A Role for Msh6 But Not Msh3 in Somatic Hypermutation and Class Switch Recombination

Stella A. Martomo, William W. Yang, and Patricia J. Gearhart

Abstract

Somatic hypermutation is initiated by activation-induced cytidine deaminase (AID), and occurs in several kilobases of DNA around rearranged immunoglobulin variable (V) genes and switch (S) sites before constant genes. AID deaminates cytosine to uracil, which can produce mutations of C:G nucleotide pairs, and the mismatch repair protein Msh2 participates in generating substitutions of downstream A:T pairs. Msh2 is always found as a heterodimer with either Msh3 or Msh6, so it is important to know which one is involved. Therefore, we sequenced V and S regions from Msh3- and Msh6-deficient mice and compared mutations to those from wild-type mice. Msh6-deficient mice had fewer substitutions of A and T bases in both regions and reduced heavy chain class switching, whereas Msh3-deficient mice had normal antibody responses. This establishes a role for the Msh2-Msh6 heterodimer in hypermutation and switch recombination. When the positions of mutation were mapped, several focused peaks were found in Msh6−/− clones, whereas mutations were dispersed in Msh3−/− and wild-type clones. The peaks occurred at either G or C in WGCW motifs (W = A or T), indicating that C was mutated on both DNA strands. This suggests that AID has limited entry points into V and S regions in vivo, and subsequent mutation requires Msh2-Msh6 and DNA polymerase.

Key words: DNA repair enzymes • cytosine deamination • class switching • mutation hotspots • B lymphocytes

Introduction

Activation-induced cytidine deaminase (AID) causes hypermutation in both immunoglobulin variable (V) and switch (S) regions and recombination in S regions (1, 2). AID deaminates cytosine to uracil on a variety of DNA substrates in vitro (3–7), and overexpression of AID causes mutations of C:G base pairs in vivo (8–13). However, in B cells, substitutions of all four bases occur at similar levels in V and S regions, indicating that other proteins are required to generate mutations at A:T base pairs. The low fidelity DNA polymerase η participates because humans deficient for the enzyme have low frequencies of A:T mutations and high levels of C:G mutations in V and S regions (14–16). The mismatch repair protein Msh2 is also engaged because Msh2-deficient mice have few A:T mutations in V and S regions (17–20). Msh2 is always associated as a heterodimer in vivo with either Msh3 or Msh6, and the two dimers have distinct and overlapping functions in mismatch repair (21). Therefore, it is important to know if Msh3 and/or Msh6 are involved in hypermutation. A paper by Wiesendanger et al. (22) implicates Msh6, but not Msh3, in V gene mutations; however, mutations in the S region and heavy chain class switching to IgG have not been examined in these mice.

How AID is directed to V and S regions is a question of intense interest. Two lines of evidence support targeting by cis DNA sequences in the locus. First, transcription complexes may bring AID to the V and S regions because mutations are found downstream of the promoters preceding V genes and S regions, and require components of the intron and 3’ enhancers (for review see reference 23). Recent studies show that histone acetylation of chromatin in the V (12) and S (24) regions precedes hypermutation and switching. Thus, DNA would be transcribed and become accessible to AID activity. Second, the DNA motifs RGYW/WRCY (R = purine, Y = pyrimidine, W = A or T) are hotspots because mutations are frequently found there in immunoglobulin genes (25). In support of hotspots, AID was shown recently to preferentially deaminate C in WRC motifs on

The online version of this article contains supplemental material.

Address correspondence to Patricia J. Gearhart, Laboratory of Molecular Gerontology, National Institute on Aging, National Institutes of Health, 5600 Nathan Shock Dr., Baltimore, MD 21224. Phone: (410) 558-8561; Fax: (410) 558-8157; email: gearhartp@grc.nia.nih.gov

Abbreviations used in this paper: AID, activation-induced cytidine deaminase; V, switch; S, variable.
Materials and Methods

Mice. Msh3−/− and Msh6−/− mice on a C57BL/6 background were provided by W. Edelmann (Albert Einstein College of Medicine, Bronx, NY), and C57BL/6 mice were purchased from Jackson ImmunoResearch Laboratories. Mice were used at 3–7 mo of age. For immunization, some mice were injected intraperitoneally with 100 μg KLH (Calbiochem) in adjuvant (RIBI Immunocen Research), boosted after 3 mo with 50 μg, and bled 4 d later.

