

# The very 5' end and the constant region of Ig genes are spared from somatic mutation because AID does not access these regions

Simonne Longerich,<sup>1</sup> Atsushi Tanaka,<sup>1</sup> Grazyna Bozek,<sup>2</sup> Dan Nicolae,<sup>3</sup> and Ursula Storb<sup>1,2</sup>

<sup>1</sup>Committee on Immunology, <sup>2</sup>Department of Molecular Genetics and Cell Biology, and <sup>3</sup>Department of Statistics, University of Chicago, Chicago, IL 60637

**Somatic hypermutation (SHM) is restricted to VDJ regions and their adjacent flanks in immunoglobulin (Ig) genes, whereas constant regions are spared. Mutations occur after about 100 nucleotides downstream of the promoter and extend to 1–2 kb. We have asked why the very 5' and most of the 3' region of Ig genes are unmutated. Does the activation-induced cytosine deaminase (AID) that initiates SHM not gain access to these regions, or does AID gain access, but the resulting uracils are repaired error-free because error-prone repair does not gain access? The distribution of mutations was compared between uracil DNA glycosylase (Ung)-deficient and wild-type mice in endogenous Ig genes and in an Ig transgene. If AID gains access to the 5' and 3' regions that are unmutated in wild-type mice, one would expect an "AID footprint," namely transition mutations from C and G in Ung-deficient mice in the regions normally devoid of SHM. We find that the distribution of total mutations and transitions from C and G is indistinguishable in wild-type and Ung-deficient mice. Thus, AID does not gain access to the 5' and constant regions of Ig genes. The implications for the role of transcription and Ung in SHM are discussed.**

## CORRESPONDENCE

Ursula Storb:  
stor@midway.uchicago.edu

Abbreviations used: AID, activation-induced cytosine deaminase; CDR, complementarity-determining region; Msh, mismatch repair; pol, polymerase; SHM, somatic hypermutation; Ung, uracil DNA glycosylase.

Somatic hypermutation (SHM) of Ig genes requires the function of activation-induced cytosine deaminase (AID; references 1 and 2). Although unidentified, AID may have an RNA substrate (3). However, *in vitro*, AID deaminates monomeric dCTP (4), single-stranded DNA (5–7), transcribed double-stranded DNA (8–11), and supercoiled double-stranded DNA (12). Moreover, studies in cells deficient in uracil DNA glycosylase (Ung) activity strongly support the idea that AID deaminates cytosines in DNA directly to produce U:G DNA mispairs (13–15). In the absence of Ung, C mutations are almost exclusively C to T (and for C deamination on the opposite strand, complementary G to A) transitions, as expected if deoxyuridines resulting from AID cytidine deamination are left unrepaired before replication. Indeed, in activated B cells, uracil excision appears to be accomplished mostly by Ung, with minimal contribution from the SMUG1 uracil glycosylase (16, 17). Although transition mutations in Ung-wild-type mammals are the most common SHMs from C and G, transversions are also frequent and are thought to arise either entirely (13, 15) or in part (18) from

Ung-mediated base excision repair. Ung-mediated excision of the uracil base creates an abasic site that, if used as a template for DNA synthesis, could lead to any mutation from C or G (for review see reference 19). Alternatively, processing of the abasic site to generate a single-stranded gap might lead to mutations when followed by error-prone DNA synthesis involving lesion-bypass polymerases (pols), of which pol  $\eta$ , pol  $\iota$ , and pol  $\zeta$  have been implicated in SHM (18). Finally, the U:G mispair could also be processed by the mismatch repair system. The SHM pattern in mismatch repair (Msh)2- and pol  $\eta$ -deficient B cells supports a model in which Msh2/6-mediated removal of the U:G mismatch, followed by pol  $\eta$  and pol  $\iota$  error-prone DNA synthesis, leads mainly to mutations from A and T (20, 21).

The mechanism by which SHM is targeted almost exclusively to Ig DNA is a mystery. Although Ig V regions accumulate mutations at an average frequency of  $10^{-4}$ – $10^{-3}$  mutations/base pair/cell generation ( $\sim 10^6$ -fold higher than the frequency of spontaneous mutations), with a few exceptions (22–25), most other genes are not mutated. Transcription of the

target DNA is required for normal SHM because (a) removing an enhancer from a murine Ig transgene decreases the SHM frequency (26); (b) mutation frequency correlates positively with the level of transcription regulated by an inducible promoter upstream of a transgene in a mutating B cell line (27); (c) the frequency of mutations is dependent on distance from the promoter (28–33); and (d) duplicating an I $\gamma$ k promoter upstream of the normally unmutated Ig C region in an I $\gamma$ k transgene results in initiation of mutations in the C region with a similar dependence on distance from the promoter (34). The latter observation led to the proposal that a “mutator factor,” perhaps an endonuclease, is recruited to transcription complexes initiating at Ig loci, travels with the complex, and gets deposited stochastically where it nicks the DNA, perhaps when the RNA pol pauses (34, 35). Error-prone repair of nicks would generate mutations. Due to a high chance of deposition within the first 1–2 kb from the promoter, the mutator factor was proposed to be absent from transcription complexes beyond 1–2 kb downstream of the transcription start, thereby protecting the C region from mutation (34, 35).

We sought to determine whether AID is a transcription-coupled mutator factor *in vivo*. *In vitro*, deamination of double-stranded linear DNA by AID requires transcription (8–11) or a putative negatively supercoiled conformation (12). Negative supercoils are generated behind an elongating transcription complex (36). However, that transcription alone creates the appropriate substrate for AID does not explain how DNA within the first ~200 bp and beyond 1–2 bp of the transcription start (i.e., Ig C regions), and most other genes that are highly transcribed in germinal center B cells, remain essentially mutation-free. To address this question, we considered two possibilities: transcription complexes at or near the transcription initiation site load and carry AID, which deaminates cytosines for 1–2 kb before becoming inactive or dissociating from the elongating transcription complex, or AID deaminates cytosines made accessible by transcription or supercoiled conformation throughout Ig genes and in other loci. However, although repair of U:G mismatches would be error-free in the very 5' and C regions of Ig loci, and in other loci (like during Ung-mediated repair of replication errors; references 37 and 38), error-prone repair could occur in V regions because error-prone repair factors (or factors that recruit them) are targeted preferentially to Ig V regions.

Whether targeting of AID or of error-prone repair causes the skewed distribution of SHM was addressed by comparing the distribution of mutations relative to the transcription start site *in vivo*, in murine B cells wild-type (Ung<sup>+</sup>) or deficient (Ung<sup>-</sup>) for Ung. Importantly, these experiments were performed under conditions of endogenous rather than overexpressed levels of AID and therefore should be most appropriate for detecting Ig-specific targeting of SHM. SHMs in endogenous Ig loci and in an unselected transgene (39) were analyzed. We reasoned that C to T (and complementary G to A) transition mutations in the absence of Ung represent AID-

catalyzed cytosine deaminations, i.e., an AID “footprint,” based on the following considerations. In the absence of Ung, uracils created by AID result in transitions from C and G when the DNA is replicated. It is unlikely that many transitions from C or G occur by the interaction of error-prone pols with mismatch repair. The major error-prone pols found to be involved in SHM are pol  $\eta$  and pol  $\iota$  (18, 20, 21, 40). Both have the highest error rates when copying T, leading to transitions from A (T) (41, 42). In the Ung<sup>-</sup> mice, where MMR would still be active, there is no increase in mutations from A or T (15 and see Figs. 1–4). Thus, the transitions from C and G are likely an AID footprint (see Discussion). We have mapped mutations relative to each transcription initiation site. Strikingly, the AID footprint is virtually absent in the first 100 transcribed nt, low between 100–200 nt, and significantly reduced beyond ~1.5 kb of the transcription start, supporting the idea that AID activity is coupled to the transition between transcription initiation and elongation from Ig promoters.

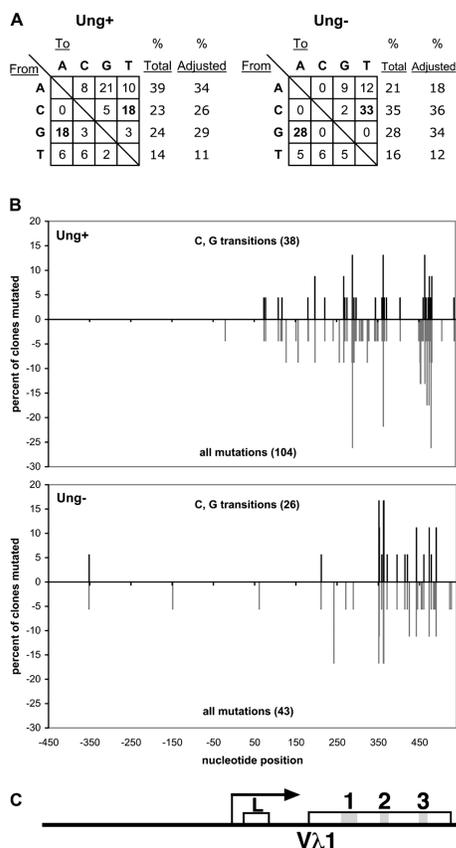
## RESULTS

To determine whether AID deamination, or post-AID processing of the deaminated cytosine, determines the mutation distribution within Ig genes, we mapped C to T (and G to A) mutations in two endogenous Ig loci, the  $\lambda$ 1 light chain locus and VJ558-rearranged IgH genes, as well as in an I $\gamma$ k transgene in germinal center B cells of immunized Ung<sup>-</sup> and Ung<sup>+</sup> control mice.

