Attracting AID to targets of somatic hypermutation

Atsushi Tanaka,[1] Hong Ming Shen,[2] Sarayu Ratnam,[2] Prashant Kodgire,[2] and Ursula Storb[1,2]

[1]Committee on Immunology and [2]Department of Molecular Genetics and Cell Biology, University of Chicago, Chicago, IL 60637

The process of somatic hypermutation (SHM) of immunoglobulin (Ig) genes requires activation–induced cytidine deaminase (AID). Although mistargeting of AID is detrimental to genome integrity, the mechanism and the cis–elements responsible for targeting of AID are largely unknown. We show that three CAGGTG cis–elements in the context of Ig enhancers are sufficient to target SHM to a nearby transcribed gene. The CAGGTG motif binds E47 in nuclear extracts of the mutating cells. Replacing CAGGTG with AAGGTG in the construct without any other E47 binding site eliminates SHM. The CA versus AA effect requires AID. CAGGTG does not enhance transcription, chromatin acetylation, or overall target gene activity. The other cis–elements of Ig enhancers alone cannot attract the SHM machinery. Collectively with other recent findings, we postulate that AID targets all genes expressed in mutating B cells that are associated with CAGGTG motifs in the appropriate context. Ig genes are the most highly mutated genes, presumably because of multiple CAGGTG motifs within the Ig genes, high transcription activity, and the presence of other cooperating elements in Ig enhancers.

CORRESPONDENCE
Ursula Storb:
stor@uchicago.edu

Abbreviations used: AID, activation-induced cytidine de-aminase; ChIP, chromatin immu-noprecipitation; eGFP, enhanced GFP; MMR, mismatch repair; SHM, somatic hypermutation.

The immune system has evolved effective mechanisms using a limited number of Ig genes to defend against a plethora of pathogens. To produce a diverse repertoire of antibodies, B cells undergo a series of genetic alterations to create a much greater variety of antibodies than specified by the number of Ig genes. Somatic hypermutation (SHM) is one of the diversification processes in activated B cells. SHM requires the activity of a mutation factor, activation-induced cytosine de-aminase (AID; Muramatsu et al. 2000), and tran-scription of the target gene (Storb et al., 2001; Barreto et al., 2005). In many studies, the mu-tation frequencies of Ig genes were found to correlate positively with the level of transcrip-tion (Bachl et al., 2001). The frequency and the range of mutations depended on the distance of the targeted sequences from the promoter (Lebecque and Gearhart, 1990; Motoyama et al., 1994; Rada et al., 1994; Rogerson, 1994; Wu and Claflin, 1998; Rada and Milstein, 2001), and initiation of transcription inside an Ig gene induced SHM at the distal constant region that is normally unmutated (Peters and Storb, 1996). SHM targets Ig genes at high rates and several other non-Ig genes, such as BCL6, at interme-diate rates (Pasqualucci et al., 1998; Shen et al., 1998; Müschen et al., 2000; Gordon et al., 2003),

and may target all other transcribed genes at very low frequencies (Liu et al., 2008).

The specific association of AID with the mu-table target genes is likely to be regulated by cis-elements within or near the target genes. The Ig enhancers are required for SHM in endogenous Ig genes. Deletion of either intronic or 3′ en-hancer regions in Ig transgenes eliminated or drastically decreased SHM (Betz et al., 1994). The Ig enhancers comprise multiple cis-elements, all of which also occur in other enhancers in one combination or another. A previous study in our laboratory showed that the presence of two CAGGTG motifs in addition to the four CAG-GTG and eight CAGCTG motifs already pres-ent in an Igκ transgene greatly enhanced the frequency of SHM without increasing transcrip-tion (Michael et al., 2003). The CAGGTG motif is present in Ig heavy and light chain enhancers that are shown to be required for Ig expression and SHM, as well as in all of the frequent non-Ig targets of SHM, such as BCL6 (Pasqualucci et al., 1998; Shen et al., 1998). Moreover, T cell

lymphomas from mice with overexpressed transgenic AID indicated that all of the mutated genes found in these tumors shared the CAGGTG motif in the enhancer and/or promoter (Kotani et al., 2005).

These findings showed that CAGGTG was an enhancer of SHM but did not test whether CAGGTG is sufficient/required to attract AID to a nearby gene. To determine whether the CAGGTG motif is sufficient/required for SHM targeting, we generated mutable GFP transgenes that contained either three CAGGTG motifs or no CAGGTG motif in the entire construct. Our findings show that SHM occurs with otherwise identical transgenes only when they contain CAGGTG motifs.

RESULTS
Transgenic DT40 cell lines
To address whether the presence of the CAGGTG motif is required for targeting of SHM, we generated a mutable enhanced GFP (eGFP) transgene containing three CAGGTG motifs (C–GFP) and a control transgene containing AAGGTG (A–GFP) but completely lacking the CAGGTG motif. The transgenes have GFP and neomycin resistance genes driven by individual CMV promoters (Fig. 1).

Because SHM in an Ig transgene context requires both an intronic enhancer/matrix attachment region and a 3′ enhancer (Betz et al., 1994), we used enhancers from a mouse Igκ gene. These enhancers contain one CAGGTG motif in the intronic enhancer and two CACCTG motifs in the 3′ enhancer, which are the reverse complements of CAGGTG. In the A–GFP transgene, these CAGGTG motifs were changed to AAGGTG or AACCTG, motifs that showed no enhancement of SHM frequency when present in an Ig transgene (Michael et al., 2003). Because one of the binding factors known to bind the CAGGTG motif is an E47 homodimer, a predominant form of E protein in the B cell lineage, we also considered much weaker binding motifs of the E47 homodimer, such as CAGCTG and CACGTG (Hsu et al., 1994). To eliminate the possibility of E47 homodimer binding to CAGCTG, which is present in the neomycin resistance gene, we changed this CAGCTG to CCGCTG. This change does not alter the amino acid coding, and thus, neomycin resistance was preserved. This makes the C–GFP construct transgene have only three CAGGTG motifs in the entire construct and no other CAG-GTG, CACCTG, CAGCTG, or CACGTG motifs, whereas the A–GFP transgene is completely devoid of these motifs.

To monitor SHM targeting, we inserted a premature stop codon in the GFP gene and a 300-bp spacer from the mutable region of the human BCL6 gene to position this stop codon at the peak of the mutable region observed in Ig genes. The spacer from BCL6 does not contain any CAG-GTG sites. We chose it because it was known not to interfere with continuous SHM in its vicinity. This seemed important because we had shown that a transcriptional promoter placed in front of the C region can initiate a new wave of mutations (Peters and Storb,1996), and we suspected that anything inhibiting transcription elongation may affect the progress of SHM further 3′. For cells transfected with these transgenes to express functional GFP, they must mutate the stop codon. Because this stop codon is also part of an AID hotspot (GYW), it should have a reasonable chance of being mutated by AID.

Both the C–GFP and A–GFP constructs were transfected into an actively mutating chicken B cell line, DT40 CL18 (Sale et al., 2001), and stably transfected clones were selected. To verify SHM in this study, GFP fluorescence must be linked directly to a mutation of the premature stop codon by DNA sequencing. In cells where multiple copies of the transgene are integrated, mutation of a premature stop codon in one of the transgene copies may escape detection by PCR cloning and sequencing analyses. Therefore, for accurate assessment of SHM activity, independent C–GFP and A–GFP clones with only a single copy of the respective transgene were selected by Southern blot analyses (unpublished data).

Because SHM depends on transcription levels as well as transcription initiation (Peters and Storb, 1996), we further selected clones with equivalent levels of transgene transcription. To assess the levels of stable transcripts as a measure of
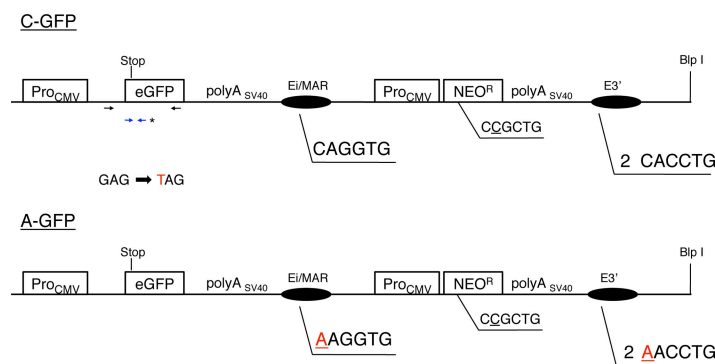
Figure 1. C–GFP and A–GFP transgenes. The C-GFP transgene contains three CAGGTG (CACCTG) motifs and the A-GFP transgene is completely devoid of the CAGGTG motif. The individual components are CMV promoter, eGFP, splice and poly A signals from SV40, intronic enhancer and matrix attachment regions from the mouse Ig κ gene, neomycin resistance gene driven by the CMV promoter, and 3′ enhancer from the mouse Ig κ gene. The Blp I site was used to linearize the plasmid. The primer pair used for sequencing is indicated with black arrows; the primer pair used for the ChIP assay is shown with blue arrows with an asterisk.

transcription of the transgenes, we measured GFP mRNA levels by quantitative real-time PCR. Five C–GFP clones and five A–GFP clones were selected that expressed very similar levels of GFP mRNA (P > 0.1; Fig. 2).

