The Biochemistry of Somatic Hypermutation

Jonathan U. Peled,[1,*] Fei Li Kuang,[1,*] Maria D. Iglesias-Ussel,[1] Sergio Roa,[1] Susan L. Kalis,[1] Myron F. Goodman,[2] and Matthew D. Scharff[1]

[1]Department of Cell Biology, Albert Einstein College of Medicine, Bronx, New York 10461

[2]Biological Sciences and Chemistry, University of Southern California, Los Angeles, California 90089; email: jpeled@aecom.yu.edu, fkuang@aecom.yu.edu, sroa@aecom.yu.edu, mdiglesi@aecom.yu.edu, skalis@aecom.yu.edu, mgoodman@usc.edu, scharff@aecom.yu.edu

Annu. Rev. Immunol. 2008. 26:481–511

First published online as a Review in Advance on December 12, 2007

The *Annual Review of Immunology* is online at immunol.annualreviews.org

This article's doi:
10.1146/annurev.immunol.26.021607.090236

0732-0582/08/0423-0481$20.00

*These authors contributed equally to this review.

Key Words

activation-induced cytidine deaminase (AID, *aicda*), mismatch repair, antibody diversity, base excision repair, error-prone repair, germinal center

Abstract

Affinity maturation of the humoral response is mediated by somatic hypermutation of the immunoglobulin (Ig) genes and selection of higher-affinity B cell clones. Activation-induced cytidine deaminase (AID) is the first of a complex series of proteins that introduce these point mutations into variable regions of the Ig genes. AID deaminates deoxycytidine residues in single-stranded DNA to deoxyuridines, which are then processed by DNA replication, base excision repair (BER), or mismatch repair (MMR). In germinal center B cells, MMR, BER, and other factors are diverted from their normal roles in preserving genomic integrity to increase diversity within the Ig locus. Both AID and these components of an emerging error-prone mutasome are regulated on many levels by complex mechanisms that are only beginning to be elucidated.

H, or IgH: immunoglobulin heavy chain

GC: germinal center

AID: activation-induced cytidine deaminase (gene symbol: *aicda*)

SHM: somatic hypermutation

CSR: class switch recombination

V, or V region: variable region of the Ig gene

C, or C region: constant region of the Ig gene

Transitions: mutations that change a pyrimidine into another pyrimidine (e.g., C to T) or a purine to the other purine (e.g., G to A)

Transversions: mutations that change a pyrimidine (C or T) into a purine (G or A) or vice versa

INTRODUCTION

Vertebrates have evolved a complex set of mechanisms to protect themselves from infections and foreign substances. The adaptive humoral response plays an important role in this process by providing antibodies that circulate throughout the body and into secretions, where they bind strongly and specifically to invading organisms and other foreign substances and dispose of them through a variety of effector functions (1). Although all vertebrates can make antibodies, species differ in the details of how they use a small amount of genetic material to generate sufficient antibody diversity to deal with all possible antigens (2). Prior to antigen exposure, mice and humans constantly recreate a highly diverse repertoire of antigen-binding sites in pro- and pre-B cells in the bone marrow through the rearrangement of germ line immunoglobulin (Ig) variable (V), diversity (D), and joining (J) elements to form the heavy (H) and light (L) chain V regions in the Ig genes (1, 3–5). These germ line–encoded IgM antibodies are of low affinity and are usually not effective in inactivating pathogenic organisms and their products. Once an antigen appears, however, cognate mature B cells are stimulated to proliferate, differentiate, and migrate to the dark zone of the germinal centers (GC) in secondary lymphoid organs, where they become centroblasts (6, 7).

In the GC microenvironment, centroblast B cells begin to express large amounts of activation-induced cytidine deaminase (AID), which initiates somatic hypermutation (SHM) of the antibody V regions that encode the antigen-binding sites (8, 9) (**Figure 1**). These point mutations result in the amino acid replacements in the H and L chain V regions that are responsible for the affinity maturation and changes in fine specificity that are required to produce effective neutralizing antibodies (1, 3–5, 10, 11). Centroblast B cells also carry out class switch recombination (CSR), which requires AID and utilizes many of the same mechanisms of mutation and repair that are involved in SHM (12). In this review, we discuss only SHM of the V region genes in mice and humans, but we also draw on some of the studies that have been done with the DT40 chicken B cell line that provide insights into the mechanisms of SHM in mice and humans (13). We focus on the more recent biochemical studies of AID and the characterization of additional genetically defective mice that provided convincing support for and extend the biochemical basis of this model. In this discussion, we draw heavily on and refer to the ideas and data summarized in a number of excellent recent reviews (1, 3–5, 10, 11) as we try to understand how the different enzymatic systems that are involved in SHM are organized and regulated. We first describe general features of SHM, followed by each of the mutation and repair systems involved in V region hypermutation, and then address the issue of the overall regulation of SHM. CSR is discussed in detail in the review by Stavnezer et al. (14) in this volume.

FEATURES OF SOMATIC HYPERMUTATION

If we consider the characteristics of SHM of the H chain V region (**Figure 1**), most of the mutations are single base changes that accumulate starting 100–200 bp from the transcription initiation site and end 1.5–2.0 kb downstream (15–19). The frequency of mutations is highest in the V(D)J coding region and the J region introns, sparing the important regulatory elements in the promoter, the intronic enhancer, and the constant (C) region genes that are responsible for effector functions (17, 19) (**Figure 1**). The frequency of V region mutation is approximately 10^{-5}–10^{-3}/base pair/generation, which can be compared to the basal level of mutation in the genome of $\sim 10^{-9}$. There are more transitions (e.g., C to T, G to A) than transversions (e.g., C to A or G; G to C or T), and many of the mutations are preferentially targeted to the deoxycytidines (dC) within WRC

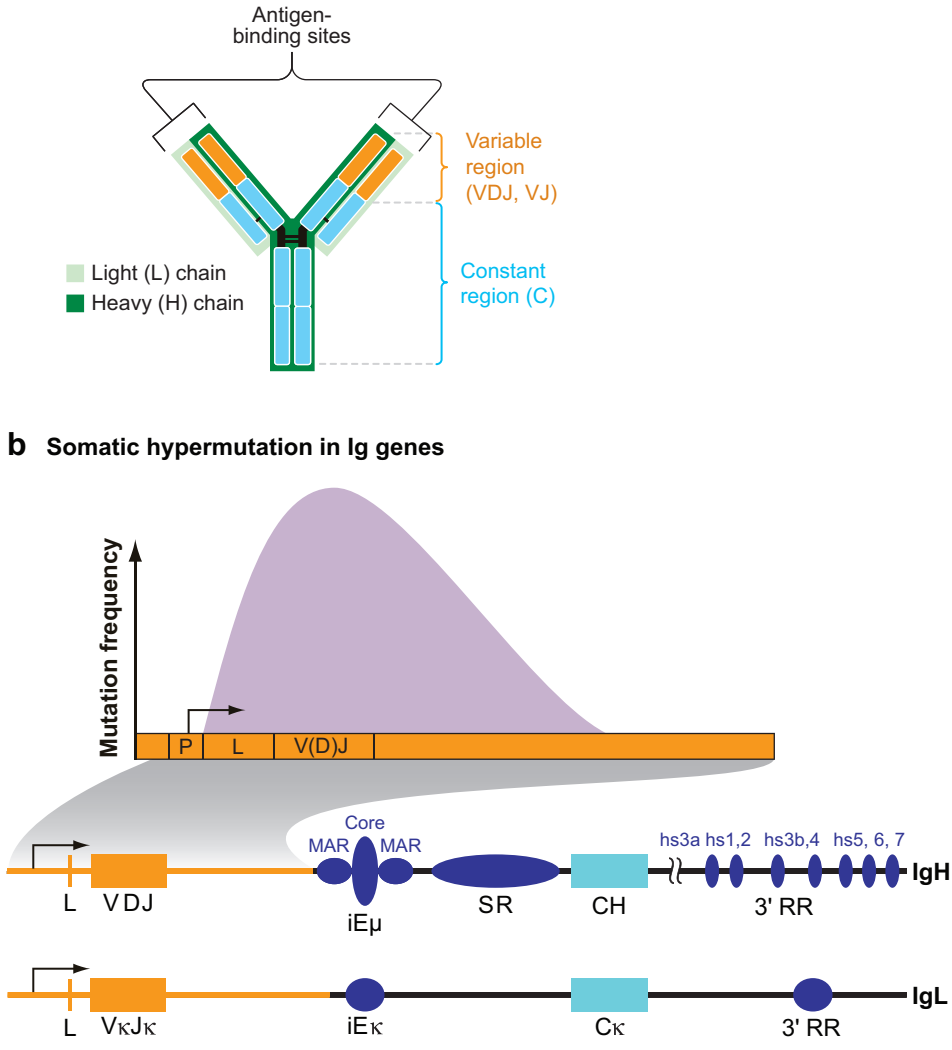

Figure 1

(*a*) Antibody structure. An antibody molecule is a dimer of heterodimers $(H+L)_2$, connected by disulfide bonds. The heavy (H) and light (L) chain V regions (*orange*) form the antigen-binding site, while the C regions (*blue*) form the effector arm. (*b*) Somatic hypermutation in Ig genes. Mutation frequency along the IgH and IgL locus is depicted. Representative IgH and Igκ genes are depicted below the graph that include the following elements: Leader (L), V region [V(D)J], intronic enhancer (iEμ and iEκ), C region (CH or Cκ), switch region (SR), and the 3′ regulatory region (3′ RR). The arrow indicates the transcription start site. The V region (*orange*) experiences significant SHM, whereas the C region (*blue*) does not. SHM is sharply delimited by the V promoter (P) at the 5′ end and starts approximately 100–200 bp from the transcription start site. Mutation frequency is maximal over the V(D)J coding exon and exponentially decays at the 3′ end at 1.5–2 kb downstream from the transcription start site. (Mutation data adapted from Reference 15; boundaries described in References 18, 19.)