### AID deaminations are essentially absent from regions upstream and within ~100 bp downstream of the transcription start site in endogenous rearranged V $\lambda$ 1 genes

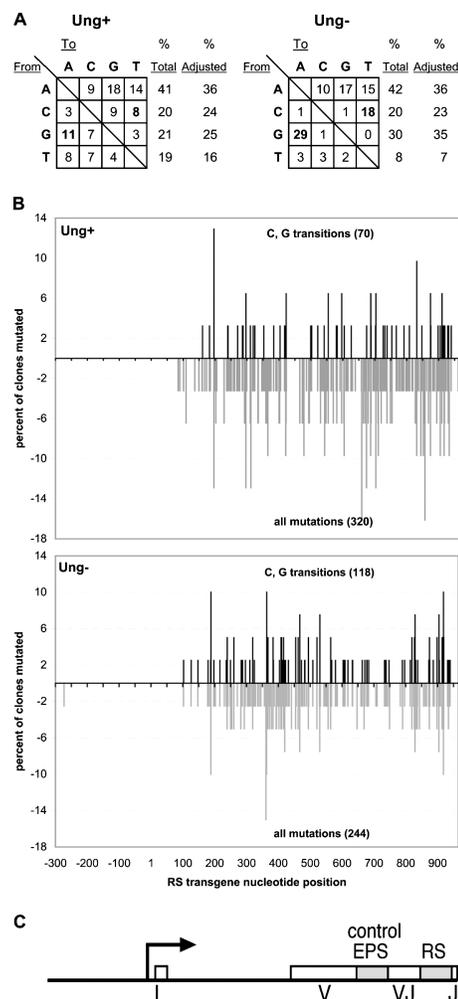
Upstream and immediately downstream of the transcription start site, SHMs have been reported to occur at very low frequencies, if at all (28–31). To determine whether this 5' region is protected or not from AID-mediated cytosine deamination, we mapped mutations in rearranged  $\lambda$ 1 light chain genes. In mice, there are only three V $\lambda$  genes (43), and ~60% of  $\lambda$ -expressing B cells derive from V $\lambda$ 1/J $\lambda$ 1 rearrangements (44). Genomic DNA from PNA<sup>high</sup> germinal center B cells of immunized Ung<sup>+</sup> or Ung<sup>-</sup> mice was used as a template to PCR amplify rearranged  $\lambda$ 1 genes. As reported previously for SHM in other Ig loci of Ung<sup>-</sup> cells (15), mutations from C and G are almost exclusively transitions in the Ung<sup>-</sup> as compared with Ung<sup>+</sup> mice (Fig. 1 A). The specific mutation pattern shift favoring C to T (and G to A) mutations in Ung<sup>-</sup> mice in all loci assayed here (see below) further substantiates that C and G transitions likely represent an AID footprint.

A 991-bp region of V $\lambda$ 1 between 453 bp upstream and 538 bp downstream (ending at the 3' region of J $\lambda$ 1) of the start of transcription of  $\lambda$ 1 was analyzed for SHM (Fig. 1 C). Both the subset of transition mutations from C and G as well as the total mutations are plotted with respect to their position in Fig. 1 B (note that all mutations from A, C, G, and T



**Figure 1. Mutations in the very 5' end of the  $\lambda 1$  gene in Ung<sup>+</sup> and Ung<sup>-</sup> mice.** In the Ung<sup>+</sup> background, a total of 29 clones were analyzed, of which 23 clones contained one or more mutations in a 991-bp region between  $-453$  upstream and  $+538$  bp downstream of the start of transcription in the rearranged V $\lambda 1$  gene. There were a total of 104 mutations, including 38 transitions from C and G, among these 23 clones. In the Ung<sup>-</sup> background, 30 clones were analyzed, of which 18 clones contained one or more mutations for a total of 43 mutations, including 26 C and G transitions. (A) Summary of mutation patterns expressed as a percentage of total mutations. In the adjusted column, variability of A:C:G:T contents in 991 nt of the analyzed region was normalized for base content. A, 27.1%; C, 22.6%; G, 19.3%; T, 31.0%. (B) Distribution of mutations from  $-450$  to  $+538$  relative to the start of transcription ( $+1$ ). Mutations equal the number of mutations at each position per the total number of mutated clones sequenced times 100. (C) The three CDRs are indicated in the diagram of the V $\lambda 1$  gene below the graph.

are plotted below the zero ordinate, whereas only the transitions from C and G are plotted above the zero ordinate). Similar to previous reports, we found only three mutations between  $-453$  and  $+100$  bp relative to the transcription start in wild-type mice. Mutations in Ung<sup>-</sup> mice, including C and G transition mutations, were almost entirely absent from  $-453$  to  $+242$  bp. In both Ung<sup>+</sup> and Ung<sup>-</sup> mice, mutations eventually reached a plateau (at about  $+200$  or  $+350$  bp, respectively). Elevated numbers of mutations at complementarity-determining region (CDR)1, CDR2, and CDR3 (Fig. 1 B) suggest that B cells expressing a functional V $\lambda 1$ -



**Figure 2. Mutations in the very 5' end of the RS transgene.** 69 clones were analyzed in the Ung<sup>+</sup> background, and 31 clones contained one or more mutations between positions  $-301$  to  $+964$  for a total of 320 mutations, including 70 transitions from C and G. In the Ung<sup>-</sup> background, 138 clones were analyzed and 40 clones contained more than one mutation for a total of 244 mutations, including 118 transitions from C and G. (A) Summary of mutation patterns in the 1,265 nt of the analyzed region, expressed as a percentage of total mutations. In the adjusted column, variability of A:C:G:T contents in 1,265 nt of the analyzed region was normalized for base content. A, 28.6%; C, 20.7%; G, 21.2%; T, 29.5%. (B) Distribution of mutations from  $-300$  to  $+964$  relative to the start of transcription ( $+1$ ). Mutations equal the number of mutations at each position per total number of mutated clones sequenced times 100. (C) Map of RS transgene designed to test certain parameters of SHM targeting (reference 39). EPS and RS are 96 nt insertions in the V and J regions, respectively.

V $\lambda 1$  gene were selected in germinal centers for their affinity to the immunogen.

The paucity of mutations upstream of the plateau is not due to a lack of AID hotspots in the upstream and very 5' regions (Table I). Therefore, because C or G transition mutations in Ung<sup>-</sup> mice, i.e., AID footprints, are virtually absent upstream and within the first 100 bp of the transcription start, AID appears not to gain efficient access to this 5' region of  $\lambda 1$ .

**Table I.** Statistics of mutation frequencies and hotspots in V $\lambda$ 1

From (bp)	To (bp)	$\lambda$				Hotspots
		Ung <sup>+</sup> <i>(no. of mutations)</i>	p-value	Ung <sup>-</sup> <i>(no. of mutations)</i>	p-value	
-400	-301	0	—	1	—	8
-300	-201	0	—	0	—	11
-200	-101	0	—	1	—	13
-100	-1	1	—	0	—	12
1	100	2	0	1	0.001	12
101	200	12	0.035	0	0.0004	16
201	300	24	0.89	6	0.27	15
301	400	21	0.72	14	0.99	11
401	500	41	0.99	18	0.99	12

**Mutations in the 5' region of the RS transgene**

B cells require surface Ig expression for survival (for review see reference 45), and mutations in Ig genes are selected for their ability to confer increased antigen affinity. Because mutations in the promoter, leader, or leader-V intron region of the endogenous  $\lambda$ 1 gene may have been selected against and might therefore be undetected, we also examined the 5' region of an unselected murine Ig transgene. The RS transgene is based on a rearranged I $\mu$ <sub>κ</sub>167 gene, including its native promoter and both intronic and 3' enhancer elements, and was reported previously to acquire 5–10-fold more SHMs than similar transgenes due to two E boxes (5' CAGGTG 3'), DNA binding elements located ~900 bp from the promoter (in the RS region; Fig. 2 C) to which E2A proteins can bind (46). Despite increasing the mutation frequency of all regions of the RS transgene, the E boxes do not increase the level of RS transgene transcription (46).