Isotype Analysis. For in vitro switching, spleen cells were treated with ACK lysing buffer (Quality Biological) to lyse red blood cells. The cells were mixed with MicroBeads conjugated to rat monoclonal anti–mouse/human CD11b and anti–mouse CD43 (Ly-48, Leukosialin; Miltenyi Biotec) to remove macrophages, T cells, and activated B cells. The mixture was incubated at 4°C for 20 min and added to a magnetic cell separating column (MACS CS; Miltenyi Biotec). The unbound population, containing resting B cells, was cultured at 106 cells/ml in RPMI 1640 supplemented with 10% heat-inactivated FBS, antibiotics, 1-glutamine, and 5 × 10−5 M β-mercaptoethanol. To induce switching, cells were stimulated with 50 μg/ml Escherichia coli LPS (Sigma–Aldrich) or LPS and 50 ng/ml recombinant mouse IL-4 (R&D Systems). The cells were harvested after 3 d, washed twice with RPMI 1640, and stained with phycoerythrin-conjugated rat monoclonal anti–mouse IgG3 or anti–mouse IgG1 (BD Biosciences) in RPMI 1640 supplemented with 1% heat-inactivated serum. Flow cytometry analysis was gated on live cells as determined forward and side scatter analysis.

For serum isotypes, an ELISA was performed to measure anti-KLH antibodies. Wells were coated with 100 ng KLH in carbonate buffer containing 0.016 M Na2CO3, 0.034 M NaHCO3, pH 9.8. Wells were blocked with 1% ELISA grade BSA and 0.1% Tween 20 in 1X PBS (blocking buffer). A 1:100 dilution of serum from immunized mice was used to bind to the antigen, and after washing with 0.1% Tween 20 in PBS (washing buffer), a 1:10 dilution of rabbit anti–mouse IgM, 1:5 of IgG1, or 1:2 of IgG3 antibody (Bio-Rad Laboratories) in blocking buffer was added to the wells. After incubation, the wells were washed twice with washing buffer, and donkey anti–rabbit Ig horseradish peroxidase–linked whole antibody (Amersham Biosciences) diluted 1:5,000 in blocking buffer was added. The wells were washed, and bound antibody was detected with o-phenylenediamine dihydrochloride (Sigma–Aldrich) followed by termination with 3M H2SO4. All incubations were done at 37°C for 30 min. The absorbance was read at 490 nm. Values were normalized to total IgM, IgG1, and IgG3 in the serum as determined by ELISA using purified immunoglobulins (BD Biosciences).

Libraries of V and S Regions for Sequencing. Cells from the Peyer’s patches of two to three mice of each genotype were stained with phycoerythrin-labeled antibody to B220 and fluorescein-labeled peanut agglutinin (PNA; E-Y Laboratories). The cells were separated by flow cytometry, and DNA was prepared from B220+PNA+ cells. For V regions, the intron region downstream of rearranged V, diversity (D), and joining (J) gene segments on the heavy chain locus was sequenced. DNA was amplified using nested 5’ primers for the third framework region of V4S58 gene segments and 3’ primers for 344 nucleotides downstream of the J14 gene segment as described previously (28). For Sμ regions, a 561-base region located upstream of the core μ S region was sequenced (29). The following sets of nested switch

Figure 1. Diminished heavy chain class switching in Msh6−/− but not Msh3−/− B cells. (A) Flow cytometry analysis. Spleen cells were stimulated with LPS or LPS plus IL-4 to induce switching, and surface immunoglobulin was measured 3 d later. The percentage of B220+ cells that switched to IgG3 or IgG1 is shown in each box. (B) Serum antibody analysis. An ELISA was used to quantify anti-KLH antibodies from immunized mice. Black bars, C57BL/6; white bars, Msh6−/−. Units are defined as the antibody titer that was normalized to the total amount of each isotype.

Focused Mutation in Variable and Switch Regions
primers were used: first set, forward (nucleotides 4,560-4,579 of Genbank/EMBL/DDBJ under accession no. J00440), 5'-AGA-TAAAATTGGACCTCAG-3'; reverse (nucleotides 5,183-5,202), 5'-TATGTTCAGCTAGCGGAGG-3'; and second set, forward (nucleotides 4,580-4,599) with XbaI addition in italics, 5'-ACTCTAGATGTTTTTTAATGTGTTTT-3'; reverse (nucleotides 5,152-5,181) with EcoRI addition in italics, 5'-ACGAAATTCATCTCGATGGGACAG-3'. 20 ng of genomic DNA was amplified with Platinum Pfx polymerase and PCR enhancer (Invitrogen) in a 50-μl volume using the first set of primers for 30 cycles of 95°C for 30 s, 55°C for 30 s, and 68°C for 1 min, followed by a final incubation at 68°C for 10 min. Nested PCR was performed with 5 μl of the first reaction and the second set of primers with an annealing temperature of 50°C for 30 s for another 30 cycles. Products were digested, cloned into pBluescript (Stratagene), and sequenced.

