Published in final edited form as:

Eur J Immunol. 2008 January ; 38(1): 227–239. doi:10.1002/eji.200737664.

Identification of murine B cell lines that undergo somatic hypermutation focused to A:T and G:C residues

Palash Bhattacharya¹, Fernando Grigera¹, Igor B. Rogozin^{2,3}, Thomas McCarty⁴, Herbert C. Morse III⁴, and Amy L. Kenter¹

¹Department of Microbiology and Immunology, University of Illinois College of Medicine, Chicago, USA

²National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, USA

³Institute of Cytology and Genetics, Novosibirsk, Russia

⁴Laboratory of Immunopathology, National Institute of Allergy and Infectious Disease, National Institutes of Health, Rockville, USA

Abstract

Activation-induced deaminase (AID) is the master regulator of class switch recombination (CSR) and somatic hypermutation (SHM), but the mechanisms regulating AID function are obscure. The differential pattern of switch plasmid activity in three IgM⁺/AID⁺ and two IgG⁺/AID⁺ B cell lines prompted an analysis of global gene expression to discover the origin of these cells. Gene profiling suggested that the IgG⁺/AID⁺ B cell lines derived from germinal center B cells. Analysis of SHM potential demonstrates that the IgV κ domains are inducibly diversified at high rate during *in vitro* culture. The mutation spectra focused to A:T base pairs, revealing a component of the hypermutation program that occurs preferentially during phase 2 of SHM. The A:T error spectra were analyzed and were not characteristic of polymerase η activity. A differential pattern of three consensus motifs used for A:T base substitutions was observed in WT and *Pol η* , *Msh2*- and *Msh6*-deficient B cells. Strikingly, mutations in our B cell lines recapitulated the mutable motif profile for *Pol η* and *Msh2* deficiency, respectively, and suggest that an additional pathway for the generation of A:T mutations in SHM is conserved in mouse and human.

Keywords

AID; B cell; Immunoglobulin; Somatic hypermutation

Introduction

Class switch recombination (CSR) promotes the expression of antibody molecules with different constant (C_H) regions and permits diversification of effector function while

Correspondence: Dr. Amy L. Kenter, Department of Microbiology and Immunology, University of Illinois College of Medicine, Chicago, IL 60612–7344, USA, Fax: +1-312-9966415, star1@uic.edu.

Conflict of interest: The authors declare no financial or commercial conflict of interest.

maintaining the original antigen binding specificity arising from V(D)J joining (reviewed in [1, 2]). Somatic hypermutation (SHM) produces primarily single base changes targeted to the expressed V genes, leading to repertoire diversification in the germinal center (GC) microenvironment [3]. CSR and SHM are both initiated by activation-induced deaminase (AID) (reviewed in [4, 5]), which most likely functions by directly deaminating dC to form dU residues in single-stranded DNA of its target genes (reviewed in [4, 6, 7]). AID-catalyzed formation of dU in DNA would yield dU:dG mismatches as obligatory intermediates and which are repaired, replicated over or resolved through subverted DNA repair to produce the remaining components of the mutation spectrum found in SHM or double-strand breaks (DSB) during CSR. Uracil DNA glycosylase (UNG) deficiency leads to a partial inhibition of CSR and perturbation of mutation spectra during SHM in mouse and man [8–10].

It has been proposed that SHM occurs in two phases (reviewed in [5]). Phase 1 is initiated by AID-dependent deamination of dC residues and leads to mutations focused on dC:dG pairs, particularly those located in the intrinsic SHM hotspot, WRC (W = A/T, R = A/G). Mutations arising during phase 2 are probably initiated by MSH2/MSH6 recognition of the dU:dG mismatch, followed by patch repair *via* polymerase (POL) η and exonuclease (EXO) 1 and results in the introduction of A:T mutations ([11, 12] and references therein). Recent studies indicate that ubiquitinated proliferating cell nuclear antigen (PCNA) is required for the recruitment of POL η to sites of DNA damage such as abasic sites [13–15] and for the generation of A:T mutations during SHM [16].

Deficiency of any one of the proteins in the MSH2/MSH6/POL η /EXO 1 complex leads to an incomplete block in A:T base mutagenesis, implying that an alternate mechanism for the introduction of A:T mutations might exist. The observation that mutations at A:T residues are essentially absent in mice doubly deficient in *Msh2-Pol η* led to the suggestion that POL η is the sole contributor of A:T errors during SHM [17]. Thus, the reduced load of A:T mutations observed during SHM in *Msh2*-deficient mice has been described as wholly dependent on POL η [17]. However, it was also possible that MSH2/MSH6 is capable of recruiting several translesional (TL) polymerases for A:T mutagenesis, which function in the absence of, or in parallel with, POL η . Analysis of mutable motif usage in A:T mutation spectra from mice with defined genetic deficits might provide information regarding the actual pathways used to generate A:T base substitutions.

Herein, we report that two murine B cell lines express a set of genes associated with GC B cells, as determined by microarray analysis. Mutation frequency analysis indicates that SHM is inducible in these cell lines and is heavily skewed toward A:T residues. Unexpectedly, the POL η signature motif, TA/TA, is not a hotspot for A:T errors in these cell lines. Analysis of base substitution spectra taken from patients with xeroderma pigmentosum variant (XP-V) disease, who are deficient in *Pol η* , allowed derivation of a new mutable motif for residual A:T mutations, and suggests that an alternate TL polymerase is active in the absence of *Pol η* . Mutable motif analysis in mutation spectra from wild-type (WT), and *Pol η* -, *Msh2*- and *Msh6*-deficient B cells indicates a differential pattern of motif usage for A:T base substitutions. Analysis of the mutable motif usage pattern in our murine B cell lines indicates similarity to that observed for *Msh2* and *Pol η* deficiencies. It is notable that mutation spectra from *Msh2*- and *Msh6*-deficient mice lacked A:T base errors at the POL η

hotspot motif. These findings indicate that the A:T mutations arising in the absence of *Msh2* and *Msh6* may derive from another pathway for A:T mutagenesis and highlight the usefulness of the newly identified murine B cell lines for the analysis of A:T mutagenesis during SHM.

Results

The IgG⁺ cell lines are related to GC B cells

1.B4.B6, CH12.LX and I.29 μ constitutively support plasmid-based CSR [18] whereas A20 and M12 must be induced to recombine switch substrates, despite their expression of AID [19], suggesting the existence of regulators of AID. We sought to identify a cohort of coordinately regulated genes that comprise the genetic signature for M12 and A20 cells, to gain a better understanding of their function and origin. The gene expression patterns were profiled on the genomic scale using oligonucleotide microarray chips. Cell lines representing pre-B cells, mature B cells and plasmacytomas (PCT) as well as LPS-activated B cells derived from BALB/c nu/nu and AID WT and KO mice were used. A total of 81 oligonucleotide microarray chips were analyzed, each containing approximately 14 000 spots representing close to 6800 unique genes, of which most are named. Global gene profiling by unsupervised two-way hierarchical clustering using the average-linkage method [20] indicates that A20 and M12 cell lines are highly related, although small differences in gene expression are also evident (Fig. 1A, Supporting Information Fig. 1).

The gene expression map identifies a cluster of 39 genes, denoted GC, that are up-regulated in the M12 and A20 cells relative to the IgM⁺/AID⁺ cell lines (1.B4.B6, CH12.LX and I29 μ) (Fig. 1A). The cluster includes the murine GC-specific transcript *MI7* (*Gcet*) [21] and genes that are associated with a proliferation signature, such as *Cdc71*, a critical regulator of the G1/S phase transition. One of the characteristic features of GC B cells is a high proliferative index. In Supporting Information Table 1, genes that were more highly expressed in M12 and A20 as compared to the IgM⁺/AID⁺ cell lines are indicated by the + symbol. To more fully chronicle the genetic program associated with M12 and A20 cells, we identified genes characteristic of these cells in two separate ways. Using a supervised cluster analysis, two clusters specific to A20 and M12 cells were found to contain genes that are genetically and functionally associated with GC B cells, including *Bcl6* [22], *Gcet* (*MI7*), [23] and *Rev3l* [24] (Fig. 1B, Supporting Information Table 1). In a second approach, 55 genes were extracted mathematically from the array data based on higher or lower than average expression in M12 and A20 cells and include both previously and newly identified genes (Supporting Information Table 1).