The transgenes with C–GFP, but not those with A–GFP, are mutable in all transfectants

From each transgenic clone, 12 subclones were derived by single-cell sorting of GFP⁻IgM⁺ cells. IgM⁺ cells were chosen to eliminate potential differences in growth efficiency between IgM⁺ and IgM⁻ cell cultures. To determine the effect of the CAGGTG motif on SHM targeting, C–GFP and A–GFP subclones were cultured individually and GFP fluorescence was measured by flow cytometry weekly for 6 wk. When assayed in this manner, all C–GFP transgenic clones showed GFP⁺ subclones and, overall, 54 out of 60 subclones showed GFP⁺ cells. On the other hand, only two of the AAGGTG clones showed any significant numbers of GFP⁺ cells and, overall, only 18 out of 60 AAGGTG subclones showed any GFP⁺ cells during the 6-wk-long experiment (Fig. 3). The differences of observed GFP⁺ cells between C–GFP subclones and A–GFP subclones were highly significant (P < 0.00001). The mean frequencies of GFP⁺ cells in 10^6 cells were also higher in C–GFP clones, ranging between 20 and 60, whereas A–GFP clones were consistently <10 GFP⁺ cells/10^6 cells (Fig. 4 B).

There were many C–GFP subclones with >100 GFP⁺ cells/10^6 cells at each week (22 C–GFP subclones; Fig. 3, green), and some had high numbers of GFP⁺ cells for 4–5 wk. In contrast, A–GFP subclones showed only a few subclones with high numbers of GFP⁺ cells (four A–GFP subclones). To verify expression of functional GFP by C–GFP and A–GFP clones, subclones with a high number of GFP⁺ cells were sorted and cultured for 1 wk. Indeed, FACS analyses showed that most of the GFP⁺ cells in C–GFP clones could be cultured (Fig. 4 A), whereas only the GFP⁺ cells of one A–GFP subclone (A2) survived (the others may have been autofluorescent dying cells).

Because the levels of transcripts of the C–GFP and A–GFP transgenes were equivalent in the 10 clones (Fig. 2), this eliminates the possibility that the difference in the occurrence of SHM was caused by varied transcription levels of the target genes. Furthermore, the AID mRNA levels were also equivalent among all 10 clones (P > 0.1; Fig. 2), indicating that the enhanced mutation frequencies of the C–GFP clones are independent of variability in AID expression.

The GFP premature stop codon is mutated

To analyze GFP sequences for mutations, sorted GFP⁺ cells from C–GFP and A–GFP subclones were cultured for 1 wk before DNA cloning and sequencing (Fig. 4 A). Although there were a few A–GFP clones with GFP⁺ cells, the A2–4 subclone (Fig. 3) was the only A–GFP subclone from which we were able to sort enough viable GFP⁺ cells to culture and expand for genomic DNA isolation. Sequencing analyses of the GFP region of C–GFP clones showed that 86% of the sequences contained a mutation at the premature stop codon (121 out of 141 sequences; Fig. 5). These were mostly G-to-C transversions at the third nucleotide of the stop codon. The SHM frequency of the GFP-expressing C–GFP clones was 1.38×10^{-3} mutations/bp. This is comparable to the SHM frequency of 3.05×10^{-3} in DT40 cells selected for loss of membrane Ig because of mutations of IgL genes (Arakawa et al., 2004). Because the latter cells overexpressed AID, the mutation frequency would be expected to be somewhat higher than in the cells we used that produce limiting amounts of AID (Arakawa et al., 2004). In accord with the SHM pattern of the IgL gene in DT40 cells, transversions were frequent.

The mutated stop codon by G-to-C transversion, TAC, is yet another hotspot for AID. Because there were several sorted GFP⁺ cells containing mutations in GFP but not in the stop codon, a fraction of the sorted cells apparently reverted TAC back to the TAG stop codon while the green cells were cultured (Fig. 4 A and Fig. 5). In addition, cells lost the green color, presumably because of internal GFP mutations. There are AID hotspots in the chromophore itself (Thr-Tyr-Gly) as well as in the rest of the GFP protein. The entire 27-kD polypeptide structure is essential for the development and maintenance of fluorescence (Olympus FluoView Resource Center at http://www.olympusfluoview.com/). Thus, many sites in
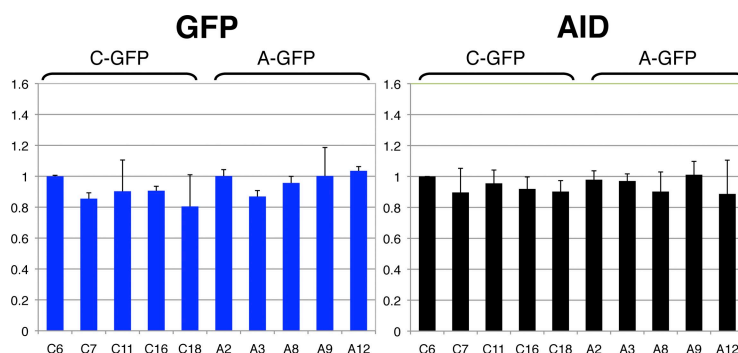

Figure 2. Relative mRNA expression of C–GFP and A–GFP transgenes. Relative mRNA levels among five single-copy clones for eGFP and AID from C–GFP and A–GFP lines were measured by real-time PCR. The data show the mean of three independent experiments. β-Actin mRNA was used to normalize eGFP and AID mRNA levels. The y axis shows mRNA levels relative to clone C6.

GFP must remain unchanged for green color but are susceptible to inactivating changes by SHM. We also sequenced the GFP transgene of unsorted cells after 6 wk of culture and found a few mutations, but none at the stop codon.

The GFP mutations depend on AID

To determine whether C-GFP transgenes were mutated by an AID-dependent mechanism, C-GFP and A-GFP constructs were transfected into AID$^{-/-}$ Cre DT40 or the parental clone, AID$^+$ Cre DT40, and six independent clones were selected for each of the four combinations (Fig. 6). There were no significant differences in GFP mRNA levels between AID$^{-/-}$ and AID$^+$ cells (Fig. S1). The AID$^+$ but not the AID$^{-/-}$ cells expressed AID mRNA (Fig. S1). The flow cytometry assay of GFP showed no GFP$^+$ cells in the AID$^{-/-}$ background in both C-GFP and A-GFP clones. However, in the AID$^+$ cells, mutations were significant for C-GFP clones. Out of the six A-GFP clones, only two showed sporadic, very low numbers of GFP$^+$ cells. We assume that, as in the experiments shown in Fig. 3, in the A-GFP clones of Fig. 6, the transgene was inserted near CAGGTG sites (see CAG-GTG motifs near... and Fig. 10). The data indicate that the

mutability differences of C-GFP and A-GFP transgenes are dependent on AID-mediated SHM (Fig. 6).