The RS transgenic mice were bred onto Ung<sup>-</sup> and Ung<sup>+</sup> backgrounds and immunized to collect activated B cells from germinal centers. SHMs were mapped in a 5' re-

gion of the RS transgene between nt positions -301 and +964 relative to the transcription start (Fig. 2 C). As in the endogenous  $\lambda$ 1 gene, Ung deficiency caused a shift in the C and G mutation pattern predominantly to transitions (Fig. 2 A). Therefore, the molecular mechanisms governing mutation patterns appear to be similar between endogenous Ig genes and the RS transgene. In the Ung<sup>+</sup> background, there were no mutations upstream of the promoter and in the first 83 bp of the transcribed region (Fig. 2 B). The numbers of mutations per 100 bp become comparable between +200 and +964 bp, such that the mutation frequency reaches an apparent plateau after +200 bp, similar to  $\lambda$ 1. In Ung<sup>-</sup> mice, there was one mutation upstream of the promoter and no mutations within the first 101 bp from the transcription start. Three mutations were seen in the next 50 bp, and an apparent plateau of mutations was reached around +180 (Fig. 2 B). Thus, the very 5' end of the RS transgene lacks mutations in both Ung<sup>+</sup> and Ung<sup>-</sup> situations.

In Ung<sup>-</sup> mice, one of the 138 clones analyzed contained eight mutations upstream of the transcription start of the RS

**Table II.** Statistics of mutation frequencies and hotspots in RS transgene

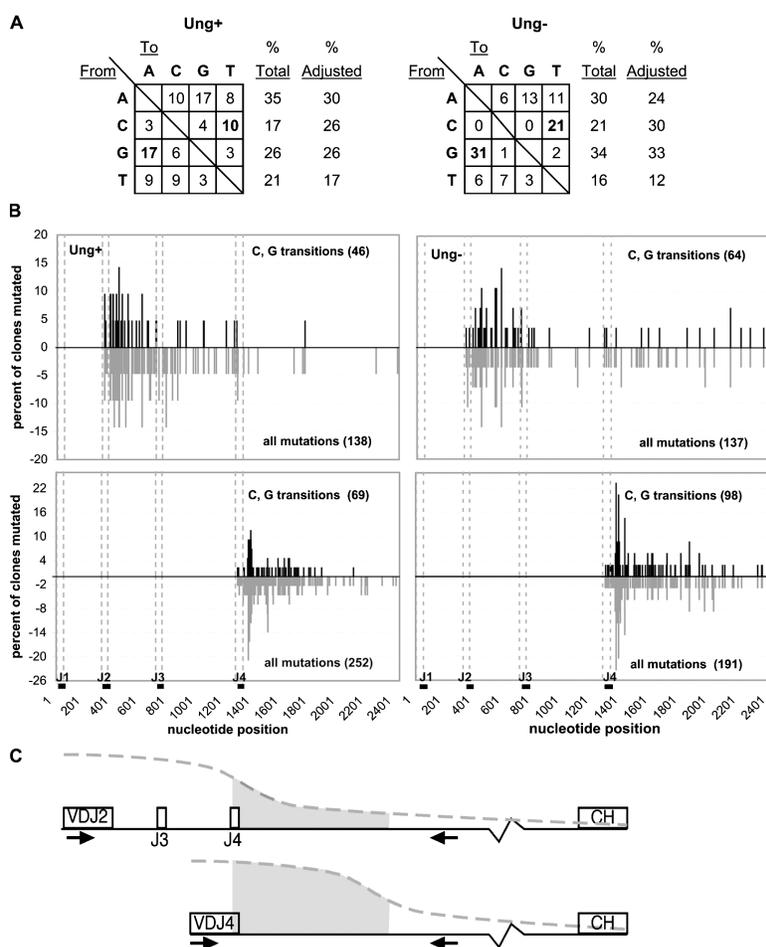
From (bp)	To (bp)	RS transgene				Hotspots
		Ung <sup>+</sup> <i>(no. of mutations)</i>	p-value	Ung <sup>-</sup> <i>(no. of mutations)</i>	p-value	
-300	-201	0	—	1	—	11
-200	-101	0	—	0	—	10
-100	-1	0	—	0	—	15
1	100	2	~0	0	~0	11
101	200	16	0.0005	14	0.0036	12
201	300	32	0.46	25	0.53	12
301	400	39	0.87	48	0.99	7
401	500	31	0.39	44	0.99	7
501	600	33	0.53	23	0.36	8
601	700	42	0.95	19	0.11	11
701	800	41	0.93	16	0.029	10
801	900	53	0.99	31	0.89	12

transgene. The mutations were all from A or T, and mostly from A to T, and therefore did not represent the AID footprint directly. The RS transgenic mouse has two copies of the RS transgene in tandem. We discovered previously that the 3'  $\kappa$  enhancer contains a promoter that can drive SHM (47). Thus, the 5' mutations in this rare clone may have gained mutations from the 3' promoter of the upstream transgene, and this clone was not included in the analysis of Ung<sup>-</sup> mice.

To rule out the possibility that lack of AID hotspots results in the mutation pattern observed in the 5' region of the RS transgene, the distribution of WRC and GYW motifs (7) was analyzed (Table II). On average, there are 11.3 hotspots per 100 bp in the 5' end of the RS gene. The upstream regions are not impoverished for hotspots, and the consistency of hotspots throughout the gene to +964 bp

does not correlate with the absence of mutations in the 5' end of the transcribed region.

Because the number of mutations in the first 200 bp is very low, the frequency of mutations in the 5' portion of the RS transgene was analyzed statistically. To test whether a very low number of mutations in the first 200 bp can result from random mutation, we defined our null hypothesis to be that mutations occur at a flat rate between +1 and +964 bp of the transcription start. For the Ung<sup>+</sup> mice, in 10,000 simulations of randomly placing a total of 318 mutations between bp +1 and +964, observing only two mutations or fewer within the first 100 bp is unlikely ( $P \cong 0$ ; Table II). For +101 to +200 bp, observing only 16 mutations or fewer is also very unlikely in this region if mutations occur randomly because there were only 5 out of 10,000 cases with 16 or



**Figure 3. Mutations in the mutation plateau and 3' thereof in the endogenous IgH chain gene.** In Ung<sup>+</sup> mice, 21 of 41 total J2-rearranged clones and 43 of 100 total J4-rearranged clones had one or more mutations. In Ung<sup>-</sup> mice, 28 of 53 total J2-rearranged clones and 34 of 64 total J4-rearranged clones had one or more mutations. (A) Summary of mutations downstream of J4 (between nt 1,284 and 2,428) in J4-rearranged genes pooled from three mice of each genotype. Numbers represent percents of the total mutations. In the adjusted column, variability of A:C:G:T contents of the analyzed region was normalized for base content. A, 29.1%; C, 16.2%;

G, 24.2%; T, 30.5%. (B) Distribution of mutations 3' of J<sub>H</sub>2 (between nt 334 to 2,428, top) and 3' of J<sub>H</sub>4 (between nt 1,284 and 2,428, bottom) in Ung<sup>+</sup> (left) and Ung<sup>-</sup> mice (right), respectively. All nt numbering is according to GenBank accession number X53774. Mutations equal the number of mutations at each position per total number of mutated clones sequenced times 100. For each graph, the number of total mutations and transitions plotted from C and G is indicated in the bottom or top-right corner. (C) Diagram of VDJ2 and VDJ4 rearrangements. The shaded area shows the staggered decline of mutations in these two rearrangements.

fewer mutations in the simulations ( $P = 0.0005$ ). For the  $Ung^{-}$  mice, the same null hypothesis was tested by placing 260 mutations randomly between nt +1 to +964 in the RS transgene. Again, observing no mutations between +1 and +100 ( $P \cong 0$ ) and 14 mutations between +101 and +200 bp ( $P = 0.0036$ ) was very unlikely, indicating that mutations are very unlikely to be distributed randomly in this region. The same conclusions are obtained when applying similar statistical analysis to the mutations in the endogenous  $\lambda 1$  gene, despite smaller numbers of mutations (Table I).

The rejection of the null hypothesis for the first 200 bp of both  $Ung^{+}$  and  $Ung^{-}$  backgrounds indicates that the mutation processes occurring within the first 200 bp are statistically different from the processes at +201 to +964. The results suggest that SHM does not operate in the first 100 bp and operates at a reduced level from  $\sim +100$  to  $\sim +200$  bp. Moreover, AID appears not to have access to the 5' end of Ig genes (see Discussion).