To independently verify this gene expression profile, we selected four genes from the array for further analysis, including *Gcet*, *Bcl6*, *Batf*, and *Rev3l*. We also studied *Aicda*, *OcaB* (*Pou2af1/Bob1*) [25], and *Idb3* [26, 27], previously shown to function in GC B cells. Using quantitative (q)RT-PCR, the cycle number at which *Gapdh* crossed the threshold was taken as a common point of reference, and the difference between it and the threshold cycle for each test gene was determined (Fig. 1B). The expression levels for *Bcl6*, *Gcet*, *Batf*, and *Rev3l* were up-regulated in A20 and M12, paralleling the results from the microarray analysis. The *OcaB* expression level was up-regulated for M12 and A20 as well as for

1.B4.B6 (Fig. 1B). The *OcaB* expression level for M12 and A20 is consistent with a genetic signature associated with GC or post-GC B cells whereas the *OcaB* level in 1.B4.B6 is consistent with an expression profile found in transitional B cells [28]. Transcripts for *Aicda* were detected at similar levels in the IgG⁺ cell lines, A20 and M12, the I.29 μ and CH12.LX cell lines (Fig. 1B) and in normal LPS-activated splenic B cells (not shown) whereas levels were intermediate for 1.B4.B6 cells and lowest for Bal17 cells (Fig. 1B). Nonetheless, 1.B4.B6 cells are capable of CSR. This observation excludes low AID transcript levels as the underlying difference between the switching, I.29 μ , CH12.LX and 1.B4.B6, cell lines and the putative GC-derived B cell lines, A20 and M12. Taken together the results of the microarray analyses and qRT-PCR analyses suggest that A20 and M12 cells are derived from GC or post-GC B cells.

M12 and A20 cells undergo inducible SHM in V κ

GC B cells actively engage in SHM whereas post-GC B cells have accumulated mutations in V genes but SHM has ceased. To screen A20 and M12 for SHM *in vitro*, we compared the extent of V κ gene diversity that had accumulated during previous expansion using approaches that have been used in many previous analyses of SHM in B cell lines [24, 29–31]. The rearranged V κ genes were PCR amplified from cDNA using degenerate primers for V κ together with a C κ oligonucleotide. Two in-frame V κ C κ rearrangements were detected in A20 cells, V κ bb1*/J κ 1 and V κ ba9/J κ 4, members of the V κ 1 and V κ 9/10 families, respectively [32] (Fig. 2A). The V κ bb1* designation was based on a “best fit” with available sequence information and differed from the published V κ bb1 germ-line gene at 31 nucleotide positions (Fig. 2B), whereas the V κ ba9 contained a single nucleotide change relative to the germ-line gene (not shown). Analysis of V κ rearrangements in M12 cells revealed an in-frame rearrangement of V κ ba9*/J κ 5, which differed from the germ-line V κ ba9 at 34 positions (Fig. 3A, B). Sequencing the cloned A20 V κ bb1*/C κ and M12 V κ ba9*/C κ PCR products from unstimulated cultures revealed little diversity as compared to the PCR control (Fig. 2C, 3C). These findings indicate that V κ bb1* and V κ ba9* are either derived from unchronicled V κ genes or from germ-line genes that have previously undergone SHM.

In a serendipitous discovery, we observed that V κ gene diversification could be induced in A20 and M12 cells. To initiate mutagenesis, cells were alternately cultured with defined or characterized FCS for 5-day periods over 2 wk and the V κ mutation frequency was determined. In A20 cells, the mutation frequency from cells induced for 2 wk was 4.36×10^{-4} mutations/bp whereas cells grown for 4 wk in uninduced conditions displayed a mutation frequency of 8.3×10^{-5} mutations/bp (Fig. 2C). When three different batches of FCS were used to sustain cell growth in non-inducing conditions, mutation frequency ranged from 8.3×10^{-5} to 1.45×10^{-4} mutations/bp (Fig. 2C and data not shown). The mutation frequency following 2 wk of growth under inducing conditions ranged from 2.8×10^{-4} to 5.24×10^{-4} mutations/bp in three independent experiments, demonstrating the reproducibility of mutation stimulation. Similarly, the mutation frequency in M12 cells grown under induced conditions for 2 wk was 1.1×10^{-3} mutations/bp, whereas cells grown uninduced for 2 wk showed a mutation frequency of 4.4×10^{-4} mutations/bp, indicating that the mutation induction protocol is appropriate for both cell lines (Fig. 3C and data not

shown). Thus, serum shock created during cell culture by alternating sources of FCS may generate genotoxic stress and initiate SHM in the A20 and M12 cell lines.

The A20 and M12 V κ sequence diversity was assessed after culture periods of 0, 2 and 4 wk of induced conditions by sequencing the products of six to nine independent PCR amplifications. The sequences indicated a progressive accumulation of mutations over time (Fig. 2C, 3C). Only unique mutations were counted, minimizing the number of presumptive independent repeats of specific nucleotide substitutions. A classic feature of SHM is that mutations accumulate predominantly in the V region but few are found in the C region. This segregation of mutations away from the C κ region is evident for both A20 and M12 cells (Fig. 2C, 3C). Sequence analysis of individual mutated templates from A20 and M12 indicated that the majority of the mutations are single base substitutions and no untemplated nucleotide insertions were found.

The likely genealogical relationships between mutated clones in the population were deduced for both the V κ bb1*/J κ 1 and V κ ba9*/J κ 5 genes (Fig. 2D, 3D). PCR artifact is unlikely to make a significant contribution to the compendium of mutations since the number of mutations is greatly in excess of that observed in control PCR amplifications (9.5×10^{-5}). Moreover, we found dynastically related clones in independent amplifications. Most frequently, the generations within a lineage differ by a single nucleotide substitution, indicating that only a small number of substitutions were introduced in each round of mutation. The position of base substitutions did not focus to the CDR1 or CDR2 regions in A20 and did not focus to CDR2 in M12 (Fig. 4A, B), indicating that the mutations are unselected. However, mutations did focus to CDR1 in M12 where 20% (23/115) mutations were found in this subregion, suggesting that there might have been some selection.

V κ mutations in A20 and M12 are biased to A:T base pairs but lack a POL η signature

The 115 and 76 unique mutations found in the 2- and 4-wk samples were scattered across the V κ bb1*/J κ 1 (A20) and V κ ba9*/J κ 5 (M12) genes, respectively (Fig. 4). In both A20 and M12, there was a preference for mutations focused to A and T residues that account for 59.2 and 64.1% of the nucleotides targeted, respectively, and these mutations were largely transitions (Fig. 4C, D). This frequency of A/T mutations is slightly elevated as compared to SHM found in mice where 46–53% of targeted residues were A and T residues [33]. Although the mutated dC/dG pairs in A20 and M12 were biased toward transitions, they were not located in or adjacent to AID (WRC) hotspots.

Based on the observation that DNA POL η is an AT-mutator in Ig V and S regions and favors transition of A→G [17, 34–39], we analyzed the context of A20 and M12 mutations for their similarity to the WA/TW mutable motif. The WA (AA/TT and TA/TA) motifs were previously found to describe a somatic mutation context at A:T bases *in vivo* and *in vitro*, for Ig V genes and POL η , respectively [33, 37, 38]. The subclass of errors at TA/TA motifs predominates over those at AA/TT in WT control spectra, whereas mutations at TA/TA and AA/TT motifs are absent and retained, respectively, in spectra from XP-V patients and Pol $\eta^{-/-}$ mice (Table 1) as previously shown [37, 39]. Surprisingly, in the A20 and M12 lines, the frequency of substitutions at TA/TA was low (Table 1) and strand bias for A:T mutations is absent (data not shown), indicating the absence of the POL η signature. However, the A20

and M12 cell lines differ with respect to A/T mutagenesis since there was a significant correlation between AA/TT mutable motifs and mutations in A20 but not in M12 (Table 1). These studies demonstrate that DNA POL η is unlikely to be involved in somatic mutagenesis in A20 or M12 cells and imply that another mechanism is responsible for the observed A:T base substitutions.

