Error-Prone DNA Repair Activity during Somatic Hypermutation in Shark B Lymphocytes

Catherine Zhu and Ellen Hsu

*J Immunol* 2010; 185:5336-5347; Prepublished online 4 October 2010;
doi: 10.4049/jimmunol.1000779

http://www.jimmunol.org/content/185/9/5336

Supplementary Material

http://www.jimmunol.org/content/suppl/2010/10/04/jimmunol.1000779.DC1

References

This article cites 76 articles, 34 of which you can access for free at:
http://www.jimmunol.org/content/185/9/5336.full#ref-list-1

Subscription

Information about subscribing to *The Journal of Immunology* is online at:
http://jimmunol.org/subscription

Permissions

Submit copyright permission requests at:
http://www.aai.org/About/Publications/JI/copyright.html

Email Alerts

Receive free email-alerts when new articles cite this article. Sign up at:
http://jimmunol.org/alerts
Error-Prone DNA Repair Activity during Somatic Hypermutation in Shark B Lymphocytes

Catherine Zhu and Ellen Hsu

Sharks are representatives of the earliest vertebrates that possess an immune system utilizing V(D)J recombination to generate Ag receptors. Their Ab repertoire diversity is based in part on a somatic hypermutation process that introduces adjacent nucleotide substitutions of 2–5 bp. We have isolated mutant nonfunctional Ig rearrangements and intronic flank sequences to characterize the nonselected, intrinsic properties of this phenomenon; changes unique to shark were observed. Duplications and deletions were associated with N additions, suggesting participation of a DNA polymerase with some degree of template independence during the repair of DNA breaks initiated by activation-induced cytidine deaminase. Other mutations were consistent with some in vitro activities of mammalian translesion DNA polymerase η: tandem base substitutions, strand slippage, and small insertions/deletions. The nature of substitution patterns shows that DNA lesions at shark Ig genes recruit DNA repair factors with a species-specific repertoire of activities. We speculate that the tandem mutations are introduced by direct sequential misinsertions and that, in shark B cells, the mispairs tend to be extended rather than proofread. Despite extensive changes undergone by some mutants, the physical range of mutational activity remained restricted to VDJ and within the first 2-kb portion of the 6.8-kb J-C intron, perhaps a self-regulating aspect of activation-induced cytidine deaminase action that is conserved in evolution. The Journal of Immunology, 2010, 185: 5336–5347.

Adaptive immunity in vertebrates is based on lymphocytes and their diverse Ag receptors. The Ag-combining site of Ig or TCR is generated by V(D)J rearrangement, a somatic recombination mechanism entailing combinatorial joining of multiple, tandemly duplicated gene segments (1). During the recombination process, the flanks of the V, D, or J gene segments are cleaved and undergo modifications such as nucleotide deletion as well as nontemplated additions by TdT. This process generates diverse sequence and sequence-length differences at the joining sites, which encode loops forming the Ag-combining site. Additional postrearrangement mechanisms as gene conversion and somatic hypermutation (SHM) further diversify the Ag receptor repertoire, and, depending upon the animal species, the repertoire is expanded in immature B lymphocytes and/or in Ag-specific cells, honing Ab affinity during an immune response (2). Both mechanisms are initiated by activation-induced cytidine deaminase (AID), which converts cytidine to uracil in the Ig gene (3–5). An AID homolog has been isolated in every jawed vertebrate class including cartilaginous fishes (6).

In tetapods, almost all of the changes resulting from SHM are point mutations. Repair of the uridine/guanosine mismatch results in mutational DNA synthesis, at the position or nearby (recently reviewed in Refs. 7, 8). Base excision repair (BER) generates an abasic site, across which REV1 inserts an untemplated nucleotide during DNA synthesis. BER and mismatch repair both can give rise to a single-stranded gap that is filled in with low fidelity by translesion DNA synthesis polymerases such as polymerase η (Pol η) and others. The polymerases each have a mutational signature but altogether produce a composite mutational pattern in which substitutions from G/C and A/T base pairs occur about equally, with a bias of ≥50% frequency of transition changes. In the past decade, the pathways acting in SHM have been variously deduced by determining the shift in mutational spectra in mice deficient in key pathways. Although major components have been identified, roles for other candidate enzymes or pathways have yet to be elucidated (reviewed in Refs. 9, 10).

Very little is known of SHM in systems outside mouse or human, but of those that have been studied, the amphibian *Xenopus* and the shark present very different substitution patterns, showing that the contributions by low-fidelity repair polymerases vary among species. The mutations in *Xenopus* Ig consist of 90% substitutions from G/C base pairs with overall 62% transition changes (11). Based on the current model for SHM in mouse, the majority of changes in *Xenopus* could have been generated due to the retention of uracil during DNA replication, causing G/C transition mutations, or by excision of the uracil, recruiting Rev1 to insert across the abasic site and generating some transversion mutations. The 10% A/T changes could arise from gap repair involving Pol η, which preferentially targets A/T base pairs and tends to generate transition mutations (T to C or the complement A to G). In *Xenopus*, the balance of the various pathways is different from mouse, perhaps with BER outcompeting mismatch repair (12).

In the shark, a representative of the earliest vertebrates with rearranging Ig/TCR gene systems, SHM in Ig and Iglike serum proteins includes a unique feature: half of the changes consist of side-by-side tandem substitutions (13, 14). The shark point mutation patterns resemble those in mouse and human Ig, possibly coincidental but indicating the participation of BER and mismatch repair pathways. The tandem substitutions of 2–5 bp appear to be
generated by another polymerase/pathway that does not favor transition mutations, although they, too, are AID-initiated because of the common bias for GRYYW/WRCY hotspots. Gene conversion would seem the most likely mechanism for simultaneous adjacent substitutions, but no sequence donation was apparent among the six genes of the L chain isotype examined (14). However, these earlier studies involved only productive rearrangements, expressed Ig that may have undergone antigenic selection. To avoid biases imposed by selective pressures and discover the intrinsic features of SHM, Storb (15) and Neuberger (16) analyzed changes occurring in passenger L chain transgenes, whereas Dörner and Lipsky (17–19) examined out-of-frame V regions isolated from single human B cells. In this study, we devised a strategy for amplifying non-productive V(D)J rearrangements from nurse shark mRNA. Mutated out-of-frame V regions would in effect serve as unselected bystanders, or reporters, for SHM that occurred in shark B cells activated by environmental Ags or deliberate immunization.

