Somatic Hypermutation Introduces Insertions and Deletions into Immunoglobulin V Genes

By Patrick C. Wilson,‡ Odette de Bouteiller,§ Yong-Jun Liu,§ Kathleen Potter,* Jacques Banchereau,‡ J. Donald Capra,* and Virginia Pascual*

From the *Molecular Immunology Center, Department of Microbiology, University of Texas Southwestern Medical Center at Dallas, Texas 75235-9140; the ‡Oklahoma Medical Research Foundation, Oklahoma City, Oklahoma 73104; §Schering-Plough Laboratory for Immunological Research, Dardilly, France; and ‡Baylor Institute for Immunological Research, Dallas, Texas 75235

Summary

During a germinal center reaction, random mutations are introduced into immunoglobulin V genes to increase the affinity of antibody molecules and to further diversify the B cell repertoire. Antigen-directed selection of B cell clones that generate high affinity surface Ig results in the affinity maturation of the antibody response. The mutations of Ig genes are typically base-pair substitutions, although DNA insertions and deletions have been reported to occur at a low frequency. In this study, we describe five insertion and four deletion events in otherwise somatically mutated V\textsubscript{H} gene cDNA molecules. Two of these insertions and all four deletions were obtained through the sequencing of 395 cDNA clones (~110,000 nucleotides) from CD38\textsuperscript{+}IgD\textsuperscript{2} germinal center, and CD38\textsuperscript{2}IgD\textsuperscript{2} memory B cell populations from a single human tonsil. No germline genes that could have encoded these six cDNA clones were found after an extensive characterization of the genomic V\textsubscript{H}4 repertoire of the tonsil donor. These six insertions or deletions and three additional insertion events isolated from other sources occurred as triplets or multiples thereof, leaving the transcripts in frame. Additionally, 8 of 9 of these events occurred in the CDR 1 or CDR 2, following a pattern consistent with selection, and making it unlikely that these events were artifacts of the experimental system. The lack of similar instances in unmutated IgD\textsuperscript{1}CD38\textsuperscript{2} follicular mantle cDNA clones statistically associates these events to the somatic hypermutation process (P = 0.014). Close scrutiny of the 9 insertion/deletion events reported here, and of 25 additional insertions or deletions collected from the literature, suggest that secondary structural elements in the DNA sequences capable of producing loop intermediates may be a prerequisite in most instances. Furthermore, these events most frequently involve sequence motifs resembling known intrinsic hotspots of somatic hypermutation. These insertion/deletion events are consistent with models of somatic hypermutation involving an unstable polymerase enzyme complex lacking proofreading capabilities, and suggest a downregulation or alteration of DNA repair at the V locus during the hypermutation process.

During the course of a T cell–dependent antibody response, B cells hone the specificity of their antibody molecules through a process of random somatic hypermutation of their V genes, followed by antigen driven selection. This is collectively referred to as affinity maturation. This process occurs within the germinal centers (GCs)\textsuperscript{1} of secondary follicles from peripheral lymphoid organs when antigen stimulated B cells receive proper signals from T and accessory cells. In the human system, GC B cells are characterized by the surface expression of CD38 and, in most cases, the loss of IgD (1–3). We have previously shown that the initiation of somatic hypermutation occurs within the CD77\textsuperscript{+} subset of these IgD\textsuperscript{−}CD38\textsuperscript{−} B cells (4). Mutated V genes can be isolated from all subsequent stages of B cell differentiation and in cells from all IgD\textsuperscript{−} and certain IgD\textsuperscript{+} B cell subsets (4, 5). The molecular process of somatic hypermutation remains elusive, primarily due to the lack of a good in vitro model until very recently (6). Much of what

\textsuperscript{1} Abbreviations used in this paper: FM, follicular mantle; FW, framework; GC, germinal center.
is known concerns: (a) localizing the somatic hypermutation process to particular B cell subsets and anatomical settings (4, 7–10); (b) delineating the limits and rates of mutational activity (11); (c) determining the minimal substrate through transgenic technology (12, 13); and (d) analyzing the mutations themselves in the context of the surrounding sequence to reveal tendencies such as strand polarity and “hotspots” of somatic hypermutation (for reviews see references 12 and 13).

Although somatic hypermutation is typically described as the generation of bp substitutions, insertions and deletions have been sporadically described. As with somatic point mutations, the analysis of these events can provide valuable information concerning somatic hypermutation itself. Analysis of human V \textsubscript{\textgamma}4 family genes generated from the amplification of cDNA from somatically mutated GC (IgD–CD38 \textsuperscript{+}) and memory (IgD–CD38 \textsuperscript{–}) B cell subpopulations led us to identify a number of cDNA clones from the mutated cell populations that contained insertions and deletions. We provide evidence that these events are linked to the somatic hypermutation process. Additionally, these events occur in a predictable fashion relative to the surrounding sequence, suggesting a model for their occurrence with implications for the molecular process of somatic hypermutation.

Materials and Methods

Isolation, Labeling, and Sorting of Tonsil B Cells. Human tonsils were obtained during routine tonsillectomy. B cell isolation and sorting for CD38 and IgD expression were performed as previously described (4, 14). In brief, human tonsillar B cells were separated into IgD–CD38– follicular mantle (FM) B cells, IgD–CD38+ GC B cells, and IgD–CD38– memory B cells to 95–98% purity as predicted by FACS \textsuperscript{®} analysis, as previously described (13).