New mutable motif in XP-V and $\text{Pol}\eta^{-/-}$ spectra is also found for A20 mutations

Despite DNA $\text{Pol}\eta$ deficiency, XP-V mutation spectra retain a measurable fraction of substitutions in A:T bases (10–20%), indicating that an additional mechanism is involved in mutagenesis [34–37, 39]. To determine whether a specific enzymatic activity is responsible for the residual A:T mutations, we tested whether these substitutions focus to definable hotspots. Consensus sequence reconstruction for small datasets involves an exhaustive analysis of all possible variants of mutable motifs in the error spectra [37, 40]. We searched for correlations between the distribution of XP-V-derived mutations in the $V\kappa$ gene [34] and in the Jh4 intronic sequence [36, 39] and the distribution of each motif in the target sequences, using a Monte Carlo procedure. Three motifs, AA/TT, ADK/MHT (D = A/G/T; H = A/C/T; K = G/T; M = A/C) and ADG/CHT (the diagonal line separates the direct and complementary variants of the motif) were found to be the best descriptors of the mutation context for A:T bases in the XP-V spectra (Table 2). Examination of mutation spectra from $\text{Pol}\eta$ -deficient mice also indicates that residual A:T mutations preferentially locate to the AA/TT and ADK/MHT motifs and confirms that this mutation distribution is directed (Table 1). To extend this analysis, we analyzed the distribution of base substitution errors in A20 and M12 $V\kappa$ genes and found a highly significant correlation between the ADK/MHT and AA/TT mutable motifs with the mutation context observed in the A20 but not the M12 spectrum (Table 1).

Integrity of the MSH2/6/POL η /EXO 1 complex determines use of the ADK/MHT motif

Studies show that POL η is an important error-prone polymerase in SHM and its function is dependent on the integrity of the MSH2/6/POL η /EXO 1 complex [5, 17]. We sought to test the proposition that introduction of mutations at ADK/MHT is dependent on the structural integrity of the MSH2/6/POL η /EXO 1 complex. Separation of function variants exist for the *Msh2* and *Msh6* genes [41, 42]. In *Msh2*, substitution by alanine for glycine at position 674 (*Msh2*^{G674A}) allows mismatch recognition but prevents mismatch repair (MMR) and leads to reduced A:T mutagenesis in SHM [42]. In *Msh6*, the threonine (T) to aspartic acid (D) mutation at position 1217 (*Msh6*^{TD/TD}) inhibits ATPase activity in the MSH2/6 heterodimer and ablates the ability of the MMR complex to repair mismatches [41]. Analysis of mutation spectra from mice bearing null or point mutations in the *Msh2* and *Msh6* [41–44] genes may clarify the mechanistic requirements for introduction of A:T mutations at classical and alternative motifs.

Residual A:T mutations from *Msh2*^{-/-} and *Msh6*^{-/-} mice [43, 44] are not located at either the TA/TA or AA/TT mutable motifs, whereas A:T base substitutions in *Msh2*^{G674A} and *Msh6*^{TD/TD} mice are present in the AA/TT mutable motif (Table 1). We find that usage of the TA/TA mutable motif is also absent in *Msh2*- and *Msh6*-deficient B cells (Table 1). Furthermore, usage of the AA/TT motif is dependent on the enzymatically inactive forms of

these proteins, MSH2^{G674A} and MSH6^{TD/TD}, suggesting that these proteins contain a scaffolding function that supports recruitment of TL polymerases. Use of the alternative ADK/MHT motif as a focus of A:T mutation was clearly evident in the *Polη*^{-/-} and XP-V mutation spectra as well as in *Msh6*^{-/-} and *Msh2*^{G674A} but not in *Msh2*^{-/-} and *Msh6*^{TD/TD} mutation spectra (Table 1). Together these findings indicate that, when the MSH2 or MSH6 proteins are absent or structurally perturbed, they lose the ability to coordinate POL η mutagenesis and have a variable ability to target the AA/TT and ADK/MHT motifs.

A:T mutations locate to the ADK/MHT motif in AID-induced T cell lymphomas

In AID-transgenic mice, T cell lymphomas arise and display the same targeting of AID-induced mutations to non-Ig genes as found in B cells, as well as to the T cell-specific genes, CD4 and CD5 [45, 46]. In the T lymphomas, mutations in the *c-myc* and *TCRb* genes showed an increase of G:C mutations at RGYW/WRCY motifs whereas substitutions in the CD4 and CD5 genes showed an increase of A:T substitutions [45, 46]. Interestingly, primary mediastinal B cell lymphomas, thought to arise from thymic medullary B cells, also show A:T-focused mutations in *Bcl6* [47]. To assess mutable motif usage as a consequence of AID expression, we analyzed the A:T mutation spectra in the CD4 and CD5 genes from T cell lymphomas arising in AID-transgenic mice and found that substitutions are located in the AA/TT and ADK/MHT motifs but not in TA/TA, the POL η fingerprint (Table 1). These studies demonstrate that A:T mutations can accumulate as a consequence of AID expression and strongly suggest that the A:T mutations observed in A20 and M12 are a direct result of AID action.

Discussion

The high expression of genes associated with GC B cells suggests that the A20 and M12 cell lines derive from either the GC or post-GC stage of differentiation. Analysis of Vκ hypermutation in A20 and M12 cell lines reveals that these cells inducibly diversify their rearranged Vκ genes during *in vitro* culture. Mutagenesis in A20 and M12 cells exhibits many of the classical features of SHM *in vivo*, including the preferential targeting of mutations to the V rather than the C genes, stepwise accumulation of single nucleotide substitutions and a strong transition bias. However, several features of Vκ SHM in A20 and M12 cells differ from the patterns of V gene diversification defined in mice and other B cell lines. In A20 and M12 cells, the incidence of unique unselected mutations focused to A:T pairs was 60% as compared to 46–53% in normal mice [33]. Strikingly, in A20 and M12 cells substitutions focused to G:C base pairs were not found in or adjacent to canonical AID hotspots. However, the strong presence of A:T-focused substitutions and the dearth of mutations in the WRC motif is not unique since these features are found in CD4 and CD5 mutation spectra from AID-induced T lymphomas [45, 46] and in the *Bcl6* gene from thymic medullary B cells [47]. These observations indicate that an alternative pathway exists for A:T mutagenesis under some circumstances.

Comparison of the POL η-dependent classical pathway with this newly recognized alternative pathway for A:T mutagenesis may provide mechanistic insights. AID initiates SHM by deaminating cytosines in Ig genes, creating U that is mispaired with G residues.

Two major DNA repair systems, base excision repair (BER) and MMR which normally engage in faithful repair of U:G mismatches are co-opted to generate mutations. BER orchestrates removal of uracil by UNG2, which generates a non-instructive abasic site that is processed to create a single-strand gap [4, 48]. During phase 1 SHM, high-fidelity replication over the uracil lesion favors C->T and G->A transitions whereas low-fidelity replication over abasic sites favors transversions. Phase 2 of SHM involves the recognition of U:G mismatches by MMR proteins including MSH2, MSH6 and the recruitment of EXO 1 to generate a gap (reviewed in [11, 12]). The excised long patch is then resynthesized predominantly by POL η that creates errors at A:T residues located at a distance from the canonical AID hotspot motifs [5]. However, in A20 and M12 mutation spectra, G:C mutations were not located at AID hotspots and A:T mutations did not distribute to the signature TA/TA mutable motif associated with POL η activity.