All germline genes encoding the IgM H chain and the NS5 L chain isotype have been defined in the nurse shark (20–24). For these studies, we isolated rearranged Ig sequences from one single-copy IgL gene and two single-copy IgH genes (Fig. 1). Mutations in V1, VDI, and the IgH intronic flanking region were analyzed. We assessed the impact of selection on shark Ig sequences by comparing in- and out-of-frame sequences and examined whether potential mechanisms such as gene conversion or DNA slippage could be participating in shark SHM. From the mutations obtained in the study, we conclude that changes were consistent with a transient misalignment process as well as error patterns generated by translesion polymerases like Pol η during in vitro DNA synthesis.

Materials and Methods

Animals

Nurse sharks (Ginglymostoma cirratum) were captured off the coast of the Florida Keys and bled from the caudal sinus into heparinized syringes. Some animals were sacrificed after capture (shark-33) and others had been immunized (shark-Y, shark-J, and shark-GR) (21). cDNA libraries were constructed from shark-33 epigonal organ and were screened for Ig sequences as previously described (23). PCR was performed on genomic DNA and total RNA extracted separately from the peripheral blood leukocytes (PBLs) of shark-J (22). Surface Ig positive (sIg+) PBLs were isolated from shark-GR using anti-shark Ig mAbs, and genomic DNA extracted from this sample was previously used for genomic Southern blotting (24).

Probes

Probes to the V and C regions of L chain NS5-2 (22) were used to screen cDNA libraries of shark-33. Probes to the VH gene segment of G2 and the intersegmental region between D2 and JH were described previously (24).

PCR assay

For L chain: oligonucleotide primers specific to the NS5 leader intron (NS5L: 5′-CTGGACAAATGCACTAAAGTAG-3′) and the NS5-2 intronic region 3′ of J gene segment (NS5J: 5′-AGAGAACAGATTCGACCT-3′) were used to amplify sequences from PBL genomic DNA. For RT-PCR, forward primers in the leader intron (NS5LI) or in framework (FR) 1 (V1: 5′-GTACATTGGCGCTTCTCG-3′) were used together with a primer in the NS5-2 C region (NS5/C: 5′-AGGACAGAGTTGCCCATC-3′), after first-strand priming with oligo(dt) (Roche Diagnostics, Mannheim, Germany). GenBank accession numbers for SPH-156, SPH2-2, SPH4, and JNS5-38, 39, 40, 41, 80, and 83 are HM171572–HM171581.

For H chain: genomic VDJ and the 3′ intron sequence, amplified by a two-step PCR protocol. Genomic DNA was extracted from sIg+–shark-GR PBL, and 50–200 ng DNA was added to 50 μl reagents (1× PCR, 200 μM dNTP, 5 U Taq per 200 μl reaction) and primers in the leader intron (‘Lint’, Fig. 1) and J-C intron [Lint[C: 5′-TCTTAATTCGTTGATGCT-3′] and G2JC2: 5′-AGCTTCAATTCGTTGAC-3′], and subjected to 20 cycles of PCR (95˚C 1 min, 46˚C 2 min, 72˚C 1 min; last cycle 72˚C 15 min). The PCR products were treated with ExoSAP-IT (USB, Cleveland, OH), and 1 μl was added to 50 μl reaction mix and subjected to a second round of PCR (15 or 20 cycles of 94˚C 1 min, 46˚C 0.5 min, 72˚C 1.5 min; last cycle 72˚C 15 min) with nested primers [LintD (5′-TCTCCAGGTGATGCTMC-3′) and G2JC2R (5′-GCATTAAGTCGTGGTCG-3′)]. The background Taq mutation in this two-stage protocol was 0.15% (15 out of 10,000 bp).

Long PCR (Roche) was performed according to manufacturer’s instructions for System 3 for a total of 35 cycles and optimized for the amount of input genomic DNA. Primers LintA (5′-AAGCTCAGAACTTTGCATATG-3′) and J-C intron [Lint[C: 5′-ATGCCTGTGGATGCT-3′] and G2JC2: 5′-AGCTTCAATTCGTTGAC-3′] were used to raise rearranged products of ~5.8 kb (G2A) and 5.1 kb (G2B). LintD and G2ACH1a (5′-AGGACAGAGTTGCCCATC-3′) were used to obtain rearranged VDJ from the region of the G2A VH to the first 38 bp of the first C region exon, ~7.4 kb for the rearranged configuration. PCR products were cloned into pGEM (Promega, Madison, WI), and plasmid samples were sequenced by Geneviz (South Plainfield, NJ). Cloned PCR products were screened by colony hybridization and selected by expected size. Amplified genomic nonrearranged L chain was anticipated to be 880 bp, distinguishable from recombined VH of ~486 bp; the latter was isolated by size fractionation. Genomic nonrearranged H chain was distinguished by hybridization of the intersegmental region between D2 and JH probe to the intersegmental D2-JH region by colony hybridization assay, whereas rearranged VHD1 hybridized only with the VH gene segment of G2 probe. GenBank accession numbers for S1, S4, S22, S23, S37, S21, S25, S57, S122, A3, C9 are HM029142–HM029150 and HM171582–HM171584.

Germline genes in outbred animals

Two Ig H chain germline genes and one IgL gene were analyzed in this study; they are single-copy genes each with one V, one J, and one C region (21–24) (Fig. 1). All rearranged H chain sequences were obtained from shark-GR, for which IgH germline genes were previously cloned for genomic single-cell studies (24). The shark-GR G2A and G2B IgH genes (accession numbers EU719631, EU719632) were identical to G2L germline sequences from other individuals (21, 23). Rearranged NS5-2 L chain sequences were obtained from three animals: shark-Y, -J, and -33. The germline NS5-2 was first isolated from a shark-Y cosmid library [AY720853 (22)], and it was the same as germline NS5-2 clones amplified from shark-J (22). Both were identical to germline NS5-2 clones isolated from shark-33 erythrocyte DNA (not shown).

Cloned rearrangements either were unmottened or contained substitutions. In every batch of cloned sequences, there were hardly any shared...
substitutions among the variants unless they also shared a common CDR3 sequence. The absence of consistently shared changes demonstrated that the substitutions were independently acquired and somatic in origin (Supplemental Fig. 1).