The mutation state of the V \textsubscript{\textgamma}4 gene cDNA clones from the various subpopulations was in agreement with our previous study (4). Clones were considered somatically mutated if they contained two or more bp substitutions, well beyond the expected error rates for the avian myeloblastosis virus reverse transcriptase (AMV-RT), Taq, and PFU polymerases used in these analyses (this mutation rate is based on our previous analyses, reference 4).

Sequencing the Ig V \textsubscript{\textgamma}4 Transcripts. Total RNA was extracted from 1–5 \times 10 \textsuperscript{6} cells using guanidinium thiocyanate–phenol–chloroform in a single step using the Ultraspec RNA isolation system (BIOTECX Laboratories, Houston, TX), and was reverse transcribed using oligo-d(T) or specific V gene constant region oligonucleotides C\textsubscript{p}12 (5′-CTGGACTTTGCAACCACGCTG-3′) for IgM transcripts or C\textsubscript{p}180 (5′-CTGCTGAGG-GAGTTAGTCC-3′) for IgG transcripts, and SuperScript II reverse transcriptase (GIBCO BRL, Gaithersburg, MD). First strand cDNA was used directly for second strand synthesis and amplification via the avian myeloblastosis virus reverse transcriptase (AMV-RT), and was used directly for second strand synthesis and amplification via the avian myeloblastosis virus reverse transcriptase (AMV-RT), Taq, and PFU polymerases used in these analyses (this mutation state could be determined as above and screened using 32P-labeled, gene-specific oligonucleotides (IgD–CD38 \textsuperscript{+}, L-4). The PCR products were purified using microcentrconetrators (Amicon, Beverly, MA), and were then kiaased and blunt-end ligated into an EcoRV-digested and dephosphorylated pBluescript plasmid (Stratagene, La Jolla, CA; Polynucleotide Kinase, T4 DNA Ligase, and EcoRV were from Boehringer Mannheim, Amsterdam, Netherlands). After transformation by electroporation into electrophoretic DH10α E. coli (GIBCO BRL) and screening with consensus internal oligonucleotides as previously described (4, 15), positive colonies were picked, plasmid mini-preparations were made, and colonies were sequenced in both directions using an automated DNA sequencer and automated sequencing protocol (ABI-377; Advanced Biotechnologies Inc., Columbia, MD). All sequences were analyzed using DNAStar (DNASTAR Inc., Madison, WI). In the first tonsil analyzed, 583 clones were picked, plasmid mini-preparations were made, and Southern blots were prepared by standard methods. These blots were screened with a set of oligonucleotides specific for the various V \textsubscript{\textgamma}4 family genes. Only those clones that screened positive with constant region probes but negative for the various V \textsubscript{\textgamma}4 complementarity-determining region (CDR)1-specific probes were sequenced (395 of 583 clones), thus enriching the somatically mutated populations analyzed, in that the CDR1 probes should anneal only to the sequences most similar to germline. The frequency of the occurrence of these events can therefore only be predicted to be between 6 out of 395 and 6 out of 583 clones (1–2%). Any sequence of interest was resequenced in both directions to ensure sequence fidelity.

Characterizing the Genomic Repertoire. Total genomic DNA was isolated from FM B cells (IgD–CD38+) using the Puregene DNA isolation kit (Genenta Systems, Inc., M Innpeolis, MN). V \textsubscript{\textgamma}4 genes were amplified using a V \textsubscript{\textgamma}4 leader-specific primer (L-4, as above) and a primer specific for all V \textsubscript{\textgamma}4 family gene heptamer–nonamer spacer regions as previously described (16). PCR products were agarose gel purified, then cloned into E. coli as described above for the cDNA clones. Clones identified in the cDNA analysis that contained insertions or deletions were used to design PCR primers to amplify both the exact sequence of clones with insertions/deletions as found and the predicted sequences based on the proposed germline counterparts. Oligonucleotides used in this analysis (Format, as is follows clone: exact/predicted) were: g645′-GGACGGG-GTTGTAGCTTGGTCC-3′/5′-GGAGGGTTGTAGGGGTAAGTC-3′; g1445′-TCTTGAAGCCGCGGTTGTGTT-3′/5′-TCTTGAGGGACGGGTTGT-3′; g1875′-CACCTCAGTG-TAAGCACC-3′/5′-CAGCTCCAGTAGTAACCA-CGCC; g1885′-GAGGGGTTGTAGGGGTAAGTC-3′/5′-GAGGGTGGT-ATGGGTGTC-3′; and g1895′-GGAGGGTTGTAGGGGTAAGTC-3′/5′-GGAGGGTTGT-AGGGGTAGTAACTCACC-3′.

Sequence Availability. All cDNA sequences with insertions or deletions, and any genomic sequences unique to the literature as described in the results section are available from EMBL/Genbank/ DDBJ under accession numbers AF013615 through AF013626. A say for Screening V \textsubscript{\textgamma}4 Gene Lengths. To facilitate the analysis of large numbers of V \textsubscript{\textgamma}4 gene transcripts for the presence of insertions or deletions, first strand cDNA produced as described above was PCR amplified using Expand high fidelity polymerase (Boehringer Mannheim) to reduce errors resulting from Taq polymerase alone. The products of this PCR amplification were cloned as described above and screened using 32P-labeled, gene-specific oligonucleotides (IgD–CD38+ L-4). Amplification products were sequenced by the dideoxynucleotide chain termination method using Sequenase (US Biochemicals, Cleveland, OH). Sequence reads were obtained from 12–20 clones per sample. Positive clones were picked, and plasmid mini-preparations were made. Sequences were analyzed with the DNAStar (DNAStar Inc., Madison, WI). All sequences were analyzed using DNAStar (DNASTAR Inc., Madison, WI). In the first tonsil analyzed, 583 clones were picked, plasmid mini-preparations were made, and Southern blots were prepared by standard methods. These blots were screened with a set of oligonucleotides specific for the various V \textsubscript{\textgamma}4 family genes. Only those clones that screened positive with constant region probes but negative for the various V \textsubscript{\textgamma}4 complementarity-determining region (CDR)1-specific probes were sequenced (395 of 583 clones), thus enriching the somatically mutated populations analyzed, in that the CDR1 probes should anneal only to the sequences most similar to germline. The frequency of the occurrence of these events can therefore only be predicted to be between 6 out of 395 and 6 out of 583 clones (1–2%). Any sequence of interest was resequenced in both directions to ensure sequence fidelity.
A) Insertion events from a single tonsil:

<table>
<thead>
<tr>
<th>VH4</th>
<th>CDR2</th>
<th>CDR3</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCTCCCACTATAATAGATACT-3'</td>
<td>GCTTCGCTACCCCTATCTACAT-3'</td>
<td></td>
</tr>
<tr>
<td>CDR1</td>
<td>CDR2</td>
<td>CDR3</td>
</tr>
<tr>
<td>GCTCCCACTATAATAGATACT-3'</td>
<td>GCTTCGCTACCCCTATCTACAT-3'</td>
<td></td>
</tr>
</tbody>
</table>

B) Deletion events from a single tonsil:

<table>
<thead>
<tr>
<th>VH4</th>
<th>CDR2</th>
<th>CDR3</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCTCCCACTATAATAGATACT-3'</td>
<td>GCTTCGCTACCCCTATCTACAT-3'</td>
<td></td>
</tr>
<tr>
<td>CDR1</td>
<td>CDR2</td>
<td>CDR3</td>
</tr>
<tr>
<td>GCTCCCACTATAATAGATACT-3'</td>
<td>GCTTCGCTACCCCTATCTACAT-3'</td>
<td></td>
</tr>
</tbody>
</table>

GCTCCCACTATAATAGATACT-3') or for analysis of Vh6 genes a 166-nucleotide fragment including the CDR1 and CDR2 of Vh6. Vh6FW1: 5'-TGCACATCTCCGGGACAGTGT-3', Vh6FW3: 5'-TGTGGTCTGGGTTGATGGTTAT-3', were sequenced in either direction. The amplification products were electrophoresed on 0.6X-TBE, 5% urea-acrylamide sequencing gels (Long R anger; J.T. Baker, Phillipsburg, NJ) and analyzed with a PhosphorImager (Molecular Dynamics, Sunnyvale, CA) using the Image Quant software supplied by the manufacturer. Clones that differed from the expected size and those clones in lanes adjacent to aberrantly migrating bands were used to produce plasmid preparations from which the inserts were sequenced in either direction.

Scoring of Insertion/Deletion Events. In the results section, insertions and deletions are scored as events per 10^4 nucleotides within the customary boundaries of CDR1 and CDR2. This unit was chosen because in the selected populations studied these events are generally only found in the CDR regions and therefore the comparison of events per total nucleotides would be misleading. In the PAGE analysis, each Vh4-39 FM clone included only the CDR1 (21 nucleotides) within a total of 230 nucleotides/clone, whereas each Vh6 FM clone was only 166 nucleotides but included both the CDR1 and CDR2 (75 CDR nucleotides). In the sequencing analysis, various B cell populations were analyzed involving a wide range of overall lengths. Comparisons of the frequency of insertions/deletions just within the CDRs allowed for a more standardized and quantitative analysis, and for more freedom in experimental design.

Baculovirus Expression System. Cloning and coexpression of clone pg86 and κ light chain FS6k in the baculovirus baculoBac system was performed as previously described (17). Recombinant A utographa californica nuclear polyhedrosis virus (AcMNPV) was cloned using the ph360NX transfer vector and expressed in Sf9 cells.

Capture ELISA for Heavy Chain, and κ Light Chains. Expression of recombinant antibodies of clone pg86 coexpressed with κ light chain FS6k were measured by capture ELISA. Wells were coated with goat anti-human IgG and incubated with supernatant of recombinant pg86/FS6k added in serial twofold dilutions. Bound antibody was detected using alkaline phosphatase-conjugated goat anti-human IgG, or goat anti-human Ck. After 1-h incubation at 37°C, phosphatase substrate was added and absorbance was measured at 405 nm in an ELISA plate reader.