Given our current understanding of the mechanism of SHM, the paucity of G:C mutations focused to AID hotspots in A20 and M12 mutation spectra could arise from a reduction of phase 1 mutagenesis by failure of some feature of BER to function in short-patch repair and/or an unusually robust phase 2 SHM which favors long-patch repair by TL polymerases. CSR is strongly dependent on the UNG2 component of the BER pathway for the generation of DSB in switch (S) regions [8, 49]. Our earlier studies of M12 and A20 cell lines indicated that endogenous and plasmid-based CSR function, demonstrating that the BER pathway is intact and expressed [19]. Even assuming that UNG2 is repressed in A20 and M12 cells during SHM, we would still expect to find transitions located at AID hotspot motifs. The absence of mutations at these motifs favors the notion that long-patch repair is preferentially expressed and might correctly repair the original uracil and then introduce errors at a distance from the original lesion. How can this happen?

During the canonical BER process (reviewed in [48]), abasic sites generated by UNG2 are recognized by apurinic/apyrimidinic endonucleases (APE) which nick the DNA backbone to create the single-strand breaks. These gaps are filled in by DNA POL β and the 5'-deoxyribose phosphate group remaining after APE action is excised by the POL β -associated lyase activity. We suggest that a TL polymerase substitutes for POL β since the presence of abasic sites will arrest DNA synthesis by high-fidelity polymerases. Interestingly, mutations were increased during SHM and CSR in POL β -deficient B cells [50].

The properties of POL ι may permit participation in SHM [51] and may account for several puzzling features associated with SHM in the A20 and M12 cell lines. In the absence of functional POL η , POL ι might contribute directly to long-patch repair with the collaborative assistance of POL ζ , a mispair extender [52–54]. POL ι misincorporates dGMP opposite template T (or U) at very high frequency, which would restore the parental genotype at the site of U:G mispairs [55, 56] and could account for the paucity of mutations at G:C pairs in A20 and M12 cells. POL ι is weakly processive and may pause following addition of a single nucleotide [53, 55, 56] followed by a switch to a POL ζ active in long-patch repair [52–54]. Previously, POL ζ was shown to contribute to SHM in human B cells [24]. We demonstrate here that Rev3l, the catalytic component of POL ζ , is up-regulated in M12 and A20 cells. POL ι - and POL ζ -dependent mutagenesis, focused on A:T base pairs in

Ig genes, has been induced in B cell lines infected with hepatitis C virus [57]. POL κ has also been implicated in Ig V gene mutagenesis of a Burkitt's lymphoma cell line, BL2 [29]. However, a naturally occurring stop codon in this enzyme is found in substrains of the 129 mouse that do not exhibit a defect in canonical SHM [58]. This model provides a plausible explanation for the absence of G:C mutations at AID hotspots in A20 and M12 mutation spectra and introduces the possibility that alternative TL polymerases could produce A:T mutations at non-POL η mutable motifs.

A second issue raised by the A20 and M12 mutation spectra was that A:T mutations did not focus to the POL η mutable motif, TA/TA. POL η -deficient mutation spectra from mouse [59, 60] and human [34, 37] retain the AA/TT, gain ADK/MHT but lose the TA/TA motifs as a focus for mutations. The observation that in SHM POL η activity requires an intact MSH2/6/POL η /EXO 1 complex [61] led us to inquire whether other constituents of this complex similarly affect mutable motif usage during A:T mutagenesis. We found that *Msh2* deficiency leads to loss of A:T mutations at all three mutable motifs, whereas substitutions in *Msh2*^{G674A} mice were focused to AA/TT, and ADK/MHT. Similarly, *Msh6* deficiency or the *Msh6*^{TD/TD} point mutation leads to alternative use of ADK/MHT or AA/TT motifs, respectively. We infer from these studies that in the absence of POL η other activities are recruited by MSH2 and MSH6 to introduce A:T mutations at AA/TT and ADK/MHT hotspots. Analysis of the error spectrum made by human POL κ *in vitro* revealed a distribution of mutations in AA/TT motifs in Ig genes [37]. However, POL κ -deficient animals displayed a normal A:T mutation pattern, indicating that this TL polymerase is not a major contributor to A:T mutagenesis in WT [62]. Nonetheless, POL κ activity may be responsible for the high incidence of errors in the AA/TT mutable motifs when the expression of *Polη*, *Msh2* or *Msh6* is perturbed.

Based on our analyses of A:T mutation spectra from *Polη*, *Msh2* and *Msh6* deficiencies, we postulate that the A:T error spectra in A20, focused on AA/TT and ADK/MHT motifs, arose due to either a *Polη* deficiency or a mutation in *Msh2*, leading to a structural alteration similar to *Msh2*^{G674A}. A:T mutations associated with M12 cells showed no focusing to any of the three mutable motifs analyzed here, a pattern that is similar to that found for the *Msh2* null mutation. However, our analysis of A:T mutagenesis is limited to mutations known to perturb phase 2 of SHM. Additional hitherto unidentified genes may also play a role in SHM and in A:T mutagenesis in A20 and M12 cells.

It has been suggested that POL η is the exclusive TL polymerase contributing to SHM and is wholly responsible for A:T substitutions, since A:T mutagenesis is completely abolished in compound *Msh2-Polη* knockout mice [17]. However, our investigation of mutable motif usage indicates that in the absence of *Msh2* and *Msh6*, A:T errors are not targeted to the POL η hotspot motif, indicating that POL η is not responsible for A:T mutagenesis. There are several possible explanations for this finding: (1) that POL η is not recruited to U:G mismatches in the absence of MSH2 or MSH6, and other TL polymerases substitute for it; (2) that POL η is recruited to U:G mismatches but targeting is focused away from the TA/TA motif; (3) that, although POL η is recruited to U:G mismatches, it is catalytically dead in the absence of MSH2 or MSH6 but can still act as a scaffolding protein and recruit alternate TL polymerases. The first situation is unlikely because in the *Msh2-Polη* knockout

mice A:T mutagenesis was abolished and was not complemented by other TL polymerases, such as POL ι and POL ζ [24, 29] interacting with MSH6. The second scenario in which POL η targeting is redirected in the absence of MMR proteins is very unlikely since its targeting preferences are the same *in vitro* and *in vivo* [37]. We favor the third explanation because POL η catalytic activity is dependent on the MSH2/MSH6/EXO 1 complex *in vitro* [61]. Thus, in the absence of the MMR proteins, POL η activity is compromised, but POL η could still be recruited to the U:G mismatch and act as a scaffold for other TL polymerases to introduce mutations. This would provide an explanation for residual A:T mutagenesis in mutation spectra from *Msh2*- and *Msh6*-deficient mice, which lack the POL η fingerprint, and the absence of A:T mutagenesis in the doubly deficient *Msh2-Pol η* mice.

Materials and methods

Mice, splenic B cell and cell line culture

B cells derived from spleens of BALB/c nu/nu mice were prepared and activated with LPS (50 μ g/mL; Sigma Chemical, St. Louis, MO), as described previously [18]. SHM in the A20 and M12 cell lines was induced by culturing the cells for alternating periods of 5 days in media containing 20% defined (Hyclone) or characterized (Atlanta) FCS. Batches of FCS that support SHM were identified by screening.