Results
Experimental design
In this report, rearrangements of one L chain gene (NS5-2) and two H chain genes (G2A and G2B) were examined for mutational changes such as substitutions, duplications, insertions, and deletions. A strategy was devised to amplify nonproductive VJ preferentially from cDNA to obtain sequences that had mutated without selection pressure. The nature of substitutions from in- and out-of-frame VJ was analyzed, and the efficacy of the PCR approach evaluated. A second source of nonselected sequence was the intronic region flanking mutated H chain rearrangements, cloned from genomic DNA of slg+ PBLs. The nature and extent of mutation in the 6.8-kb J-C intron was investigated and the 3′ boundary of SHM at shark IgH determined.

Isolating nonproductive Ig rearrangements
The first murine L chain sequence containing a nonproductive rearrangement also carried an unspliced leader intron (25). This observation suggested out-of-frame V regions could be preferentially amplified by using a forward primer targeting the leader intron, and the single-copy NS5-2 gene (22) was selected to test this idea. Table I shows L chain sequences generated using a forward primer in the leader intron compared with those amplified with a primer in FR1 (Fig. 1, NSSLI and V3, respectively). With the latter, V3, the VJ are mostly in-frame (25 out of 31, 81%; Table I). Of 52 sequences raised with the leader intron primer, 39 (75%) contained nonproductive rearrangements; 24 out of 39 clones contained somatic mutations, and six mutants carried insertions or deletions (indels). The association of indels only with nonproductive mutants is statistically significant among the 126 clones shown in Table I. Indels occurred in mutant sequences JNS5/83, -38, -39, -46, -48, -61, -80, and SPH2, -4, and -156 (nonproductive). Tally: 0 indels in 54 nonmutant sequences and 10 indels in 72 mutant sequences (0.005, Fisher’s exact test). These changes are examined in a separate section below.

To get a better idea of the rearrangements present in the B cell population, VJs were amplified from genomic DNA. Of the 43 SPH sequences, only 35% were in frame, and overall 37% were mutated. Thus, nonproductive L chain rearrangements are targeted bystanders in lymphocytes undergoing SHM at transcribing Ig genes.

Selection in shark Ig
In a previous study (14), we could not show that shark L chain mutants undergo selection according to the criteria traditionally applied in mammalian systems: replacement (R) to synonymous (S) values. Nucleotide changes in mouse Ig mutants are scored according to a ratio of R to S changes calculated for the FR and CDRs. Having in hand two distinct Ig sequence groups, one clearly not functional, the shark Ig mutation patterns were reassessed.

In Table II, 24 in-frame V3 sequences with 183 mutations are compared with 18 nonproductive JNS5 sequences with 169 mutations. The selected sequences do not contain indels. Because the PCR primer targets the S′ area of FR1 in the V3 series, this is also left out of the JNS5 group. To avoid the frameshift complication, only the first six codons of CDR3 are used. The RS values for nonproductive cDNA are 4.8 for FR and 5.3 for CDR and for in-frame cDNA are 1.9 for FR and 4.2 for the CDR. These values are consistent with the nature of the rearrangements: nonfunctional mutants, FR and CDR values are similar, whereas in functional mutants, there is greater disparity. Region by region, the extent of change is at the same frequency in both groups except in FR2, where the clear difference in the productive (0.8%) versus the nonproductive group (5.2%) demonstrates selection against changes in FR2 of functional Ig sequences.

Nature of L chain substitutions
In Table III, substitutions from four groups of L chain mutants are characterized. The first set consists of 29 sequences from shark-Y PBLs obtained by RT-PCR and from shark-33 epigonal cDNA library (23). The second and third sets are, respectively, from in-frame V3 mutants and nonproductive JNS5/SPH mutants from shark-J. Sequences with indels are not included. As noted previously (14), there is a bias for mutated positions associated with the RGYWRNYC motif (26). The motifs make up 81 bp (19.5%) out of the 415 bp of the leader, VL, and JL. In 15 cDNA sequences isolated from the shark-33 library, there were 141 substitutions, of which 65 (46%) were in or partly in the motif. For point mutations, all three sets show a transition frequency of 50%, both for G/C and A/T base pairs. The transition bias in A/T could reflect activity of Pol η, which tends to incorporate Gs opposite Ts, leading to T to C or A to G transition changes (27). Tandem mutations usually occurred as di- or trinucleotide substitutions. In the nonproductive sequences, the transition frequency in tandem changes is even lower, 28%, and is closer to the expected randomly acquired changes, 30%. At this point, we asked whether the frequencies of transitions among tandem mutations in the first two sets, 37% and 40%, might have been skewed upwards by selection. We extended this study to H chain flanking intronic sequence because the J-C intron of NS5-2 L chain gene contains sequencing blockages.

Hypermutation at IgH
A stock of genomic DNA extracted from shark-GR slg+ PBL was characterized previously (24). In preliminary studies using this DNA, 21 G2 VDJ sequences of 500 bp were cloned (not shown).

Table I. L chain rearrangements amplified from shark-J PBL DNA and RNA

<table>
<thead>
<tr>
<th>Source</th>
<th>Primer Pair</th>
<th>Total</th>
<th>Productive</th>
<th>Nonproductive</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Total Clones [No. (%)] Mutants [No. (%)] Indel</td>
<td>Total Clones [No. (%)] Mutants [No. (%)] Indel</td>
<td></td>
</tr>
<tr>
<td>cDNA</td>
<td>Intron-C_L</td>
<td>52</td>
<td>13 (25)</td>
<td>5 (38)</td>
</tr>
<tr>
<td>cDNA</td>
<td>FR1-C_L</td>
<td>31</td>
<td>25 (81)</td>
<td>24 (96)</td>
</tr>
<tr>
<td>Genomic</td>
<td>Intron-J_L</td>
<td>43</td>
<td>14 (33)</td>
<td>5 (36)</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>34</td>
<td>1</td>
<td>38</td>
</tr>
</tbody>
</table>

*Indels occurred in mutant sequences JNS5/83 (in-frame); JNS5/38, -39, -46, -48, -61, -80, and SPH2, -4, and -156 (nonproductive). Tally: 0 indels in 54 nonmutant sequences and 10 indels in 72 mutant sequences (p = 0.005, Fisher’s exact test); 0 out of 36 indels in nonproductive nonmutants compared with 8 out of 37 nonproductive mutants; 0 out of 18 indels in productive nonmutants compared with 1 out of 35 productive mutants.
Ten sequences were nonproductive, and five carried 3–52 substitutions; 11 sequences were in-frame and one mutant carried two changes. Most of the clones (18 out of 21) were G2A, although the primers also detect the G2B isotype. Similar results were obtained in the three experiments described below, so that the mutants tallied in this study are almost all nonproductively rearranged G2A sequences.