Online Supplemental Material. Figs. S1–S3 present the sequences of JH4 introns from C57BL/6, Msh3−/−, and Msh6−/− clones, and Figs. S4–S6 show the sequences of Sp4 regions from the three sets of clones. Online supplemental material is available at http://www.jem.org/cgi/content/full/jem.20040691/DC1.

Results

Msh6−/− B Cells Had Diminished IgG Switching. To see if the Msh3 or Msh6 protein affected heavy chain class switching in vitro, the ability of B cells to switch isotypes in cell culture was measured by flow cytometry. Spleen cells from C57BL/6, Msh3−/−, and Msh6−/− mice were stimulated with LPS to induce IgG3 switching and cell culture was measured by flow cytometry. Spleen cells switching in vitro, the ability of B cells to switch isotypes in the absence of T cells and environmental antigens in the gut. However, among the mutated sequences, the Msh6−/− clones had a significantly lower frequency of mutation (5 × 10−3 mutations/bp, P < 0.001) compared with C57BL/6 and Msh3−/− clones (1 × 10−2 mutations/bp for both). In Msh6−/− clones, there were strikingly fewer mutations of A and T nucleotides (P < 10−6), and correspondingly more mutations of G and C bases than in C57BL/6 and Msh3−/− clones (Fig. 2, B and C). Most of the G:C mutations were transitions. These data support a role for the Msh2-Msh6 heterodimer in causing A:T substitutions in the V region (22).

For S regions, we focused on mutations in Sp4 because Msh6-deficient B cells had impaired switching to IgG.

Figure 2. Fewer A:T substitutions in JH4 introns from Msh6−/− clones. (A) The total number of clones analyzed is shown in the center of each circle. Segments represent the proportion of clones that contain the indicated number of mutations. (B) The sequence contains 26% A, 32% T, 28% G, and 14% C; values were corrected to represent a sequence with equal amounts of the four nucleotides. Mutations are shown from the nontranscribed strand. (C) Total mutations for each nucleotide are grouped.

By May 29, 2017 Downloaded from
561-bp region located upstream of the core S region has been reported to accumulate mutations in B cells stimulated in culture (29, 30). To obtain a distribution profile that more closely resembles mutations occurring in vivo, we sequenced this region in PNA\(^{+}\) B cells from Peyer’s patches. The following number of clones and mutations were identified: C57BL/6 mice, 31% of 87 clones were mutated at a frequency of \(1.3 \times 10^{-3}\) mutations/bp with 63 substitutions and four deletions (5, 15, >100, and >100 bp); Msh3\(^{−/−}\) mice, 19% of 81 clones were mutated at 0.6 \(\times 10^{-3}\) mutations/bp with 29 substitutions and one deletion (>100 bp); and Msh6\(^{−/−}\) mice, 29% of 76 clones were mutated at 1.4 \(\times 10^{-3}\) mutations/bp with 59 substitutions and two deletions (1 and 27 bp; Fig. 3 A). Among the mutated sequences, there was no significant difference between the frequencies of mutation in all three genotypes (3–5 \(\times 10^{-3}\) mutations/bp). The large deletions may be signatures of internal recombination in the S\(\mu\) region (29).

Only clones with unique mutations were analyzed for the types of substitutions. As shown in Fig. 3 (B and C), all three types of mice had more mutations of C than G on the nontranscribed strand. The Msh6\(^{−/−}\) clones had a dramatic decrease in the frequency of A and T mutations (P < \(10^{-3}\)) compared with G and C mutations, with G:C transitions being the major category of mutations. Differences between the clones in individual categories, such as C to A, were not statistically significant. Thus, Msh6 participates in hypermutation of A:T nucleotides in the S\(\mu\) region.