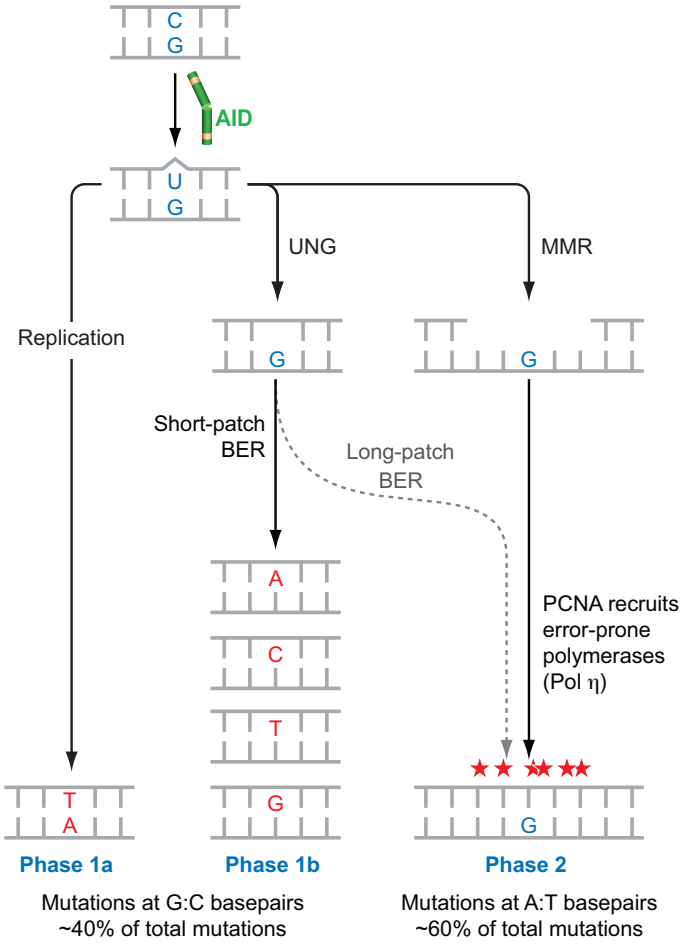

Figure 2

Model of somatic hypermutation. AID deaminates a cytidine residue, creating a uridine:guanosine (U:G) mismatch that is resolved by several pathways that may compete with one another. AID deaminates single-stranded DNA formed during transcription of both strands of the DNA (not shown). The subsequent steps, however, might not occur equally on both strands. (*Left*) The general replication machinery can interpret the U as if it were a deoxythymidine (T). One of the daughter cells will acquire a C-to-T transition mutation. (*Center*) UNG can remove the uracil, leaving behind an abasic site. Short-patch base excision repair (BER) can fill the gap with error-prone polymerases, which can insert any nucleotide in place of the U, leading to transitions and transversions at G:C bases. (*Right*) Mismatch repair (MMR) can recognize the U:G mismatch. The U-bearing strand is excised and, at loci that undergo SHM, monoubiquitylated PCNA (proliferating cell nuclear antigen) recruits error-prone polymerases to fill the gap, leading to transition and transversion mutations at A:T bases as well as at neighboring G:C bases. (*Dashed line*) Long-patch BER can also be a source of mutations at A:T bases and may compete with MMR.

motifs that are hot-spots for SHM (20–22). High rates of transcription are required for SHM to occur, and the frequency of mutation is roughly proportional to the rate of transcription (23–25). Both the transcribed (template, bottom) and the nontranscribed (nontemplate, top) strands undergo AID-induced deamination of dC to deoxyuridine (dU) to produce C-to-T mutations at the same frequency (26). Transcription alone is not sufficient, as many transcribed genes in GC B cells are not targeted for SHM (27, 28).

A general model for the enzyme systems involved in SHM includes a first phase that depends on the mutagenic activity of AID and a second phase that depends on the error-prone repair of the AID-induced mutations (**Figure 2**). Even before the discovery of AID, Rada et al. (29) proposed two distinct phases of SHM because mice deficient in DNA mismatch repair (MMR) exhibited a selective loss of mutations at A:T bases but retained the ability to mutate G:C bases. Mutations at G:C bases were termed phase 1 mutations, with MMR-dependent mutations at A:T bases occurring in phase 2. Further studies (30) demonstrated that mutations detected in the antibody V regions require the propagation of the C-to-U mutation through replication, generating a C-to-T transition, or, in the opposite strand, a G-to-A transition (phase 1a). Phase 1b accounts for transversions at G:C bases, which depend on one of the uracil DNA glycosylase enzymes, UNG, to trigger base excision repair (BER), remove uracil from DNA, and create an abasic site that can be filled in with any of the four bases. Long-patch BER and MMR are responsible for the large number of mutations at A and T bases (3) generated in phase 2.

Although the BER and MMR enzymes normally maintain genomic stability, they become error prone in the GC centroblast B cells when they are acting on the Ig genes and thus greatly increase the number of mutations that accumulate in antibody V regions. A critical conclusion from this model is that in mice and humans ~60% of the mutations

that arise as the result of SHM are in A:T bases, and about half of those are transversions rather than transitions (5). Thus, more than half of the mutations in the V region are not the result of the direct biochemical action of AID, but rather depend on the error-prone BER and MMR of the AID-induced mutation (**Figure 2**). This error-prone repair is contributed by members of a family of low-fidelity translesional DNA polymerases (31). These translesional polymerases, including Pol η, θ, ι, ζ, λ, and REV1, are ubiquitously expressed and appear to have a shallower and less constrained binding pocket than the high-fidelity polymerases. This allows them to recognize DNA lesions and bypass them in newly replicated or mutated DNA by inserting bases opposite the lesion (32). As a consequence of their more permissive binding sites, they are also error prone. The discovery that AID introduces dU mutations at a high frequency in the V regions (8, 9) led us to propose that these translesional polymerases were responsible for the error-prone repair of the dU introduced by AID (33). The discovery of an important role for BER and MMR led to an expansion and refinement of this DNA deamination model by Neuberger and his colleagues (3, 30, 34) (**Figure 2**) that serves as the working model for most investigators in the field. BER and MMR (and perhaps AID itself) participate in large complexes of proteins, the nature of which are under intense investigation. The extent to which these proteins interact with each other and other factors is important because this is one way in which SHM may be regulated. The **Supplemental Table** lists the proteins reported to be involved in SHM and selected interactions among them (follow the Supplemental Material link from the Annual Reviews home page at **http://www.annualreviews.org**).

BIOCHEMICAL BASIS OF AID ACTIVITY

When AID was first discovered, investigators proposed that it worked as an RNA-editing enzyme on an mRNA that encodes a yet to be discovered endonuclease that may initiate SHM (8, 9, 35). Many subsequent studies, however, have revealed that AID initiates the process of V region hypermutation by deaminating dC to dU on single-stranded DNA (ssDNA) (1, 3–5, 10, 11). The role of AID as the initiating factor in SHM was confirmed by studies showing that genetically engineered mice that lack AID and patients with type II hyper-IgM immunodeficiency syndrome (HIGM-2) who have inactivating mutations in AID were unable to carry out SHM (36, 37).

Although there is still much to be learned about the biochemistry and targeting of AID, a critical observation came from the first biochemical studies: In the absence of cofactors, AID deaminates dC to dU on ssDNA, but not on double-stranded DNA (dsDNA), DNA:RNA hybrids, or RNA in any form (38–40). Furthermore, in this cell-free system, semipurified AID preferentially deaminates dC in WRC hot-spot motifs, while exhibiting much lower activity for dC in SYC cold-spot motifs (39, 40). As has been observed in vivo, not every dC in a hot-spot is deaminated, and some non-hot-spot motifs undergo frequent mutation in vitro. AID will only deaminate dC in dsDNA in vitro if there is ongoing transcription at the site or if single-stranded bubbles are otherwise introduced (40–46). Thus, the requirement for active transcription to initiate SHM (and CSR) is explained by the generation of the ssDNA substrate for AID within a moving transcription bubble. These observations provided direct evidence that the substrate of AID was DNA and suggested that the AID protein itself has the inherent information to preferentially deaminate dC in the context of certain sequences such as WRC and to reduce deamination in other sequences such as SYC (38–42, 44–48).

A second important finding from the studies with AID purified from insect cells was that AID has high processivity on ssDNA substrates. After binding ssDNA in vitro,

BER: base excision repair

MMR: mismatch repair

Translesional polymerase: a DNA polymerase that can bypass a bulky DNA lesion, sometimes with the introduction of non-Watson-Crick base pairings

HIGM-2: type II hyper-IgM immunodeficiency syndrome

Hot-spot: a short sequence of DNA where AID-induced point mutations are preferentially found: WRC, where W is either of the weakly hydrogen-bound bases A or T, and R is either of the purines G or A

Cold-spot: a short sequence of DNA where AID-induced point mutations are rare: SYC, where S is either of the strongly hydrogen-bound bases G or C, and Y is either of the pyrimidines C or T

Processive: a processive enzyme catalyzes multiple reactions on a single substrate prior to acting on a different substrate

AID causes multiple deaminations prior to dissociation (39, 40, 49). Although it appears that the processive behavior of AID would allow it to translocate along with a transcription bubble in vitro, generating dC deaminations principally in hot-spot motifs (38–42, 44, 47), evidence is lacking that AID catalyzes multiple deaminations during transcription in vivo because only a few V-gene mutations appear to occur per cell division. An alternative possibility, however, is that AID induces many mutations per cell division but that the majority are rapidly and discreetly repaired in an error-free fashion (see Conclusions).

Another perhaps more compelling teleological rationale for AID processivity is that the enzyme is globally targeted to regions of ssDNA, but not to specific target dC motifs. AID binds with roughly equal affinity to ssDNA that contains either WRC motifs or SYC motifs and even to DNA with no C residues or with only U residues (M.F. Goodman, unpublished data). Therefore, AID would have to remain bound to the DNA to catalyze deamination when it does encounter a dC, perhaps while tracking along with a moving transcription bubble (5, 10, 50). Once bound to ssDNA, AID catalyzes deamination in what appears to be a "hit-and-miss" process of the sort described for restriction enzymes rather than for polymerases (39, 48, 51). For example, deamination patterns on individual DNA clones exposed to AID in vitro contain small clusters of deaminations, often separated by lengthy regions where dC is left intact even though there are numerous WRC hot-spot motifs. Each DNA clone exhibits a different deamination pattern (39, 40, 51), and, similarly, clonal patterns have been reported for different B cell clones bearing the same Ig transgene (52). When deaminations do occur, WRC hot-spot motifs are favored over SYC cold-spots by about 6:1. Multiple deaminations in vitro tend to congregate near WRC sites, suggesting that after a hot-spot deamination, AID can slide to and attack a proximal C residue and then jump to another region on the same ssDNA strand. The biochemical data suggest that AID binds randomly to ssDNA and performs a bidirectional random search for C residues by jumping and sliding along the DNA backbone (**Figure 3**). This behavior of AID reflects special properties of the enzyme because it differs significantly in this cell-free system from one of its homologs, APOBEC-3G (Apo3G), which favors deamination toward the 5′ region of ssDNA (51). It is unclear how the processive properties of AID are used in vivo, where directional deamination might be imposed upon it by the 5′ to 3′ motion of RNA polymerase II (RNA Pol II), which has been reported to associate with AID (53) (see the **Supplemental Table**). An important biochemical challenge will be to recapitulate the salient features of transcription-dependent deamination with purified mammalian enzymes.