No correlation between mutation frequency and DNase I hypersensitive sites

Between rearranged Ig alleles, which are mutable, and unrearranged Ig alleles, there ware dramatic differences in DNase I hypersensitive sites (Storb et al., 1986; Thompson and Neiman, 1987). To investigate factors leading to differences in C-GFP and A-GFP mutation frequencies, the overall accessibility of DNA between C-GFP and A-GFP clones within the transgene-specific region was examined by treating nuclei with DNase I (Fig. 7). The digestibility of the two C-GFP clones appears greater than that of the A-GFP clones (Fig. 7 A). However, the ethidium bromide–stained gel showed that there was less DNA in the C-GFP clones, and overall, the DNA was more digested with lower amounts of DNA (Fig. 7 B). Thus, the digestability is equal. This is also supported by re-probing the Southern blot with an ovalbumin probe: the ovalbumin gene is also more digested in the C-GFP clones (Fig. 7 C). The ovalbumin gene is not expressed in DT40; therefore, one does not see DNase I hypersensitive sites but an overall

C-GFP

	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6
C6-1	0.0061	0	0	0.0041	0	0
C6-2	0.0056	0	0.016	0.021	0.024	0.017
C6-3	0	0.004	0.009	0.018	0	0
C6-4	0	0	0	0	0	0
C6-5	0	0	0	0	0.0035	0.004
C6-6	0	0	0	0	0	0.015
C6-7	0.0031	0	0	0	0	0.0079
C6-8	0	0	0	0.0042	0.0068	0
C6-9	0	0	0	0	0.007	0
C6-10	0	0	0	0	0	0.0038
C6-11	0	0	0	0	0.0035	0.0079
C6-12	0.0045	0	0	0	0	0
C7-1	0	0	0	0.025	0.0071	0
C7-2	0	0.0038	0.0059	0	0.0034	0.0041
C7-3	0	0.0036	0.0055	0	0.0032	0.0041
C7-4	0	0	0.0059	0	0	0
C7-5	0.0049	0	0.015	0.0043	0	0.0039
C7-6	0	0	0	0	0	0.0039
C7-7	0	0	0.015	0.0041	0.0073	0.085
C7-8	0	0	0.0058	0	0	0
C7-9	0	0	0	0	0	0.026
C7-10	0	0.026	0.032	0.034	0.061	0.048
C7-11	0	0.0071	0.0088	0	0.003	0
C7-12	0.0056	0.024	0.024	0.03	0.02	0.0084
C11-1	0.011	0	0	0	0	0
C11-2	0	0.0028	0	0.017	0.0067	0.0038
C11-3	0	0	0.0062	0	0	0
C11-4	0.014	0	0.0059	0	0	0.0084
C11-5	0	0	0	0	0	0
C11-6	0	0	0	0.0041	0	0
C11-7	0	0	0	0	0	0.0038
C11-8	0	0	0	0	0	0
C11-9	0.006	0	0.003	0	0	0
C11-10	0.016	0	0	0	0	0
C11-11	0	0	0	0	0.0031	0
C11-12	0	0.0035	0	0	0.003	0.004
C16-1	0.019	0	0.0092	0.0042	0.0033	0.02
C16-2	0	0	0	0	0	0
C16-3	0	0	0.0092	0.0042	0	0
C16-4	0	0.024	0.013	0.0043	0.049	0.0081
C16-5	0	0	0	0	0	0.0081
C16-6	0	0	0	0.004	0	0
C16-7	0.0097	0	0	0.0028	0.013	0.0059
C16-8	0	0	0	0.0042	0	0
C16-9	0.0052	0	0	0	0	0
C16-10	0.0042	0	0	0	0	0
C16-11	0	0.0037	0.0031	0.0041	0	0
C16-12	0	0	0	0	0	0
C18-1	0	0	0.0033	0	0	0.0041
C18-2	0	0	0	0	0.023	0.0084
C18-3	0	0	0.0039	0.0094	0.0041	0
C18-4	0.01	0	0	0	0	0.0043
C18-5	0	0.016	0.0031	0	0	0
C18-6	0.005	0	0	0	0	0.0085
C18-7	0	0	0	0.0031	0	0
C18-8	0	0	0.0032	0.0042	0.013	0
C18-9	0	0	0.003	0.0041	0.017	0.016
C18-10	0	0	0	0.0041	0.0031	0.029
C18-11	0	0	0	0	0	0
C18-12	0.0049	0.0037	0	0.0043	0	0

A-GFP

	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6
A2-1	0	0	0	0	0	0
A2-2	0	0	0	0	0	0
A2-3	0	0	0	0	0	0
A2-4	0	0.0092	0.01	0.021	0.0072	0.008
A2-5	0	0	0	0	0	0
A2-6	0	0	0.0096	0.0042	0	0.0038
A2-7	0	0	0	0	0	0.0042
A2-8	0	0	0	0	0	0
A2-9	0	0	0	0	0	0.0041
A2-10	0	0	0	0	0	0
A2-11	0	0	0	0.0045	0	0
A2-12	0	0	0.0031	0	0	0
A3-1	0	0	0	0	0	0
A3-2	0	0	0	0	0	0
A3-3	0	0.005	0	0.0046	0	0
A3-4	0	0	0	0	0	0
A3-5	0	0	0	0	0	0
A3-6	0	0	0	0	0	0
A3-7	0	0	0	0	0	0
A3-8	0	0	0	0	0	0
A3-9	0	0	0	0.0045	0	0
A3-10	0	0	0	0	0	0
A3-11	0	0	0	0	0	0
A3-12	0	0	0	0	0	0
A8-1	0	0	0	0	0	0.0038
A8-2	0	0	0	0	0	0.0041
A8-3	0	0	0	0	0	0
A8-4	0	0	0	0	0	0
A8-5	0	0	0	0	0	0
A8-6	0	0	0	0	0	0
A8-7	0	0	0	0	0	0
A8-8	0	0	0	0	0	0
A8-9	0	0	0	0	0	0.0037
A8-10	0	0	0.0072	0	0	0
A8-11	0.0008	0	0	0	0	0
A8-12	0	0	0.0036	0	0	0
A9-1	0	0	0	0	0	0
A9-2	0	0	0	0	0	0
A9-3	0	0	0	0	0	0
A9-4	0	0	0	0	0	0
A9-5	0	0	0	0	0	0
A9-6	0.0063	0	0	0	0	0
A9-7	0	0	0	0	0	0
A9-8	0	0	0	0	0	0
A9-9	0	0	0	0	0	0
A9-10	0	0	0	0	0	0
A9-11	0	0	0	0	0	0
A9-12	0	0	0	0	0	0
A12-1	0	0	0	0	0	0
A12-2	0	0	0	0	0	0
A12-3	0	0	0	0.0047	0	0.02
A12-4	0	0	0	0	0	0
A12-5	0	0	0	0	0	0
A12-6	0	0	0	0	0	0
A12-7	0	0	0	0	0	0
A12-8	0	0.0048	0	0.011	0	0
A12-9	0	0	0	0	0	0
A12-10	0	0	0	0	0	0
A12-11	0	0	0	0	0	0
A12-12	0	0	0.0035	0	0	0

Figure 3. Percentage of GFP$^+$ cells from C–GFP and A–GFP clones. 60 subclones each from C–GFP and A–GFP lines were monitored for GFP$^+$ cells by flow cytometry every week for 6 wk. Green boxes are subclones with numbers of GFP$^+$ cells ≥0.01% or 100 GFP$^+$ cells/10^6 cells. Bolded borders indicate subclones that were sorted for GFP$^+$ cells; red (black) borders indicate sorted subclones that did (did not) expand or survive for subsequent analyses.

increasing digestion with increasing DNase I. Among C–GFP clones, there was no variability in hypersensitive sites despite the differences in integration sites (Fig. 7 D). This was true among A-GFP clones as well. However, contrary to our expectation, DNase I hypersensitive sites for C-GFP and A-GFP clones also showed identical patterns (Fig. 7 D). Thus, it appears that the C-GFP and A-GFP transgenes exist in equally accessible chromatin.

Protein binding to CAGGTG

To study the possibility that a protein bound to CAGGTG may mediate an increase in SHM targeting, each of the three CAGGTG and AAGGTG motifs in the transgenes was assayed for protein binding by EMSA. Nuclear extracts of parental DT40 (CL18) cells showed a specific protein binding band with all probes containing the CAGGTG motif that was not seen with the AAGGTG motif (Fig. 8 A). This protein binding specific to the CAGGTG motif was not affected by the presence of an unlabeled AAGGTG competitor probe

(50×), but it was diminished by an unlabeled CAGGTG competitor probe (unpublished data). Because one of the known factors binding to CAGGTG are homodimers of E47, we included anti-E47 or IgG control antibodies in the binding reaction. Clearly, inhibition of the protein–DNA complexes was observed with anti-E47 antibody for all three CAGGTG probes (Fig. 8 B). We also used nuclear extracts of DT40 cells expressing a human E47 transgene and observed inhibition of the protein–DNA complexes with anti–human E47 antibody as well (unpublished data). These results suggest that the predominant protein binding specifically to CAGGTG motifs, but not to AAGGTG motifs, is likely to be E47. If the binding of E47 to the CAGGTG motif leads to targeting of SHM to the E47-bound gene, it may be involved in a direct recruitment of AID and SHM factors. However, in our EMSA conditions, we did not observe differences in protein–DNA binding patterns between AID[+] and AID null nuclear extracts (Fig. S2).