Results

Insertions and Deletions into Immunoglobulin Vh Genes. In a large scale analysis of Vh genes from both the IgM and IgG compartments of B cell subpopulations separated from a single human tonsil, six clones that contained DNA insertions or deletions were isolated. These insertions and deletions were apparently selected in that they involved nucleotide triplets or multiples of nucleotide triplets, leaving the cDNAs (transcripts) in frame, and they were localized to the CDR1 and CDR2 (Fig. 1, A and B). The six clones with insertions or deletions were identified from the sequencing of 395 cDNA clones (~110,000 nucleotides) from GC and memory B cell subpopulations, resulting in a frequency of <2% of clones analyzed (~1 event/18,000 nucleotides). All six events were in IgG transcripts. Two events were obtained from IgD-CD38+ GC and four events from IgD-CD38- memory cell populations. None of the IgM Vh cDNAs analyzed from this tonsil had insertions or deletions, although we have observed such events in IgM transcripts in the past and in subsequent analyses, as described below.
The Insertions and Deletions Are Not Germline Encoded. The analysis described above focused on the VH4 gene family, which consists of 10-14 member genomes, varying slightly between individuals (16, 18). As shown in Fig. 2, the major difference between VH4 genes involves the length of CDR1. Because genomic diversity between VH4 family members resembles the events described in this paper we had to rule out possible alternative explanations for these events, such as (a) different alleles of the detected genes (b) rarely expressed or otherwise unknown VH4 gene family members (c) hybrids between known and detected VH4 genes and/or other artifacts of the experimental system. To address these issues, both the expressed and genomic repertoires from this tonsil were characterized. As indicated in Table 1, 2 out of 118 VH4-4-39, 2 out of 49 VH4-4-31, 1 out of 87 VH4-4-34, and 1 out of 45 VH4-4-59 cDNA clones contained insertions/deletions. cDNA clones were judged as unique isolates based on CDR3 analysis, and the few isolates that appeared to be clonally related differed in their patterns of somatic mutation beyond the level explainable by reverse transcription and PCR errors (maximum: >1 mutation/500 nucleotides of VH4 gene sequence as previously described [4]). To characterize the genomic repertoire of the initial tonsil, 80 germline VH4 gene clones were isolated and sequenced (Table 1), which encompassed all 14 VH4 family members or alternate alleles represented in the 446 cDNA clones analyzed from all of the tonsillar B cell subsets. In the course of this study, we isolated the germline counterpart of a novel VH4 gene segment for which transcripts had been found. In addition, germline genes corresponding to two apparently functional VH4 genes not found as cDNA clones in this analysis were isolated, as well as one nonfunctional VH4 gene and a divergent polymorphism of a known VH4 pseudogene. The proposed germline counterparts of each of the VH4 genes containing insertion/deletion events were isolated from 4 to 11 times (Table 1). 8 independent genomic isolates of VH4-4-31 and of VH4-4-39 were cloned. VH4-4-34 and VH4-4-59 were isolated 11 and 4 times, respectively. No germline genes were isolated that could have encoded the insertion/deletion events described.

Further to be certain that the insertion/deletion events described herein were not germline encoded, two sets of PCR primers were designed to specifically recognize: (a) the exact sequence of the events; (b) the predicted, unmutated, germline sequence corresponding to the cDNAs containing insertion and deletion events. These primers were used to amplify genomic DNA from this individual, yielding negative results (data not shown). The unique nature of these events relative to both the expressed and genomic repertoire and our inability to amplify genomic counterparts for these events by PCR suggest that they are not germline encoded.

The Proposed Insertion/Deletion Events Are Not the Result of V(H/V) Recombination. As in most V gene repertoire analyses, we detected hybrid VH4 sequences that could be the result of either PCR splicing by overlap extension artifacts, or reciprocal homologous recombination between unarranged V genes (19). However, none of these likely artefactual events were altered in size such that they resembled the insertion or deletion of DNA described above. A number of artifacts of this type had been isolated in the cDNA analysis as well; such artifacts are common to V gene analyses (20). The cDNA isolates with deletion and insertion events were stringently compared to all germline and cDNA isolates and were found to be unique relative to both the expressed and germline VH4 gene repertoires of this individual, supporting a somatic origin for their occurrence.

The Insertions and Deletions Are Associated with Somatic Hypermutation. To determine whether or not these inser-
tion/deletion events were associated with somatic hypermutation, we analyzed their occurrence in unmutated FM transcripts. This was done using either direct sequencing or PCR amplification of portions of the VH genes spanning the CDRs, followed by size comparisons on polyacrylamide gels (Fig. 3). Any clones that ran aberrantly, and the clones in adjacent lanes, were sequenced (75 out of the 485 clones). None of these 75 clones were related based on CDR3 homology. To ensure that the remaining 410 FM clones were polyclonal, the CDR3s were PCR amplified and loaded on the sequencing gels simultaneously to the VH gene amplification products for size comparisons (Fig. 3A). The size distribution of these CDR3s was similar to that of ~500 VH gene sequences analyzed in this study (Fig. 3B), providing evidence that our FM sample is polyclonal.

The six events detected from a single tonsil were isolated from 395 mutated cDNA clones (25,482 CDR nucleotides), corresponding to a frequency of 2.35 events/10^4 CDR nucleotides. This is significantly different (p = 0.014 by a one-sided χ² test) from the analysis of unmutated FM-derived clones (25,515 CDR nucleotides) that yielded no insertions or deletions (Table 2).

In the course of the analysis described above, we isolated one IgM clone containing a 6-nucleotide insertion into framework (FW)3 (see below). We believe that this clone is part of the mutated GC or memory repertoire because it contained 4 bp substitutions in addition to the insertion. In this study, the B cell populations analyzed were 95–98% pure, and the FM B cell subpopulation could therefore include between 2 and 5% contaminating clones, that is, IgM-expressing cells not from the naïve population that can therefore be somatically mutated. However, none of

![Figure 3](https://example.com/figure3.png)

**Figure 3.** Polyacrylamide gel assay to identify insertions or deletions into VH genes. (A) Phosphorimage of a polyacrylamide gel: each lane contains the hot-PCR products (32P-labeled) of the VH gene and the CDR3 of an individual clone. (B) A comparison of the distribution of CDR3 sizes of the 485 CDR3s assayed to the distribution of 500 CDR3s observed in sequences from this report indicates that the clones assayed by electrophoresis were a polyclonal population. CDR3 sizes were measured from the most 3’ Tyr residue (common to all V genes analyzed) to the most 5’ Cµ or Cγ residue. CDR3 lengths for those assayed by electrophoresis were extrapolated based on sequencing of 75 out of the 485 clones assayed. The x-axis is the number of amino acids greater than the shortest CDR3 found.