Cloning of rearranged V κ genes and PCR protocols for V κ gene mutation frequency

Amplification of V κ in recombined L κ genes was carried out using the universal primer V κ 5'-GATATTGTGATGACCC-AGTCT-3' and a reverse C κ primer 5'-ACACTCATTCCCT-GTTGAAGCTCTT-3'. A two-step PCR reaction with Expand Long Template Polymerase (Roche) was performed in 50 μ L with 3 cycles at 94°C (1 min, 1 s), 56°C (2 s), 25°C (50 s) and 72°C (11 min, 30 s) and 35 cycles at 94°C (1 min, 30 s), 62°C (2 s), 57°C (40 s) and 72°C (1 min, 16 s). The rearranged V κ ba9*/J κ 5 gene found in A20 cells was PCR amplified with the specific primers V κ ba9*F 5'-TGACCCAGTCTCCAC-TCTCC-3' and J κ 5R 5'-CGTTTGACTTCCAGCTTGGT-3', and amplicon length was 341 bp. The rearranged gene V κ bb1*/J κ 1 found in M12 cells was PCR amplified using V κ bb1*F 5'-TGACCCAGTCTCCATCTAC-3' and J κ 1R 5'-CGTTTCACC-TCCACCTTGGT-3', and amplicon length was 313 bp. PCR was carried out using Pfu Turbo (Stratagene) in 50 μ L with a two-step touchdown PCR. The first step was for 10 cycles at 94°C (30 s), 60°C with a 1°C reduction per cycle (30 s) and 72°C (1 min, 15 s). The second step was for 20 cycles at 94°C (30 s), 55°C with a 1°C reduction per cycle (30 s) and 72°C (1 min, 15 s). The C κ region was amplified using the primers C κ .2 5'-GTTAACATCTGGAGGTGCCTCAGTCG-3' and C κ R 5'-GAAGCTCTTGACAATGGGTGA-3', and amplicon length was 258 bp. PCR was done with Pfu Turbo in a 50- μ L reaction with 30 cycles at 94°C (30 s), 55°C (30 s) and 72°C (1 min, 15 s). The PCR error rate was measured by amplifying a region of a BAC clone containing the S γ 3-C γ 3 locus (a gift from Dr. F. Alt, Harvard University). The fragment amplified was 846 bp, which corresponds to the NCBI DNA sequence (Accession number D78343 and is located between residues 5494 and 6337. A 752-bp subregion was used for all mutation frequency calculations. PCR was done using primers Dedalus3 5'-AAGATTCAAGGAGGGCTGAGGTC-3' and C γ 3R 5'-CTCAG-G-G-A-A-G-T-

AGCCTTTGAC-3', and Pfu Turbo in a 50-μL reaction with 30 cycles at 94°C (30 s), 55°C (30 s) and 72°C (2 min, 30 s). All PCR products were cloned using the TOPOTA cloning kit (Invitrogen) and the clones were submitted to automated DNA sequence analysis. Template lengths for Vκba9*/Jκ5, Vκbb1*/Jκ1, and the Cκ region used for mutation frequency analysis were 301, 273 and 211 bp, respectively.

qRT-PCR

RT-PCR for *Aicda* and *Gapdh* was carried out as described [63–65]. Additional primers for RT-PCR were designed using PrimerExpress (ABI) and were *Batf* F 5'-ATCAAACAGCTCACC-GAGGAG-3', R 5'-AGGCATGGGCAC-TGTATACCA-3'; *Bcl6* F 5'-AAGCCGTACCCCTGTGAAATC-3', R 5'-TCGCAGTTGGCTT-TTGTGAC-3'; *Gcet* F 5'-GGGCTGGAGT-TGTTGTCACA-3', R 5'-GATGGCCTTCCCCTGACAG-3'; *Idb3* F 5'-GAGCTCACTCC-GGAAGTTGTG-3', R 5'-AGGGCTGGGTAA-GATCGAAG-3'; *Ocab* F 5'-GCTACACAGTGGTGGGACCC-3', R 5'-GGCCACGG-GAAATAAGTGAG-3'; *Rev3* F 5'-GCGGAAAGGCACTATCTC-TCA-3', R 5'-CGGGATATGTCGGTCAAAGG-3'. Quantitative RT-PCR assays for specific primer pairs using 2 × SYBR Green PCR Master Mix (ABI) according to the manufacturer's instructions and PCR was carried out using an iCycler iQ (Bio-Rad). Gene expression levels were compared with *Gapdh* levels and cDNA were synthesized so that all GAPDH amplifications crossed the threshold at 20 cycles (±2 cycles). Amplified PCR products were DNA sequence verified. In Fig. 1, each data point represents two independent cDNA analyzed in duplicate or triplicate and the results averaged.

Microarray analysis and generation of gene expression profiles

Microarray chips were printed by the Microarray Research Facility, NIAID (<http://madb.niaid.nih.gov>), and comprised approximately 7400 mouse gene targets represented by 70-mer oligonucleotides purchased from Compugen (Jamesburg, NJ). Microarray chips, total RNA and the reference RNA were prepared as described [66]. The dual-color labeling strategy to create cDNA probes for hybridization with microarray chips, the chip hybridization and scanning have been previously described [66]. Log-transformed, median-centered and normalized microarray data were analyzed by a binary, agglomerative hierarchical clustering process, using an average-linkage method, with the one minus (centered) correlation as a similarity metric. In supervised clustering, a phenotype set (A20 and M12 cells) and a control set of samples is pre-defined. To define a set of genes that optimally discriminate between these sets, 1501 genes that were changed 1.4-fold with respect to expression ($p < 0.001$) in M12 and A20 cells were extracted using the two-way *t*-test and then submitted to hierarchical clustering analysis using the entire panel of cell samples. An independent mathematical selection of genes preferentially expressed in M12 and A20 cells was carried out as follows. The 1501 genes that are up- or down-regulated in M12 and A20 were further mathematically extracted by a two-way *t*-test using an iterative comparison to each of the other B cell categories (PCT, pre-B, mature B and to LPS activated splenic B cells), and 52 unique genes were identified (Supporting Information Table 1).

Statistical analysis of mutation spectra and derivation of mutable motifs

The significance of correlations between the distribution of mutable motifs and mutations along a target sequence was measured by a Monte Carlo procedure (the CONSEN program), as previously described [37, 67].

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by AI/GM52400 (to A.L.K.) and by the National Institutes of Health (to H.C.M.) We wish to thank Drs. J. Stavnezer, P. Gearhart and C.-A. Reynaud for helpful discussions and Dr. Stavnezer for the critical reading of this manuscript. We are grateful to Drs. M. Scharff, A. Martin, C. Rada, M. Neuberger, and C.-A. Reynaud for providing mutation spectra.

Abbreviations

AID	activation-induced deaminase
BER	base excision repair
CSR	class switch recombination
DSB	doublestrand breaks
EXO	exonuclease
GC	germinal center
MMR	mismatch repair
PCT	plasmacytoma
POL	polymerase
q	quantitative
SHM	somatic hypermutation
TL	transleisional
UNG	uracil DNA glycosylase
XP-V	xeroderma pigmentosum variant

References

1. Stavnezer J. Molecular processes that regulate class switching. *Curr Top Microbiol Immunol.* 2000; 245:127–168. [PubMed: 10533321]
2. Manis JP, Tian M, Alt FW. Mechanism and control of class-switch recombination. *Trends Immunol.* 2002; 23:31–39. [PubMed: 11801452]
3. Kelsoe G. The germinal center reaction. *Immunol Today.* 1995; 16:324–326. [PubMed: 7576065]