Histone acetylation of the GFP transgene

One of the known functions of E47 is the recruitment of the SAGA histone acetyltransferase complex (Massari et al., 1999). To determine whether there were differences in the histones associated with C-GFP and A-GFP, chromatin immunoprecipitation (ChIP) assays were performed using antibodies against acetylated H3 and H4 histones. ChIP results from both histones H3 and H4 showed no significant differences between C-GFP and A-GFP clones (Fig. 9; primers for ChIP PCR are shown in Fig. 1). This result suggests that the preferential targeting of SHM to the C-GFP transgene is not caused by hyperacetylation of histones H3 and H4 when E47 can interact with the transgene.

Collectively with the DNase I hypersensitivity data, these findings suggest that the difference in SHM targeting triggered
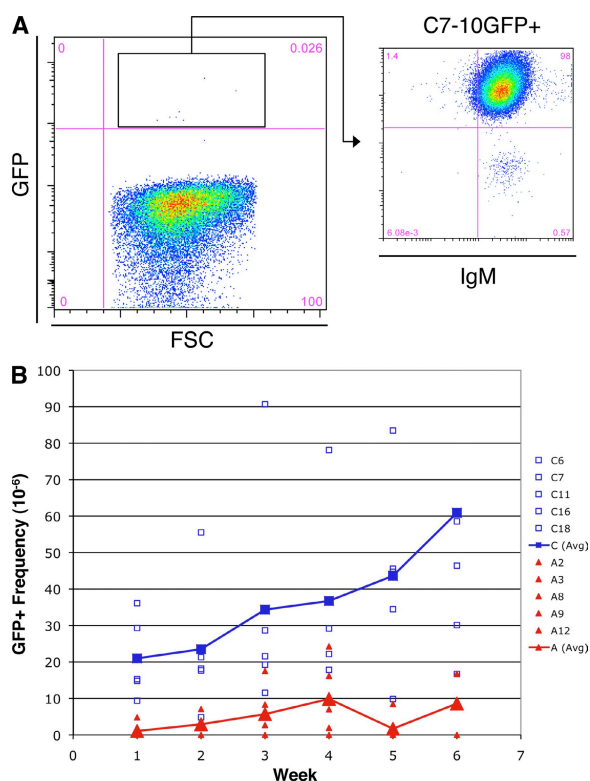
Figure 4.   FACS–sorted GFP[+] cells from a C–GFP clone and mean frequency of GFP[+] cells. (A) GFP[+] cells were sorted and cultured for 1 wk to verify functional GFP expression (percentages are shown). (B) Mean numbers of GFP[+] cells/10[6] cells. Each open square indicates a clonal mean of 12 subclones for C-GFP; filled triangles show A-GFP. The blue line with closed squares and the red line with triangles show means of all C-GFP and A-GFP clones, respectively. The entire experiment was performed twice with similar results: once without selection of cells with a single-copy transgene (not depicted) and the second time in a more stringent comparison with all clones having only a single-copy transgene, as reported in this paper. FSC, forward scatter.
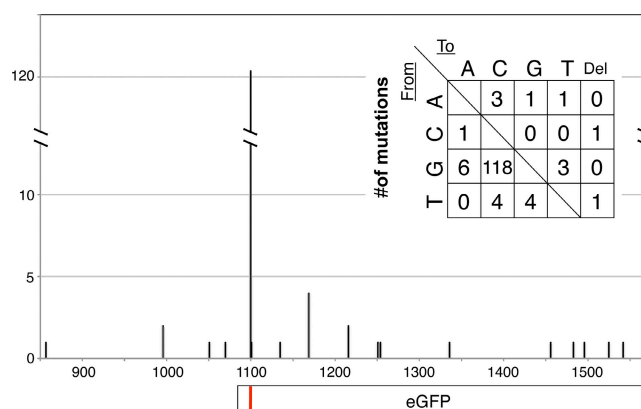
Figure 5.   Mutation analysis of sorted GFP[+] C–GFP clones. Sorted GFP[+] cells were cultured for 1 wk, and 736 bp of the eGFP gene were cloned for sequencing analyses. From a total of 141 sequences, 121 contained a mutation at the G nucleotide (bp 1,099) of the premature stop codon. The x and y axes indicate the basepair of the transgene and the number of mutations, respectively. These were mostly G-to-C transversions. GFP[+] cells were selected after either 2 or 4 wk of culture.

by the presence or absence of the CAGGTG cis-element is not caused by changes in accessibility of the target gene; rather, it may be caused by a specific recruitment of AID and other SHM machinery mediated by E47 protein bound to the CAGGTG cis-element.

CAGGTG motifs near the integration site of mutated A–GFP clones

Sequencing of sorted GFP[+] cells from A2 subclones showed mutations at the premature stop codon. Because the hexamer motif, CAGGTG, is in theory present once every ∼4 kb of the genome, we sought to determine whether a CAGGTG motif is present in the neighborhood of the A–GFP transgene integration site in the mutated A2 clone. Cloning of a fragment containing the integration site showed that the A2 clone has two CACCTG motifs ∼1 kb downstream from the 3′ end of the transgene (Fig. 10). The presence of these motifs might have induced the targeting of SHM to the A–GFP transgene in this particular clone. Because the chicken genome sequence is incomplete and a contig for this particular integration site is unavailable, we were only able to match this integration site with a cDNA library from chicken. Therefore, the chromosome and nature of this genomic region is uncertain.

To determine whether unmutated A-GFP clones lack the CAGGTG motif nearby, we attempted to sequence integration sites from these clones. Surprisingly, the A9 clone had an identical integration site as the A2 clone. Although we did not observe a significant number of GFP[+] cells in the A9 subclones, upon random sequencing of GFP in unsorted A9-1 cells grown for 6 wk, we did observe two transition mutations upstream of the GFP gene out of eight sequences (Fig. 10). This is consistent with the idea that the CAGGTG cis-element can positively influence SHM targeting of a nearby gene.

We have searched for other Ig enhancer motifs near the integration site of the A2 and A9 clones that may have enhanced

the mutability of the GFP transgene. We found two potential PU.1 binding sequences; however, unlike PU.1 motifs found in proximity to CAGGTG in Ig enhancers, they were far from the CACCTG motifs in the A2 and A9 integration site. This finding and the enhanced mutability seen in transgenic mice with CAGGTG in the J region without other nearby motifs found in Ig enhancers (Michael et al., 2003) suggest that the presence of CAGGTG in a context other than an Ig enhancer may aid the targeting of SHM to a nearby gene (see Discussion).

DISCUSSION

The findings reported in this paper show that SHM occurs in transgenes containing CAGGTG but not those without CAGGTG. The only difference between the C–GFP and A–GFP clones is a C-to-A mutation in three CAGGTG motifs in the Ig enhancers. The C–GFP and A–GFP transgenic constructs are transcribed into identical GFP mRNA sequences and their only differences, CAGGTG versus AAGGTG, lie after the 3′ poly
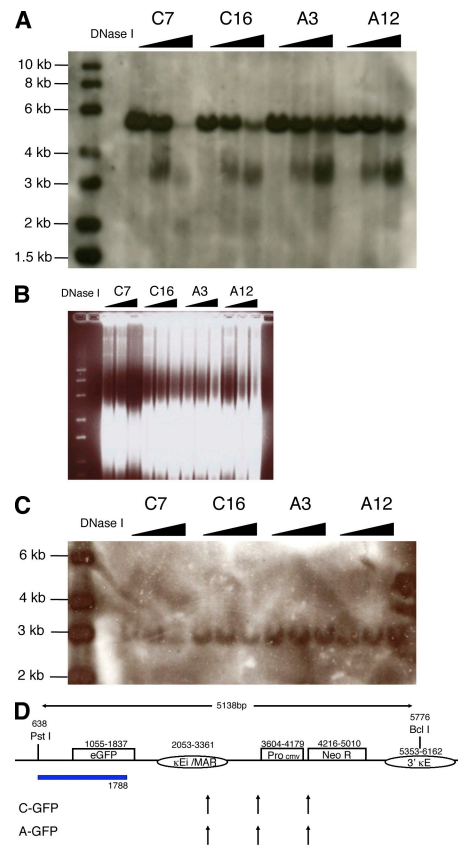

Figure 7. DNase I hypersensitive sites of GFP transgenes. (A) Southern blot of two C-GFP and two A-GFP subclones. DNA from DNase I–digested cells was digested with Pst I and Bcl I and hybridized with a GFP probe. 0, 10, and 50 U DNase I are indicated by the filled horizontal arrows. (B) Ethidium bromide–stained gel of DNA fragments before hybridization for A. (C) An ovalbumin probe was used to reprobe the blot in A. (D) Vertical arrows indicate DNase I hypersensitive sites observed with GFP for C-GFP and A-GFP clones. The horizontal blue box indicates the probe. The experiment was repeated with DNase I digestion of different C-GFP and A-GFP cell clones showing no difference in DNase hypersensitivity.