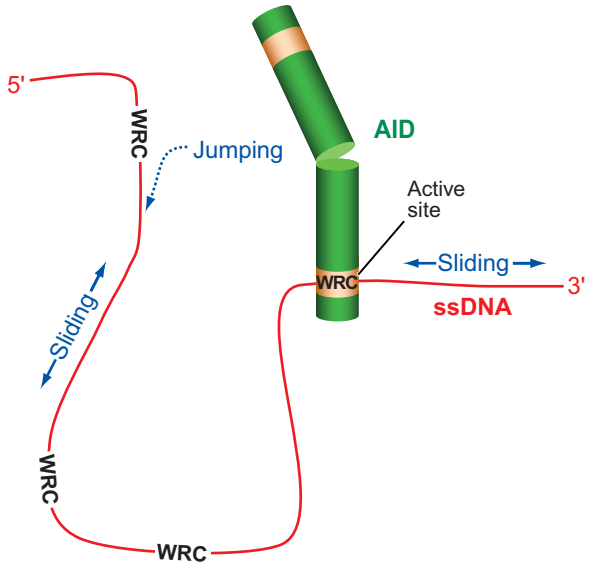

Figure 3

Model describing processive C-to-U deaminations. AID is depicted as a dimer on ssDNA. Current biochemical data suggest that AID binding occurs randomly, and enzyme motion, for example sliding and jumping, occurs in either direction along the ssDNA substrate. Deamination by AID occurs processively (making multiple deaminations per substrate molecule) and equally in 5′ and 3′ directions, with preferential targeting to WRC motifs. Notably, there is no external energy source present, for example ATP or GTP hydrolysis.

Strand-Specific Targeting

On the basis of the biochemical studies described above, we might expect that only the nontranscribed strand would be attacked by AID. In fact, a single cytosine in the nontranscribed strand is sufficient to recruit AID-induced deamination and allow the mutation of upstream and downstream A and T residues, whereas in one experimental system this does not seem to happen when the cytosine is in the transcribed strand (54). The analysis of this strand asymmetry in vivo has revealed that error-prone MMR preferentially targets dU in the nontranscribed strand, suggesting different post-AID repair of the two strands. However, mutations at dC sites in cells lacking both BER and MMR occur with roughly equal frequencies on both DNA strands (3, 55, 56). Assuming that the deamination events that trigger mutation occur during V-gene transcription, one possibility is that both strands can undergo transcription to provide ssDNA on which AID can act, and this has now been observed in Ramos cells, a human centroblast-like B cell line that constitutively undergoes SHM (57). As far back as 1992, antisense promoters were identified in another part of the Ig locus, the Ig μ-switch regions that are associated with *c-myc* in a Burkitt's lymphoma cell line (58), and antisense transcription has been reported in other regions of the Ig gene (59, 60). However, if bidirectional transcription is not occurring in vivo, then how C-to-T mutation occurs on both strands is unclear. One proposal is that DNA supercoiling may generate transient single-stranded regions on the transcribed strand (5, 47, 50, 61). Although the use of T7 polymerase has revealed only occasional deamination of the nontranscribed strand (48), deamination of both strands has been reported using *Escherichia coli* polymerase in vitro (44) and in the *E. coli* chromosomal *rpoB* locus when AID is overexpressed (30). The addition of AID to a mammalian transcription system may provide a clearer answer to how both strands are targeted in vivo.

AID Protein Structure

A sketch of the AID domain structure is shown in **Figure 4**. HIGM-2 disease-causing mutations are found throughout the protein

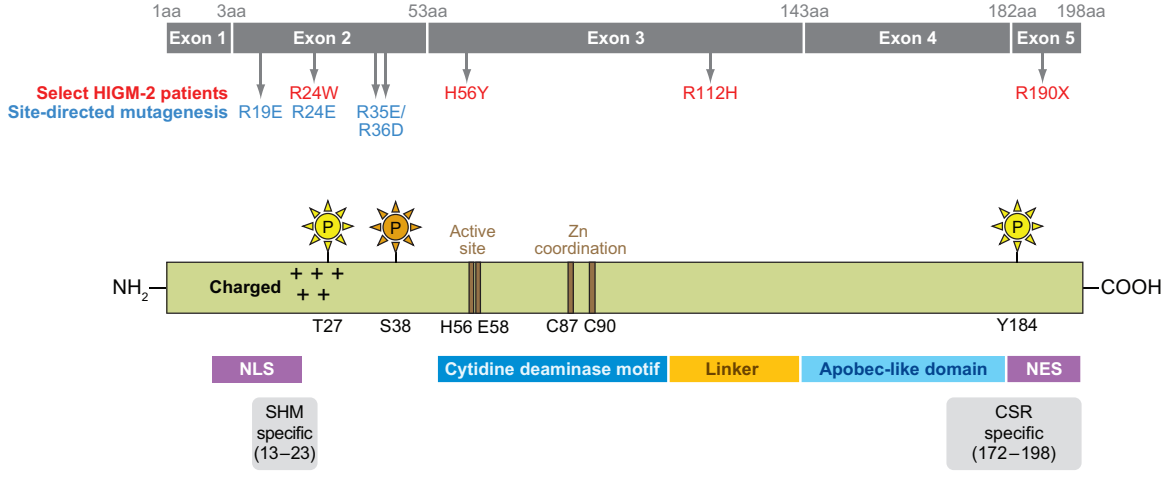

Figure 4

Inactivating mutations and functional domains in AID. The exon domains of AID are depicted (aa, amino acids). Select inactivating mutations found in HIGM-2 patients (*red*) or generated by site-directed mutagenesis (*blue*) that result in changes in amino acids are indicated with arrows off of the exon domain diagram. The AID protein is depicted in green, with critical residues indicated and functional domains drawn below. Known phosphorylation sites are represented by yellow or orange sunbursts. The model is not to scale.

(**http://bioinf.uta.fi/AICDAbase**), and most result in loss of both SHM and CSR (62). Some of these mutations (shown in **Figure 4**) and complementary studies using site-directed mutagenesis have provided clues to the functional domains of AID (39, 63, 64). The C-terminal portion of AID is required for CSR but not for SHM (63–65). This region is not required for deamination: Biochemical analysis of the mutation spectrum generated by a C-terminal deletion mutant of AID revealed close resemblance to that generated by wild-type AID (39). Thus, some aspect of AID targeting to switch regions versus V regions resides within these last few amino acids, perhaps via the presence of CSR-specific cofactors (63, 64).

The N-terminal region of AID contains a remarkably high concentration of basic amino acid residues, resulting in a +11 net positive charge (39). An AID double mutant (R35E/R36D, **Figure 4**), in which the N-terminal charge has been reduced to +7, exhibits reduced processivity in vitro (39), which is not surprising if it binds with lower affinity to the negatively charged ssDNA backbone. Yet this double mutant also shows a change in specificity in that the highest deamination rate occurs in a non-WRC motif (39). This modified deamination specificity is surprising because the R35E/R36D N-terminal mutations are far removed from the catalytic region.

Our understanding of the functional properties of AID would be helped significantly by structural data. Although there is as yet no crystal structure for AID, a member of the APOBEC family of enzymes, APOBEC-2 (APO2), has been crystallized and shown to form a rod-shaped tetramer (66). Owing to a significant degree of sequence similarity, APO2 can be used as a surrogate for AID to make structure-based predictions. First, AID complexed with Zn is necessary for deamination activity. The APO2 structure suggests that Zn hydroxylation may regulate substrate access and enzyme activity. A site-directed mutation at the equivalent site in AID (R19E, **Figure 4**) was predicted to alter Zn hydroxylation and, when tested in vitro, resulted in a ~fivefold reduction in AID deamination activity (66). Second, AID has been identified as a dimer (63, 71, 162), and mutations predicted to disrupt a putative dimer interface (F46A/Y48A) resulted in a fourfold loss of activity (66). Finally, a residue in APO2 is predicted to help stabilize a β1′-hairpin conformation and thus allow access to nucleic acid substrate. When the corresponding residue is mutated in AID (R24E), deamination activity is completely abrogated. The residues of AID that are mutated in HIGM-2 syndrome, causing impaired production of high-affinity antibodies, are well conserved in APO2. Notably, the aforementioned R24 residue is mutated in some HIGM-2 patients, whereas additional HIGM-2 mutations occur on the predicted surface of an AID monomer (66). Despite biochemical and physical (66) evidence that AID likely works principally as a dimer in vivo, we cannot rule out that an AID monomer or even higher polymers may also be active (41).

Posttranslational Modifications

AID isolated from stimulated primary B cell nuclei is phosphorylated at multiple sites, including Ser38 (68–70) (**Figure 4**). Only about 10% of the protein contains phosphate at this residue, and, interestingly, the phosphorylated form is enriched in the chromatin fraction (69). AID expressed in *E. coli*, which is not expected to be phosphorylated, deaminates transcribed dsDNA in vivo (43, 46), and protein partially purified from *E. coli* deaminates dC residues on ssDNA in vitro (43–45). Thus, phosphorylation per se is not necessary for AID activity.

It has been reported that AID enriched from B cells must be phosphorylated at Ser38 in order to deaminate linear dsDNA undergoing transcription with T7 RNA polymerase in a reaction that also requires replication protein A (RPA), a factor that binds ssDNA (71). Human AID obtained from insect cells is also phosphorylated at Ser38, yet it is able

to deaminate ssDNA in a T7-based transcription assay in the absence of added RPA (39, 40). Unphosphorylated AID expressed in *E. coli* also appears to be active in the setting of in vitro transcription assays without added RPA (43, 44). Therefore, uncertainties concerning the role of phosphorylation at Ser38 and perhaps at other sites remain an outstanding issue. At present, phosphorylation at Ser38 appears to play a role in the efficiency of SHM. In activated B cells deficient for AID, a Ser38Ala mutant exhibits significantly delayed appearance of SHM (69). Other residues in AID may be phosphorylated as well (**Figure 4**). A thorough biochemical comparison using phosphorylated and nonphosphorylated AID could shed light on how AID interacts with ssDNA, RPA, and perhaps transcription factors that may serve to target AID to DNA undergoing transcription.

AID Splice Variants

Another aspect of AID that needs to be examined in more detail is the role, if any, of the various isoforms that have been observed. Alternative transcripts of AID have been reported in asthmatic patients (72), human B cell non-Hodgkin's lymphomas (B-NHL) (73), chronic lymphocytic leukemia (74–77), and normal B cells stimulated with CD40L and IL-4 (74, 75). The constitutive expression of AID and its splice variants may contribute to B-NHL formation (73). Five different transcript variants of AID have been detected: (*a*) the full-length AID transcript (36), which is the most prevalent AID transcript in healthy and neoplastic B cells (73); (*b*) a variant that lacks the first 30 bp of exon 4 (72, 74, 75); (*c*) a variant that lacks all of exon 4 (72–75); (*d*) a variant in which intron 3 is retained (73, 76, 77); and (*e*) a variant that includes a short neo-exon located in intron 3, but lacks exons 3 and 4 (73). The biochemical examination of these isoforms would provide useful structural and functional information and may also provide some insights into the role of AID in B cell malignancies.

BASE EXCISION REPAIR

BER is a DNA repair pathway in which altered bases are removed by a DNA glycosylase, followed by subsequent steps to repair the lesion. Approximately 60% of the mutations that accumulate in vivo in mice and humans are in A:T bases and are not caused by the direct action of AID. As illustrated in **Figure 2**, once AID has mutated dC to dU in DNA, the uracil may be either replicated or excised by UNG to create an abasic site. This intermediate can be converted into a single-stranded break by apurinic/apyrimidic endonucleases (APEs) (not shown in **Figure 2**) that can in turn be repaired by error-prone polymerases. This UNG-dependent pathway can generate both transition and transversion mutations, whereas replication yields only transitions (78, 79).