4. Kenter AL. Class-switch recombination: After the dawn of AID. *Curr Opin Immunol.* 2003; 15:190–198. [PubMed: 12633669]
5. Neuberger MS, Di Noia JM, Beale RC, Williams GT, Yang Z, Rada C. Somatic hypermutation at A.T pairs: Polymerase error *versus* dUTP incorporation. *Nat Rev Immunol.* 2005; 5:171–178. [PubMed: 15688043]
6. Barreto VM, Ramiro AR, Nussenzweig MC. Activation-induced deaminase: Controversies and open questions. *Trends Immunol.* 2005; 26:90–96. [PubMed: 15668124]
7. Longerich S, Storb U. The contested role of uracil DNA glycosylase in immunoglobulin gene diversification. *Trends Genet.* 2005; 21:253–256. [PubMed: 15851057]
8. Rada C, Williams GT, Nilsen H, Barnes DE, Lindahl T, Neuberger MS. Immunoglobulin isotype switching is inhibited and somatic hypermutation perturbed in UNG-deficient mice. *Curr Biol.* 2002; 12:1748–1755. [PubMed: 12401169]
9. Imai K, Slupphaug G, Lee WI, Revy P, Nonoyama S, Catalan N, Yel L, et al. Human uracil-DNA glycosylase deficiency associated with profoundly impaired immunoglobulin class-switch recombination. *Nat Immunol.* 2003; 4:1023–1028. [PubMed: 12958596]
10. Di Noia J, Neuberger MS. Altering the pathway of immunoglobulin hypermutation by inhibiting uracil-DNA glycosylase. *Nature.* 2002; 419:43–48. [PubMed: 12214226]
11. Li Z, Woo CJ, Iglesias-Ussel MD, Ronai D, Scharff MD. The generation of antibody diversity through somatic hypermutation and class switch recombination. *Genes Dev.* 2004; 18:1–11. [PubMed: 14724175]
12. Martomo SA, Gearhart PJ. Somatic hypermutation: Subverted DNA repair. *Curr Opin Immunol.* 2006; 18:243–248. [PubMed: 16616477]
13. Kannouche PL, Wing J, Lehmann AR. Interaction of human DNA polymerase eta with monoubiquitinated PCNA: A possible mechanism for the polymerase switch in response to DNA damage. *Mol Cell.* 2004; 14:491–500. [PubMed: 15149598]
14. Watanabe K, Tateishi S, Kawasuji M, Tsurimoto T, Inoue H, Yamaizumi M. Rad18 guides poleta to replication stalling sites through physical interaction and PCNA monoubiquitination. *EMBO J.* 2004; 23:3886–3896. [PubMed: 15359278]
15. Plosky BS, Vidal AE, Fernandez de Henestrosa AR, McLenigan MP, McDonald JP, Mead S, Woodgate R. Controlling the subcellular localization of DNA polymerases iota and eta *via* interactions with ubiquitin. *EMBO J.* 2006; 25:2847–2855. [PubMed: 16763556]
16. Langerak P, Nygren AO, Krijger PH, van den Berk PC, Jacobs H. A/T mutagenesis in hypermutated immunoglobulin genes strongly depends on PCNAK164 modification. *J Exp Med.* 2007; 204:1989–1998. [PubMed: 17664295]
17. Delbos F, Aoufouchi S, Faili A, Weill JC, Reynaud CA. DNA polymerase eta is the sole contributor of A/T modifications during immunoglobulin gene hypermutation in the mouse. *J Exp Med.* 2007; 204:17–23. [PubMed: 17190840]
18. Shanmugam A, Shi MJ, Yauch L, Stavnezer J, Kenter A. Evidence for class specific factors in immunoglobulin isotype switching. *J Exp Med.* 2000; 191:1365–1380. [PubMed: 10770803]
19. Ma L, Wortis H, Kenter AL. Two new isotype specific switching factors detected for Ig class switching. *J Immunol.* 2002; 168:2835–2846. [PubMed: 11884453]
20. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA.* 1998; 95:14863–14868. [PubMed: 9843981]
21. Christoph T, Rickert R, Rajewsky K. M17: A novel gene expressed in germinal centers. *Int Immunol.* 1994; 6:1203–1211. [PubMed: 7981148]
22. Shaffer AL, Yu X, He Y, Boldrick J, Chan EP, Staudt LM. BCL-6 represses genes that function in lymphocyte differentiation, inflammation, and cell cycle control. *Immunity.* 2000; 13:199–212. [PubMed: 10981963]
23. Shaffer AL, Rosenwald A, Hurt EM, Giltnane JM, Lam LT, Pickeral OK, Staudt LM. Signatures of the immune response. *Immunity.* 2001; 15:375–385. [PubMed: 11567628]
24. Zan H, Komori A, Li Z, Cerutti A, Schaffer A, Flajnik MF, Diaz M, Casali P. The translesion DNA polymerase zeta plays a major role in Ig and *bcl-6* somatic hypermutation. *Immunity.* 2001; 14:643–653. [PubMed: 11371365]

25. Schubart DB, Rolink A, Kosco-Vilbois MH, Botteri F, Matthias P. B-cell-specific coactivator OBF-1/OCA-B/Bob1 required for immune response and germinal centre formation. *Nature*. 1996; 383:538–542. [PubMed: 8849727]
26. Quong MW, Harris DP, Swain SL, Murre C. E2A activity is induced during B-cell activation to promote immunoglobulin class switch recombination. *EMBO J*. 1999; 18:6307–6318. [PubMed: 10562543]
27. Sayegh CE, Quong MW, Agata Y, Murre C. E-Proteins directly regulate expression of activation-induced deaminase in mature B cells. *Nat Immunol*. 2003; 4:586–593. [PubMed: 12717431]
28. Hess J, Nielsen PJ, Fischer KD, Bujard H, Wirth T. The B lymphocyte-specific coactivator BOB.1/OBF.1 is required at multiple stages of B-cell development. *Mol Cell Biol*. 2001; 21:1531–1539. [PubMed: 11238890]
29. Faili A, Aoufouchi S, Gueranger Q, Zober C, Leon A, Bertocci B, Weill JC, Reynaud CA. AID-dependent somatic hypermutation occurs as a DNA single-strand event in the BL2 cell line. *Nat Immunol*. 2002; 3:815–821. [PubMed: 12145648]
30. Sale JE, Neuberger MS. TdT-accessible breaks are scattered over the immunoglobulin V domain in a constitutively hypermutating B cell line. *Immunity*. 1998; 9:859–869. [PubMed: 9881976]
31. Zhang W, Bardwell PD, Woo CJ, Poltoratsky V, Scharff MD, Martin A. Clonal instability of V region hypermutation in the Ramos Burkitt's lymphoma cell line. *Int Immunol*. 2001; 13:1175–1184. [PubMed: 11526098]
32. Thiebe R, Schable KF, Bensch A, Brensing-Kuppers J, Heim V, Kirschbaum T, Mitlohner H, et al. The variable genes and gene families of the mouse immunoglobulin kappa locus. *Eur J Immunol*. 1999; 29:2072–2081. [PubMed: 10427969]
33. Milstein C, Neuberger MS, Staden R. Both DNA strands of antibody genes are hypermutation targets. *Proc Natl Acad Sci USA*. 1998; 95:8791–8794. [PubMed: 9671757]
34. Zeng X, Winter DB, Kasmer C, Kraemer KH, Lehmann AR, Gearhart PJ. DNA polymerase eta is an A-T mutator in somatic hypermutation of immunoglobulin variable genes. *Nat Immunol*. 2001; 2:537–541. [PubMed: 11376341]
35. Zeng X, Negrete GA, Kasmer C, Yang WW, Gearhart PJ. Absence of DNA polymerase eta reveals targeting of C mutations on the nontranscribed strand in immunoglobulin switch regions. *J Exp Med*. 2004; 199:917–924. [PubMed: 15051760]
36. Faili A, Aoufouchi S, Weller S, Vuillier F, Stary A, Sarasin A, Reynaud CA, Weill JC. DNA polymerase eta is involved in hypermutation occurring during immunoglobulin class switch recombination. *J Exp Med*. 2004; 199:265–270. [PubMed: 14734526]
37. Rogozin IB, Pavlov YI, Bebenek K, Matsuda T, Kunkel TA. Somatic mutation hotspots correlate with DNA polymerase eta error spectrum. *Nat Immunol*. 2001; 2:530–536. [PubMed: 11376340]
38. Pavlov YI, Rogozin IB, Galkin AP, Aksenova AY, Hanaoka F, Rada C, Kunkel TA. Correlation of somatic hypermutation specificity and A-T base pair substitution errors by DNA polymerase eta during copying of a mouse immunoglobulin kappa light chain transgene. *Proc Natl Acad Sci USA*. 2002; 99:9954–9959. [PubMed: 12119399]
39. Mayorov V, Rogozin IB, Adkison LR, Gearhart PJ. DNA Polymerase h contributes to strand bias of mutations of A *versus* T in immunoglobulin genes. *J Immunol*. 2005; 174:7781–7786. [PubMed: 15944281]
40. Rogozin IB, Pavlov YI. Theoretical analysis of mutation hotspots and their DNA sequence context specificity. *Mutat Res*. 2003; 544:65–85. [PubMed: 12888108]
41. Li Z, Zhao C, Iglesias-Ussel MD, Polonskaya Z, Zhuang M, Yang G, Luo Z, et al. The mismatch repair protein Msh6 influences the *in vivo* AID targeting to the Ig locus. *Immunity*. 2006; 24:393–403. [PubMed: 16618598]
42. Martin A, Li Z, Lin DP, Bardwell PD, Iglesias-Ussel MD, Edelmann W, Scharff MD. Msh2 ATPase activity is essential for somatic hypermutation at A-T basepairs and for efficient class switch recombination. *J Exp Med*. 2003; 198:1171–1178. [PubMed: 14568978]
43. Rada C, Ehrenstein MR, Neuberger MS, Milstein C. Hot spot focusing of somatic hypermutation in MSH2-deficient mice suggests two stages of mutational targeting. *Immunity*. 1998; 9:135–141. [PubMed: 9697843]