In the first experiment, the VDJ and 3' flanking regions of 1.1 kb were amplified and cloned. G2A/B genes were amplified equally, although G2B tended to be in germline or partially rearranged configuration (screening results not shown). The 5% nucleotide and indel differences in the two isotypes allowed us to evaluate the extent of in vitro crossing over, if any. Fig. 2 shows the frequency of mutation over the VDJ and 1 kb of flanking sequence and Table II.

### Table II. Comparing substitution patterns between in-frame and nonproductive V regions

<table>
<thead>
<tr>
<th>Scored Codons</th>
<th>FR1 (21 Codons)</th>
<th>CDRL (13 Codons)</th>
<th>FR2 (15 Codons)</th>
<th>CDRL (11 Codons)</th>
<th>FR3 (34 Codons)</th>
<th>CDRL (6 Codons)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>In (24 sequences)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tallied</td>
<td>504</td>
<td>312</td>
<td>360</td>
<td>264</td>
<td>816</td>
<td>144</td>
<td>2400</td>
</tr>
<tr>
<td>Affected [no. (%)]</td>
<td>10 (2.0)</td>
<td>32 (10.3)</td>
<td>3 (0.8)</td>
<td>17 (6.4)</td>
<td>34 (4.2)</td>
<td>17 (11.8)</td>
<td>113</td>
</tr>
<tr>
<td>Mean ± SD</td>
<td>0.42 ± 0.93</td>
<td>1.33 ± 1.61</td>
<td>0.13 ± 0.34</td>
<td>0.71 ± 1.12</td>
<td>1.42 ± 1.74</td>
<td>0.71 ± 0.81</td>
<td></td>
</tr>
<tr>
<td>Replacements</td>
<td>7</td>
<td>23</td>
<td>2</td>
<td>15</td>
<td>22</td>
<td>15</td>
<td>84</td>
</tr>
<tr>
<td>Replacement: synonymous</td>
<td>2.3</td>
<td>2.5</td>
<td>2</td>
<td>7.5</td>
<td>1.8</td>
<td>7.5</td>
<td></td>
</tr>
<tr>
<td>Out (18 sequences)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tallied</td>
<td>378</td>
<td>234</td>
<td>270</td>
<td>198</td>
<td>612</td>
<td>108</td>
<td>1800</td>
</tr>
<tr>
<td>Affected [no. (%)]</td>
<td>6 (1.6)</td>
<td>23 (9.8)</td>
<td>14 (5.2)</td>
<td>11 (5.5)</td>
<td>32 (5.2)</td>
<td>16 (14.8)</td>
<td>102</td>
</tr>
<tr>
<td>Mean ± SD</td>
<td>0.33 ± 0.77</td>
<td>1.28 ± 1.71</td>
<td>0.78 ± 1.22</td>
<td>0.61 ± 1.24</td>
<td>1.78 ± 1.80</td>
<td>0.89 ± 1.18</td>
<td></td>
</tr>
<tr>
<td>Replacements</td>
<td>5</td>
<td>18</td>
<td>10</td>
<td>9</td>
<td>28</td>
<td>15</td>
<td>85</td>
</tr>
<tr>
<td>Replacement: synonymous</td>
<td>5</td>
<td>3.6</td>
<td>2.5</td>
<td>4.5</td>
<td>7</td>
<td>15</td>
<td></td>
</tr>
</tbody>
</table>

*a* In, in-frame rearrangements: 24 mutants from V3 series (Table I); 183 substitutions.

*b* Out, nonproductive rearrangements: 18 mutants from JNS5 series that do not contain in/dels (Table I); 169 substitutions.

### Table III. Nature of substitutions in shark L chain

#### In-Frame Shark-Y/Shark-33 Mutants

<table>
<thead>
<tr>
<th>All Substitutions</th>
<th>In Frame Shark-Y/Shark-33 Mutants</th>
<th>Tandem Substitutions</th>
<th>Single Base Substitutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>From G A T C</td>
<td>From G A T C Total</td>
<td>From G A T C Total</td>
<td>From G A T C Total</td>
</tr>
<tr>
<td>Total 249 changes</td>
<td>Transitions: 46%</td>
<td>Total 126 changes (55 groups)</td>
<td>Transitions: 37% (21/26)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total 123 changes</td>
<td>Transitions: 55% (34/34)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>25</td>
<td>15</td>
<td>10</td>
</tr>
<tr>
<td>T</td>
<td>32</td>
<td>33</td>
<td>24</td>
</tr>
<tr>
<td>C</td>
<td>28</td>
<td>30</td>
<td>25</td>
</tr>
<tr>
<td>Total</td>
<td>85</td>
<td>88</td>
<td>85</td>
</tr>
</tbody>
</table>

#### In-Frame Shark-J Mutants

<table>
<thead>
<tr>
<th>All Substitutions</th>
<th>In Frame Shark-J Mutants</th>
<th>Tandem Substitutions</th>
<th>Single Base Substitutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>From G A T C</td>
<td>From G A T C Total</td>
<td>From G A T C Total</td>
<td>From G A T C Total</td>
</tr>
<tr>
<td>Total 183 changes</td>
<td>Transitions: 46%</td>
<td>Total 87 changes (36 groups)</td>
<td>Transitions: 40% (22/13)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total 96 changes</td>
<td>Transitions: 51% (25/4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>22</td>
<td>15</td>
<td>10</td>
</tr>
<tr>
<td>T</td>
<td>10</td>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td>C</td>
<td>23</td>
<td>17</td>
<td>11</td>
</tr>
<tr>
<td>Total</td>
<td>45</td>
<td>38</td>
<td>22</td>
</tr>
</tbody>
</table>

#### Nonproductive Shark-J Mutants

<table>
<thead>
<tr>
<th>All Substitutions</th>
<th>Nonproductive Shark-J Mutants</th>
<th>Tandem Substitutions</th>
<th>Single Base Substitutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>From G A T C</td>
<td>From G A T C Total</td>
<td>From G A T C Total</td>
<td>From G A T C Total</td>
</tr>
<tr>
<td>Total 282 changes</td>
<td>Transitions: 39%</td>
<td>Total 168 changes (71 groups)</td>
<td>Transitions: 28% (20/72)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total 114 changes</td>
<td>Transitions: 54% (22/41)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>20</td>
<td>20</td>
<td>18</td>
</tr>
<tr>
<td>T</td>
<td>15</td>
<td>15</td>
<td>13</td>
</tr>
<tr>
<td>C</td>
<td>43</td>
<td>43</td>
<td>37</td>
</tr>
<tr>
<td>Total</td>
<td>80</td>
<td>80</td>
<td>70</td>
</tr>
</tbody>
</table>

*a* V and J gene segments nucleotide content: 101 G, 78 A, 94 T, 96 C; 369 total. To normalize GC and AT base pair ratios, the derived ratio is multiplied by 0.87. For example, in the top set of point mutations, the GC/AT is 2.08 (83 changes from GC/40 changes from AT in shark-Y/33). This is multiplied by 172 (A+T)/197 (G+C), or 0.87. The result, 1.81, is the normalized GC and AT ratio.