#### The frequency of AID deaminations decreases with increasing distance downstream of transcription initiation in endogenous IgH chain genes

To analyze the decline in mutation frequency in 3' regions beyond 1–2 kb from the transcription start, mutations were mapped in the intron between  $J_H4$  and  $C\mu$  in  $V_HJ558$ -rearranged IgH genes. IgH chain genes rearrange using one of four possible  $J_H$  segments, such that the intronic DNA immediately downstream of  $J4$  is closer to the start of transcription in  $D-J4$  rearrangements compared with  $D-J$  rearrangements with upstream  $J_H$  segments,  $J1$ ,  $J2$ , or  $J3$ , due to deletion of the DNA between the recombined D and J segments. This feature allows comparison of the frequency of SHM in an identical, endogenous segment of DNA when located at different distances from the start of transcription (Fig. 3 C). Mutation frequencies between  $J3$  and  $J4$  rearrangements reportedly differ 3.6-fold ( $4.0 \times 10^{-3}$  vs.  $14.4 \times 10^{-3}$ , respectively) in a 314-bp segment encompassing two thirds of the 3' end of  $J4$  and the immediate 3' intronic flank (33).

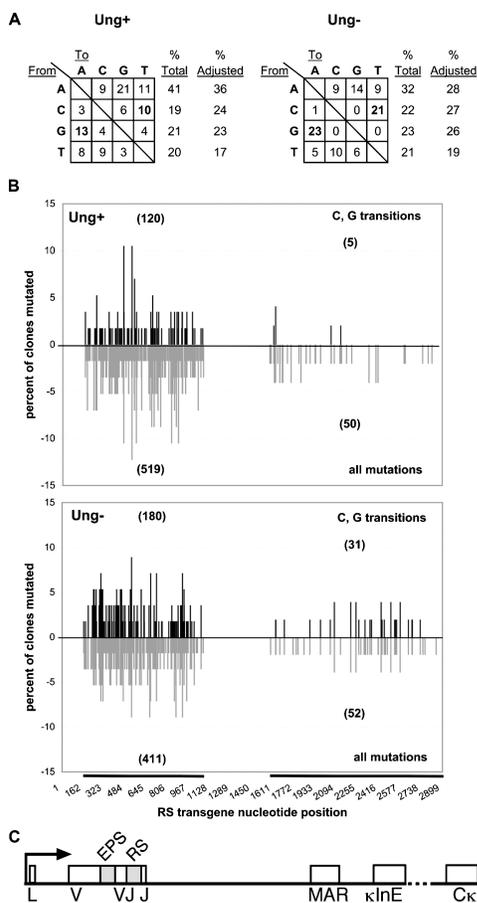
We performed a similar analysis of mutations in the  $J4$ -flanking intron segment in immunized  $Ung^{+}$  and  $Ung^{-}$  mice to determine whether AID gains equal access to this region regardless of the distance between this DNA segment and the start of transcription. (We did not consider switch region mutations that would be located further 3' and are probably induced independently of VDJ region mutations.) The distribution of mutations in  $J2$ - and  $J4$ -rearranged genes is shown in Fig. 3 B, top and bottom, respectively. For both  $J2$ - and  $J4$ -rearranged genes, the sequence was analyzed between the 3' end of the rearranged J segment and the 3' primer used for PCR amplification (located 1,108 bp downstream of  $J4$ ; Fig. 3 B). Data in Fig. 3 represent mutations pooled from three mice, either  $Ung^{+}$  or  $Ung^{-}$ , and as reported previously for this locus (15), mutations from C or G in  $Ung^{-}$  mice are predominantly transitions (Fig. 3 A). Moreover, the distribution of mutations in the  $J4-C\mu$  intron region is similar between

$Ung^{+}$  and  $Ung^{-}$  mice. In both  $Ung^{+}$  and  $Ung^{-}$  mice, all mutations are more frequent within and near the rearranged J segment than further downstream (Fig. 3 B). Finally, in the segment 3' of  $J4$  present in both  $J2$  and  $J4$  rearrangements, but located  $\sim 950$  bp further from the transcription start in  $J2$ -rearranged genes, mutations are  $\sim 12$ - and 5-fold higher in  $J4$ -rearranged genes versus  $J2$ -rearranged genes in  $Ung^{+}$  and  $Ung^{-}$  mice, respectively (Fig. 3 B). The subset of C and G transition mutations differs  $\sim 33$ - and 6-fold in  $Ung^{+}$  and  $Ung^{-}$  mice, respectively, for the same region. Thus, independent of DNA sequence, high levels of cytosine deamination by AID occur only up to  $\sim 600$ –700 bp downstream of the 3' end of the rearranged  $J_H$  gene. Further 3', mutations and transitions from C and G also decrease drastically in the absence of  $Ung$ .

There appear to be more mutations in the 3' region (beyond +1284) in the  $Ung^{-}$  mice as compared with the  $Ung^{+}$  mice. We speculate that this might be due to the technical difficulty of PCR amplifying long regions of contiguous genomic Ig DNA, which produces a significant fraction of hybrid PCR products as observed previously for the IgH locus using the same PCR primers (33). We excluded all detectable hybrid products from this analysis using polymorphisms in the analyzed region (see Materials and methods), but of course hybrids between DNA molecules derived from the same allele would not be detectable. As a consequence, some of the mutations downstream of  $J2$  might be derived from hybrids (for example, derived from a  $J3$ - or  $J4$ -rearranged segment with mutations after +1284), possibly leading to an overestimate of the mutations 3' of  $J4$  in clones scored as  $J2$ -rearranged. Conversely, hybrid PCR artifacts would lead to an underestimation of the mutations after +1,284 in  $J4$ -rearranged clones via similar events. As a consequence, we expect that the observed difference in mutation frequency between  $Ung^{+}$  and  $Ung^{-}$  mice beyond +1,284 is not significant.

#### Mutations in the mutation plateau and 3' regions of the RS transgene

Just as for the  $\lambda$  light chain analysis, mutations in the heavy chain J-C intron may have been selected against because of unknown functional consequences (e.g., RNA splicing, IgH intron enhancer function, etc.) and thus would be undetected. Therefore, we also examined the 3' distribution of mutations in the unselected RS transgene. To examine whether there is a decline in mutation frequency with increasing distance from the start of transcription as is seen in endogenous IgH genes, two sections in the RS transgene were analyzed (Fig. 4, B and C). As in the endogenous IgH locus (Fig. 3 B), mutations are frequent between +162 and +1,078 and 10–13-fold less frequent 500 bp further downstream between +1,580 and +2,900. A similar decline in mutation frequency was observed in the subset of transition mutations from C and G, both in the presence and absence of  $Ung$  (e.g., a sevenfold decrease in  $Ung^{-}$ ). The paucity of mutations in the 3' region beyond 1.5 kb of the RS trans-



**Figure 4. Mutations in the mutation plateau and 3' thereof in the RS transgene.** Two regions of the RS transgene were PCR amplified to map mutations between +162 to +1,078 and +1,580 to +2,900 bp downstream of the transcription start (at position +1). (A) Summary of mutation patterns in the region between nt 162 and 1,078 in Ung<sup>+</sup> and Ung<sup>-</sup> mice. Numbers represent percents of the total mutations. In the adjusted column, variability of A:C:G:T contents of the analyzed region was normalized for base content. A, 28.14%; C, 20.07%; G, 23.01%; T, 28.79%. (B) Distribution of mutations in two regions of the RS transgene downstream of transcription. Mutations equal the number of mutations at each position per total number of clones sequenced times 100. The total number of mutations and transitions from C and G in each region is indicated in parentheses at the bottom or top of each plot. (C) Map of RS transgene (refer to legend to Fig. 2).

gene is particularly striking when considering the large mutation load of some clones (e.g., in Ung<sup>-</sup> mice, the maximum number of mutations per clone in the region between +162 and +1,078 was 55, whereas a maximum of only 9 mutations per clone was found between +1,580 and +2,900), suggesting that not only the 5' but also the 3' boundary of mutations is rather defined in the RS transgene. Moreover, DNA elements sufficient to define both the transcriptional correlation of SHM, as well as the mutation pattern, must be contained within the elements present in the RS transgene.

The WRC trinucleotide in which the C (or G in the complementary GYW sequence) that is mutated is a hotspot for AID targeting in vitro (7). To determine whether this is also true in vivo, we compared C and G mutations in hotspots and nonhotspots in Ung<sup>+</sup> versus Ung<sup>-</sup> mice in the RS transgene between +162 and +1078 bp. Interestingly, 59 and 64% of mutations from C and G were found in hotspots in Ung<sup>+</sup> and Ung<sup>-</sup> mice, respectively, suggesting that the WRC/GYW hotspots are also preferred AID substrates in vivo, and that targeting of these hotspots is determined by AID rather than post-AID repair steps.