	Week 1	Week 2	Week 3	Week 4
C-GFP Cre 1	0	0.016	0.00117	0
C-GFP Cre 2	0.00221	0.015	0	0.013
C-GFP Cre 3	0	0.021	0	0.00824
C-GFP Cre 4	0	0.013	0.023	0.00859
C-GFP Cre 5	0.00205	0	0.00211	0.00991
C-GFP Cre 6	0	0.012	0	0.00956
A-GFP Cre 1	0	0	0	0
A-GFP Cre 2	0	0	0	0
A-GFP Cre 3	0	0	0	0
A-GFP Cre 4	0	0	0.00237	0.00163
A-GFP Cre 5	0	0	0	0
A-GFP Cre 6	0.00191	0	0	0.00119
C-GFP AIDko 1	0	0	0	0
C-GFP AIDko 2	0	0	0	0
C-GFP AIDko 3	0	0	0	0
C-GFP AIDko 4	0	0	0	0
C-GFP AIDko 5	0	0	0	0
C-GFP AIDko 6	0	0	0	0
A-GFP AIDko 1	0	0	0	0
A-GFP AIDko 2	0	0	0	0
A-GFP AIDko 3	0	0	0	0
A-GFP AIDko 4	0	0	0	0
A-GFP AIDko 5	0	0	0	0
A-GFP AIDko 6	0	0	0	0

Figure 6. C–GFP and A–GFP clones in an AID–deficient background. Six transgenic clones each from Cre DT40 or AID[−/−] Cre DT40 lines were monitored for GFP[+] cells by flow cytometry. The 24 subclones were analyzed by FACS weekly for 4 wk. Green boxes are as in Fig. 3.

A site. Because the mRNA terminates 5′ of the CAGGTG/AAGGTG sites, it is extremely unlikely that the GFP mRNAs from these transgenes have different structures or stability. Therefore, finding the same transgene mRNA levels in the C-GFP and A-GFP transgenes should reflect equal mRNA transcription efficiency. Thus, the effect on mutability seems independent of transcription activity, chromatin acetylation, or overall target gene activity. It seems unlikely that within the C-GFP pools only the cells that mutated were only at that time in more accessible chromatin than the rest of the cells; there is no precedence where a well-transcribed gene is in a heterochromatic state most of the time. The results show that the CAGGTG cis-element in the context of Ig enhancers is sufficient and likely required to attract AID to a target gene.

There are several studies showing a requirement for Ig enhancers for SHM induction (for review see Longerich et al., 2006). It was also clearly shown that CAGGTG, an element of Ig enhancers, enhances SHM (Michael et al., 2003). However, it was not known whether CAGGTG was sufficient/required for SHM. CAGGTG is a binding site for E2A proteins. In previous studies, the IgL enhancer of DT40 was not required for SHM (Yang et al., 2006); however, an element 3′ of the enhancer that contains five predicted E2A binding sites was required when the enhancer was deleted (Kothapalli et al., 2008; Blagodatski et al., 2009). This supported the idea that E2A proteins may be required for SHM. In the present study, the EMSA assays (Fig. 8) show that there are clear differences in protein binding to CAGGTG- versus AAGGTG-containing probes. Furthermore, the proteins binding to CAGGTG sequences were inhibited by addition of anti-E47 antibody, suggesting that the E47 protein may be the one that leads to increased targeting of SHM to the GFP transgene. Besides E47, there may be other known or unknown helix-loop-helix proteins that are involved or can substitute in experiments where low levels of

SHM still existed in cells with E2A deletion (Schoetz et al., 2006; Kitao et al., 2008).

The finding that only transgenes with CAGGTG motifs attract AID shows that the other cis-elements of Ig enhancers alone cannot attract the SHM machinery. However, proteins that bind to other motifs in Ig enhancers may be cofactors in SHM. The avid targeting of Ig genes compared with non-Ig genes may be caused by the combined action of CAGGTG binding factors with factors that bind other Ig enhancer motifs. Mutation of PU.1 or NF-EM5 motifs in the 3′ κ enhancer was found to reduce the mutation rate in a κ transgene (Kodama et al., 2001). The available evidence on enhanced mutability by additional CAGGTG motifs in a J region of RS-transgenic mice without other adjacent Ig enhancer motifs (Michael et al., 2003; Fig. S3), as well as finding low mutagenesis in A2 and A9 clones with CACCTG elements present within 1 kb of the transgene outside of enhancers (Fig. 10), may indicate that the CAGGTG motif and other cis-elements need not be at the same site. The CAGGTG motif and other potentially required motifs for SHM targeting may act independently or coordinately over some distance. Furthermore, mutations of the Bcl6 gene and other non-Ig genes imply that the presence of CAGGTG in or near a gene that has the appropriate conditions, such as transcription activity and chromatin configuration similar to Ig genes, is able to attract the SHM machinery.

A known downstream effect of E2A binding is the recruitment of p300 and SAGA histone acetyltransferase complexes (Massari et al., 1999). Acetylated histones H3 and H4 were suggested to allow SHM induction in the V region of mutating BL2 cells (Woo et al., 2003) or the IgL locus in DT40 cells
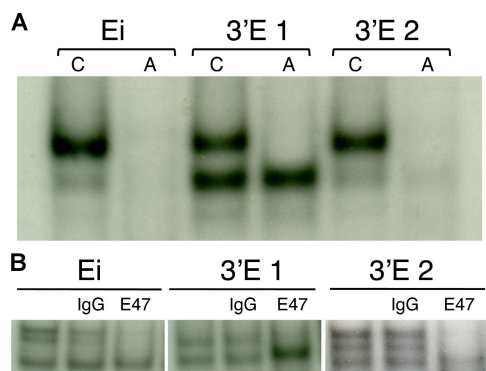

Figure 8. EMSA on CAGGTG binding protein. (A) Each CAGGTG and AAGGTG motif present in Ei and 3′E of the transgenes was assayed for protein binding. Nuclear lysates from DT40 CL18 cells were used. (B) EMSA in the presence of antibodies. 15 µg of nuclear extract and 1 µg anti-E47 or IgG control antibodies were used. EMSA assays were repeated at least three times.
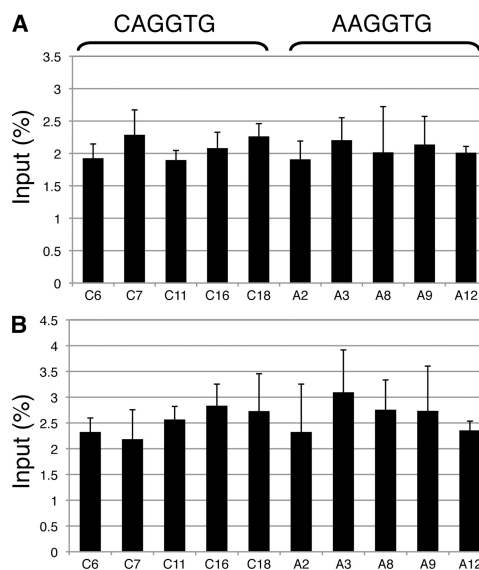

Figure 9. ChIP assays of the GFP transgenes. (A and B) ChIP with antibodies against acetylated histone H3 (A) and H4 (B). Vertical bars indicate mean percentages of immunoprecipitated eGFP DNA over input DNA by quantitative PCR. The data represent means and SDs of three independent experiments.
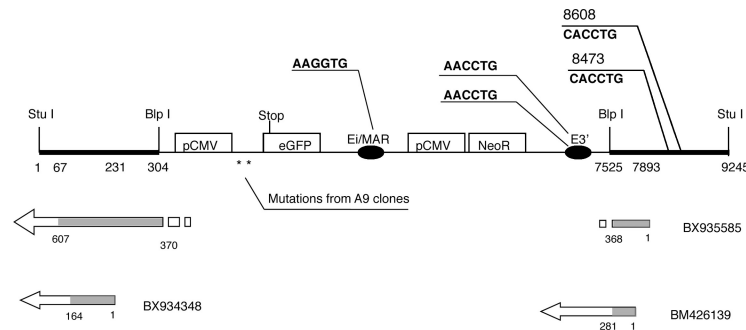
Figure 10. Integration site of mutated A–GFP clones. From A2 and A9 clones (Fig. 3), Stu I–digested DNA fragments containing the transgene and adjacent integration sites were isolated and sequenced. Bold lines indicate genomic chicken DNA. Nearby CACCTG motifs in the genomic DNA are indicated. Arrows indicate chicken cDNA sequences and the shaded portions align with the isolated Stu I fragments. Asterisks indicate mutations found in the unsorted A9-1 subclone after 6 wk of culture.