UNG is primarily responsible for initiating BER of the Ig genes in centroblast B cells, and its genetic inactivation causes a profound defect in both SHM and CSR in mice and humans (34, 80), whereas the genetic inactivation of other uracil DNA glycosylases has little or no effect (3, 81, 82). The dominance of UNG in SHM (and CSR), compared with the many other uracil DNA glycosylases that are available (83, 84), illustrates a common theme in the SHM of the Ig genes: Even in the presence of considerable redundancy, often one particular enzyme and its downstream protein partners are hijacked by the centroblast B cell to repair the antibody V regions in an error-prone manner.

A great deal has been learned in recent years about high-fidelity BER that may be relevant to the error-prone BER of Ig genes. For example, it is now clear that the abasic sites generated by UNG and the single-stranded breaks created by the action of APE1 can be repaired by either of two alternative pathways: short-patch repair of one base or long-patch repair that involves the excision of 2–8 or

Short-patch repair: BER subpathway where the length of the repair patch is exactly one nucleotide

Long-patch repair: BER subpathway where the length of the repair patch is 2–8 nucleotides or possibly more

possibly more bases and their replacement (84, 85) (**Figure 2**). The repair of AID-induced mutations by BER can in principle be limited to mutations at single dC bases after the dU has been removed by UNG, in which case it is called short-patch repair. Because replication over the uracil that arises from the direct action of AID results only in transition mutations, transversion mutations at C or G could arise as a result of such short-patch BER. In animal model systems, the absence of transversion mutations at C and G residues is used to identify defects in this aspect of BER in SHM. Because C-to-G transversions are lost in mice that are genetically defective in REV1, this translesional enzyme is thought to be primarily responsible for the short-patch BER of abasic sites in SHM (78). As with UNG, REV1 is yet another example of one of the many potential enzymes that is preferentially recruited to carry out error-prone repair of AID-induced mutations in the Ig V regions. Recent studies suggest that the switch from high-fidelity polymerases to error-prone polymerases is mediated by the monoubiquitylation of proliferating cell nuclear antigen (PCNA) (86–89) (see below) so that modified PCNA may be required to recruit REV1 to repair the abasic sites or the ssDNA breaks that arise in short-patch BER.

Barring the existence of an A:T deaminase, mutations at A:T bases must be brought about by excision of bases surrounding the initially targeted dC. One important pathway that carries this out is MMR (discussed below). However, MMR-deficient mice still accumulate some A:T mutations. Interestingly, all A:T mutations disappear in MSH2-UNG (90) and MSH6-UNG double-deficient mice (56), suggesting that a UNG-dependent, BER pathway also contributes to A:T mutagenesis. Since short-patch BER only acts at the initially targeted dC (Figure 2), it cannot be responsible for those residual A:T mutations seen in MMR-deficient mice. Therefore, it is likely that long-patch BER contributes to A:T mutagenesis, although the exact enzymatic players are not as well understood as they are in MMR. In addition, there is some evidence that the MRN (MRE11-RAD50-NBS1) complex, normally involved in double-stranded break repair, can also participate in this process, perhaps in concert with or downstream from BER (84, 92, 93).

Once UNG has removed the uracil, it is also unclear what regulates the recruitment of short-patch and long-patch BER during SHM. Although uracils that arise during replication more often undergo long-patch repair at least in some cell lines (91), this may not be true of AID-induced mutations in centroblast B cells. In general, the presence of PCNA at a DNA lesion seems to favor long-patch over short-patch repair, whereas Pol β tends to favor short-patch repair (85). We now know that monoubiquitylated PCNA (see below) is involved in SHM and likely facilitates A:T mutagenesis in the Ig gene via both MMR and long-patch BER. (86–89; S. Roa & M.D. Scharff, unpublished data).

Another emerging principle is that the relative abundance of the various players makes a critical difference in determining error-prone versus error-free repair of AID-induced mutations. This is nicely illustrated in the case of the BER of V region mutations where Pol β, which usually mediates the high-fidelity postreplicative short-patch repair of abasic sites, is downregulated in SHM-proficient BL2 cells, a centroblast-like human B cell line (94). This appears to allow the error-prone polymerases, especially REV1 and perhaps Pol η or other translesional polymerases, to replace Pol β and allow for error-prone repair in SHM. In Pol β–deficient B cells under certain experimental conditions, CSR can be more proficient, although the levels of Pol β are not modulated during CSR in wild-type mice (78, 95).

Taken together, all these studies suggest that both short-patch and long-patch BER play a critical role in the SHM of mouse and human V regions. Because high-fidelity postreplicative BER is also important in maintaining genome stability, it is unclear how the centroblast B cell targets error-prone

BER to the Ig. One possibility, which is discussed below, is that AID mutations and their error-prone repair may be restricted to the nonreplicative G_1 and G_2 phases of the cell cycle. It is probably also important to understand how BER and MMR compete to repair the dU mutations produced by AID. That competition could be between UNG and MSH2-MSH6 or for factors such as RPA, PCNA, and the error-prone translesional polymerases that are used in both BER and MMR. Because monoubiquitylated PCNA can recruit error-prone polymerases, the state of modification of PCNA could determine whether there is long- or short-patch BER of V region mutations and whether it is error free or error prone. Most of the translesional error-prone polymerases are not processive and add only one, or a few, bases. The repair of longer patches of excised DNA either in BER or MMR (see below) requires that a more processive DNA polymerase like Pol ζ and/or high-fidelity polymerases like Pol δ or ε assist the translesional enzyme in replacing the excised DNA (32). One report, for instance, has suggested that Pol η may replace the first nucleotide, followed by extension of the new strand by Pol θ (79). There could also be competition for these translesional enzymes between long-patch BER and MMR. Because these different BER processes are mediated by different complexes (96), it would be interesting to compare the protein complexes that are present in centroblast B cells at different stages of the cell cycle to resolve some of these questions.

MISMATCH REPAIR

As with BER, the mutagenic role that MMR plays in SHM is surprising, considering its critical role in maintaining the integrity of the genome. MMR is a complex process requiring the sequential action of many proteins to increase the overall fidelity of DNA replication and the repair of genotoxic damage (97, 98). Postreplicative MMR is conserved from prokaryotes to mammals, and in eukaryotic organisms it is mediated by the following proteins: (1) the MSH2-MSH6 heterodimer (called MutSα) that recognizes single mismatched base pairs, or MSH2-MSH3 (MutSβ) that recognizes larger mismatches or loops resulting from deletions or insertions; (2) following the initial binding of MSH2-MSH6 or MSH2-MSH3 to the mismatch, there are a series of ATP-dependent events that lead to the recruitment of MLH1 and PMS2 (or MLH1-MLH3) that recruit other downstream elements and introduce a single-stranded nick near the mismatch (99); (3) exonuclease 1 (EXO1), which excises the mismatch and a yet to be determined stretch of the surrounding DNA strand; (4) Pol δ and ε, high-fidelity polymerases that resynthesize the excised DNA stand; and (5) DNA ligase I, which ligates the ends. Other proteins that are also involved in long-patch BER, such as RPA, replication factor chaperone-like complex (RFC), and PCNA, play an important role in MMR (97, 98) (see **Supplemental Table**).

Like BER, the centroblast B cell has hijacked the MMR process and made it error prone when it encounters AID-induced G:U mismatches in the Ig V regions (**Figure 5**). Error-prone MMR accounts for more than 50% of all the mutations and for most of the transversion mutations at A:T bases (3, 100). The genetic inactivation of both *Ung* and either *Msh2* or *Msh6* leads to a loss of virtually all mutations at A:T bases in V regions and a profound loss of CSR (55, 56). In confirmation of a direct role for MMR in SHM and CSR, chromatin immunoprecipitation showed that EXO1 and MLH1 are associated with mutating V regions in the BL2 cell line and that MSH6 is associated with the switch regions involved in CSR in primary B cells (101, 102). As described by Stavnezer et al. (14), MMR is also important in CSR.

Because the mutations introduced by AID create single-base G:U mismatches and because short-patch BER generates G:abasic site mismatches, it is not surprising that the absence of Msh3, which recognizes larger

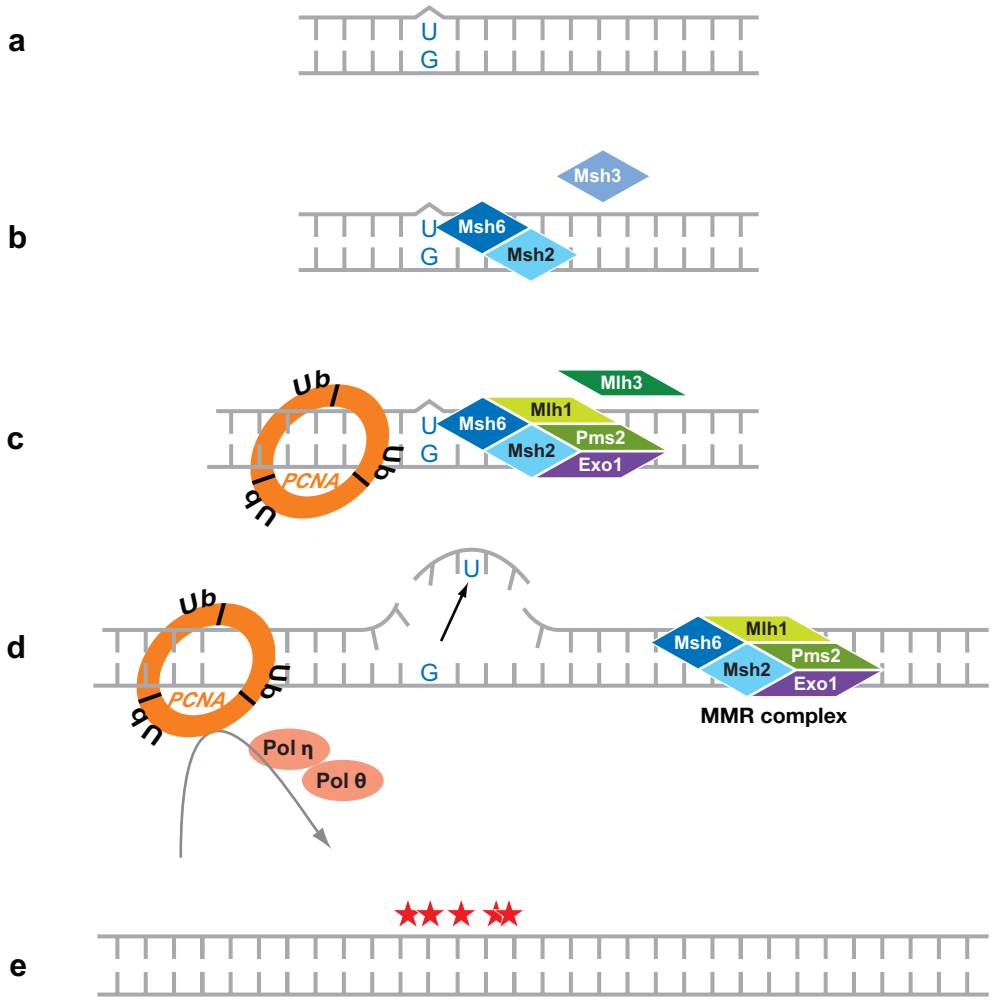

Figure 5

Mismatch repair in SHM. A model of phase II SHM mediated by the mismatch repair pathway and error-prone polymerases. (*a*) A uridine (U) created by AID deamination of a cytidine creates a mismatch. (*b*) The Msh2-Msh6 complex recognizes the mismatch. Msh3 competes with Msh6 for Msh2. (*c*) Other components of the mismatch repair pathway are recruited, including Pms2, Mlh1, and Exo1. Each subunit of the homotrimeric PCNA can be monoubiquitylated (Ub). Mlh3 competes with Pms2 for Mlh1. The exact nature of the MMR complex is unknown. (*d*) Exo1 excises the U-bearing strand, and monoubiquitylated PCNA recruits error-prone polymerases, especially Pol η, to resynthesize the gap. (*e*) The region near the deaminated residue acquires point mutations (*red stars*), including transitions and transversions at A:T bases. The newly transcribed strand is ligated in place (not shown).

mismatches, has little effect on SHM (103–105). Although both Msh2 and Msh6 bind DNA, only Msh6 bears a conserved Phe-X-Glu motif that directly interacts with mismatched nucleotides and is inserted into the minor groove of the helix at the mismatch site, bending the DNA some 60° (98, 106). In addition to triggering the recruitment of the other proteins required to excise and repair the mismatch (**Figure 5**), the binding of the

MSH2-MSH6 heterodimer also informs the cell of the presence of damaged DNA and can either directly or indirectly activate the DNA damage response pathway, including the ATR kinase and the CHK1/CHK2 G_2/M checkpoint proteins, causing cell cycle arrest and signaling for apoptosis.