*b* Total 249 changes: 159 from RT-PCR of shark-Y PBL and 90 from shark-33 cDNA library.
generally resembles that published by Lebecque and Gearhart (28) for mouse mutants. Multiple indels were present, as depicted below the drawing of the VDJ gene, but these were tallied separately from substitutions. The indel pattern parallels that of the substitutions and so is part of the same process.

Two sets of clones with longer 3' flank sequence were obtained and the extent of their tandem mutations shown in Fig. 3. Occasional point mutations were found scattered throughout the intronic region, and some are probably in vitro changes (Fig. 3 legend). We thus prefer to distinguish the area where tandem substitutions occurred because they are the distinctive feature of shark hypermutation. The brackets encompass the 5'-most and 3'-most tandem groups (doublets or triplets) of the clone; all substitutions are tallied at the left-hand side. The mutations in clone A3 at the 3.5-kb region of the reference sequence are actually 1995 bp from its own JH, the consequence of the large 1.4-kb deletion decreasing distance from the VH promoter. In clone C9, 255 changes occurred in the first part of the gene (Fig. 3, heavy brackets), but an isolated group of seven substitutions appeared downstream over an area of 400 bp (Fig. 3, light brackets). The "s" series in Fig. 3 contains the ~7.4 kb of sequence from the leader intron to the first C exon of G2A. A total of 103 VDJs were sequenced, of which 34 were mutants and represented 15 unique rearrangements. Nine clones in Fig. 3 confirm that hypermutation mostly takes place within a 2.5-kb region encompassing the VDI. In clone s1, the 27 mutations occurred over the same area as the 58 mutations in s57 or the 178 in s122.

Comparison of mutants from all PCR sets showed that A2 and s122 shared ancestry. Although the CDR3 were mutated, there are common N region additions and multiple shared mutations. In Fig. 4, this pair of nonfunctional VDJ from long-diverged daughter clones demonstrates one point clearly: that the early (shared) mutations occurred in the same region as the later ones. There are 184 mutations in A2 and 178 in s122, not counting nucleotides in insertions/duplications. Of these, 37 shared mutations (Fig. 4, highlighted in purple) are distributed over an area of 1.7 kb that overlaps with the subsequent 147 and 141 changes occurring in A2 and s122, respectively. This is diagrammed in Fig. 3 (asterisked s122, A2, and the ancestral A2/122).

The distribution of changes in the ancestral clone as well as in A2 and s122 suggests that there are roughly two regions of ~1–1.1 kb, with the one proximal to the promoter experiencing more changes. The decrease in mutation frequency is apparent in the distal 1 kb; the difference is not due to any disproportionate distribution of

FIGURE 3. The boundaries of mutation in the shark H chain gene. The fully rearranged VDDJ is shown, together with a scale marking distance from the JH gene segment to the first C exon (6.8 kb). Genomic G2A rearrangements cloned with 5357 bp (C9, A2, A3) or 6848 bp (s series) of 3' flanking sequence were scored for total substitutions (left column). The extent of the 5'-most and 3'-most tandem changes is depicted as a bracket for each clone. The lighter bracket in C9 includes an isolated group of seven mutations distributed in the 3'-most 400 bp. Some of the mutants contained extensive deletion, and the true distance from the JH border is indicated in parentheses if there is a discrepancy of >100 bp. For example, the 3'-most doublet in A3 is at position 3496 bp according to the J-C intron reference sequence, but it is located 1995 bp from its JH border. Asterisks mark the related clones A2 and s122; the hypothetical ancestral mutant A2/122 is discussed in the text. The error for the Expand Long Template enzyme mix (Roche) is 0.11% (4 substitutions/3696 bp) as scored in nonrearranged clones.
hotspots (not shown). Although there is considerable difference in AT to GC content in the VH gene segment (1.04) compared with 2 kb of the J-C intron (1.967), this does not explain the mutation pattern. The first 700 bp of the J-C intron, which is included in the proximal 1.1 kb, is heavily mutated, and this has the same nucleotide representation as the downstream portion.

**Nature of changes in the J-C intron**

Nucleotide substitutions in the J-C intron were scored in s1, s4, s22, s23, s37, s25, and s57 because the substitutions are sufficiently sparse for most to be deemed independent events. The overall pattern of nucleotide substitution in the intron (Table IV) is similar to that in L chain mutants; about half the changes were in tandem groups (77 out of 150), and the frequency of transition changes is higher for point mutations (56%) compared with tandem mutations (36%).

Short (1–3 bp) insertions or 1-bp deletions were frequent in mutated J-C intron sequence but were not scored in Table IV. Some examples are marked with asterisks and number signs in Fig. 4. In Fig. 5, three kinds of insertions are shown in the top sequence (Fig. 5A, clone G20-88). One is a 1-bp deletion that is replaced by 2-bp mutation, the second is a 7-bp insertion that appears to be a (mutated) inverted repeat of the adjacent 5′ flank, and the third is a 2-bp AG insertion. The 7-bp insertion could be T nucleotides, which are inserts that are direct or inverted repeats of flank sequence (29) and coupled to repair after a dsDNA break.