**Table 2.** A analysis of unmutated FM cDNA clones for insertion or deletion events

<table>
<thead>
<tr>
<th>Clone type</th>
<th>Clones assayed</th>
<th>CDR nucleotides</th>
<th>Events observed</th>
<th>Frequency</th>
<th>Expected (events/10^4 CDR nucleotides)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mutated VH4 clones (GC and memory B cells)</td>
<td>395</td>
<td>25,482</td>
<td>6</td>
<td>2.35 events/10^4 mn</td>
<td></td>
</tr>
<tr>
<td>Unmutated clones:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VH4-FM, CDR1*</td>
<td>265</td>
<td>5,565</td>
<td>0</td>
<td>0</td>
<td>1.31</td>
</tr>
<tr>
<td>VH6 IgM FM VH4 genes*</td>
<td>220</td>
<td>16,500</td>
<td>0</td>
<td>0</td>
<td>3.38</td>
</tr>
<tr>
<td>VH4 family FM sequences</td>
<td>51</td>
<td>3,450</td>
<td>0</td>
<td>0</td>
<td>0.81</td>
</tr>
<tr>
<td>Total unmutated values</td>
<td></td>
<td>25,515</td>
<td>0</td>
<td>0</td>
<td>2.35 events/10^4 CDR nucleotides</td>
</tr>
</tbody>
</table>

* Clones analyzed by hot-PCR/PAGE assay as described in the text.

* CDR nucleotides are those within the customary bounds of the CDR1 and CDR2. (See Materials and Methods for a more detailed explanation of this unit).

* Events per 10^4 CDR nucleotides.

* Expected frequency (events/10^4 CDR nucleotides) derived from sequencing data: 6 events in 25,482 CDR nucleotides; 6/(25,482 CDR nucleotides/10^4) = 2.35.

* Statistical analysis: χ² test for independence.
A. Insertion events:

\[
\begin{align*}
V_6: & \quad \text{GTG TCT AGC} \quad \text{AAC AGT GCT} \\
\text{hap-2:} & \quad \text{GTG TCT AGC AGC AAC AGT GCT} \\
V_6-39: & \quad \text{TAC TAC AAC} \quad \text{CGG TCC TCT} \\
\text{g144:} & \quad \text{TAC TAC ACC AAC CCG TCC TCT} \\
V_6-39: & \quad \text{AGT TAC TAC} \quad \text{TGG GGC TGG} \\
\text{g192:} & \quad \text{AGT TAC TAC TAC TGG GGC TGG} \\
V_6: & \quad \text{TCC AAG AAC} \quad \text{--- CAG TTC} \\
\text{tm121:} & \quad \text{TCC AAG AAC AAG AAC CAG TTC}
\end{align*}
\]

B. Deletion events:

\[
\begin{align*}
V_6-31: & \quad \text{GGG AGC ACC} \quad \text{TAC TAC AAC CGG} \\
\text{g64:} & \quad \text{GGA ACC Aag} \quad \text{TAC AAC CGG} \\
V_6-31: & \quad \text{TCC ATG AGC AGT} \quad \text{GGT GGT TAC} \\
\text{g187:} & \quad \text{TCC ATG AGC} \quad \text{--- GGG GCT TAC} \\
V_6-59: & \quad \text{TAC AGT GGG AGC ACC AAC TAC} \\
\text{g188:} & \quad \text{TAC AGG GGG} \quad \text{--- tcc AAC TAC} \\
V_6-34: & \quad \text{TAC TAC TGG TGG TGG TAC CCG} \\
\text{g801:} & \quad \text{agc TAC} \quad \text{--- TGG TAC CCG}
\end{align*}
\]

C. Long duplication/insertion:

\[
\begin{align*}
\text{V_6-31:} & \quad \text{GGG AGC ACC TAC TAC AAC CGG} \\
\text{g64:} & \quad \text{GGA ACC Aag TAC AAC CGG} \\
\text{V_6-31:} & \quad \text{TCC ATG AGC AGT GGT GGT TAC} \\
\text{g187:} & \quad \text{TCC ATG AGC --- GGG GCT TAC} \\
\text{V_6-59:} & \quad \text{TAC AGT GGG AGC ACC AAC TAC} \\
\text{g188:} & \quad \text{TAC AGG GGG --- tcc AAC TAC} \\
\text{V_6-34:} & \quad \text{TAC TAC TGG TGG TGG TAC CCG} \\
\text{g801:} & \quad \text{agc TAC --- TGG TAC CCG}
\end{align*}
\]

The unmutated FM clones analyzed had insertions or deletions.

Other insertions and deletions into V_{\text{H}} genes. We have observed similar instances of insertions and deletions into the coding regions of apparently functional immunoglobulin V genes, including: (a) a V_{\text{H}} 6 IgM isolate containing a triplet duplication/insertion into the CDR1 in addition to several bp substitutions (Fig. 4A, clone tm121); (b) a 6-nucleotide insertion into the FW3 region of a mutated IgM V_{\text{H}} 6 gene, representing the only insertion or deletion observed outside of the CDRs (Fig. 1C and 4A, clone tm121); and (c) an 18-nucleotide duplication/insertion into a human plasma cell cDNA transcript at the boundary between the FW1 and CDR1 (Figs. 1C and 4C), doubling the length of this hypervariable loop. The viability of clone pg86 was tested by expressing it in the baculovirus system in association with a k light chain encoding construct (FS-6\_k; Fig. 5). The efficient expression, secretion, and pairing with light chain in the baculovirus system suggest that the product of clone pg86 is a functional heavy chain despite the large duplication/insertion.