44. Martomo SA, Yang WW, Gearhart PJ. A role for Msh6 but not Msh3 in somatic hypermutation and class switch recombination. *J Exp Med*. 2004; 200:61–68. [PubMed: 15238605]
45. Okazaki IM, Hiai H, Kakazu N, Yamada S, Muramatsu M, Kinoshita K, Honjo T. Constitutive expression of AID leads to tumorigenesis. *J Exp Med*. 2003; 197:1173–1181. [PubMed: 12732658]
46. Kotani A, Okazaki IM, Muramatsu M, Kinoshita K, Begum NA, Nakajima T, Saito H, Honjo T. 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*. 2005; 102:4506–4511. [PubMed: 15767564]
47. Malpeli G, Barbi S, Moore PS, Scardoni M, Chilosi M, Scarpa A, Menestrina F. Primary mediastinal B-cell lymphoma: Hypermutation of the BCL6 gene targets motifs different from those in diffuse large B-cell and follicular lymphomas. *Haematologica*. 2004; 89:1091–1099. [PubMed: 15377470]
48. Friedberg, EC.; Walker, GC.; Siede, W. DNA repair and mutagenesis. ASM Press; Washington, DC: 2006.
49. Schrader CE, Linehan EK, Mochevova SN, Woodland RT, Stavnezer J. Inducible DNA breaks in Ig S regions are dependent on AID and UNG. *J Exp Med*. 2005; 202:561–568. [PubMed: 16103411]
50. Wu X, Stavnezer J. DNA polymerase beta is able to repair breaks in switch regions and plays an inhibitory role during immunoglobulin class switch recombination. *J Exp Med*. 2007; 204:1677–1689. [PubMed: 17591858]
51. Kunkel TA, Pavlov YI, Bebenek K. Functions of human DNA polymerases eta, kappa and iota suggested by their properties, including fidelity with undamaged DNA templates. *DNA Repair (Amst)*. 2003; 2:135–149. [PubMed: 12531385]
52. Guo D, Wu X, Rajpal DK, Taylor JS, Wang Z. Translesion synthesis by yeast DNA polymerase zeta from templates containing lesions of ultraviolet radiation and acetylaminofluorene. *Nucleic Acids Res*. 2001; 29:2875–2883. [PubMed: 11433034]
53. Tissier A, Frank EG, McDonald JP, Iwai S, Hanaoka F, Woodgate R. Misinsertion and bypass of thymine-thymine dimers by human DNA polymerase iota. *EMBO J*. 2000; 19:5259–5266. [PubMed: 11013228]
54. Johnson RE, Washington MT, Haracska L, Prakash S, Prakash L. Eukaryotic polymerases iota and zeta act sequentially to bypass DNA lesions. *Nature*. 2000; 406:1015–1019. [PubMed: 10984059]
55. Zhang Y, Yuan F, Wu X, Wang Z. Preferential incorporation of G opposite template T by the low-fidelity human DNA polymerase iota. *Mol Cell Biol*. 2000; 20:7099–7108. [PubMed: 10982826]
56. Bebenek K, Tissier A, Frank EG, McDonald JP, Prasad R, Wilson SH, Woodgate R, Kunkel TA. 5' Deoxyribose phosphate lyase activity of human DNA polymerase iota *in vitro*. *Science*. 2001; 291:2156–2159. [PubMed: 11251121]
57. Machida K, Cheng KT, Sung VM, Shimodaira S, Lindsay KL, Levine AM, Lai MY, Lai MM. Hepatitis C virus induces a mutator phenotype: Enhanced mutations of immunoglobulin and protooncogenes. *Proc Natl Acad Sci USA*. 2004; 101:4262–4267. [PubMed: 14999097]
58. McDonald JP, Frank EG, Plosky BS, Rogozin IB, Masutani C, Hanaoka F, Woodgate R, Gearhart PJ. 129-derived strains of mice are deficient in DNA polymerase iota and have normal immunoglobulin hypermutation. *J Exp Med*. 2003; 198:635–643. [PubMed: 12925679]
59. Delbos F, De Smet A, Faili A, Aoufouchi S, Weill JC, Reynaud CA. Contribution of DNA polymerase {eta} to immunoglobulin gene hypermutation in the mouse. *J Exp Med*. 2005; 201:1191–1196. [PubMed: 15824086]
60. Martomo SA, Yang WW, Wersto RP, Ohkumo T, Kondo Y, Yokoi M, Masutani C, et al. Different mutation signatures in DNA polymerase eta- and MSH6-deficient mice suggest separate roles in antibody diversification. *Proc Natl Acad Sci USA*. 2005; 102:8656–8661. [PubMed: 15939880]
61. Wilson TM, Vaisman A, Martomo SA, Sullivan P, Lan L, Hanaoka F, Yasui A, et al. MSH2-MSH6 stimulates DNA polymerase eta, suggesting a role for A:T mutations in antibody genes. *J Exp Med*. 2005; 201:637–645. [PubMed: 15710654]
62. Schenten D, Gerlach VL, Guo C, Velasco-Miguel S, Hladik CL, White CL, Friedberg EC, et al. DNA polymerase kappa deficiency does not affect somatic hypermutation in mice. *Eur J Immunol*. 2002; 32:3152–3160. [PubMed: 12555660]

63. Wuerffel RA, Ma L, Kenter AL. NF- κ B p50-dependent *in vivo* footprints at Ig Sc3 DNA are correlated with I $\rightarrow$ C3 switch recombination. *J Immunol.* 2001; 166:4552–4559. [PubMed: 11254712]
64. Snapper C, Zelazowski P, Rosas F, Kehry F, Tian M, Baltimore D, Sha W. B cells from p50/NF κ B knockout mice have selective defects in proliferation, differentiation, germline CH transcription and Ig class switching. *J Immunol.* 1996; 156:183–191. [PubMed: 8598461]
65. Muramatsu M, Sankaranand VS, Anant S, Sugai M, Kinoshita K, Davidson NO, Honjo T. 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.* 1999; 274:18470–18476. [PubMed: 10373455]
66. Shin MS, Fredrickson TN, Hartley JW, Suzuki T, Agaki K, Morse HC 3rd. High-throughput retroviral tagging for identification of genes involved in initiation and progression of mouse splenic marginal zone lymphomas. *Cancer Res.* 2004; 64:4419–4427. [PubMed: 15231650]
67. Rogozin IB, Kolchanov NA. Somatic hypermutagenesis in immunoglobulin genes. II. Influence of neighbouring base sequences on mutagenesis. *Biochim Biophys Acta.* 1992; 1171:11–18. [PubMed: 1420357]

Figure 1.