(Kitao et al., 2008). On the other hand, the ChIP data reported in this paper from DT40 cells and by others from mouse B cells (Odegard et al., 2005) suggested that changes of acetylated histones do not play a major role in inducing SHM. The highly transcribed Ig loci in mature B cells as well as developing B cells seem to be occupied by acetylated histones (McMurry and Krangel, 2000; Nambu et al., 2003; Osipovich et al., 2004). It is possible that the highly active transcription from the CMV promoter in the present study may have caused equally high acetylation of histones H3 and H4 in C-GFP and A-GFP cells; nevertheless, mutability was not rescued in A-GFP cells.

The CAGGTG motif is present in all of the frequent targets of SHM. Diffuse large B cell lymphomas undergo SHM of Ig genes as well as certain non-Ig genes. Searching non-Ig genes in diffuse large B cell lymphoma sequences revealed CAGGTG motifs in both mutated and several unmutated genes (unpublished data; Pasqualucci et al., 2001). However, some of the unmutated genes did not contain any CAGGTG motif within the coding regions, whereas all of the mutated genes observed contained at least one CAGGTG motif. Because this hexamer motif, theoretically, is present every ~4 kb, it is unlikely that all CAGGTG motifs in the genome of activated B cells are associated with DNA binding proteins. Therefore, whether CAGGTG motifs are associated with E2A proteins may determine whether AID associates efficiently with the respective gene.

An extensive analysis of dozens of genes expressed in mutating B cells has shown that essentially all of these genes undergo some level of SHM (Liu et al., 2008). This was revealed by checking mutations in mice with deletions of both Ung and Msh2. Because most of the genes had no significant mutations in wild-type mice, the authors concluded that most non-Ig genes in mutating B cells are accessed by AID but protected from mutation by restriction of error-prone repair to Ig genes and some non-Ig genes, such as BCL6. Other genes would suffer C deamination by AID, but the uracil would be repaired error free by base excision repair and mismatch repair (MMR).

An important finding of this report by Liu et al. (2008) is that most of the genes that can interact with AID have CAGGTG motifs nearby within ~2 kb. It remains to be determined whether non-Ig genes that show no significant mutations in wild-type mice possess a special mechanism that prevents error-prone repair. Alternatively, the lack of mutations in most non-Ig genes may actually reflect what is also going on in Ig genes that have been targeted by AID (Storb et al., 2009). In Ig genes of Ung/MMR null mice there are about twice as many transitions at C/G than in wild-type mice despite about the same mutation frequencies overall (Rada et al., 2004; Shen et al., 2006; Storb et al., 2009). C/G transitions in these null mice are footprints of AID. Thus, a conservative calculation shows that in wild-type mice, 39–47% or more of the uracils created in Ig genes by AID are repaired error free (Storb et al., 2009). In the study by Liu et al. (2008), the non-Ig genes show mutations in Ung/MMR null mice at very low frequencies. If, indeed, in Ung/MMR wild-type mice one half or more of AID-induced mutations in non-Ig genes are also repaired error free, they are difficult or impossible to detect because of the background of PCR errors. Furthermore, even the groups of genes with the least mutations (groups II and III in Liu et al. [2008]) in wild-type mice have 35% of their rare mutations at A/T and 20% of transversions at C/G; thus, they did undergo error-prone repair in wild-type mice. Therefore, it seems that uracils in non-Ig genes targeted by AID can also be processed by error-prone repair. It will be very interesting to learn if and how the same genes can be treated either by high-fidelity or error-prone repair, or whether there simply are situations where the high-fidelity repair capacity is overwhelmed by excess uracils in the genome.

The important take-home message from the findings of Pasqualucci et al. (2001) and Liu et al. (2008) are that AID may attack most non-Ig genes, at least those that are expressed, and that CAGGTG may be required to target AID to any gene, at least under conditions of limited AID levels as they exist during SHM. Parsa et al. (2007) and Wang et al. (2004) found that transfected non-Ig genes inserted at various places in the genome can also be mutated without Ig enhancers. These non-Ig transgenes contained multiple CAGGTG as well as CAGCTG sites. Thus, these publications agree with our finding in this paper that CAGGTG is clearly

an enabler of SHM without leading to increased transcription or chromatin activation of the target gene. We postulate that AID may be directed to expressed genes that are associated with CAGGTG elements through the interaction of E-box proteins, such as E47, bound to CAGGTG, either with AID and the transcription complex or with the transcription complex alone, thus somehow enabling AID access to the transcribed gene. Ig genes may be the most highly mutated genes because of the coincidence of several conditions: very active transcription, a high proportion of CAGGTG sites that are located within the gene, and SHM-favoring combinations of other cis–elements present in the enhancers. Many mouse and human Ig genes contain CAGGTG motifs besides the several present in the enhancers (unpublished data). The enhanced mutability by additional CAGGTG motifs in a J region of RS transgenic mice (Michael et al., 2003) indicates that CAGGTG motifs located within genes can increase SHM targeting. The distance of CAGGTG to the target may also play a role. The A2 and A9 A–GFP clones with two distant CACCTG sites ∼1 kb 3′ of the A-GFP transgene have a 5.6 times lower mutation frequency than the average C-GFP transgene (based on the number of GFP[+] cells). The relative roles of numbers and distance from the target of CAGGTG motifs and their vicinity to enhancers are not known. It will be important to determine whether CAGGTG alone can suffice to attract AID and, if not, which other motifs of Ig enhancers are essential.

As suggested by a transcription model of SHM (Peters and Storb, 1996; Michael et al., 2003), E-box proteins bound to CAGGTG motifs in Ig enhancers may associate with the transcription complex and facilitate binding of AID to the RNA polymerase. This and other models implicating CAGGTG but not requiring the tethering of AID to the transcription complex remain to be investigated.

MATERIALS AND METHODS

Cell culture and transgenic clones. DT40 CL18 cells derived from avian leukosis virus–induced chicken bursal B cells were a gift of H. Arakawa and J.M. Buerstedde (Institute of Molecular Radiology, Neuherberg, Germany; Arakawa et al., 2002). The cells were cultured in RPMI 1640 with 1% penicillin/streptomycin, 1% L-glutamine (Invitrogen), 1% chicken serum, and β-mercaptoethanol (Sigma-Aldrich) at 39.5°C with 5% CO_2. C-GFP and A-GFP constructs were linearized with Blp I and transfected as previously described (Arakawa et al., 2004). After 12 h, transfected cells were treated with 2 mg/ml G418 for selection of neomycin resistance and single clones were isolated by subsequent limiting dilutions. Single clones from limiting dilutions were expanded to collect their genomic DNA for PCR analysis using a 5′ primer from the CMV promoter and a 3′ primer from the eGFP gene. For analyzing copy numbers of the transgene, Pst I, which cuts inside the transgene, as well as Stu I and Spe I, noncutters of the transgene, were used to digest genomic DNA. A radioactive probe for eGFP was used for Southern blot analysis.

Flow–cytometric analysis and cell sorting. Five C-GFP and five A-GFP DT40 clones with a single copy of the transgene were stained with PE-conjugated anti–chicken IgM antibody (Santa Cruz Biotechnology, Inc.) and sorted for sIgM[+]GFP[−] single cells on a cell sorter (FACSAria; BD) at the University of Chicago Flow Cytometry Facility. 60 C-GFP and 60 A-GFP subclones derived from sorting were monitored for GFP expression of 50,000

live cells on an LSR II (BD) using CL18 cells and GFP[+] pseudo-V KO (a gift of H. Arakawa and J.M. Buerstedde; Arakawa et al., 2004) as gating controls.