Since error-prone MMR is responsible for so many of the mutations that accumulate in the V region, it is important to understand how this process is regulated. How does error-prone MMR continue to maintain the integrity of most of the genome of the rapidly dividing centroblast B cell, while simultaneously contributing many of the mutations in the V region? Not surprisingly, the proteins that are involved in MMR are highly expressed in rapidly dividing centroblast B cells that are undergoing SHM (107). There is probably tight control of the relative abundance of each of the MMR proteins because the inactivation of MLH3, which can also form a dimer with PMS2 (**Figure 5**), results in a change in the spectrum of mutations in vivo (108), suggesting that in wild-type mice it is competing with PMS2 to form complexes that are acting on the Ig gene. Studies with genetically defective mice and patients with the variant form of Xeroderma Pigmentosum, in whom Pol η is mutated, clearly demonstrate a critical role for Pol η in the introduction of transversion mutations at A:T bases that are the hallmark of the MMR-dependent second phase of SHM (109, 110) (**Figures 2** and **5**). Moreover, the complete absence of mutations at A:T bases in MSH2-Pol η double-deficient mice indicates that the residual A:T mutagenesis in the single MMR-deficient mice is contributed by Pol η (110). Additional studies in Pol θ–defective mice suggest that it works in concert with Pol η and other more processive, high-fidelity DNA polymerases to fill in the gap created by EXO1 (11, 79, 111–113).

The question then becomes how these error-prone polymerases are recruited to the repair of the AID-induced mutations in the Ig V region and not to the postreplicative repair of uracils that are introduced during the normal course of replication. The beginning of an answer comes from studies in yeast showing that the recruitment of these error-prone polymerases is mediated by PCNA that has been monoubiquitylated at residue 164 (32, 114–116). Although polyubiquitylation is a well-known mechanism for targeting proteins for degradation in the proteasome, the addition of a single ubiquitin monomer is a posttranslational modification that recruits and activates various repair pathways. PCNA is a homotrimer that is central to all forms of DNA replication and serves as a sliding platform that recruits polymerases and other factors. These interactions are competitive and allow PCNA to serve as a central organizing factor, or traffic cop, for DNA repair and replication (117). In eukaryotes, the monoubiquitylation of PCNA at residue 164 is mediated primarily by the Rad6-Rad18 ubiquitin ligase pathway, and the levels of monoubiquitylated PCNA are adjusted by deubiquitylation, primarily by USP1 (118). During normal DNA replication, if replication forks stall because of bulky lesions, PCNA becomes monoubiquitylated so that it can recruit translesional enzymes that bypass those lesions. In addition, PCNA plays a critical role in the resynthesis of the excised strand in MMR and seems to be involved at some level in the recognition of MMR enzymes; it physically interacts with MSH6, MSH3, MLH1, and EXO1 as well as with the high-fidelity and low-fidelity polymerases (**Supplemental Table**).

Very recently, a few reports addressed the role of PCNA in recruitment of error-prone polymerases by using modified DT40 chicken B cell lines that have been rendered defective in the Rad6-Rad18 pathway or that express PCNA that cannot be ubiquitylated because of a lysine-to-arginine mutation at 164 (86–88). Although DT40 cells do not seem to target A:T bases for mutation, their inability to monoubiquitylate PCNA is associated with a decrease in SHM (87, 88). These studies have now been extended to mice, in which the expression of PCNA with a lysine-to-arginine

mutation at 164 has resulted in a significant loss of mutations at A:T bases, which is comparable to that seen in mice that are genetically defective in Pol η (89; S. Roa & M.D. Scharff, unpublished data). Although many details of this modification in mammalian B cells must still be worked out, the recruitment of monoubiquitylated PCNA to V region repair provides a possible mechanism for how error-prone polymerases can be recruited to Ig V regions (see **Figure 5**). This may be another example of the hijacking of a normal process by the centroblast B cell to generate SHM. Nevertheless, it remains unclear how PCNA, or any other protein, could distinguish AID-induced mutations from mutations that arise during normal DNA replication and repair. This is discussed below in the context of the targeting of SHM.

TARGETING OF SHM

Historically, most of the studies on the targeting of SHM have been conceptually organized around how AID itself is targeted to the V regions of Ig genes in centroblast B cells. However, because more than half of the mutations are dependent on error-prone BER and MMR, it is also important to determine whether the G:U created by AID is sufficient to recruit those processes and mediate their switch from error-free to error-prone repair or if other factors are involved. The targeting of SHM probably occurs at multiple levels and in multiple layers. The critical and recurring question now is how the mutagenic properties of AID combine with error-prone BER and MMR to target parts of the Ig V region for very high rates of mutation, whereas some other genes accumulate fewer mutations, and C regions and most of the genome are spared or protected.

Expression of AID

As AID is the initiating factor in SHM, its expression is, not surprisingly, an important regulatory step. In mice and humans, AID is highly expressed in centroblast B cells in the GC microenvironment (9). This ensures that high levels of SHM of Ig V regions will occur during a brief stage of B cell development and will then be turned off in the memory and plasma cell stages of B cell differentiation. However, AID expression and lower rates of SHM have been observed in T-independent responses and in immature B cells (119, 120), and AID protein has been detected in vivo in extrafollicular B cells (121). Recent studies with AID indicator mice (122) confirm the original observations that AID is primarily restricted to the GC stage of B cell development. When AID is artificially expressed in non-GC B cells (such as hybridomas) or in non-B cells (such as CHO cells), the mutation rate is at least tenfold lower (123, 124). Surprisingly, mice engineered to constitutively express AID in all cells succumb to T cell but not B cell malignancies (125). These observations suggest that other factors must synergize with AID to produce effective SHM in the B cell, and there must be protective measures in place in B cells to prevent their malignant transformation.

At a cellular level, most of the AID protein in centroblast B cells is cytoplasmic, and it shuttles back and forth from the cytoplasm to the nucleus, with only 10%–15% located in the nucleus (68, 126, 127). Although there are both nuclear localization and nuclear export motifs in AID (**Figure 4**), the subcellular localization of AID is likely regulated by posttranslational modifications such as phosphorylation (see above) and associated proteins. Once modified AID and its presumed associated proteins have entered the nucleus, they must be recruited to the Ig V regions and largely restricted from acting on other sites. In normal individuals several genes outside of the Ig locus, including *Bcl6*, *CD79*, and *CD95*, undergo SHM, albeit at much lower levels than in the Ig V region (128–130). In malignant B cells, several other loci, including *Bcl6*, *Pim1*, *Myc*, *RhoH/TTF*, and *Pax5*, can undergo aberrant SHM (131, 132). AID will even mutate many different types of reporter genes if they are highly expressed in cultured

B and non-B cells and in vivo (124, 125, 133). Yet not all highly transcribed genes accumulate AID-induced mutations in centroblast B cells. These findings and the AID-inactivating mutations in patients (**Figure 4**) indicate that there must be other *cis* or *trans* factors that are responsible for the accumulation of mutations in the Ig V region and for the lower rates of mutation in genes like *Bcl6* and *c-myc* and in reporter genes that are distributed throughout the genome (134, 135). The importance of these other putative factors is also consistent with the surprising finding that when naive splenic B cells are stimulated in short-term culture to express high levels of AID and undergo high rates of CSR, the V regions of those same activated B cells are not mutated (136). Another surprising finding that suggested other cellular factors is that AID conditionally overexpressed in B cells leads to fewer mutations, indicating mechanisms that specifically negatively regulate AID (137).

The many detailed studies addressing the mechanism of AID targeting in SHM are the subject of recent reviews (5, 50, 138, 139). Several explanations have been explored, including (*a*) that the V region is made more accessible to AID than is the C region or other genes in centroblast B cells; (*b*) that there are *cis*-acting sequences in or around the V region, and perhaps a few other genes, that either recruit DNA-binding proteins which in turn recruit AID or form macromolecular DNA structures that serve as a nidus for the recruitment of AID; (*c*) that there are proteins that associate with AID especially in B cells that both target it to the V region and perhaps restrict it from acting on most other parts of the genome; and (*d*) that AID-induced mutations and/or error-prone repair occur only during the nonreplicative stages of the cell cycle, temporally segregating SHM from replication and error-free repair, and there are subnuclear domains in which the targeted Ig genes and SHM enzymes are compartmentalized. Importantly, many of these mechanisms could also be used to regulate or target error-prone BER and MMR. We offer a brief synopsis of evidence supporting each of these models; however, these hypotheses are not mutually exclusive.

Accessibility

Regulation of transcription, repair, and replication of particular genes is often associated with epigenetic changes such as changes in DNA methylation and modifications of histones. These changes regulate the accessibility of those genes to the protein complexes responsible for these DNA transactions. Investigators have surmised that a similar paradigm applies to SHM: The selective targeting of AID and error-prone BER and MMR could be facilitated by increased accessibility of the V region in SHM (and of the switch regions in CSR). Recent studies in transgenic mice have suggested that DNA demethylation of cytosines early in B cell development may play a role in targeting the active kappa light chain allele for SHM (140). These studies showed that SHM occurs with a tenfold preference on the unmethylated allele compared with an identical methylated allele of the Igκ gene. Earlier studies in the IgH gene and the Igλ gene found no difference between V regions versus C regions with respect to DNA methylation (141), but those studies did not address specific alleles. How DNA methylation can affect targeting of AID is unclear, as both methylated and demethylated alleles have similar rates of transcription in vivo, and the presence of a methylated CpG does not have a negative effect on AID deamination activity on nearby cytosine residues (142).