Hot Spot Focusing of Mutations in V and S Regions from Msh6-deficient Mice. Mutations were plotted along V and S sequences to determine if certain positions were targeted. In J\(_{\mu}\)4 introns, three peaks were observed in Msh6-deficient clones, but not in Msh3-deficient and C57BL/6 clones (Fig. 4). Approximately 50% of the Msh6\(^{−/−}\) mutations were in these peaks, compared with only 8% of the mutations from each of the Msh3\(^{−/−}\) and C57BL/6 groups. The positions are nucleotides 57–58, 62–63, and 253–254 (clones shown in Figs. S1–S3, available at http://www.jem.org/cgi/content/full/jem.20040691/DC1) and consist of adjacent G and C bases in WGCW motifs as recorded from the nontranscribed strand. In S\(\mu\), five major peaks were seen in Msh6\(^{−/−}\) clones that contained 50% of the total mutations, whereas these positions had only 16% mutations in the Msh3\(^{−/−}\) and C57BL/6 clones (Fig. 5). The positions are nucleotides 216–217, 282–283, 392–393, 452–453, and 462–463 (clones shown in Figs. S4–S6, available at http://www.jem.org/cgi/content/full/jem.20040691/DC1), which are G-C bases in WGCW motifs.
Martomo et al. The number of mutations of G and C in WGCW sequences from Msh6 clones is listed in Table I. Mutations in individual clones occurred at either G or C, but not in tandem. In the JH4 intron region containing 344 bp, only three WGCW motifs were found, and they all had mutations at either the G or C positions. Other motifs such as RGCW or WGCY had far fewer mutations of G or C. In the S/H9262 region containing 561 bp, 14 WGCW sequences were found; 12 of them had substitutions in either the G or C positions. Fewer mutations were located in RGCW or WGCY motifs. Therefore, the WGCW subset of the canonical RGYW/WRCY motif was highly targeted, whereas other variants were avoided.

Discussion

Msh6, But Not Msh3, Affects Heavy Chain Class Switching.
As reported for Msh2-deficient mice (31–33), Msh6-deficient mice had diminished switching to IgG1 and IgG3, compared with Msh3-deficient and wild-type cells. This indicates that the Msh2-Msh6 heterodimer, but not Msh2-Msh3, is involved in S recombination. Mismatch repair proteins could be involved in switching by two potential mechanisms. First, they could be involved in the recombination process itself by recruiting nucleases to process cleavage intermediates (34). Msh2-, Pms2-, and Mlh1-deficient mice all had reduced switching and different lengths of microhomology at the S junctions compared with wild-type cells (31, 32, 35, 36). In particular, Msh2−/− clones had short lengths of overlapping nucleotides, and Min et al. (37) postulated that the protein is needed for handling long staggered ends before recombination. Although S junctions in Msh6−/− mice were not analyzed in this work due to the low frequency of μ−γ switching, it is possible that the diminished switching is due to inefficient recombination.

Second, Msh6 may facilitate nicking at many positions to create more DNA breaks. Breaks would occur when cytosine on single-stranded DNA is deaminated to uracil, uracil is removed by uracil glycosylase, and the abasic site is cut by an abasic endonuclease (3–7, 13, 38, 39). The restricted focusing of mutations observed in Msh6-deficient clones suggests that Msh6 is needed to expose more cytosines to deamination. This may require an exonuclease to digest DNA from the nick, which would expose C on the opposite strand for AID to generate a double strand break. In support of this model, Bardwell et al. (40) have shown that mice deficient for exonuclease 1 have reduced heavy chain class switching. Because Msh2 binds to exonuclease 1 (41, 42), the Msh2-Msh6 dimer could recruit it to the uracil lesion. Thus, reduced switching in the absence of Msh2-Msh6 could be due to both limited processing of recombination intermediates and creation of fewer nicks.

Msh6 Is Necessary for A-T Mutations in V and S Regions.
We analyzed mutations in the intron region downstream of rearranged V-D-JH4 genes and the Sμ region upstream of the core repeats for types of substitutions. For the V region, the mutational frequency in the mutated sequences was significantly lower in Msh6−/− B cells compared with C57BL/6 and Msh3−/− cells; a lower frequency in this region has also been reported in earlier studies on Msh6 and Msh2 (18, 19, 22). However, for the S region, the frequency of mutation for the mutated sequences was the same for all three genotypes. We propose that in the absence of Msh6, mutations are targeted to WGCW motifs, and because there are fewer of these in the V region versus the S region, there will be fewer mutations in the V region.