RT-PCR analysis. Total RNA was made from DT40 cells with RNA STAT-60 (Tel-Test Inc.), recovered in 50 μl, and stored at −80°C. Equal amounts of RNA were used for making cDNA by the SuperScript III First-Strand Synthesis System for RT-PCR (Invitrogen). Real-time PCRs were run and analyzed on an MYiQ system with SYBR Green SuperMix (both from Bio-Rad Laboratories). Primers used were eGFP Fwd4 (5′-CCTACCAGGGATCCACCGGTCG-3′), eGFP Rev3 (5′-GATCGCGCTTCTCGTTGGGGTCTTTGCTCA-3′), ggActin1 (5′-CCCCAAGCTTACTCCCACAGCCAGCCATGG-3′), ggActin2 (5′-GGCTCTAGATAGTCCGTCAGGTCACGGCCA-3′), ggAID1 (5′-GTTTCTGTGCACCAGAGGGCTGAACA-3′), and ggAID4 (5′-CTCCTTTCTTGGCTGGGTGAGAGGTC-3′). PCR conditions were 95°C for 30 s, 64°C for 45 s, and 72°C for 60 s for 40 cycles. The data from chicken actin were used as a reference for the relative quantification of eGFP or AID cDNA levels using the Pfaffl method (Pfaffl, 2001).

Identification of somatic mutations. Mutations in a transgene were detected by PCR cloning using Pfu polymerase (Agilent Technologies), and eGFP primers Fwd2 (5′-TTTCTGCTGCTGCTTGCGTACGG-3′) and Rev6 (5′-GCCGTTCTTCTGCTTGTCGGCCATGATATAG-3′) were used for PCR cloning with Pfu polymerase at 95°C for 30 s, 58°C for 30 s, and 72°C for 90 s for 24 cycles and cloned into a PCR cloning kit (Zero Blunt TOPO; Invitrogen). DNA sequencing were performed by the University of Chicago Cancer Research Center DNA Sequencing Facility.

Mapping DNase I hypersensitive sites. Nuclei from DT40 cells were isolated and digested as previously described (Sambrook, 2001) with modified lysis buffer containing 0.05% NP-40. 10^7 DT40 cells were used for each digestion with 0, 10, and 50 U DNase I (Applied Biosystems) in 250 μl of solution. Isolated DNA was digested with Pst I and Bcl I (New England Biolabs, Inc.), and a Sac I/BsrG I fragment from eGFP was used to make a probe using a random-priming synthesis kit (Roche) for Southern blot analysis.

ChIP. ChIP assays for acetylated histones H3 and H4 were performed as previously described (Agata et al., 2001) with few modifications. For shearing chromatin, 10 U MNase (Sigma-Aldrich) was used in buffer containing 50 mM Tris, 1 mM $CaCl_2$, and protease inhibitors (Roche) at 37°C. Antibodies used were anti–acetyl histone H3 or anti–acetyl histone H4 (both from Millipore). Primers for the GFP region were eGFP Fwd4 (5′-CCTACCAGGGATCCACCGGTCG-3′) and eGFP Rev1 (5′-TGCCGGTGGTGCAGATGAACTTCAGGGTCA-3′) to amplify a 185-bp fragment. For quantification, the quantitative RT-PCR conditions using SYBR green described in RT-PCR analysis were used.

EMSA. Nuclear extracts were prepared as described by Dignani et al. (1983). The EMSA was performed as previously described (Bain et al., 1993; Spaulding et al., 2007). Oligonucleotides used for EMSA assays were 22mers: Ei (5′-CCAGGCAGGTGGCCCAGATTAC-3′), A-Ei (5′-CCAGGAAGGTGGCCCAGATTAC-3′), 3′E1 (5′-CAAAGCCT-CATACACCTGCTCC-3′), A-3′E1 (5′-CAAAGCCTCATAAACCT-GCTCC-3′), 3′E2 (5′-TACCCCAGCACCTGGCCAAGGC-3′), and A-3′E2 (5′-TACCCCAGAACCTGGCCAAGGC-3′). These oligonucleotides were annealed to the complementary strand and labeled with γP32-ATP using polynucleotide kinase. Approximately 0.1 ng of labeled probe was added to 10 μg of nuclear extract in binding buffer containing 1 μg BSA/0.8 μg poly-d(I)d(C). Competitor oligonucleotides were added to protein for 10 min before labeled probe, and the mixture was incubated for a further 20 min at room temperature. Antibodies used were anti-E47 and control IgG (Santa Cruz Biotechnology, Inc.).

Online supplemental material. Fig. S1 shows relative mRNA expression for GFP, AID, and β-actin. Fig. S2 depicts an EMSA to test whether AID associates with the CAGGTG motif. Fig. S3 shows the location of

two transgene-specific CAGGTG motifs in the J region of the RS trans-gene. Online supplemental material is available at http://www.jem.org/cgi/content/full/jem.20090821/DC1.

REFERENCES

Agata, Y., T. Katakai, S.K. Ye, M. Sugai, H. Gonda, T. Honjo, K. Ikuta, and A. Shimizu. 2001. Histone acetylation determines the developmentally regulated accessibility for T cell receptor γ gene recombination. *J. Exp. Med.* 193:873–880. doi:10.1084/jem.193.7.873

Arakawa, H., J. Hauschild, and J.M. Buerstedde. 2002. Requirement of the activation-induced deaminase (AID) gene for immunoglobulin gene conversion. *Science.* 295:1301–1306.

Arakawa, H., H. Saribasak, and J.M. Buerstedde. 2004. Activation-induced cytidine deaminase initiates immunoglobulin gene conversion and hypermutation by a common intermediate. *PLoS Biol.* 2:e179. doi:10.1371/journal.pbio.0020179

Bachl, J., C. Carlson, V. Gray-Schopfer, M. Dessing, and C. Olsson. 2001. Increased transcription levels induce higher mutation rates in a hyper-mutating cell line. *J. Immunol.* 166:5051–5057.

Bain, G., S. Gruenwald, and C. Murre. 1993. E2A and E2-2 are subunits of B-cell-specific E2-box DNA-binding proteins. *Mol. Cell. Biol.* 13:3522–3529.

Barreto, V.M., A.R. Ramiro, and M.C. Nussenzweig. 2005. Activation-induced deaminase: controversies and open questions. *Trends Immunol.* 26:90–96. doi:10.1016/j.it.2004.12.004

Betz, A.G., C. Milstein, A. González-Fernández, R. Pannell, T. Larson, and M.S. Neuberger. 1994. Elements regulating somatic hyper-mutation of an immunoglobulin kappa gene: critical role for the intron enhancer/matrix attachment region. *Cell.* 77:239–248. doi:10.1016/0092-8674(94)90316-6

Blagodatski, A., V. Batrak, S. Schmidl, U. Schoetz, R.B. Caldwell, H. Arakawa, and J.M. Buerstedde. 2009. A cis-acting diversification activator both necessary and sufficient for AID-mediated hypermutation. *PLoS Genet.* 5:e1000332. doi:10.1371/journal.pgen.1000332

Dignani, J.D., R.M. Lebovitz, and R.G. Roeder. 1983. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.* 11:1475–1489. doi:10.1093/nar/11.5.1475

Gordon, M.S., C.M. Kanegai, J.R. Doerr, and R. Wall. 2003. Somatic hypermutation of the B cell receptor genes B29 (Igbeta, CD79b) and mb1 (Igalpha, CD79a). *Proc. Natl. Acad. Sci. USA.* 100:4126–4131. doi:10.1073/pnas.0735266100

Hsu, H.L., L. Huang, J.T. Tsan, W. Funk, W.E. Wright, J.S. Hu, R.E. Kingston, and R. Baer. 1994. Preferred sequences for DNA recognition by the TAL1 helix–loop–helix proteins. *Mol. Cell. Biol.* 14:1256–1265.

Kitao, H., M. Kimura, K. Yamamoto, H. Seo, K. Namikoshi, Y. Agata, K. Ohta, and M. Takata. 2008. Regulation of histone H4 acetylation by transcription factor E2A in Ig gene conversion. *Int. Immunol.* 20:277–284. doi:10.1093/intimm/dxm140

Kodama, M., R. Hayashi, H. Nishizumi, F. Nagawa, T. Takemori, and H. Sakano. 2001. The PU.1 and NF-EM5 binding motifs in the Igkappa 3′ enhancer are responsible for directing somatic hypermutations to the intrinsic hotspots in the transgenic Vkappa gene. *Int. Immunol.* 13:1415–1422. doi:10.1093/intimm/13.11.1415

Kotani, A., I.M. Okazaki, M. Muramatsu, K. Kinoshita, N.A. Begum, T. Nakajima, H. Saito, and T. Honjo. 2005. A target selection of somatic hypermutations is regulated similarly between T and B cells upon activation-induced cytidine deaminase expression. *Proc. Natl. Acad. Sci. USA.* 102:4506–4511. doi:10.1073/pnas.0500830102

Kothapalli, N., D.D. Norton, and S.D. Fugmann. 2008. Cutting edge: a cis-acting DNA element targets AID-mediated sequence diversification to the chicken Ig light chain gene locus. *J. Immunol.* 180:2019–2023.