**Indels and T/N nucleotides**

The duplications and deletions in shark Ig are often accompanied by nucleotides of unknown origin. The single instance involving an in-frame VJ (Fig. 5H, JNS5-83) was a nondisruptive 3-bp insertion in a loop region, CDR3. Many events shown in Fig. 5 could have resulted from a staggered double-stranded break (DSB) from nearby lesions on either strand. Depending on the nature of the breaks, deletion from DNA end resection or duplication from filling in may result. There are simple deletion events, as in JNS5-46 or JNS5-61(1), but in others, sequence was added. In JNS5-80 (Fig. 5G), the CGCC appears to be nontemplated. Similarly, some duplications are simple (Fig. 5J, SPH2-2), but in others, there are accompanying additions [JNS5-61, GGGG; SPH-156, AG; SPH4 (1), G]. The insertions in duplications and deletions tend to be GC-rich, although not in all cases.

The events designated with arrows in Fig. 5 are interpreted as T nucleotides because the presumed template is in complementary orientation to the insert or removed from the mutated site by 2–49 bp [Fig. 5H, JNS5-83, I(2) SPH-156, K(2) SPH4]. The T nucleotide copies tend to contain mismatches, including indels, like in Fig. 5C (JNS5-39, marked by “i”) (30). However, some complex events like the incomplete duplication (missing final A) in JNS5-61 (Fig. 5F) and the extra nucleotide at one end of the duplication in SPH4 (Fig. 5K, 1) cannot easily be explained by a staggered DSB and fill-in.

**FIGURE 4.** Comparison of related IgH mutants. The reference germline (GL) gene G2A is shown without intergenic sequence in VH, D1, D2, and JH gene segments (labeled in boldface; gaps indicated by slashes). The start of the VH gene segment is indicated over the splice acceptor site and is divided into FRs and CDRs. The beginning of the J-C intron is indicated over the splice donor site. Clones A2 and s122 are aligned with respect to the GL, and the PCR primers are italicized. Differences from the GL are highlighted in blue and those shared between the mutants are highlighted in purple. Changes in CDR3 are not marked. Although the rearrangement is in-frame, the less mutated A2 CDR3 contains two stops. The VDJ was probably never functional, and its first mutations (in CDR1, AT to GA generated TGA) were deleterious. Dashes indicate gaps. Duplicated sequences are highlighted and their template underlined. Lowercase “n” is placed where point deletion was introduced during PCR slippage. Asterisks shows sites of inserted nt, crosshatches deletions. The numbering of the reference sequence includes leader intron and coding VH gene segments (1–474) and 3′ flank (475–2289).
from the VH gene segment, which is 27% G, 27% A, 24% T, and 22% C out of 308 bp.

Homologous position in another IgH or is a longer tract derived from a non-GCC mutation of clone 20-6 would have to be a 3 nt gene conversion event, none for the homologous positions (not shown). For instance, the transition frequency closer to the anticipated random one. We inspected the J-C intron, expecting that slippage could occur more frequently in a region of relatively repetitive DNA (35). In Fig. 6, the TC tandem mutation at positions 887 to 888 in clone G20-74 could have resulted from a misalignment occurring when an upstream T on the primer strand assumed an extrahelical position (Fig. 6, bottom panel, step 2). After synthesis of TCT, the error-prone polymerase dissociates, and the primer strand realigns with the template, shifting the 3’-most T to position 889 from position 888 (Fig. 6, bottom panel, step 4). There are other substitutions possibly generated this way via transient misalignment of the primer or the template strand with one to three unpaired bases, but it does not explain many other changes. If tandem substitutions arose from transient misalignment events, we would expect to find mostly A to T and T to A changes in the A/T-rich regions, but this is generally not the case.

In Fig. 6, top panel, the germline G/C are highlighted, and these are potential sites for AID action on one of the strands. Substitutions can take place at or at least 1–6 nt away from these sites. The maximal distance is derived from the longest stretch of A/T at positions 604–617, where point and tandem changes occur 6 nt from the nearest cytosine in clones G20-2 and G20-13. Taking clone G20-6 as an example, the GCC substitution at positions 875, 887, or 890 is isolated and unlikely to be the result of more than one AID-initiated event, so that the lesion could have been generated at positions 875, 887, or 890. A single-stranded gap is created, extending at least 7 nt downstream of position 875 or 7 nt upstream of position 887 or on the complementary strand 10 nt downstream of position 890. Subsequent gap repair with error-prone polymerases generate 3 mismatching nt GCC.

The region depicted in Fig. 6 is within 500 bp of the 3’ flank of JH of G2A. We inspected the same region of five other IgH genes—G2B, G1, G3, G4A, and G5—looking for donor templates in a gene conversion mechanism for tandem substitutions. There were none for the homologous positions (not shown). For instance, the GCC mutation of clone 20-6 would have to be a 3 nt gene conversion tract copied from upstream or downstream sequence or from a non-homologous position in another IgH or is a longer tract derived from an unknown source.

**Discussion**

The Ig genes in shark and other chondrichthyan fishes have an early alternative organization of multiple mini-loci (36) where each one consists of a few gene segments individually undergoing intralocus V(D)J recombination (Fig. 1) and SHM in B cells (21, 24). AID has been isolated from nurse shark and found in adult lymphoid tissue by RT-PCR (E. Hsu, unpublished observations), and the biased occurrence of all mutations at RYGW hotspots suggests their common initiation by AID (14). The unique pattern of adjacent mutations in shark Ig and Ig-like genes was first observed in the Ig New Antigen Receptor (37) and fully characterized in traditional Ig L chains (14). In this study, we isolated nonproductive Ig rearrangements and 3’ flanking sequence to define the intrinsic properties of this unusual hypermutation process. We first compared in- and out-of-frame L chain VJ mutants to reassess the standard criteria, RS values, for selection. These ratios showed that the nonfunctional rearrangements were unselected, but, at the same time, the low differential in FR/CDR RS values for in-frame mutants suggests that antigenic selection in shark L chains is not as strong as in mammals. For example, data from an unselected human B cell population (18) show a greater differential between nonproductive (2.1 for FR versus 2.5 for CDR) and productive (1.4 FR versus 7.8 CDR) rearrangements in the κ Vκ12/O2 genes (38). Only the relative absence of replacement mutations in FR2 indicated that expressed mutants were selected at least on the basis of structural integrity.

Germinal centers do not exist in cartilaginous fishes or in the amphibian Xenopus (reviewed in Ref. 39), but affinity maturation has been demonstrated in their humoral responses (40, reviewed in Ref. 41). Dooley and Flajnik (40) studied the nurse shark Ab response to hen egg lysozyme and found that, of the 7S and 19S IgM present in serum, it was the monomeric species that increased in affinity. They speculated that cells producing the monomer are from a T cell-dependent lineage that is responsible for the specificity of the response. SHM in sharks is an Ag-driven process (14, 42, 43); there are no H or L chain mutants in neonates. However, antigenic selection may not need to be comparably effective as in mammals (41), and variant sequences resulting from the remarkably efficient shark SHM could also serve to diversify the existing repertoire (44).

