Zachau.** 1981. Differences between germline and rearranged immunoglobulin V_K coding sequences suggest a localized mutation mechanism. Nature (London) **291**:668–670.
31. **Reanney, D. C.** 1984. Genetic noise in evolution? Nature (London) **307**:318–319.
32. **Reanney, D. C.** 1986. Genetic error and genome design. Trends Genet. **2**:41–46.
33. **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**:1022–1026.
34. **Sablitzky, F., and K. Rajewsky.** 1984. Molecular basis of an isogenic anti-idiotypic response. EMBO J. **3**:3005–3012.
35. **Sablitzky, F., G. Wildner, and K. Rajewsky.** 1985. Somatic mutation and clonal expansion of B cells in an antigen-driven immune response. EMBO J. **4**:345–350.
36. **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 (London) **286**:676–683.
37. **Saiki, R. K., D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis, and H. A. Erlich.** 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science **239**:487–491.
38. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
39. **Selsing, E., and U. Storb.** 1981. Somatic mutation of immunoglobulin light-chain variable-region genes. Cell **25**:47–58.
40. **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**:757–770.
41. **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**:667–673.

42. **Thurston, C. F., C. R. Perry, and J. W. Pollard.** 1988. Electrophoresis of RNA denatured with glyoxal or formaldehyde, p. 13–25. *In* J. Walker (ed.), Methods in molecular biology. The Humana Press, Clifton, N.J.
43. **Tindall, K. R., and T. A. Kunkel.** 1988. Fidelity of DNA synthesis by the *Thermus aquaticus* DNA polymerase. Biochemistry **27**:6008–6013.
44. **Weiss, S., and G. E. Wu.** 1987. Somatic point mutations in unrearranged immunoglobulin gene segments encoding the variable region of lambda light chains. EMBO J. **6**:927–932.