Identification of genes differentially expressed in IgG⁺ M12 and A20 cells. (A, B) Cluster analysis was performed on 81 microarrays of 38 samples derived from splenic B cells (nu/nu Balb/c, AID WT, AID KO) activated with LPS for 72 h. The pre-B cell lines are 230.37, 8A5.4A5.II.88, and 70Z/3. The mature B cell lines are CH12.LX, I.29j, 1.B4.B6 cells and Bal17. The plasma cell stage is represented by nine PCT lines. The expression data are presented as a heatmap matrix in which the rows represent individual genes and the columns represent individual mRNA samples ordered by hierarchical clustering. The bar below the sample dendrogram is color coded according to the category of mRNA sample studied (see key at bottom). The relative level of gene expression is depicted according to the color scale

shown at the bottom. Gray squares indicate missing or excluded data. (A) Unsupervised hierarchical clustering of gene expression profiles from LPS-activated splenic B cells and B cell lines shows 2609 selected genes. The sample dendrogram at the top provides a measure of gene relatedness in each sample. The black horizontal bar below the sample dendrogram indicates the presence of the AID transcript as assessed by RT-PCR, respectively. The cluster of genes associated with the GC genotype is indicated by the bar on the right of the heatmap. (B) Supervised hierarchical cluster identified 63 signature genes that were more highly expressed in M12 and A20 cells than in other B cell lines or LPS-activated B cells. (C) qRT-PCR SYBR Green PCR assays were carried out for the GC-specific genes, as indicated. Gapdh transcript levels served as a point of reference, and the differences between the threshold cycle for Gapdh and the test transcripts are shown.

Figure 2.

Vκ sequence diversity is induced in the A20 cell line. (A) In-frame Vκ rearrangements, Vκbb1*/Jκ1 and Vκba9/Jκ4, are found in A20 cells. (B) The Vκbb1* designation was based on a “best fit” with available sequence information and differed from the published Vκbb1 germ-line gene at 31 nucleotide positions. (C) Sequence diversity in the rearranged Vκbb1* or the Cκ gene in the A20 cell line is shown as pie charts. The rearranged Vκ genes were PCR amplified and cloned. The number of clones sequenced is indicated in the center of the pie. The segment sizes depict the proportion of sequences that contain 0, 1, 2 *etc.* mutations differing from the consensus. A20 cells were cultured as described in the text. The mutation frequency in the PCR control was determined from cloned Vκbb1*/Jκ1 spiked into genomic DNA. The mutation distribution in the wk 4 sample was broken down by nucleotide and is shown in the lower panel. (D) Presumed dynastic relationship of Vκbb1*/Jκ1 mutations identified in the induced A20 culture. Each circle represents a distinct sequence with the number of mutations accumulated indicated within the circle. Individual mutations are annotated by nucleotide position in the A20 Vκbb1* gene (numbered as in B) and indicated by the number outside the circle.

Figure 3.

Vκ sequence diversity is induced in the M12 cell line. (A) An in-frame Vκ rearrangements, Vκba9*/Jκ5, is found in M12 cells. (B) The Vκba9* designation was based on a “best fit” with available sequence information and differed from the published Vκba9 germ-line gene. (C) Sequence diversity in the M12 cell line is shown as pie graphs. The rearranged Vκ genes were PCR amplified and cloned. The number of clones sequenced is indicated in the center of the pie. The segment sizes depict the proportion of sequences that contain 0, 1, 2 *etc.* mutations differing from the consensus. M12 cells were cultured as described in the text. The mutation frequency in the PCR control was determined from cloned Vκba9*/Jκ5 spiked into genomic DNA. The mutation distribution in the wk 4 sample was broken down by nucleotide and is shown in the lower panel. (D) Presumed dynastic relationship of Vκba9*/Jκ5 mutations identified in the induced M12 culture. Each circle represents a distinct sequence with the number of mutations accumulated indicated within the circle. Individual mutations are annotated by nucleotide position in the M12 Vκba9* gene (numbered as in B) and indicated by the number outside the circle.

Figure 4. Distribution of nucleotide substitutions in A20 and M12 V κ genes. Independently occurring base substitutions are indicated at each nucleotide position. The locations of CDR1 and CDR2 are indicated. Nucleotide positions are numbered from the 3' end of the sequencing primer, with nucleotide position 1 corresponding to the first base in codon 1. Mutations are indicated in italics. (A) Substitutions found in the V κ bb1*/J κ 1 gene in A20. (B) Substitutions found in the V κ ba9*/J κ 5 gene from M12. A detailed mutation characterization

from A20 (C) and M12 (D) is shown. Identical mutations were counted only once in each independent experiment.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Mutations in different mutable motifs

Motifs	Increase in mutations (<i>p</i> values) ^{d)}									
	A20 ^{b)}	MI2 ^{b)}	Control ^{c)}	XP-Y ^{c)}	POL η ^{-/-d)}	Msh2 ^{-/-e)}	Msh2G674A,f)	Msh6 ^{-/-g)}	MSH6 ^{TD/TD,h)}	CD4 + CD5 ^{h)}
AA	<i>1.8 (0.03)</i>	1.4 (0.4)	<i>1.5 (0.05)</i>	<i>2.0 (0.01)</i>	<i>3.4 (0.001)</i>	0.8 (0.61)	<i>7.5 (0.02)</i>	1.8 (0.26)	<i>2.6 (0.03)</i>	<i>4.7 (0.002)</i>
T $\overline{A}$ /TA	0.7 (0.90)	1.1 (0.78)	<i>2.2 (0.04)</i>	1.2 (0.42)	1.3 (0.38)	0.7 (0.80)	1.2 (0.62)	–	1.1 (0.41)	0.7 (0.91)
ADK/MHT	<i>1.5 (0.04)</i>	1.3 (0.33)	1.1 (0.43)	<i>1.9 (0.02)</i>	<i>1.9 (0.02)</i>	1.6 (0.12)	<i>4.4 (0.03)</i>	<i>2.3 (0.05)</i>	1.4 (0.18)	<i>2.9 (0.007)</i>
A:T Mutations	65	48	96	47	77	43	16	10	24	38

^{a)}The values listed represent the fold increase/decrease in occurrence of mutations at mutable motifs above the average occurrence of mutations at other A/T sites. The number of mutations in mutable motifs was calculated for the underlined bases. Bold italicized values represent a statistically significant correlation p ($W = W_{\text{random}} - 0.05$) between a mutable motif and the distribution of mutations, as revealed by using a Monte Carlo procedure [37, 67]. The letter designations for alternate nucleotides were: D (A, G or T), K (G or T), M (A or C).

^{b)}This paper
^{c)}Refs. [34–37, 39]
^{d)}Refs. [59, 60]
^{e)}Ref. [43]
^{f)}Ref. [42]
^{g)}Ref. [44]

Table 2

Significance of the correlation $p(W, W_{\text{random}})$ between various mutable motifs and somatic mutations from XP-V patients^{a)}

Spectrum	Number of A:T mutations	Mutable motifs					Reference
		AA/TT	ADK/MHT	ADG/CHT	CAG/CTG	AG/CT	
Jh4_int1	47	<u>0.05</u>	<u>0.02</u>	<u>0.01</u>	0.14	0.2	[39]
Jh4_int2	11	<u>0.01</u>	0.21	0.43	0.7	0.95	[36]
Vh6	42	0.06	<u>0.03</u>	0.33	0.71	0.74	[35]
Score Sc		0.04	0.09	0.26	0.52	0.66	

^{a)} All possible combinations of pentanucleotides NNANN, replacing each N by all possible bases, including the ambiguous descriptions R (A or G), Y (C or T), W (A or T), S (G or T), M (A or C), D (A, G or T), H (A, C or T), V (A, C or G) were considered [37]. Bold italicized values represent a statistically significant correlation $p(W, W_{\text{random}})$ between a mutable motif and the distribution of mutations, as revealed by using a Monte Carlo procedure [67]. The sum of $p(W, W_{\text{random}})$ for the four XP-V spectra divided by the number of spectra where a motif was observed (some motifs were not found in some spectra) defined the score Sc for a particular motif.