**Boundaries of hypermutation at shark Ig genes**

The boundaries of hypermutation have been studied extensively at murine IgH and IgL loci and in a variety of transgene constructs in which the relative positions of the cis-acting elements were shifted. The upstream boundary is characterized as an abrupt increase in mutation frequency 180 bp after the transcription start site (45, 46), as if determined by a fixed point from the promoter. That correlation between hypermutation and transcription was further demonstrated by Peters and Storb (47), who inserted the Vκ promoter 5’ of the Cε exon and elicited hypermutation over a similar distance downstream as the endogenous promoter. In contrast, the downstream boundary is not a clear-cut one, and the

---

**Table IV. Shark-GR J-C intron mutations**

<table>
<thead>
<tr>
<th>From</th>
<th>G</th>
<th>A</th>
<th>T</th>
<th>C</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>21</td>
<td>25</td>
<td>16</td>
<td>62</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>9</td>
<td>10</td>
<td>2</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>3</td>
<td>10</td>
<td>24</td>
<td>36</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>9</td>
<td>6</td>
<td>26</td>
<td>41</td>
<td></td>
</tr>
<tr>
<td>Total: 150 changes</td>
<td>9</td>
<td>21</td>
<td>37</td>
<td>10</td>
<td></td>
</tr>
</tbody>
</table>

**All Tandem Substitutions**

<table>
<thead>
<tr>
<th>From</th>
<th>G</th>
<th>A</th>
<th>T</th>
<th>C</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>10</td>
<td>17</td>
<td>4</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>6</td>
<td>2</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>2</td>
<td>8</td>
<td>4</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>7</td>
<td>3</td>
<td>14</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>Total: 77 changes</td>
<td>9</td>
<td>21</td>
<td>37</td>
<td>10</td>
<td></td>
</tr>
</tbody>
</table>

**All Single-Base Substitutions**

<table>
<thead>
<tr>
<th>From</th>
<th>G</th>
<th>A</th>
<th>T</th>
<th>C</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>11</td>
<td>8</td>
<td>12</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>9</td>
<td>4</td>
<td>0</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>1</td>
<td>2</td>
<td>9</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>2</td>
<td>3</td>
<td>12</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>Total: 73 changes</td>
<td>12</td>
<td>16</td>
<td>24</td>
<td>21</td>
<td></td>
</tr>
</tbody>
</table>

Nucleotide composition of first 2 kb J-C intron: 19.1% G, 28.3% A, 38% T, and 14.6% C. To normalize GC/AT, multiply ratio by 1.97. The composition differs considerably from the VH gene segment, which is 27% G, 27% A, 24% T, and 22% C out of 308 bp.

Data from clones s1, s4, s22, s23, s25, s37, s57.
area targeted by SHM is characterized as extending 1 to 2 kb (45) or rarely beyond 1.5–2 kb (46) from the promoter or occurring “with a high frequency for almost 1 kb before tapering off” (48). These descriptions also fit observations in shark IgH.

We interpret the shark IgH mutation distribution curve as consisting of two distinct regions of high and low substitution frequency. Any place within a ∼2-kb stretch can be mutated, at any time, although the frequency is higher for the upstream 1.1 kb. The related clones A2 and s122 give us three snapshots: the deduced ancestor with 37 mutations and its two descendants with 150 subsequent changes. The mutation spread is the same in all three and suggests, as in the murine system (46), there exists a limiting element no matter what the mutation load. Models for AID-targeting couple the mutator with transcription; for example, one idea is that the mutator is associated with transcription initiation/elongation factors (mutasome), slides along the DNA, and initiates the mutation event with greater probability at the beginning of the V segment (49). What restricts the area affected could be the 1/2 of the mutator, which becomes exhausted or disengages from the transcription complex (49, 47).

**FIGURE 5.** Insertions and deletions in Ig sequence. **A**, Portion of the J-C intron from clone G20-88 containing three different insertion events. **B–K**, Ten L chain sequences with indels from experiments shown in Table I. B, JNS-38, nonproductive, insertion in CDR1. C, JNS-39, nonproductive, insertion in JL. The letter “T” over I indicates insertion in inversely duplicated sequence. D, JNS-46, nonproductive, deletion in FR3. E, JNS-48, nonproductive, deletion in CDR1. F, JNS-51, nonproductive, deletion in FR3. G, JNS-50, nonproductive, insertion in FR3. The sequences are compared with the J-C intron or NS5 reference sequence and identities indicated by vertical bars. Mutations are shown in lowercase. Duplicated nucleotides are shaded, and the template sequence is underlined. In those instances that could involve T nucleotides, both insert and flank templates are marked with arrows indicating their relative orientation.
Figure 6. Changes in A/T-rich areas in the J-C intron. Top panel, Clones with substitutions in two regions of the J-C intron of the G2A gene are shown, with numbering according to the reference sequence in Fig. 4. G/C nucleotides are highlighted in yellow and mutations in blue. An inserted duplication (AAACA) is shown for clone A2 at position 640. Dashes indicate deleted sequence. Lowercase “n” is placed where point deletion was introduced during PCR slippage.

Bottom panel, Hypothetical pathway by which CTs at position 887–888 in clone G20-88 are mutated to TC through transient misalignment during gap repair. Step 1: dissociation of DNA polymerase causes fraying, followed by step 2: strand misalignment. Step 3: distributive DNA synthesis. Polymerase dissociates, with attendant fraying. Step 4: primer and template realign. Step 5: extension of primer stabilizes mismatches. Tandem mutations CT to TC established.