As described in the review on CSR by Stavnezer et al. in this volume (14), a number of studies in primary B cells have shown that the histones associated with AID-targeted switch regions that undergo CSR are hyperacetylated compared with switch regions that do not undergo CSR (53, 143, 144). As occurs in the switch regions, the acetylation of the H3 and H4 histones associated with the H chain V region increases when compared with

Chromatin modifications: posttranslational modifications of histone residues, such as acetylation, that can change the local accessibility of genes to various protein complexes

***cis*-acting elements:** sequences that recruit protein factors in a sequence-specific manner or that form higher-order structures (e.g., stem-loops) that recruit protein factors

the C region in BL2 cells (101). When BL2 cells were treated with trichostatin A (TSA), a drug that globally inhibits histone deacetylases, mutations were observed to accumulate in the C region and within the first 100 bases downstream from the transcription start site. Because mutations are not usually observed in these two regions of the Ig gene (**Figure 1**), this effect of TSA suggests that changes in histone acetylation can affect the targeting of AID. The global hyperacetylation effects of TSA, however, make it difficult to demonstrate that histone modifications are required for proper targeting of SHM.

Hyperacetylation of H3 and H4 histones in the H chain V region compared with the C region has also been observed in primary B cells in vivo (141). These differences in chromatin modifications preceded the expression of AID, suggesting that if they play a role in the targeting of SHM, they may be part of a broader program of regulation that prepares the V region so it can be targeted by AID. In this study, there was no difference in histone acetylation between the V and C region of the λ light chains from these same mice, but this particular C region is only 1.6 kb away from the promoter and still within the domain that can be targeted for SHM. Because histone H3 and H4 hyperacetylation are also marks of transcribed genes, it is difficult to know whether this chromatin modification pattern is a reflection of transcription per se or important for SHM targeting in this context.

In the same study, phosphorylation of histone H2B-Ser14 (H2B-pSer14) was associated with V regions of both L chain and H chain genes, but not with C regions. On the one hand, this chromatin modification pattern was dependent on AID, as AID-deficient mice did not exhibit these changes, which suggests that H2B-pSer14 is unlikely to be responsible for the initial recruitment of AID. On the other hand, this chromatin modification could play a role in the recruitment of error-prone BER or MMR, a possibility that remains to be explored. Thus, there is circumstantial but not definitive evidence that specific epigenetic changes in V regions translate into increased accessibility to SHM enzymes and play a role in targeting SHM. In addition, it is likely that a combination of chromatin modifications work in parallel to target the action of SHM.

Cis-Acting Sequences

Analogous to transcription, in which particular DNA motifs recruit specific transcription factors, *cis*-acting sequences may recruit AID, BER, and MMR to the Ig V regions in centroblast B cells. The Ig genes are rich in promoters and enhancers that have been well studied over the years for their roles in transcription and activation of the locus (**Figure 1**). Because deletion or mutation of these elements often compromises transcription rates, it has been difficult to distinguish between the loss of specific SHM targeting that leads to decreases in mutation levels and loss of SHM owing to decreases in transcription. Nevertheless, many studies of the endogenous Ig locus or of ectopically located transgenes in cell lines and mice have attempted to identify *cis*-acting sequences that might recruit SHM enzymes to the Ig V region (reviewed in 138, 145). Although the results of these studies are conflicting, likely owing to variability within transgenic mice and within or between B cell lines, in general the studies of L chain genes reveal that the SHM machinery could be activated at the correct time during B cell differentiation and could be recruited to almost any ectopically located reporter gene, as long as it is flanked by a strong transcriptional promoter and carries both the intronic and 3′ L chain enhancers. These observations suggest that the V region coding exon itself does not contain sequences that are necessary for the targeting of AID but that important motifs must reside in the promoter and/or enhancer regions. However, any strong promoter seems able to target SHM to the V region and, at least in cell lines, even viral enhancers are sufficient (146, 147). Recent studies in the chicken B cell line, DT40, in which it is easier to examine different

modifications of promoters or enhancers, have suggested that the situation may be more complicated. In DT40 B cells, the human EF1-α promoter, which induces as much transcription as the beta-actin or the endogenous L chain promoters, did not support high levels of SHM in the endogenous L chain (148).

With respect to the H chain gene, a variety of promoters were able to target SHM to the V region or to other genes that were inserted in place of the V region. Studies in which the core Eμ intronic H chain enhancer was deleted resulted in mice that had a small but not significant effect on SHM, suggesting that it does not play an important role in the targeting of SHM (149). Similarly, deletion of various known parts of the endogenous 3′ regulatory regions of the H chain did not affect SHM in mice (150), but the roles of the recently described elements hs5, 6, and 7 have yet to be determined (151). However, some recent studies with cell lines suggest that the situation may be more complicated. To circumvent the problem of maintaining high levels of transcription in the absence of the H chain enhancer, we have used Sp6 hybridoma cell lines in which the endogenous intronic enhancer is in fact deleted but the H chain continues to be stably transcribed at high levels through variegated expression (152). In addition, the 3′ regulatory region is insulated by the insertion of a *gpt* gene, which allowed us to examine whether the intronic enhancer is required for SHM without any compensating assistance from the 3′ regulatory region. Because the Sp6 hybridoma does not express AID, we stably transfected the cells with AID. These studies revealed that the intronic core Eμ enhancer was required for SHM of the endogenous H chain V region if the matrix attachment regions (MARs) were still present but that if both the MARs and core Eμ were deleted, then SHM was restored to wild-type levels. We concluded that at least in this hybridoma system, *cis*-acting sequences associated with these transcriptional regulatory regions were affecting the targeting of SHM, even if they were no longer required to control transcription. In addition, these results suggested to us that the MARs and core Eμ elements can singly or in combination impose positive and negative regulation of SHM.

This general idea is supported by studies in which the intronic enhancer has been manipulated in other ways (153). For example, in Ramos cells, when the core Eμ enhancer is deleted in an ectopic H chain gene, there is a twofold reduction in SHM. If part of the 3′ regulatory region is added, mutations in the V region of the transgene occur primarily in A:T bases rather than G:C bases. In the same cells, the mutations in the endogenous H chain V region continue to be primarily in G:C. This suggests that *cis*-acting sequences could also regulate the recruitment or activity of error-prone BER or MMR because the spectrum of mutations is altered in a *cis*-acting manner.

In vivo, the presence of *trans* factor–binding sites, such as E-box motifs for E2A, PU.1, and NF-EM5, alters the rate of SHM in H and/or L chain V regions (52, 154). Because these motifs are not found in all H and L chain genes and are present in many nonmutating genes, they are unlikely to be the sole regulators of targeting of SHM. In addition, these *trans* factors alter the rate of transcription in Ig genes, so it is difficult to know if the changes in SHM are primary or secondary to changes in rates of transcription. However, these binding sites and presumably other *cis*-acting sequences and DNA-binding proteins are likely to play an as yet to be defined role in the targeting process.

Another type of *cis*-acting sequence that might recruit the SHM enzymes to Ig regions undergoing high rates of mutation is ssDNA structures. For example, G-loops have been observed in transcribed switch regions, and they might offer AID a stable ssDNA target in the nontranscribed strand (155), whereas collapsed R-loops may provide the same in the transcribed strand (156). To probe for ssDNA associated with mutating V regions, we have crosslinked the nucleic acids and proteins in intact B cell nuclei and used sodium bisulfite

to identify regions of ssDNA. We found more patches of ssDNA in crosslinked chromatin in genes that are undergoing SHM than in highly transcribed genes that are not targeted by AID (57). In addition, a number of investigators have used computer programs to identify potential stem-loop structures in the V and switch regions that may stabilize ssDNA in regions that are frequently mutated (157, 158).

Trans-Acting Protein (Co-)Factors

As discussed above in the AID section, the location of some of the inactivating mutations in AID (**Figure 4**) and the importance of its phosphorylation sites suggest that there are interacting proteins that could influence AID targeting to either the V or switch regions (64, 65). One of the first AID cofactors identified was RNA Pol II (53). Additionally, the 5′ boundaries of SHM may reflect the transition between the initiating and elongating forms of RNA Pol II, with AID activity or binding requiring the elongating form (50). If true, this requirement for elongation might explain the relatively protected 100–200 bp from the transcription start site. Similarly, stochastic dissociation of AID from RNA Pol II has been invoked to explain the exponential decay of SHM after the first 1.5–2 kb (15, 19). Direct evidence for this model is lacking, and it does not explain how other highly transcribed genes, which also use RNA Pol II, would be protected from AID.

Other identified protein cofactors include RPA; MDM2, which is a ubiquitin ligase 3 that modifies P53; protein kinase A (PKA), which presumably phosphorylates serine 38 and/or perhaps other residues in AID; and DNA PKcs. All these cofactors have been reported to associate with AID (70, 71, 159, 160) (**Supplemental Table**). Considering that these proteins are ubiquitously expressed and perform general cellular functions, it seems unlikely that they are responsible for mediating the preference of AID for the Ig genes or for specific targeting to V and switch regions.

Cell Cycle and Subnuclear Restriction

In the discussion of BER and MMR, we emphasized the important role of error-prone repair in SHM and CSR, whereas error-free repair is an essential characteristic of DNA replication during the S phase of the cell cycle. The recent crystal structures of human Msh2-Msh6 bound to various types of mismatches (106) suggest that the conformation of the heterodimer is identical, whether it is bound to a typical postreplicative G:T mismatch, the G:U mismatch that is the result of AID deamination, or even bulkier adducts caused by genotoxic agents. Therefore, the nature of the DNA lesion itself is unlikely to explain the recruitment of error-prone repair to AID-induced lesions.

By restricting SHM and CSR to the nonreplicative phases of the cell cycle, the centroblast B cell could potentially separate the error-free repair needed during DNA replication from the error-prone repair of antibody V and switch regions that have undergone AID-induced mutations. This seemingly straightforward mechanism might work because error-free repair is an essential characteristic of DNA replication during S phase. Only a few studies have been done that are relevant to this possibility. In BL2 cells in which SHM was induced by incubation with anti-IgM, anti-CD19, and anti-CD21, V region mutations were detected during G_1 and G_2 but not during S phase (161). Similarly, during CSR the colocalization of the IgH gene with proteins involved in nonhomologous end joining (NHEJ) was observed in G_1 (136). Immunohistochemical studies on the subcellular location of AID also suggest that there may be differences in localization at various stages of the cell cycle (121, 162). Clearly, a great deal is unknown in this area, including when during the cell cycle AID is itself expressed and has access to the Ig genes.

Two other ways that SHM may be targeted and restricted are through subnuclear localization of those genes that will undergo

SHM and through the restriction of AID and the various other enzymes that are involved to that region of the nucleus. A number of studies have reported changes in the subnuclear localization of the H chain gene around the time of V(D)J rearrangement (163), as well as colocalization of the locus with proteins involved in NHEJ during isotype switching (164). However, neither colocalization of the Ig genes with AID nor the proteins involved in BER and MMR has been reported. Particularly intriguing is the recent finding that the IgH and IgL genes colocalize with *c-myc* in the same "transcription factories" in primary resting B cells (165), but it remains to be determined if this colocalization is also present during SHM or CSR and whether AID and its associated cofactors are in such transcription factories.