## DISCUSSION

### Relationship of SHM to transcription

Evidence for a correlation between SHM and transcription is abundant (for review see references 48 and 49). Yet, how SHM and transcription are related mechanistically is unknown. Recent in vitro studies of the function of AID suggest that transcription creates a substrate for AID, perhaps single-stranded DNA (for review see reference 49) or negatively supercoiled double-stranded DNA (12). However, in B cells undergoing SHM, how mutations are restricted mostly to V regions of Ig DNA in vivo, where Ig C-region DNA and many other genes are transcribed, is not understood.

C to T (and the corresponding G to A) transition mutations in the absence of Ung likely represent AID-catalyzed mutations (also see below). If AID initiates the SHM process, and AID (rather than error-prone repair) is directed to the target by transcription, then the frequency of AID-promoted transition mutations from C and G in Ung<sup>-</sup> mice should follow the distribution frequency of typical SHM in an Ung<sup>+</sup> mouse. Precisely this pattern was observed both in endogenous Ig loci and in the RS transgene in vivo, where C and G transition mutations overlapped with the distribution of all mutations in both Ung<sup>+</sup> and Ung<sup>-</sup> mice. Moreover, the fact that the promoter and general region upstream of the transcription start are essentially devoid of C or G transition mutations in Ung<sup>-</sup> mice suggests that targeting of AID requires transcribed DNA.

Transcription coupling of AID activity might be accomplished by association of AID with the transcription complex during transcription elongation. (AID-RNA pol complexes have been observed so far only in experiments in which AID was epitope-tagged and overexpressed [50]). Whether physiological levels of AID bind RNA pol has not been demonstrated. Alternatively, transcribed V-region DNA might be in a special conformation that allows access to AID that is not associated with RNA pol. AID might be tethered to Ig-specific enhancers (or to cis elements specific to other genes that undergo SHM) and may thus be available in the vicinity of target DNA.

Within transcribed DNA, the first ~100 transcribed bp are essentially mutation-free, indicating that these bp are not deaminated by AID. There are several possible reasons for the lack of access by AID to the very 5' end. First, although

AID may interact with the transcription complex (34), it may do so only when RNA pol II is in the processive mode. In this case, AID may not be loaded at the promoter, but rather after the pol has switched from the abortive to the processive mode (51, 52). These two modes are distinguished by the phosphorylation state of the pol. In the abortive phase, the COOH-terminal domain of the largest subunit of RNA pol II is phosphorylated at serine-5 of the heptad repeats. Gradually, these heptad repeats acquire a different phosphorylation pattern, such that in the processive phase, serine-5 phosphorylation has been removed and serine-2 is phosphorylated. Although the switch to the processive mode likely begins before +100 bp from the transcription start, an AID-compatible phosphorylation state may not be present until after +100 bp. AID would be expected to access DNA that is associated with RNA pol, or, more likely, in negative supercoils arising behind the pol (see below). This scenario implies that AID travels with the elongating transcription complex.

Second, the lack of mutations in the upstream region might be due to a lack of negative supercoils in that region. Based on the possibility that AID gains access to DNA in negatively supercoiled regions accumulating behind the RNA pol (12), transcription-associated negative supercoils may need to reach a certain threshold level before AID can interact with the DNA. It is conceivable that negative supercoil propagation in the upstream direction is restricted to the transcribed portion of the gene due to occupancy of the promoter by multiple factors that anchor the gene in a particular position in the nucleus. This scenario could apply whether or not AID was associated with the transcription complex.

Finally, an “inhibitory” nucleosome might be positioned somewhere over the first +146 bp downstream of the promoter. Promoters are often flanked by positioned nucleosomes due to the occupancy of the promoter itself by transcription factors that may exclude nucleosomes (53). In this case, AID would begin to gain access to the DNA after the dyad center of the nucleosome (at  $\sim +73$  bp). Finding this result would mean that transcription can ignore such a nucleosome but AID cannot.

In contrast to the 5' end, the clear decrease of mutations beyond  $\sim 1$  kb, in  $\text{Ung}^-$  mice to the same degree as in  $\text{Ung}^+$  mice, is explained most easily by the hypothesis that AID is associated with the elongating transcription complex. AID may act stochastically on the DNA during elongation and, as a result of interacting with DNA, be unable to remain bound to the transcription complex. Alternatively, AID may need to be deposited on the DNA by a conformational change of the transcription complex, e.g., when the RNA pol pauses. If AID cannot reload after deposition, again, mutations would stochastically decrease in frequency toward the 3' end of the gene.

Alternative hypotheses for the lack of mutations in the 3' ends of Ig genes are as follows. First, AID might instead be tethered in the vicinity of the Ig genes, for example, at spe-

cific enhancers from where it prefers to target transcribed DNA without associating with the transcription complex. However, if so, why would AID mutations rise and fall in a relatively constant pattern from the 5' to the 3' region of the gene? When the intronic  $\text{Ig}\kappa$  enhancer was moved to a position 3' of the C region and an Ig promoter placed upstream of the C region, the V region was still mutated in the normal pattern (34). Second, chromatin differences between the V- and C-region DNA in heavy chain genes have been observed (54). However, in murine light chain genes, both the V and C regions were found to be in an active chromatin conformation in vivo (55). Thus, chromatin differences between V- and C-region DNA may not explain restriction of mutations to V-region DNA. Finally, it is possible that during SHM, certain rounds of transcription of the Ig genes abort somewhere in the middle of the gene. Clearly, full-length mRNAs are produced in  $\text{PNA}^{\text{high}}$  B cells and mutating B cell lines, but short transcripts may also be made, although they would be predicted to be unstable. (We have been unable to find evidence for short transcripts; unpublished data). For AID to associate only with abortive transcription complexes, the rounds of transcription resulting in short transcripts would have to be differentiable from rounds resulting in full-length transcripts. Thus, the original hypothesis (34, 56) that a mutator factor, i.e., AID, travels with the RNA pol during Ig gene transcription seems more likely. That is, if all bases in a length of DNA are equal targets for AID, but AID always begins acting from the same point (just downstream of the transcription start) and travels in only one direction, cytosines at the start of AID's path have a greater chance to become deaminated than those further downstream if AID is deposited at its site of action.

#### Ung and mismatch repair in SHM

A predominance of transition mutations from C and G is not only a hallmark of SHM in  $\text{Ung}^-$  mice, but also of mice lacking proteins of the mismatch repair system (16, 57–59). In addition, mutations at A and T bases are reduced in  $\text{Msh2}^-$  and  $\text{Msh6}^-$  deficient mice, leading to the proposal that both  $\text{Ung}$  and  $\text{Msh2}$  can recognize the U:G mismatch generated by AID, although the SHM pattern would be different depending on the DNA repair enzymes involved. That is,  $\text{Ung}$ -mediated uracil removal and subsequent replication over the abasic site, or nicking and removal of the abasic site followed by error-prone DNA synthesis, were proposed to generate mutations from C or G almost exclusively. In contrast,  $\text{Msh2}$ -mediated removal of a single-stranded DNA patch that extends beyond the original U:G mismatch could lead to mutations from any base, including A or T bases, during resynthesis of the removed strand. In this scenario,  $\text{Msh2}$ -mediated SHM would necessarily require error-prone synthesis.  $\text{pol}\eta$  has been implicated in this role, especially because (a) the pattern of synthesis errors made in vitro by  $\text{Pol}\eta$  (60), which appears to be stimulated by  $\text{Msh2}$  (61), resembles the SHM pattern from A and T, and (b) SHMs from A and T are re-

duced in Pol  $\eta$ -deficient humans and mice (21, 61). However, if considering Msh2 and Ung as independent paths to generate SHMs from a common U:G mismatch, the SHM patterns in mice lacking one of the repair systems are difficult to explain. For example, why are mutations from A or T in Ung<sup>-</sup> mice unaffected? One would expect them to be increased if, in the absence of Ung, the Msh2–Pol  $\eta$ /Pol  $\tau$  pathway predominates. Importantly, mismatch repair appears not to be limiting in germinal center B cells (62). Also, why are transitions from C and G increased in Msh2-deficient mice (57–59)? In the absence of Msh2 in this model, Ung would have been expected to gain more access, thereby still generating transversion mutations from C and G.