The insertions and deletions are related to the surrounding DNA sequence. As shown in Fig. 4, the insertions involve repetitions of the immediately adjacent sequence. (A) The insertions are deletions of tandem repeats. (C) The 18-base insertion in clone pg86 is a duplication of the adjacent sequence. (C) The insertions and deletions are related to the surrounding DNA sequence. (A) The insertions involve repetitions of the immediately adjacent sequence. (B) The deletions are deletions of tandem repeats. (C) The 18-base insertion in clone pg86 is a duplication of the adjacent sequence. 

Figure 4. The insertions and deletions are related to the surrounding DNA sequence. (A) The insertions involve repetitions of the immediately adjacent sequence. (B) The deletions are deletions of tandem repeats. (C) The 18-base insertion in clone pg86 is a duplication of the adjacent sequence. 

Discussion

Somatic modification of V genes encoding immunoglobulin and T cell receptors recapitulates most mechanisms observed in the evolutionary diversification of DNA: (a) V gene recombination, including imprecise junctions, P nucleotides, and untemplated N nucleotide addition; (b) gene conversion; and (c) bp substitutions in Ig somatic hypermutation. The insertion and deletion of nucleotides is another means for the evolutionary diversification of DNA, and has been proposed as an explanation for unusual V gene sequences in the past (Table 3). In this study, we show that insertions and deletions are associated with the somatic hypermutation process.
Complements of the Analysis of Insertions and Deletions into VH Genes. The formal characterization of these events has been a daunting task because of their low frequency, and the complexity of the germline VH repertoire. According to our study, these events occur in <2% of somatically mutated clones. As shown in Fig. 2, the primary variability between VH 4 family members is 3-6-bp size variances in the CDR1s, which is comparable to the short insertions and deletions that we attribute to somatic hypermutation (in selected B cell populations). The similarity between evolutionary diversity and somatic diversification was expected, as the molecules are likely subject to the same functional and structural constraints. This has made it difficult to determine whether these events were generated somatically, versus germline encoded, or if they were artifacts of the experimental system: they could result from homologous recombination between alternate alleles or imperfect recombination between identical alleles, or they could have occurred during B cell replication independent of somatic hypermutation. In fact, VH genes may exhibit particularly unstable sequence characteristics evolved to help support both germline diversity and the generation of somatic mutations, as suggested by the identification of intrinsic hotspots of somatic hypermutation within the CDR s of V genes (25, 26). Perhaps the area of greatest contention in this complex system remains the possibility that these low frequency events are artifacts of the experimental manipulations performed, the AMV-RT, Taq, or PFU polymerases, and/or the cloning in E. coli.

The Insertion/Deletion Events Are the Result of the Somatic Hypermutation Process. Our system addresses several key issues that associate the occurrence of insertions and deletions to the somatic hypermutation process. (a) Six of the nine insertions/deletions were identified within the VH 4 gene repertoire of a single tonsil, providing an experimental system that could be characterized extensively as described below. (b) All of the insertion/deletion events reported involved triplets or multiples of triplets, leaving the transcripts in frame and therefore functional, and eight of nine events reported were localized to the CDR s. As with somatic point mutations, no insertions or deletions were observed in the 80 to 120 nucleotides of constant region (Cμ or Cγ) DNA sequenced with each cDNA clone. These hallmarks of somatic hypermutation and selection argue strongly that these events are not artifacts. (c) The B cells analyzed were processed and separated into highly pure, mutated B cell populations including GC (IgD-CD38+), and memory (IgD-CD38+) B cells, and an unmutated FM B cell population (IgD-CD38-), making it possible to focus our analysis on the mutated populations and use the unmutated population as a negative control, which in turn allows the statistical association of the observed insertion and deletions to the somatic hypermutation process (P = 0.014). In addition, the isolation of four of the insertion/deletion events from memory B cells provides evidence that these events did not result from artifacts related to contamination from endonucleolytically cleaved DNA from the apoptotic GC cells. (d) Seven of nine events reported in this study involved γ heavy chains that contain nearly twice the mutations of μ heavy chains (4), further correlating the events described here to somatic hypermutation. (e) As discussed below, the insertion/deletion events described tended to involve sequence motifs resembling previously described hotspots of somatic hypermutation, providing evidence that these events occur by the same process. (f) Finally, we extensively analyzed the VH 4 gene family of the tonsil donor at both the expressed and genomic levels, facilitating the assignment of the insertions/deletions as somatic rather than germline encoded. 6 of the clones with insertions and deletions were unique among 395 VH 4 cDNA clones sequenced from a single tonsil, including many independent isolates of each of the VH 4 genes expressed (Table 1). In addition, we were unable to isolate genomic templates for any of the insertion or deletion events either by PCR or through the extensive characterization of the genomic VH 4 repertoire of the tonsil donor (Table 1). Templating of these events from any other VH gene family can also be ruled out as members of the seven human VH gene families differ significantly in the CDR sequences where the events described had occurred.