CONCLUSIONS

Although a great deal of genetic and biochemical evidence supports the model of SHM depicted in **Figure 2**, the discussions of BER and MMR in the preceding sections reveal many unresolved issues and raise a number of provocative questions. Even if one focuses on just V region mutation, it is unclear how the B cell manages the ordered recruitment of the many different components of each system to carry out the SHM of antibody V regions (110, 166). Although AID may initiate this process by converting dC to dU, we know that a whole program of gene expression changes occurs as B cells enter the GC and become centroblasts (167, 168). These changes include the transcriptional activation of AID, the increased expression of enzymes involved in MMR and BER, changes in the chromatin to make the Ig V regions accessible to AID (101, 141), and suppression of the DNA damage response (169, 170). Although AID may deaminate dC on both strands at about the same frequency, a strand bias for mutations in A:T bases has been recognized (26), which suggests that the targeting of AID and of the repair processes that it recruits may be different. A very recent paper suggests that error-prone MMR preferentially targets dU in the nontranscribed strand (54).

Dividing SHM into two phases has been convenient, but whether they are really distinct is unclear. We have assumed that the presence of dU is responsible for recruiting BER, whereas the G:U and probably the G:abasic sites create mismatched bases that recruit MMR (1, 3–5, 10). However, the AID protein, the molecules that are associated with it, and/or some as yet unidentified protein or transcription complex may also play a role in recruiting and managing the competition between BER and MMR. For example, a dominant-negative mutant of MSH6 affects the targeting of AID to particular hot-spots, suggesting a connection between the first and second phase (102). This may merely reflect that the accumulation of 10–15 mutations, which is often seen in a single V region, requires multiple cycles of mutation and repair and that mutations introduced by BER and MMR introduce new mismatches that must be repaired. But phase 1 and phase 2 may also be connected by an organizing molecule such as PCNA, or, perhaps, B cell–specific complexes or mutasomes contain both phase 1 and phase 2 enzymes and associated proteins (171), which coordinately regulate SHM.

Even more perplexing is how a high rate of mutation can be selectively targeted to small regions of the antibody genes, whereas other regions of the Ig genes and non-Ig genes that are highly expressed in centroblast B cells do not undergo such high rates of mutation. Although the number of mutations that accumulate in antibody V regions is clearly much higher than in other genes in AID-expressing B cells, mutations have been observed in other genes. Many of the genes that appear to be mistargeted by AID are proto-oncogenes, and the mutations in them are observed in B cell tumors (131). However, high mutation rates have been reported in normal primary cells in genes that do not contribute to malignancies (128). In addition, reporter genes that are integrated throughout the genome also

accumulate mutations at frequencies that are lower than the V region but are still very high when compared with the genome-wide mutation rate (134, 135). This suggests that AID-induced mutations are not as restricted as previously believed and raises the question of how much DNA damage is actually being produced throughout the genome in B cells undergoing SHM and CSR. There is accumulating evidence that the amount of damage is substantial because centroblast B cells appear to have developed mechanisms to ignore or deal with such damage (137, 169). There is already evidence that many of the mutations in the V region are repaired and not scored as mutagenic events (161). This raises the possibility that AID is causing mutations in many genes but that the mutations in non-Ig genes are effectively repaired in an error-free way.

A third and related question is how the centroblast B cell organizes the mutation of the antibody V regions so that the BER and MMR of the V region is error prone, whereas the repair of other genes is error free. Approximately 70–250 dUs are normally introduced during each cycle of DNA replication, and all or most of these are repaired by BER, MMR, and other mechanisms with high fidelity (172). How are these postreplicative lesions distinguished from the AID-induced lesions, and, if AID is not restricted to the antibody genes, how are the AID-induced lesions in non-Ig genes distinguished from those in the V regions, or are they? These are important issues because the B cell needs to create genomic instability in the V region while still maintaining sufficient stability in the rest of its genome so that it will survive to produce clonal progeny that can be positively selected for high-affinity antibodies. These affinity-matured B cells must then either differentiate into plasma cells that secrete useful antibodies or into memory B cells that can meet future challenges against the same antigen. Notably, the loss of control of this process may be responsible for the many B cell lymphomas that arise from GC B cells.

To address these issues, it will be necessary to identify the proteins that interact with AID and contribute to its differential targeting to Ig genes and to the V region and different switch regions, to learn more about when in the cell cycle SHM is occurring, and to identify mutation and repair complexes that participate in error-prone and error-free repair. We will also have to gain a better understanding of how the levels of these SHM-participating proteins and their high-fidelity, error-free competitors are regulated in centroblast B cells and identify the signal transduction pathways that are responsible for controlling these events. Such studies not only will reveal how SHM is regulated and targeted but also will lead to new insights into the basic mechanisms of mutation and repair.

SUMMARY POINTS

1. Affinity maturation of the humoral response occurs through diversification of Ig genes by AID-induced somatic hypermutation (SHM) of the Ig V regions, followed by selection of high-affinity B cells.
2. AID initiates SHM by deaminating cytidine residues in ssDNA during transcription of the Ig locus.
3. AID is recruited to specific regions by a variety of targeting mechanisms that may include increased accessibility, *cis*-acting elements, proteins that associate with AID, and regions of stabilized ssDNA.

4. The resulting uridine is either replicated over or processed by a complex series of enzymes, each of which paradoxically functions outside the Ig locus in high-fidelity DNA repair and maintains genome stability.
5. Mismatch repair, base excision repair, and DNA polymerases are diverted from their typical cellular roles of preserving genomic integrity to process uridines and surrounding sequences in an error-prone fashion, which leads to significant diversification of the Ig locus.
6. The regulation of this process occurs at many levels that likely include posttranslational modification of AID and PCNA, downregulation of components of high-fidelity repair, subcellular localization and trafficking of AID, chromatin changes in the targeted loci, differences in protein interactions and complex formation at different phases of the cell cycle, and global changes that occur in GC B cells in transcription programs and the DNA damage response.

FUTURE ISSUES

1. How is AID targeted to specific genomic loci?
2. How are mismatch repair and base excision repair induced to recruit error-prone repair to sites of AID action while performing high-fidelity repair elsewhere in the genome?
3. What is the crystal structure of AID and what is its mechanism of deamination and targeting?
4. Are the chromatin modifications that are observed in the areas of AID activity required for AID targeting? What other chromatin modifications are involved?
5. What are the *cis*-acting elements that recruit AID and the error-prone mutasome?
6. What are the other components of the error-prone mutasome, which protein-protein interactions are critical for its regulation, and which are the proteins that interact with AID?
7. How do posttranslational modifications of AID regulate its activity and targeting, and what role, if any, do alternative isoforms of AID play in SHM?
8. Are there AID-induced mutations within or outside of the Ig locus that are repaired in an error-free fashion, and if so how does the centroblast B cell target error-free or error-prone repair to the sites of AID activity?

DISCLOSURE STATEMENT

The authors are not aware of any biases that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We thank Phuong Pham and Janet Stavnezer for thoughtful comments on the manuscript. F.L.K. and J.U.P. are supported by the Medical Scientist Training Program T32 GM 007288

at Albert Einstein College of Medicine. S.R. is supported by a fellowship from the Spanish Ministry of Education and Science EX-2006-0732. S.L.K. is supported by the Immunology and Immuno-oncology Training Program T32 CA 09173 at Albert Einstein College of Medicine. M.F.G. acknowledges funding from NIH ES013192 and NIH R37GM21422. M.D.S. is supported by RO1CA72649 and R01CA102705.

LITERATURE CITED

1. Maizels N. 2005. Immunoglobulin gene diversification. *Annu. Rev. Genet* 39:23–46
2. Diaz M, Flajnik MF. 1998. Evolution of somatic hypermutation and gene conversion in adaptive immunity. *Immunol. Rev.* 162:13–24
3. Di Noia JM, Neuberger MS. 2007. Molecular mechanisms of antibody somatic hypermutation. *Annu. Rev. Biochem.* 76:1–22
4. Li Z, Woo CJ, Iglesias-Ussel MD, Ronai D, Scharff MD. 2004. The generation of antibody diversity through somatic hypermutation and class switch recombination. *Genes Dev.* 18:1–11
5. Longerich S, Basu U, Alt F, Storb U. 2006. AID in somatic hypermutation and class switch recombination. *Curr. Opin. Immunol.* 18:164–74
6. Kelsoe G. 1996. The germinal center: a crucible for lymphocyte selection. *Semin. Immunol.* 8:179–84
7. MacLennan IC. 2005. Germinal centers still hold secrets. *Immunity* 22:656–57
8. Muramatsu M, Kinoshita K, Fagarasan S, Yamada S, Shinkai Y, Honjo T. 2000. Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell* 102:553–63
9. Muramatsu M, Sankaranand VS, Anant S, Sugai M, Kinoshita K, et al. 1999. Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. *J. Biol. Chem.* 274:18470–76
10. Teng G, Papavasiliou FN. 2007. Immunoglobulin somatic hypermutation. *Annu. Rev. Genet.* 41:107–20
11. Casali P, Pal Z, Xu Z, Zan H. 2006. DNA repair in antibody somatic hypermutation. *Trends Immunol.* 27:313–21
12. Chaudhuri J, Alt FW. 2004. Class-switch recombination: interplay of transcription, DNA deamination and DNA repair. *Nat. Rev. Immunol.* 4:541–52
13. Bachl J, Caldwell RB, Buerstedde JM. 2007. Biotechnology and the chicken B cell line DT40. *Cytogenet. Genome Res.* 117:189–94
14. Stavnezer J, Guikema JEJ, Schrader CE. 2008. Mechanism and regulation of class switch recombination. *Annu. Rev. Immunol.* 26:261–92
15. Lebecque SG, Gearhart PJ. 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.* 172:1717–27
16. Both GW, Taylor L, Pollard JW, Steele EJ. 1990. Distribution of mutations around rearranged heavy-chain antibody variable-region genes. *Mol. Cell. Biol.* 10:5187–96
17. Longerich S, Tanaka A, Bozek G, Nicolae D, Storb U. 2005. The very 5′ end and the constant region of Ig genes are spared from somatic mutation because AID does not access these regions. *J. Exp. Med.* 202:1443–54
18. Rada C, Yelamos J, Dean W, Milstein C. 1997. The 5′ hypermutation boundary of kappa chains is independent of local and neighbouring sequences and related to the distance from the initiation of transcription. *Eur. J. Immunol.* 27:3115–20