Mutations in both V and S regions from the Msh6−/− clones showed a dramatic decrease in substitutions of A and T bases, and a rise in mutations of G and C bases, compared with Msh3−/− and C57BL/6 clones. There were more substitutions of C on the nontranscribed strand from

---

Figure 5. Location of mutations in Sμ shows targeting in Msh6−/− clones. Nucleotides are numbered with 1 corresponding to 4,600 of Genbank/EMBL/DDBJ under accession no. J00440. Data represent the number of mutations at a position divided by the total number of clones. Bars below the abscissa represent WGCW motifs.

---

Published July 6, 2004
all three genotypes in the S region, which suggests that secondary structures in this region, such as stable R-loops, may enhance deamination of single-stranded C (16). These data demonstrate that the Msh2-Msh6 heterodimer is involved in generating mutations of A:T bp in V and S regions. Because exonuclease 1- and DNA polymerase ε-deficient B cells also have fewer A:T mutations, all of these proteins may interact at the uracil lesion. For example, Msh2-Msh6 could bind to a U:G mismatch, and recruit the other proteins to extend mutation past the original C target to include A:T pairs.

### AID May Be Initially Targeted to WGCW Sites

Focused peaks of mutation were observed in the JH4 intron in Msh6<sup>−/−</sup>/H11002/H11408/H11002 clones, which exactly coincided with the peaks noted for Msh2<sup>−/−</sup>/H11002/H11408/H11002 clones (Table I; references 18 and 19). Peaks of mutation were also identified in the S<sub>H9262</sub> region in Msh6<sup>−/−</sup>/H11002/H11408/H11002 clones, but not in Msh3<sup>−/−</sup> or C57BL/6 clones (Fig. 5). The peaks occurred exclusively at WGCW motifs, Table I. Mutations Are Focused in WGCW Motifs in JH4 Intron and S<sub>H9262</sub> Regions in Msh6<sup>−/−</sup> and Msh2<sup>−/−</sup> Clones

<table>
<thead>
<tr>
<th>Nucleotide position&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Sequence</th>
<th>Target sequence</th>
<th>No. mutations at G/C</th>
<th>Msh6&lt;sup&gt;−/−&lt;/sup&gt;</th>
<th>Msh2&lt;sup&gt;−/−&lt;/sup&gt;&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>JH4 intron</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>56</td>
<td>TTCTGAGCATTTGCA</td>
<td>WGCW</td>
<td>7/1</td>
<td>7/3</td>
<td></td>
</tr>
<tr>
<td>61</td>
<td>AGCATGAGCAGCTA</td>
<td>WGCW</td>
<td>0/5</td>
<td>2/5</td>
<td></td>
</tr>
<tr>
<td>252</td>
<td>GAAAGATGCTAAAC</td>
<td>WGCW</td>
<td>3/3</td>
<td>1/4</td>
<td></td>
</tr>
<tr>
<td>269</td>
<td>CCTATGCGTGGAGG</td>
<td>RGCW</td>
<td>1/0</td>
<td>ND</td>
<td></td>
</tr>
<tr>
<td>98</td>
<td>GAGCCGGCTGAGAG</td>
<td>RGCW</td>
<td>0/0</td>
<td>1/0</td>
<td></td>
</tr>
<tr>
<td>235</td>
<td>TTGAGGCTCATTTT</td>
<td>RGCW</td>
<td>0/0</td>
<td>0/0</td>
<td></td>
</tr>
<tr>
<td>287</td>
<td>GTGGGAGCTTAGT</td>
<td>RGCW</td>
<td>0/0</td>
<td>ND</td>
<td></td>
</tr>
<tr>
<td>94</td>
<td>GAGGGAGGCGGCTG</td>
<td>WGCY</td>
<td>0/0</td>
<td>0/0</td>
<td></td>
</tr>
<tr>
<td>139</td>
<td>CTCAGAGGCCTTTAG</td>
<td>WGCY</td>
<td>0/0</td>
<td>1/0</td>
<td></td>
</tr>
<tr>
<td>S&lt;sub&gt;H9262&lt;/sub&gt; region</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>391</td>
<td>CATAAAGCTGCTG</td>
<td>WGCW</td>
<td>5/4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>215</td>
<td>GACAGTGCTTAGAT</td>
<td>WGCW</td>
<td>2/4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>451</td>
<td>TAAAAGACAGCT</td>
<td>WGCW</td>
<td>3/3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>461</td>
<td>AGCTGAGCTGAGAT</td>
<td>WGCW</td>
<td>3/3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>281</td>
<td>GCCACAGCTGTACA</td>
<td>WGCW</td>
<td>3/2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>487</td>
<td>CTGAGTGCTCTAA</td>
<td>WGCW</td>
<td>2/2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>529</td>
<td>AGGTAAGCAAAGCT</td>
<td>WGCW</td>
<td>2/1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>534</td>
<td>AGCAAAGCTGGGCT</td>
<td>WGCW</td>
<td>2/0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>456</td>
<td>AGCACAGCTGAGCT</td>
<td>WGCW</td>
<td>1/1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>TTGGAAGCTAATTT</td>
<td>WGCW</td>
<td>0/1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>400</td>
<td>TGCTAGCACAATAAT</td>
<td>WGCW</td>
<td>1/0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>312</td>
<td>AGACCTGCAAGTTGA</td>
<td>WGCW</td>
<td>0/1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>325</td>
<td>AGGCAAGCAAGCTG</td>
<td>WGCW</td>
<td>0/0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>395</td>
<td>AAGCTGCTAGCA</td>
<td>WGCW</td>
<td>0/0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>475</td>
<td>GGTTGGGCTTCTCT</td>
<td>RGCW</td>
<td>0/3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>56</td>
<td>ACCAGGCTAAGAA</td>
<td>RGCW</td>
<td>0/0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>65</td>
<td>AAGAAGCACAATCCT</td>
<td>RGCW</td>
<td>0/0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>248</td>
<td>ACAGGAGCTGGGCT</td>
<td>RGCW</td>
<td>0/0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>333</td>
<td>AGGTCAGCTGACCT</td>
<td>RGCW</td>
<td>0/0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>441</td>
<td>AGCAGAGCTCTAA</td>
<td>RGCW</td>
<td>0/0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>166</td>
<td>GAAGTGCCAGTAA</td>
<td>WGCY</td>
<td>0/0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>545</td>
<td>GTTTGAGGCACAAAT</td>
<td>WGCY</td>
<td>0/0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>All of the WGCW, RGCW, and WGCY sequences in both regions are listed. The position of the first base in the underlined sequence is taken from Figs. S1 and S4, and shown in Figs. 4 and 5.