Structural and Functional Considerations of Insertions and Deletions into VH Genes. The events involving the insertion or deletion of a single amino acid from the CDR 1 or CDR 2 would not be expected to profoundly alter the backbone structure of these molecules, as the CDR s are the most malleable portions of antibodies. The clone g80 has two of the five amino acids that are customarily considered its CDR 1 deleted, leaving only three amino acids to form this hypervariable loop (Fig. 1B). Thus, this is one of the shortest CDR 1s reported to date. The clone tm121 has two amino acids inserted into the FW 3 region. The portion of the FW 3 where this insertion occurred is believed to be solvent exposed and corresponds to the region where the B cell superantigen staphylococcal protein A binds to most VH 3-encoded Ig molecules (28); therefore, it is likely that the insertion into this VH 6 clone can be tolerated as a loop or bulge on the molecule’s surface. The most complex structural change observed in our study involved clone pg86, with a six amino acid insertion at the FW 1/CDR 1 junction that would presumably double the length of this hypervariable loop and require dramatic structural accommodation. However, we were able to express this heavy chain and found it paired with light chain, indicating that it is likely functional (Fig. 5). The clone HBP2, containing a triplet insert into its CDR 1, is particularly interesting because it has a known specificity. This VH 6 gene was isolated from a human B cell hybridoma with anti-Babesia microti specificity (21, 22). Clone HBP2 has also been expressed in the baculovirus system and is fully functional. We are currently performing mutational analysis of this heavy chain molecule to determine if the additional inserted amino acid plays a role in the affinity and/or specificity of this antibody.

Analysis of Insertions and Deletions Reported in the Literature. Various groups have reported a number of insertion and deletion events (Table 3). Virtually all of the insertions...
Table 3. Insertions and Deletions into Somatically mutated V Genes Reported in the Literature

<table>
<thead>
<tr>
<th>Name</th>
<th>Source</th>
<th>Ins/Del (position)</th>
<th>Relation to surrounding sequence</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Selected populations or coding regions:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L4-1e</td>
<td>Human VH4-34 (4.21)</td>
<td>ACC insert (within CDR2)</td>
<td>4-34: AGC ACC AAC (RT)</td>
<td>38</td>
</tr>
<tr>
<td>3B62</td>
<td>Murine VH 186.2</td>
<td>GTT deletion (CDR2)</td>
<td>VH186.2: AGT GGT GGT ACT (RT)</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Unselected populations or untranslated regions:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3B62</td>
<td>Murine VH 186.2/D/JH2 to JH4</td>
<td>ACT deletion (3’ untranslated)</td>
<td>GL: GTG ACT ACT TTG (RT)</td>
<td>40</td>
</tr>
<tr>
<td>3B62</td>
<td>Murine VH 186.2</td>
<td>4 single-base deletions (leader intron)</td>
<td>VH186.2: GCC, GGT (RT)</td>
<td>39, 40</td>
</tr>
<tr>
<td>M167</td>
<td>Murine VH 107/DFL16/JH1</td>
<td>2 single-base insertions (leader intron)</td>
<td>GL: ATAC AAGTATTAGTAG (RT)</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3’ untranslated)</td>
<td>M167: ATAGTAAGTATTAGTAG (3’ untranslated)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>M167: TTTGAGGTGATGAAAGGA (3’ untranslated)</td>
<td>29, 41</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>M167: GCCTTTG TGTA...CCCAGAAAAAGA (IR)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>M167: GCTTTTT TGTA...CCCGG AAAAAA (IR)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>M167: CTCTCTCTT (RT)</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>M167: AGATTTTAC (IR)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>M167: TCAGTG (RT)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>M167: GTGACTACTTTTACTG (RT)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>M167: GT ACTACTTTTACTG</td>
<td></td>
</tr>
</tbody>
</table>

Continued
<table>
<thead>
<tr>
<th>Name</th>
<th>Source</th>
<th>Ins/Del (position)</th>
<th>Relation to surrounding sequence</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>M603</td>
<td>Murine V(_\alpha)S10 7/DFL16.1/JH1</td>
<td>TA deletion (3')</td>
<td>None found (possible hotspot)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>GTGT deletion (leader intron)</td>
<td>GL: TCTGTGTGTGTGTAT (RT) M67: TCTGTGT GTAT</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>M603: TTTTCTGTCTTTCTTTTT (RT) M603: TTTTCTGTCTTTCTTTTT</td>
<td>42, 41</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AAAT deletion (3' Ei/MAR)</td>
<td>GL: GCATTCTAAAATAAGTGAGGA (IR) M603 GCATTCTAAAATAAGTGAGGA</td>
<td></td>
</tr>
<tr>
<td>MC101</td>
<td>Murine V(<em>\alpha)Q52/D/J(</em>\nu) 3</td>
<td>GG deletion (3' untranslated)</td>
<td>GL: AACCGGGAATC (RT) MC101: AAAC GAATC</td>
<td>42, 43</td>
</tr>
<tr>
<td>M511</td>
<td>Murine Vk167/Jk5</td>
<td>GAA deletion (3' untranslated)</td>
<td>GL: TTTGAAGATAAA (RT) M511: TTT GAATAA</td>
<td>42, 44</td>
</tr>
<tr>
<td>H37-65</td>
<td>Murine Vk Vk21E/Jk1-Jk2</td>
<td>11 base deletion (Jk1/Jk2 intron)</td>
<td>GL: AGGGACACCAGTGTGTACGAC (IL) H37: AGGGc GTACAC</td>
<td>45</td>
</tr>
<tr>
<td>296.4C11, 253.12D3</td>
<td>Murine JkC intron</td>
<td>7 base deletion and a 154 base deletion</td>
<td>No good relationship to surrounding sequence</td>
<td>46</td>
</tr>
<tr>
<td>2G7</td>
<td>Murine transgene</td>
<td>single-base deletion, and a 49 nucleotide deletion</td>
<td>CTTTGAAGAT... (N 30)... CAGATCAAG (R repeats form ends of deleted “loop”) (IL)</td>
<td>47</td>
</tr>
<tr>
<td>85k</td>
<td>Human myeloma Vk genes</td>
<td>single-base (T) insertions into the CDR1/FW2 junction rendering genes out-of-frame</td>
<td>No relation, however, event followed the proposed hotspot motif TAC</td>
<td>48</td>
</tr>
<tr>
<td>H-F-1(^{2}) clone A6: Human lymphoma (J(_\nu) untranslated)</td>
<td>AG insertion into V(_\nu)3' untranslated region</td>
<td>Consensus: GGGGCAG GGC (RT) clone A6: GGGGCAGAGGGG</td>
<td>49</td>
<td></td>
</tr>
</tbody>
</table>