To explain this conundrum, we propose that Msh2 and Ung are factors in codependent systems in SHM. Msh2-mediated mismatch “repair” is expected not to be strand-specific for the U-containing strand, presumed here to be either the transcribed or the nontranscribed strand, because in Ung<sup>-</sup>/Msh2<sup>-</sup> double knockout mice, C to T and G to A mutations occur in roughly equal proportions (no strand bias for AID deamination; reference 16). As such, Msh2 and its cofactors are expected to remove either the U- or the G-containing DNA strand, as observed in *Escherichia coli* for a U:G mismatch in the absence of Ung (63). Removing the U-containing strand, followed by DNA synthesis to replace it, could result in mutations made by an error-prone pol, such as pol  $\eta$  or pol  $\tau$ . Removing the G-containing strand would result in a similar scenario, except that in addition, the nonexcised strand contains a uracil base (Fig. 5). Interestingly, in vitro, human Ung is able to excise U from single-stranded DNA 2.5-fold (and a predicted 40-fold in vivo) better than from double-stranded DNA (17, 64). Thus, the single-stranded region produced by mismatch repair might provide a preferred substrate for Ung. A corollary of this hy-

pothesis is that the ability of Ung to excise AID-deaminated cytosines is hampered by loss of Msh2. Ung inhibition results in an increase in C and G transitions, precisely what is seen in both Ung<sup>-</sup> and Msh2-deficient cells, and to a higher degree in the former as expected if Ung plays the major role in removing uracils. In Ung<sup>-</sup> mice, 94–98% of the mutations from C or G are transitions (15 and this paper), whereas in Msh2<sup>-</sup> mice, only 84–88% of mutations from C or G are transitions (16, 62). Thus, in the scenario described, all of the transition mutations from C and G observed in both Msh2- and Ung-deficient situations would be due to AID deamination events left unrepaired by Ung. Therefore, we expect that the great majority of C and G transitions mapped here in Ung<sup>-</sup> mice represent an accurate in vivo AID footprint.

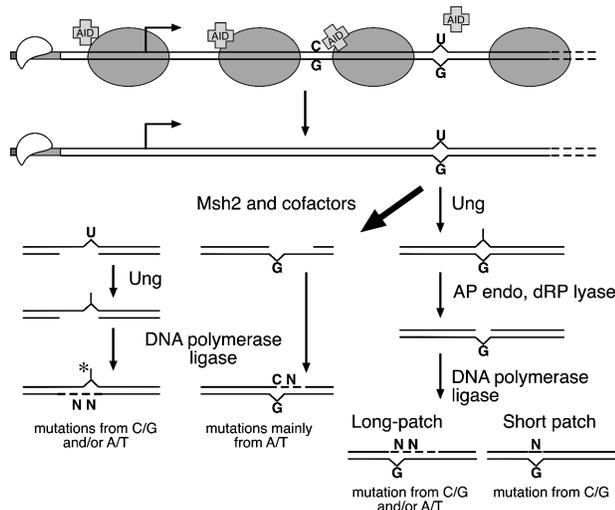
We found that the footprint of AID overlaps with the distribution of SHMs in Ung<sup>-</sup> mice overall. Interestingly, the overlap implies that during Msh2-mediated mismatch repair at a U:G mismatch in Ig V regions during SHM, the length of single-stranded DNA removed is only a few bp, compared with the estimated length of as much as 1 kb during mismatch repair outside of SHM (65). Removal of a long tract of DNA by Msh2 and its cofactors during SHM, initiated by an AID-generated mismatch, would have been expected to expand the distribution of SHMs beyond the AID footprint because error-prone pols could potentially participate in resynthesis of the excised strand anywhere along the length of the gap.

Thus, the interactions of various DNA repair systems and AID appear rather complex. Finding that AID does not gain access to the complete Ig gene elucidates one important aspect of regulating AID, a potentially hazardous mutator factor.

## MATERIALS AND METHODS

**Mice, immunizations, and cell sorting.** The Ung<sup>-</sup> mice on a hybrid C57BL/6J-129SV background (37) were provided by D. Barnes and T. Lindahl (Imperial Cancer Research Fund, Clare Hall Laboratories, UK) and maintained in our facility by breeding with C57BL/6J or RS transgenic mice, also on a C57BL/6J background. Ung<sup>-</sup> mice were genotyped by PCR using two primer sets: UngF (5' GTGAATGCAGGGCTCACTTAAGTC 3') and UngR2 (5' CAGTGCCTATAACTTCAGCTCC) produce a 493-bp product in Ung<sup>+</sup> mice and a 1.6-kb product in Ung<sup>-</sup> mice; NeoF (5' GTGAATGCAGGGCTCACTTAAGTC 3') and UngR2 produce an ~600-bp product only in Ung<sup>-</sup> mice. PCR conditions were 1 cycle at 96°C for 4 min, 96°C for 30 s, and 58–55°C for 30 s (touchdown annealing), 7 cycles at 72°C for 1.5 min, followed by 30 identical cycles except using a constant annealing temperature of 55°C, and a final extension cycle at 72°C for 7 min. The RS transgenic mice have been described previously (39). The RS  $\kappa$  transgene does not encode a functional protein due to a stop codon in the leader exon and is therefore not selectable.

To immunize mice,  $2 \times 10^8$  PBS-washed sheep red blood cells were injected intraperitoneally on days 1 and 22. B cells were isolated on day 25 from the spleen and Peyer's patches of immunized mice aged 4–10 mo. Cells were stained with anti-mouse B220/CD45-PE (BD Biosciences), anti-mouse PNA-FITC (Sigma-Aldrich), and anti-mouse GL7-FITC (BD Biosciences) antibodies and sorted on a Mo-Flo or FACSAria (BD Biosciences) cell sorter at the Immunology Core Facility at the University of Chicago. PNA<sup>low</sup>B220<sup>+</sup> and PNA<sup>high</sup>B220<sup>+</sup> cells were collected for DNA extraction using DNeasy columns (QIAGEN).



**Figure 5.** Model of codependence of Ung and mismatch repair activities in SHM.

The Animal Care and Use Committee of the University of Chicago has approved the animal studies.

**PCR amplification and sequencing.** To amplify a 991-bp segment of the Ig $\lambda$  gene with a V $\lambda$ 1-J $\lambda$ 1 rearrangement, primers were designed to hybridize at -453 bp upstream of the transcriptional start for the V $\lambda$ 1 gene (5' TCCCAGATTTAGACTCATTACTACTACACAC) and +538 bp downstream in the J $\lambda$ 1 region (5' CACGGACAGGATCCTAGGACAGTCAGTTTG). PCRs were performed with DNA template from 12,000 splenic PNA<sup>high</sup> B cells using Pfu pol (Stratagene) for 25 cycles at 98°C for 45 s, 54°C for 45 s, and 72°C for 2 min.

VJ558-rearranged IgH genes were amplified as described previously (33, 66) using ~10,000–20,000 cell equivalents of template DNA from splenic germinal center B cells, Pfu turbo pol (Stratagene), and PCR using 1 cycle at 95°C for 4 min, 95°C for 40 s, and 64–58°C for 40 s (touchdown annealing), 13 cycles at 72°C for 4 min, followed by 27 cycles at 95°C for 40 s, 57°C for 40 s, 72°C for 4 min, and a final extension at 72°C for 7 min. J1, J2, J3, and J4 rearrangements were resolved approximately by agarose gel electrophoresis, bands were excised, and DNA was purified using a gel extraction kit (Qiaquick; QIAGEN). The actual J rearrangement was ultimately revealed by DNA sequencing. Clones derived from hybrid PCR products identified by polymorphisms located throughout the region as analyzed by sequencing (33) were excluded.

To amplify a 1,265-bp segment of the RS transgene between -301 to +964 bp relative to the transcription start, the following primers were used: the forward primer (5' GTTACTTTTGTCTCCTTGTCATTACAGG), which anneals to 300 bp upstream of the transcription start in the RS transgene, and the reverse primer (5' TGTAGCACCTGTCCATGGTTAGCA), which anneals to the RS insert, a sequence that is unique to the RS transgene.

Two sets of primers were used to amplify the two downstream sections of the RS transgene from PNA<sup>high</sup> Peyer's patch B cells: JRH1F (5' GCAGATGTAGATTCAGGTGCT) and 950R (5' GCCCTCTCCATTTTCTCAA-GAT, unique to the RS transgene), which amplify a region between +141 and +1100 relative to the transcription start; and RsendF (5' CTACCA-GAATTCAGTTCTCACGTTC, unique to the RS transgene) and mIgkJCR (5' CTAATGGTTTGAACCACATGGGAC), which amplify the region between +950 and +2,982 nt relative to the transcription start. PCR was performed using 1 cycle at 95°C for 4 min, 27 cycles at 95°C for 30 s, 55°C for 30 s, and 72°C for 3 min, followed by 1 cycle at 72°C for 7 min.

For all DNA, PCR products were gel purified using the Qiaquick gel extraction kit and cloned using the Zero Blunt Topo PCR Cloning Kit (Invitrogen). Individual colonies were picked for automated DNA preparation and sequencing using M13F (5' GTAAAACGACGGCCAGT) and M13R (5' CACACAGGAAACAGCTATGACCAT) primers at the University of Chicago DNA Sequencing Core Facility. Additional sequencing primers used were J<sub>C</sub>intronRseq (5' CAGTTCTGAATAGGGTATGAGAGAGCC), which hybridizes to nt 2,120 to 2,146 (relative to GenBank/EMBL/DDBJ under accession no. X53774) between the end of J4 and the 3' primer used for PCR, which is complementary to nt 2,429–2,458, located 3' of the IgH intron enhancer (66); mIgkJcseqF (5' GTGGGATAGCAGAGTTGAGTGAGC); and mIgkJCR to sequence the furthest 3' portion of the RS transgene.