1. 5'-ATTT 3' - TAAAGAACTT
   DNA polymerase performing gap repair dissociates

2. 5'-AT 3' - TAAAGAACTT
   Unpaired base in primer strand, misalignment

3. 5'-ATTTCT 3' - TAAAGAACTT
   DNA synthesis, polymerase dissociates

4. 5'-ATTTTCT 3' - TAAAGAACTT
   Realignment

5. 5'-ATTTTCTGAA 3' - TAAAGAACTT
   Further DNA synthesis stabilizes mismatches

FIGURE 6. Changes in A/T-rich areas in the J-C intron. Top panel, Clones with substitutions in two regions of the J-C intron of the G2A gene are shown, with numbering according to the reference sequence in Fig. 4. G/C nucleotides are highlighted in yellow and mutations in blue. An inserted duplication (AAACA) is shown for clone A2 at position 640. Dashes indicate deleted sequence. Lowercase “n” is placed where point deletion was introduced during PCR slippage. Bottom panel, Hypothetical pathway by which CTs at position 887–888 in clone G20-88 are mutated to TC through transient misalignment during gap repair. Step 1: dissociation of DNA polymerase causes fraying, followed by step 2: strand misalignment. Step 3: distributive DNA synthesis. Polymerase dissociates, with attendant fraying. Step 4: primer and template realign. Step 5: extension of primer stabilizes mismatches. Tandem mutations CT to TC established.
In contrast, a larger area is affected by SHM in the switch regions [4 to 5 kb (50, 51)] and a smaller one in an in vitro transcription-based assay system of AID activity [500 bp before “declining precipitously” (52)]. The differences from in vivo SHM at the VDJ may be respectively attributed to a different content of the mutant at the I exon promoters (50) and suboptimal input in vitro components.

Consideration of TdT and/or Pol μ.

Duplication and deletion occur during SHM in 43% of out-of-frame and 4% of in-frame human VDJ (53), but there are no examples of nontemplated additions longer than 1 bp or any inserts that are inverted repeats like those shown in Fig. 5. The only instance in which DSB repair in Ig was accompanied by >1 bp insertions was from experiments in Ramos Burkitt, which constitutively mutates its Ig genes, after transfection with a construct with TdT under its Ig genes, after transfection with a construct with TdT under a β-globin promoter and IgH enhancer (54). In that case, duplications and deletions, usually simple ones when the Ig sequences are isolated from organ tissue (53, 55), acquired G/C-rich nontemplated additions.

We investigated whether the nurse shark Ig insertions could have been introduced by TdT, looking for evidence of TdT expression outside of the thymus and bone marrow equivalent (epigonal organ). From a battery of organ samples, TdT RT-PCR signal consistently coincided with RAG2 and only sometimes with AID or RAG1 (C. Zhu and E. Hsu, unpublished observations). At this point, the probable explanation is that TdT and RAG2 signals in nonprimary lymphoid tissue originated from newly mature B cells in the process of downregulating both. It would have been surprising to find TdT at a stage well beyond the completion of V(D)J rearrangement, although TdT expression in activated B cells does not appear to affect lymphocyte viability in TdT-transgenic mice (56, 57). Pol μ, in contrast, is more likely to be present at all times; the zebrafish homolog is expressed in all tissues examined (58) and increased in lymphoid tissues, like in mouse and human (59). It is difficult to distinguish a duplication with N addition from a T insertion, but some of the mutations in shark indels can be explained as inverted repeats or repeats of nearby but nonabutting flank sequence, which could have involved a polymerase tending to slippage, like Pol μ (30). What is clear is that some of the duplications in shark Ig copied mutated nucleotides (Fig. 5F, 5K), showing that DNA breakage and repair took place during or after SHM events; the strict correlation of indels with mutants in Table I (see footnote) suggests breakage was AID initiated. In conclusion, the polymerases/polymerase for DSB repair are different for shark and human B cells during SHM.

TdT and Pol μ are members of family X polymerases that contribute to BER and nonhomologous end joining (60). Although overexpressed Pol μ in the Ramos cell line introduced mutations that can be explained by a transient misalignment process (61), Pol μ and Pol λ-deficient mice display no effects on SHM (62), and a role of family X polymerases in this process has yet to be established (9). This, however, does not rule out such a function existing in B lymphocytes of species other than mouse or human beings.

Candidate polymerases and pathways.

Following deamination of cytidine by AID, the resultant uracil is removed and a single-stranded gap may be generated, perhaps extending ≤29 nt upstream and/or ≤18 nt downstream of the lesion in mouse systems (63). In shark, the gap is at least 6 nt, and perhaps the process supporting nuclease accessibility leads to some DNA fraying. We hypothesize that several error-prone polymerases are available in shark B cells, some of which can add ≥2 mismatching nt before dissociating. The extension efficiency depends on the nature of the terminus mispairing, and repair may switch to another polymerase (64) or one specifically targeting and extending aberrant primer-terminal base pair (65, 66). Once the gap is filled, strand ligation completes the repair.

The detection of an unexpected DNA polymerase activity during DSB repair raises some speculation as to its involvement in generating tandem mutations, which are unique to sharks or elasmobranchs. If we hypothesize that a template-independent polymerase adds nucleotides, both this model and one for transient misalignment (34, 67) require some unwinding of the primer terminus. Although there is no direct evidence for this occurring during processing of the uracil, models for AID-initiated gene conversion postulate single-strand invasion into the homologous duplex DNA template (68, 69), and a single-stranded gap or nick has been demonstrated as the intermediate (70, 71).

An alternative idea is that the low-fidelity polymerases consecutively misincorporate ≥2 nt. Gene conversion may appear energetically more palatable, but as observed in the Results section, the conversion tract would have to be very short and initiated with minimum homology. Moreover, the donor sequences would be randomly acquired because identical and identical positioned tandem mutations have seldom been observed outside of clonally unrelated sequences. Although it seems unlikely one mismatch would be extended with another (72), in vitro experiments monitoring base-substitution fidelity of mammalian Pol η and yeast Pol ζ showed that these polymerases are capable of direct tandem misinsertions (27, 73).

Pol η can produce adjacent substitutions

Whereas a role of Pol ζ in SHM has not been fully elucidated, that of Pol η was demonstrated in humans or mice deficient for the enzyme (74, 75). In the classical studies that established the specificity preference of human and murine Pol η for placing G opposite T, a high frequency of 2- and 3-bp substitutions was also observed (27). These tandem mutations occurred at a rate of 10% of point mutations. However, the substitution spectrum is different from shark. The human Pol η in vitro mutations, both single and
tandem, are strongly A and T biased (normalized GC/AT ratio 0.2 for tandem substitutions; Table V) with an overwhelming preference for T to C changes; the single and tandem changes roughly parallel. In contrast, the shark tandem mutations display some A/T bias (GC/AT ratio 0.7; Table III) and no transition preference. Inasmuch as the shark point mutations from A/T have a preference for T to C changes; the single and tandem changes strongly A and T biased (normalized GC/AT ratio 0.5346).

References