RT, repetitive tract; IR, inverted repeat (loop with local DNA); IL, internal loop. \(^{1}\) Secondary structure reported by Golding et al. (29). \(^{2}\) This study is difficult to interpret in the context of the current report as the genomic J\(_\nu\) locus was not available. The 10 clones were only 80\% homologous to the closest J\(_\nu\) locus reported in the literature with most alterations being similar between all of the isolates. Therefore, only 2 of the 26 proposed insertions/deletions can be attributed to somatic mutation with certainty, as they were unique to the consensus of the individual clones.
Somatic Hypermutation. The evidence for the involvement of the template strand in the repetitive element or in the daughter strand, or a deletion event reported in our study resembled one of these hotspots (AGC, TAC, and AAC; references 12 and 27; Figs. 1 and 4). The analysis of selected populations may have influenced this tendency because seven out of eight of these events occurred in the CDRs where it has been shown that hotspot motifs are preferentially found (25, 26). Furthermore, only a weak correlation to hotspots could be found for the previously reported insertions/deletions involving unselected regions of V loci (Table 3). However, the single event found in this analysis that occurred outside of the CDRs in 

A Model for the Occurrence of Insertions and Deletions during Somatic Hypermutation. The evidence for the involvement of DNA secondary structure in the production of insertion or deletion mutations during somatic hypermutation, as suggested in 1986 by Golding et al. (29), now seems unequivocal. The insertions and deletions described in our study, and those illustrated in Table 3, occur in a predictable fashion, involving sequence motifs that could form loop intermediates reminiscent of the replication slippage model of Streisinger et al. (30) and Ripley (31). The same model can account for both (A) insertions and (B) deletions.

Figure 6. Proposed mechanism causing insertion/deletion events polymerase slippage. This figure is based on model a of Streisinger et al. (30) and Ripley (31). The model can account for both (A) insertions and (B) deletions.
same advantage, i.e., the alteration of a single complex, would be the alteration of a DNA repair system such as transcription-coupled repair to be the somatic mutator, as suggested in recent studies (13). Alternatively, the insertions and deletions might result solely from a downregulation of postreplicative mismatch repair at the V locus in the rapidly proliferating centroblasts that are undergoing somatic hypermutation or due to a polymerase enzyme with such a high fault rate as to overwhelm any repair.

All currently accepted models of somatic hypermutation, whether related to DNA excision-repair–like systems or transcription-repair, or to DNA polymerization or reverse transcription, involve transcriptional activation involving cis-factors in the V locus (enhancers, etc.) followed by the activity of unknown polymerase enzymes of some type. This analysis does not refute or corroborate any of these models directly, but it does provide further characterization of the polymerization system involved, based on the types of mutations observed and on the molecular biology that is known to cause such mutations. This analysis and the model presented here provide further information or criteria to be contemplated as the various possible polymerase systems involved are considered.

Conclusions. Insertions and deletions into immunoglobulin V\(_5\) genes during somatic hypermutation are additional means by which the immunoglobulin repertoire can be diversified. These events display characteristics supporting models of somatic hypermutation involving a particularly unstable or error-prone polymerase to allow the introduction of mutations, and involving the downregulation of DNA repair to allow the perpetuation of these mutations. Additionally, we show that these events tend to involve sequence motifs resembling intrinsic hotspots of somatic hypermutation, suggesting that the polymerase complex is destabilized in a sequence-specific manner to allow preferential mutation at these sequence elements.

We are grateful to Yucheng Li, Fang Zhao, Steve Scholl, Carol Williams, Shirley Hall, and Robin Wray, all of whom provided unprecedented technical assistance for various aspects of this work. We also thank Kimberly Frazer for excellent discussions.

These studies were supported by a grant from the National Institutes of Health (AI-12127). J. Donald Capra holds the Edwin L. Cox Distinguished Chair in Immunology and Genetics.

Address correspondence to Dr. Virginia Pascual, Molecular Immunology Center, Department of Microbiology, UT Southwestern Medical Center, 6000 Harry Hines Blvd., Dallas, Texas 75235-9140. Phone: 214-648-1918; Fax: 214-648-1915; E-mail: vpascu@mednet.swmed.edu

Received for publication 14 July 1997 and in revised form 10 October 1997.

References

17. Potter, K.N., Y.C. Li, and J.D. Capra. 1994. The cross-react-


49. Wu, H.Y., and M. K arnteen. 1995. The somatic hypermuta-