Lebecque, S.G., and P.J. Gearhart. 1990. Boundaries of somatic mutation in rearranged immunoglobulin genes: 5′ boundary is near the promoter, and 3′ boundary is approximately 1 kb from V(D)J gene. *J. Exp. Med.* 172:1717–1727. doi:10.1084/jem.172.6.1717

Liu, M., J.L. Duke, D.J. Richter, C.G. Vinuesa, C.C. Goodnow, S.H. Kleinstein, and D.G. Schatz. 2008. Two levels of protection for the B cell genome during somatic hypermutation. *Nature.* 451:841–845. doi:10.1038/nature06547

Longerich, S., U. Basu, F. Alt, and U. Storb. 2006. AID in somatic hyper-mutation and class switch recombination. *Curr. Opin. Immunol.* 18:164–174. doi:10.1016/j.coi.2006.01.008

Massari, M.E., P.A. Grant, M.G. Pray-Grant, S.L. Berger, J.L. Workman, and C. Murre. 1999. A conserved motif present in a class of helix-loop-helix proteins activates transcription by direct recruitment of the SAGA complex. *Mol. Cell.* 4:63–73. doi:10.1016/S1097-2765(00)80188-4

McMurry, M.T., and M.S. Krangel. 2000. A role for histone acetylation in the developmental regulation of VDJ recombination. *Science.* 287:495–498. doi:10.1126/science.287.5452.495

Michael, N., H.M. Shen, S. Longerich, N. Kim, A. Longacre, and U. Storb. 2003. The E box motif CAGGTG enhances somatic hyper-mutation without enhancing transcription. *Immunity.* 19:235–242. doi:10.1016/S1074-7613(03)00204-8

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. doi:10.1084/jem.179.2.395

Muramatsu, M., K. Kinoshita, S. Fagarasan, S. Yamada, Y. Shinkai, and T. Honjo. 2000. Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell.* 102:553–563. doi:10.1016/S0092-8674(00)00078-7

Müschen, M., D. Re, B. Jungnickel, V. Diehl, K. Rajewsky, and R. Küppers. 2000. Somatic mutation of the CD95 gene in human B cells as a side-effect of the germinal center reaction. *J. Exp. Med.* 192:1833–1840. doi:10.1084/jem.192.12.1833

Nambu, Y., M. Sugai, H. Gonda, C.G. Lee, T. Katakai, Y. Agata, Y. Yokota, and A. Shimizu. 2003. Transcription-coupled events associating with immunoglobulin switch region chromatin. *Science.* 302:2137–2140. doi:10.1126/science.1092481

Odegard, V.H., S.T. Kim, S.M. Anderson, M.J. Shlomchik, and D.G. Schatz. 2005. Histone modifications associated with somatic hypermutation. *Immunity.* 23:101–110. doi:10.1016/j.immuni.2005.05.007

Osipovich, O., R. Milley, A. Meade, M. Tachibana, Y. Shinkai, M.S. Krangel, and E.M. Oltz. 2004. Targeted inhibition of V(D)J recombination by a histone methyltransferase. *Nat. Immunol.* 5:309–316. doi:10.1038/ni1042

Parsa, J.Y., W. Basit, C.L. Wang, J.L. Gommerman, J.R. Carlyle, and A. Martin. 2007. AID mutates a non-immunoglobulin transgene independent of chromosomal position. *Mol. Immunol.* 44:567–575. doi:10.1016/j.molimm.2006.02.003

Pasqualucci, L., A. Migliazza, N. Fracchiolla, C. William, A. Neri, L. Baldini, R.S. Chaganti, U. Klein, R. Küppers, 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. doi:10.1073/pnas.95.20.11816

Pasqualucci, L., P. Neumeister, T. Goossens, G. Nanjangud, R.S. Chaganti, R. Küppers, and R. Dalla-Favera. 2001. Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. *Nature.* 412:341–346. doi:10.1038/35085588

Peters, A., and U. Storb. 1996. Somatic hypermutation of immunoglobulin genes is linked to transcription initiation. *Immunity*. 4:57–65. doi:10.1016/S1074-7613(00)80298-8

Pfaffl, M.W. 2001. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res*. 29:e45. doi:10.1093/nar/29.9.e45

Rada, C., and C. Milstein. 2001. The intrinsic hypermutability of antibody heavy and light chain genes decays exponentially. *EMBO J*. 20:4570–4576. doi:10.1093/emboj/20.16.4570

Rada, C., A. González-Fernández, 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. doi:10.1002/eji.1830240632

Rada, C., J.M. Di Noia, and M.S. Neuberger. 2004. Mismatch recognition and uracil excision provide complementary paths to both Ig switching and the A/T-focused phase of somatic mutation. *Mol. Cell*. 16:163–171. doi:10.1016/j.molcel.2004.10.011

Rogerson, B.J. 1994. Mapping the upstream boundary of somatic mutations in rearranged immunoglobulin transgenes and endogenous genes. *Mol. Immunol*. 31:83–98. doi:10.1016/0161-5890(94)90081-7

Sale, J.E., D.M. Calandrini, M. Takata, S. Takeda, and M.S. Neuberger. 2001. Ablaion of XRCC2/3 transforms immunoglobulin V gene conversion into somatic hypermutation. *Nature*. 412:921–926. doi:10.1038/35091100

Sambrook, J., and D.W. Russell. 2001. Molecular Cloning. A Laboratory Manual. Third edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. 17.18–17.22.

Schoetz, U., M. Cervelli, Y.D. Wang, P. Fiedler, and J.M. Buerstedde. 2006. E2A expression stimulates Ig hypermutation. *J. Immunol*. 177:395–400.

Shen, H.M., 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. doi:10.1126/science.280.5370.1750

Shen, H.M., A. Tanaka, G. Bozek, D. Nicolae, and U. Storb. 2006. Somatic hypermutation and class switch recombination in Msh6(−/−)Ung(−/−) double-knockout mice. *J. Immunol*. 177:5386–5392.

Spaulding, C., E.J. Reschly, D.E. Zagort, Y. Yashiro-Ohtani, L.J. Beverly, A. Capobianco, W.S. Pear, and B.L. Kee. 2007. Notch1 co-opts lymphoid enhancer factor 1 for survival of murine T-cell lymphomas. *Blood*. 110:2650–2658. doi:10.1182/blood-2007-04-084202

Storb, U., K.A. Ritchie, R. O'Brien, B. Arp, and R. Brinster. 1986. Expression, allelic exclusion and somatic mutation of mouse immunoglobulin kappa genes. *Immunol. Rev*. 89:85–102. doi:10.1111/j.1600-065X.1986.tb01474.x

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. doi:10.1098/rstb.2000.0743

Storb, U., H.M. Shen, and D. Nicolae. 2009. Somatic hypermutation: processivity of the cytosine deaminase AID and error-free repair of the resulting uracils. *Cell Cycle*. 8:3097–3101.

Thompson, C.B., and P.E. Neiman. 1987. Somatic diversification of the chicken immunoglobulin light chain gene is limited to the rearranged variable gene segment. *Cell*. 48:369–378. doi:10.1016/0092-8674(87)90188-7

Wang, C.L., R.A. Harper, and M. Wabl. 2004. Genome-wide somatic hypermutation. *Proc. Natl. Acad. Sci. USA*. 101:7352–7356. doi:10.1073/pnas.0402009101

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. doi:10.1016/S1074-7613(03)00261-9

Wu, P., and L. Claflin. 1998. Promoter-associated displacement of hypermutations. *Int. Immunol*. 10:1131–1138. doi:10.1093/intimm/10.8.1131

Yang, S.Y., S.D. Fugmann, and D.G. Schatz. 2006. Control of gene conversion and somatic hypermutation by immunoglobulin promoter and enhancer sequences. *J. Exp. Med*. 203:2919–2928. doi:10.1084/jem.20061835