<sup>b</sup>Data from reference 18.
which are a subset of RGYW/WRCY hotspots: in the J14 intron, there are three motifs with three corresponding peaks of mutation; and in the S4 region, there are 14 motifs with major peaks at five of them. Substitutions were found at either the G or C bases in WGCW, but were rarely found in other variants of RGYW such as GGCT, GGCA, AGCC, and TGCC. Recombination junctions are also focused at the frequent WGCW sequences (AGCT) in S regions from Msh2-deficient mice, but not from wild-type mice (31). AID has been shown to preferentially deaminate C in WRC on single-stranded DNA in vitro (26, 27), indicating that the protein has high affinity for this sequence. Thus, WGCW represents overlapping WGC motifs on both strands, which suggests that C is targeted by AID on either strand. W may be necessary on both ends because A:T has a lower melting temperature than G:C, which would be R or Y. Thus, unwinding of WGCW in DNA during transcription could expose C on both strands as a transcription bubble is being formed. Crystal structures of transcription of DNA by RNA polymerase II confirm that the strands that make up the DNA strands at the edges of the transcription bubble, particularly at the trailing edge where RNA is extruded (43).

Subsequent dispersal of mutations from the entry points in wild-type cells could be caused by a secondary round of AID attacks, where Msh2-Msh6 may facilitate the progressive deamination of cytosines in other sequence contexts. For example, exonuclease 1 could generate long stretches of single-strand DNA that resemble the DNA substrates used in vitro, which show that C is deaminated in other WRC motifs (26, 27). Dispersed mutations could also result from error-prone synthesis by DNA polymerase η, which can synthesize in a gapped substrate. It remains to be seen how Msh2-Msh6 and pol η are recruited to uracil lesions, and how they function in the presence of other players, such as uracil glycosylase, abasic endonuclease, and exonuclease 1, in this error-prone DNA repair scenario.

We thank W. Edelmann for generously providing mice; D. Winter, A. Martin, N. Joshi, B. Wersto, J. Chest, and C. Morris for help in various aspects of the work; D. Wilson, R. Wood, and A. Gnatt for discussions; and V. Bohr for support. This work was supported by the National Institutes of Health intramural research program.

Submitted: 7 April 2004
Accepted: 28 May 2004


References