**Mutation analyses.** For all mutation analyses shown, all duplicate mutations were included in the totals to purposefully overestimate mutations found in the mutation-poor 5' and 3' regions of the genes examined. Nevertheless, we found only a few duplicated clones, and the data therefore are not significantly altered by retaining them (not depicted).

**Statistical analyses.** Simulations were performed using the R statistical software (R Foundation: <http://www.r-project.org>). To test the null hypothesis of a flat mutation rate, simulated datasets were created by generating random positions along the sequence. P-values were calculated as twice the percentage of simulations that had fewer mutations in the region investigated than in the collected data (e.g., a p-value of 0.0004 indicates that 4 simulated datasets out of 10,000 had a lower number of mutations in the region).

We thank B. Eisfelder, R. Duggan, and J. Marvin for skillful flow cytometric cell sorting, and W. Buikema and C. Hall for DNA sequencing. We thank T. Martin, S. Ratnam, P. Engler, and K. Padjen for critical reading of the manuscript.

The work was supported by National Institutes of Health grants AI47380 and AI053130. S. Longrich was supported by an AAUW International Ph.D. fellowship.

The authors have no conflicting financial interests.

Submitted: 8 August 2005

Accepted: 18 October 2005

## REFERENCES

- Muramatsu, M., K. Kinoshita, S. Fagarasan, S. Yamada, Y. Shinkai, and T. Honjo. 2000. Class switch recombination and somatic hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell*. 102:553–563.
- Revy, P., T. Muto, Y. Levy, F. Geissman, A. Plebani, O. Sanal, N. Catalan, M. Forveille, R. Dufourcq-Lagelouse, A. Gennery, et al. 2000. Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of hyper-IgM syndrome (HIGM2). *Cell*. 102:565–575.
- Begum, N.A., K. Kinoshita, M. Muramatsu, H. Nagaoka, R. Shinkura, and T. Honjo. 2004. De novo protein synthesis is required for activation-induced cytidine deaminase-dependent DNA cleavage in immunoglobulin class switch recombination. *Proc. Natl. Acad. Sci. USA*. 101:13003–13007.
- Muramatsu, M., V. Sankaranand, S. Anant, M. Sugai, K. Kinoshita, N. Davidson, and T. Honjo. 1999. Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. *J. Biol. Chem.* 274:18470–18476.
- Bransteitter, R., P. Pham, M. Scharff, and M. Goodman. 2003. Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. *Proc. Natl. Acad. Sci. USA*. 100:4102–4107.
- Dickerson, S., E. Market, E. Besmer, and F.N. Papavasiliou. 2003. AID mediates hypermutation by deaminating single stranded DNA. *J. Exp. Med.* 197:1291–1296.
- Pham, P., R. Bransteitter, J. Petruska, and M. Goodman. 2003. Processive AID-catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation. *Nature*. 424:103–107.
- Sohail, A., J. Klapacz, M. Samaranyake, A. Ullah, and A. Bhagwat. 2003. Human activation-induced cytidine deaminase causes transcription-dependent, strand-biased C to U deaminations. *Nucleic Acids Res.* 31:2990–2994.
- Chaudhuri, J., M. Tian, C. Khuong, K. Chua, E. Pinaud, and F.W. Alt. 2003. Transcription-targeted DNA deamination by the AID antibody diversification enzyme. *Nature*. 422:726–730.
- Chaudhuri, J., C. Khuong, and F.W. Alt. 2004. Replication protein A interacts with AID to promote deamination of somatic hypermutation targets. *Nature*. 430:992–998.
- Ramiro, A., P. Stavropoulos, M. Jankovic, and M. Nussenzweig. 2003. Transcription enhances AID-mediated cytidine deamination by exposing single-stranded DNA on nontemplate strand. *Nat. Immunol.* 4:452–456.
- Shen, H., and U. Storb. 2004. Activation-induced cytidine deaminase (AID) can target both DNA strands when the DNA is supercoiled. *Proc. Natl. Acad. Sci. USA*. 101:12997–13002.
- Petersen-Mahrt, S.K., R.S. Harris, and M.S. Neuberger. 2002. AID mutates E. coli suggesting a DNA deamination mechanism for antibody diversification. *Nature*. 418:99–103.
- Di Noia, J., and M.S. Neuberger. 2002. Altering the pathway of immunoglobulin hypermutation by inhibiting uracil-DNA glycosylase. *Nature*. 419:43–48.
- Rada, C., G. Williams, H. Nilsen, D. Barnes, T. Lindahl, and M. Neuberger. 2002. Immunoglobulin isotype switching is inhibited and somatic hypermutation perturbed in UNG-deficient mice. *Curr. Biol.* 12:1748–1755.
- Rada, C., J. Di Noia, and M. Neuberger. 2004. Mismatch recognition and uracil-excision provide complementary paths to both immunoglobulin switching and the second (dA:dT-focused) phase of somatic mu-

- tation. *Mol. Cell.* 16:163–171.
17. Kavli, B., S. Andersen, M. Otterlei, N. Liabakk, K. Imai, A. Fischer, A. Diurandy, H. Krokan, and G. Slupphaug. 2005. B cells from hyper-IgM patients carrying UNG mutations lack ability to remove uracil from ssDNA and have elevated genomic uracil. *J. Exp. Med.* 201: 2011–2021.
  18. Diaz, M., and C. Lawrence. 2005. An update on the role of translesion synthesis DNA polymerases in Ig hypermutation. *Trends Immunol.* 26: 215–220.
  19. Storb, U., and J. Stavnezer. 2002. Immunoglobulin genes: generating diversity with AID and UNG. *Curr. Biol.* 12:R725–R727.
  20. Zeng, X., D. Winter, C. Kasmer, K. Kraemer, A. Lehmann, and P. Gearhart. 2001. DNA polymerase eta is an A-T mutator in somatic hypermutation of immunoglobulin variable genes. *Nat. Immunol.* 2:537–541.
  21. Delbos, F., A. De Smet, A. Faili, S. Aoufouchi, J.-C. Weill, and C.-A. Reynaud. 2005. Contribution of DNA polymerase  $\eta$  to immunoglobulin gene hypermutation in the mouse. *J. Exp. Med.* 201:1191–1196.
  22. Shen, H., A. Peters, B. Baron, X. Zhu, and U. Storb. 1998. Mutation of *BCL-6* gene in normal B cells by the process of somatic hypermutation of Ig genes. *Science.* 280:1750–1752.
  23. Pasqualucci, L., A. Migliazza, N. Fracchiolla, C. William, A. Neri, L. Baldini, R. Chaganti, U. Klein, R. Kueppers, K. Rajewsky, and R. Dalla-Favera. 1998. *BCL-6* mutations in normal germinal center B cells: evidence of somatic hypermutation acting outside Ig loci. *Proc. Natl. Acad. Sci. USA.* 95:11816–11821.
  24. Gordon, M., C. Kanegai, J. Doerr, and R. Wall. 2003. Somatic hypermutation of the B cell receptor genes B29 (Ig-beta, CD79b) and mb1 (Ig-alpha, CD79a). *Proc. Natl. Acad. Sci. USA.* 100:4126–4131.
  25. Müschen, M., D. Re, B. Jungnickel, V. Diehl, K. Rajewsky, and R. Kueppers. 2000. Somatic mutation of the CD95 gene in human B cells as a side-effect of the germinal center reaction. *J. Exp. Med.* 192:1833–1839.
  26. Betz, A., C. Milstein, R. Gonzalez-Fernandes, R. Pannell, T. Larson, and M. Neuberger. 1994. Elements regulating somatic hypermutation of an immunoglobulin K gene: critical role for the intron enhancer/matrix attachment region. *Cell.* 77:239–248.
  27. Bachl, J., C. Carlson, V. Gray-Schopfer, M. Dessing, and C. Olsson. 2001. Increased transcription levels induce higher mutation rates in a hypermutating cell line. *J. Immunol.* 166:5051–5057.
  28. Motoyama, N., T. Miwa, Y. Suzuki, H. Okada, and T. Azuma. 1994. Comparison of somatic mutation frequency among immunoglobulin genes. *J. Exp. Med.* 179:395–403.
  29. Rada, C., A. Gonzalez-Fernandez, J.M. Jarvis, and C. Milstein. 1994. The 5' boundary of somatic hypermutation in a V $\kappa$  gene is in the leader intron. *Eur. J. Immunol.* 24:1453–1457.
  30. Rogerson, B. 1994. Mapping the upstream boundary of somatic mutations in rearranged immunoglobulin transgenes and endogenous genes. *Mol. Immunol.* 31:83–98.
  31. Wu, P., and L. Clafin. 1998. Promoter-associated displacement of hypermutations. *Int. Immunol.* 10:1131–1138.
  32. Lebecque, S., and P. Gearhart. 1990. Boundaries of somatic mutation in rearranged immunoglobulin genes: 5' boundary is near the promoter, and 3' boundary is  $\sim$ 1 kb from V(D)J gene. *J. Exp. Med.* 172: 1717–1727.
  33. Rada, C., and C. Milstein. 2001. The intrinsic hypermutability of antibody heavy and light chain genes decays exponentially. *EMBO J.* 20: 4570–4576.
  34. Peters, A., and U. Storb. 1996. Somatic hypermutation of immunoglobulin genes is linked to transcription initiation. *Immunity.* 4:57–65.
  35. Storb, U., A. Peters, E. Klotz, N. Kim, H.M. Shen, J. Hackett, B. Rogerson, and T.E. Martin. 1998. Cis-acting sequences that affect somatic hypermutation of Ig genes. *Immunol. Rev.* 162:153–160.
  36. Liu, L.F., and J. Wang. 1987. Supercoiling of the DNA template during transcription. *Proc. Natl. Acad. Sci. USA.* 84:7024–7027.
  37. Nilsen, H., I. Rosewell, P. Robins, C. Skjelbred, S. Andersen, G. Slupphaug, G. Daly, H. Krokan, T. Lindahl, and D. Barnes. 2000. Uracil-DNA glycosylase (UNG)-deficient mice reveal a primary role of the enzyme during DNA replication. *Mol. Cell.* 5:1059–1065.
  38. An, Q., P. Robins, T. Lindahl, and D.E. Barnes. 2005. C to T mutagenesis and gamma-radiation sensitivity due to deficiency in the Smug1 and Ung DNA glycosylases. *EMBO J.* 24:2205–2213.
  39. Michael, N., T.E. Martin, D. Nicolae, N. Kim, K. Padjen, P. Zhan, H. Nguyen, C. Pinkert, and U. Storb. 2002. Effects of sequence and structure on the hypermutability of immunoglobulin genes. *Immunity.* 16:123–134.
  40. Faili, A., S. Aoufouchi, E. Flatter, Q. Gueranger, C.A. Reynaud, and J.C. Weill. 2002. Induction of somatic hypermutation in immunoglobulin genes is dependent on DNA polymerase iota. *Nature.* 419:944–947.
  41. Kunkel, T., Y. Pavlov, and K. Bebenek. 2003. Functions of human DNA polymerases eta, kappa, and iota. *DNA Repair (Amst.)* 2:135–149.
  42. Tissier, A., J.P. McDonald, E.G. Frank, and R. Woodgate. 2000. poliota, a remarkably error-prone human DNA polymerase. *Genes Dev.* 14:1642–1650.
  43. Storb, U., D. Haasch, B. Arp, P. Sanchez, P. Cazenave, and J. Miller. 1989. Physical linkage of mouse I genes by pulsed-field gel electrophoresis suggests that the rearrangement process favors proximate target sequences. *Mol. Cell. Biol.* 9:711–718.
  44. Sun, T., and U. Storb. 2001. Insertion of phosphoglycerine kinase (PGK)-neo 5' of J $\lambda$ 1 dramatically enhances VJ $\lambda$ 1 rearrangement. *J. Exp. Med.* 193:699–712.
  45. Fuentes-Panana, E.M., G. Bannish, and J.G. Monroe. 2004. Basal B-cell receptor signaling in B lymphocytes: mechanisms of regulation and role in positive selection, differentiation, and peripheral survival. *Immunol. Rev.* 197:26–40.
  46. Michael, N., H. Shen, S. Longereich, N. Kim, A. Longacre, and U. Storb. 2003. The E box motif CAGGTG enhances somatic hypermutation without enhancing transcription. *Immunity.* 19:235–242.
  47. Shen, H.M., A. Peters, D. Kao, and U. Storb. 2001. The 3' I $\kappa$  enhancer contains RNA polymerase II promoters: implications for endogenous and transgenic kappa gene expression. *Int. Immunol.* 13:665–674.
  48. Storb, U., H.M. Shen, N. Michael, and N. Kim. 2001. Somatic hypermutation of immunoglobulin and non-immunoglobulin genes. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 356:13–19.
  49. Barreto, V.M., A.R. Ramiro, and M.C. Nussenzweig. 2005. Activation-induced deaminase: controversies and open questions. *Trends Immunol.* 26:90–96.
  50. Nambu, Y., M. Sugai, H. Gonda, C. Lee, T. Katakai, Y. Agata, Y. Yokota, and A. Shimizu. 2003. Transcription-coupled events associating with immunoglobulin switch region chromatin. *Science.* 302:2137–2140.
  51. Cho, E., M. Kobor, M. Kim, J. Greenblatt, and S. Buratowski. 2001. Opposing effects of Ctk1 kinase and Fcp1 phosphatase at ser2 of RNA polymerase II C-terminal domain. *Genes Dev.* 15:3319–3329.
  52. Price, D. 2000. P-TEFb, a cyclin-dependent kinase controlling elongation by RNA polymerase II. *Mol. Cell. Biol.* 20:2629–2634.
  53. Reinke, H., and W. Hoerz. 2003. Histones are first hyperacetylated and then lose contact with the activated PHO5 promoter. *Mol. Cell.* 11:1599–1607.
  54. Woo, C.J., A. Martin, and M.D. Scharff. 2003. Induction of somatic hypermutation is associated with modifications in immunoglobulin variable region chromatin. *Immunity.* 19:479–489.
  55. Odegard, V.H., S.T. Kim, S.M. Anderson, M. Shlomchik, and D. Schatz. 2005. Histone modifications associated with somatic hypermutation. *Immunity.* 23:101–110.
  56. Storb, U., A. Peters, E. Klotz, N. Kim, H.M. Shen, K. Kage, and B. Rogerson. 1998. Somatic hypermutation of immunoglobulin genes is linked to transcription. *Curr. Top. Microbiol. Immunol.* 229:11–19.
  57. Rada, C., M. Ehrenstein, M. Neuberger, and C. Milstein. 1998. Hot spot focusing of somatic hypermutation in MSH2-deficient mice suggests two stages of mutational targeting. *Immunity.* 9:135–141.
  58. Frey, S., B. Bertocci, F. Delbos, L. Quint, J. Weill, and C. Reynaud. 1998. Mismatch repair deficiency interferes with the accumulation of mutations in chronically stimulated B cells and not with the hypermutation process. *Immunity.* 9:127–134.
  59. Phung, Q., D. Winter, A. Cranston, R. Tarone, W. Bohr, R. Fishel, and P. Gearhart. 1998. Increased hypermutation at G and C nucleotides in immunoglobulin variable genes from mice deficient for the MSH2 mismatch repair protein. *J. Exp. Med.* 187:1745–1751.

60. Rogozin, I., Y. Pavlov, K. Bebenek, T. Matsuda, and T. Kunkel. 2001. Somatic mutation hotspots correlate with DNA polymerase  $\epsilon$  error spectrum. *Nat. Immunol.* 2:530–536.
61. Martomo, S.A., W.W. Yang, R.P. Wersto, T. Ohkumo, Y. Kondo, M. Yokoi, C. Masutani, F. Hanaoka, and P.J. Gearhart. 2005. Different mutation signatures in DNA polymerase  $\epsilon$ - and MSH6-deficient mice suggest separate roles in antibody diversification. *Proc. Natl. Acad. Sci. USA.* 102:8656–8661.
62. Bertocci, B., L. Quint, F. Delbos, C. Garcia, C.-A. Reynaud, and J.-C. Weill. 1998. Probing Ig gene hypermutation with microsatellites suggests a nonreplicative short patch DNA synthesis process. *Immunity.* 9:257–265.
63. Shenoy, S., K.C. Ehrlich, and M. Ehrlich. 1987. Repair of thymine.guanine and uracil.guanine mismatched base-pairs in bacteriophage M13mp18 DNA heteroduplexes. *J. Mol. Biol.* 197:617–626.
64. Eftedal, I., P.H. Guddal, G. Slupphaug, G. Volden, and H.E. Krokan. 1993. Consensus sequences for good and poor removal of uracil from double stranded DNA by uracil-DNA glycosylase. *Nucleic Acids Res.* 21:2095–2101.
65. Friedberg, E., G. Walker, and W. Siede. 1995. DNA Repair and Mutagenesis. A.S.F. Microbiology, Washington, D.C. 698 pp.
66. Jolly, C.J., N. Klix, and M.S. Neuberger. 1997. Rapid methods for the analysis of Ig gene hypermutation: application to transgenic and gene targeted mice. *Nucleic Acids Res.* 25:1913–1919.