19. Rada C, Milstein C. 2001. The intrinsic hypermutability of antibody heavy and light chain genes decays exponentially. *EMBO J.* 20:4570–76
20. Golding GB, Gearhart PJ, Glickman BW. 1987. Patterns of somatic mutations in immunoglobulin variable genes. *Genetics* 115:169–76
21. Smith DS, Creadon G, Jena PK, Portanova JP, Kotzin BL, Wysocki LJ. 1996. Di- and trinucleotide target preferences of somatic mutagenesis in normal and autoreactive B cells. *J. Immunol.* 156:2642–52
22. Rogozin IB, Diaz M. 2004. Cutting edge: DGYW/WRCH is a better predictor of mutability at G:C bases in Ig hypermutation than the widely accepted RGYW/WRCY motif and probably reflects a two-step activation-induced cytidine deaminase-triggered process. *J. Immunol.* 172:3382–84
23. Peters A, Storb U. 1996. Somatic hypermutation of immunoglobulin genes is linked to transcription initiation. *Immunity* 4:57–65
24. Fukita Y, Jacobs H, Rajewsky K. 1998. Somatic hypermutation in the heavy chain locus correlates with transcription. *Immunity* 9:105–14
25. Bachl J, Carlson C, Gray-Schopfer V, Dessing M, Olsson C. 2001. Increased transcription levels induce higher mutation rates in a hypermutating cell line. *J. Immunol.* 166:5051–57
26. Milstein C, Neuberger MS, Staden R. 1998. Both DNA strands of antibody genes are hypermutation targets. *Proc. Natl. Acad. Sci. USA* 95:8791–94
27. Shen HM, Michael N, Kim N, Storb U. 2000. The TATA binding protein, c-Myc and survivin genes are not somatically hypermutated, while Ig and BCL6 genes are hypermutated in human memory B cells. *Int. Immunol.* 12:1085–93
28. Storb U, Peters A, Kim N, Shen HM, Bozek G, et al. 1999. Molecular aspects of somatic hypermutation of immunoglobulin genes. *Cold Spring Harb. Symp. Quant. Biol.* 64:227–34
29. Rada C, Ehrenstein MR, Neuberger MS, Milstein C. 1998. Hot spot focusing of somatic hypermutation in MSH2-deficient mice suggests two stages of mutational targeting. *Immunity* 9:135–41
30. Petersen-Mahrt SK, Harris RS, Neuberger MS. 2002. AID mutates *E. coli* suggesting a DNA deamination mechanism for antibody diversification. *Nature* 418:99–103
31. Tippin B, Goodman MF. 2001. A new class of errant DNA polymerases provides candidates for somatic hypermutation. *Philos. Trans. R. Soc. London B Biol. Sci.* 356:47–51
32. Lehmann AR, Niimi A, Ogi T, Brown S, Sabbioneda S, et al. 2007. Translesion synthesis: Y-family polymerases and the polymerase switch. *DNA Repair* 6:891–99
33. Poltoratsky V, Goodman MF, Scharff MD. 2000. Error prone candidates vie for somatic mutation. *J. Exp. Med.* 192:F27–30
34. Di Noia J, Neuberger MS. 2002. Altering the pathway of immunoglobulin hypermutation by inhibiting uracil-DNA glycosylase. *Nature* 419:43–48
35. Muramatsu M, Nagaoka H, Shinkura R, Begum NA, Honjo T. 2007. Discovery of activation-induced cytidine deaminase, the engraver of antibody memory. *Adv. Immunol.* 94:1–36
36. Revy P, Muto T, Levy Y, Geissmann F, Plebani A, et al. 2000. Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the hyper-IgM syndrome (HIGM2). *Cell* 102:565–75
37. Durandy A, Peron S, Fischer A. 2006. Hyper-IgM syndromes. *Curr. Opin. Rheumatol.* 18:369–76
38. Bransteitter R, Pham P, Scharff MD, Goodman MF. 2003. Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. *Proc. Natl. Acad. Sci. USA* 100:4102–7

39. Bransteitter R, Pham P, Calabrese P, Goodman MF. 2004. Biochemical analysis of hypermutational targeting by wild type and mutant activation-induced cytidine deaminase. *J. Biol. Chem.* 279:51612–21
40. Pham P, Bransteitter R, Petruska J, Goodman MF. 2003. Processive AID-catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation. *Nature* 424:103–7
41. Chaudhuri J, Tian M, Khuong C, Chua K, Pinaud E, Alt FW. 2003. Transcription-targeted DNA deamination by the AID antibody diversification enzyme. *Nature* 422:726–30
42. Dickerson SK, Market E, Besmer E, Papavasiliou FN. 2003. AID mediates hypermutation by deaminating single stranded DNA. *J. Exp. Med.* 197:1291–96
43. Sohail A, Klapacz J, Samaranayake M, Ullah A, Bhagwat AS. 2003. Human activation-induced cytidine deaminase causes transcription-dependent, strand-biased C to U deaminations. *Nucleic Acids Res.* 31:2990–94
44. Besmer E, Market E, Papavasiliou FN. 2006. The transcription elongation complex directs activation-induced cytidine deaminase-mediated DNA deamination. *Mol. Cell. Biol.* 26:4378–85
45. Larijani M, Petrov AP, Kolenchenko O, Berru M, Krylov SN, Martin A. 2007. AID associates with single-stranded DNA with high affinity and a long complex half-life in a sequence-independent manner. *Mol. Cell. Biol.* 27:20–30
46. Ramiro AR, Stavropoulos P, Jankovic M, Nussenzweig MC. 2003. Transcription enhances AID-mediated cytidine deamination by exposing single-stranded DNA on the nontemplate strand. *Nat. Immunol.* 4:452–56
47. Shen HM, Storb U. 2004. Activation-induced cytidine deaminase (AID) can target both DNA strands when the DNA is supercoiled. *Proc. Natl. Acad. Sci. USA* 101:12997–3002
48. Pham P, Chelico L, Goodman MF. 2007. DNA deaminases AID and APOBEC3G act processively on single-stranded DNA. *DNA Repair* 6:689–94
49. Goodman MF, Scharff MD, Romesberg FE. 2007. AID-initiated purposeful mutations in immunoglobulin genes. *Adv. Immunol.* 94:127–55
50. Storb U, Shen HM, Longerich S, Ratnam S, Tanaka A, et al. 2007. Targeting of AID to immunoglobulin genes. *Adv. Exp. Med. Biol.* 596:83–91
51. Chelico L, Pham P, Calabrese P, Goodman MF. 2006. APOBEC3G DNA deaminase acts processively 3′–>5′ on single-stranded DNA. *Nat. Struct. Mol. Biol.* 13:392–99
52. Michael N, Shen HM, Longerich S, Kim N, Longacre A, Storb U. 2003. The E box motif CAGGTG enhances somatic hypermutation without enhancing transcription. *Immunity* 19:235–42
53. Nambu Y, Sugai M, Gonda H, Lee CG, Katakai T, et al. 2003. Transcription-coupled events associating with immunoglobulin switch region chromatin. *Science* 302:2137–40
54. Unniraman S, Schatz DG. 2007. Strand-biased spreading of mutations during somatic hypermutation. *Science* 317:1227–30
55. Rada C, Di Noia JM, Neuberger MS. 2004. Mismatch recognition and uracil excision provide complementary paths to both Ig switching and the A/T-focused phase of somatic mutation. *Mol. Cell* 16:163–71
56. Shen HM, Tanaka A, Bozek G, Nicolae D, Storb U. 2006. Somatic hypermutation and class switch recombination in Msh6$^{-/-}$Ung$^{-/-}$ double-knockout mice. *J. Immunol.* 177:5386–92
57. Ronai D, Iglesias MD, Fan M, Li Z, Martin A, Scharff MD. 2007. Single-stranded DNA-protein complexes in regions targeted for somatic hypermutation. *J. Exp. Med.* 2007:181–90

58. Apel TW, Mautner J, Polack A, Bornkamm GW, Eick D. 1992. Two antisense promoters in the immunoglobulin mu-switch region drive expression of c-myc in the Burkitt's lymphoma cell line BL67. *Oncogene* 7:1267–71
59. Bolland DJ, Wood AL, Afshar R, Featherstone K, Oltz EM, Corcoran AE. 2007. Antisense intergenic transcription precedes *Igh* D-to-J recombination and is controlled by the intronic enhancer Eμ. *Mol. Cell. Biol.* 27:5523–33
60. Bolland DJ, Wood AL, Johnston CM, Bunting SF, Morgan G, et al. 2004. Antisense intergenic transcription in V(D)J recombination. *Nat. Immunol.* 5:630–37
61. Shen HM, Ratnam S, Storb U. 2005. Targeting of the activation-induced cytosine deaminase is strongly influenced by the sequence and structure of the targeted DNA. *Mol. Cell. Biol.* 25:10815–21
62. Durandy A, Peron S, Taubenheim N, Fischer A. 2006. Activation-induced cytidine deaminase: structure-function relationship as based on the study of mutants. *Hum. Mutat.* 27:1185–91
63. Ta VT, Nagaoka H, Catalan N, Durandy A, Fischer A, et al. 2003. AID mutant analyses indicate requirement for class-switch-specific cofactors. *Nat. Immunol.* 4:843–48
64. Barreto V, Reina-San-Martin B, Ramiro AR, McBride KM, Nussenzweig MC. 2003. C-terminal deletion of AID uncouples class switch recombination from somatic hypermutation and gene conversion. *Mol. Cell* 12:501–8
65. Shinkura R, Ito S, Begum NA, Nagaoka H, Muramatsu M, et al. 2004. Separate domains of AID are required for somatic hypermutation and class-switch recombination. *Nat. Immunol.* 5:707–12
66. Prochnow C, Bransteitter R, Klein MG, Goodman MF, Chen XS. 2007. The APOBEC-2 crystal structure and functional implications for the deaminase AID. *Nature* 445:447–51
67. Deleted in proof
68. Basu U, Chaudhuri J, Alpert C, Dutt S, Ranganath S, et al. 2005. The AID antibody diversification enzyme is regulated by protein kinase A phosphorylation. *Nature* 438:508–11
69. McBride KM, Gazumyan A, Woo EM, Barreto VM, Robbiani DF, et al. 2006. Regulation of hypermutation by activation-induced cytidine deaminase phosphorylation. *Proc. Natl. Acad. Sci. USA* 103:8798–803
70. Pasqualucci L, Kitaura Y, Gu H, Dalla-Favera R. 2006. PKA-mediated phosphorylation regulates the function of activation-induced deaminase (AID) in B cells. *Proc. Natl. Acad. Sci. USA* 103:395–400
71. Chaudhuri J, Khuong C, Alt FW. 2004. Replication protein A interacts with AID to promote deamination of somatic hypermutation targets. *Nature* 430:992–98
72. Noguchi E, Shibasaki M, Inudou M, Kamioka M, Yokouchi Y, et al. 2001. Association between a new polymorphism in the activation-induced cytidine deaminase gene and atopic asthma and the regulation of total serum IgE levels. *J. Allergy Clin. Immunol.* 108:382–86
73. Greeve J, Philipsen A, Krause K, Klapper W, Heidorn K, et al. 2003. Expression of activation-induced cytidine deaminase in human B-cell non-Hodgkin lymphomas. *Blood* 101:3574–80
74. Oppezzo P, Dumas G, Lalanne AI, Payelle-Brogard B, Magnac C, et al. 2005. Different isoforms of BSAP regulate expression of AID in normal and chronic lymphocytic leukemia B cells. *Blood* 105:2495–503
75. Oppezzo P, Vuillier F, Vasconcelos Y, Dumas G, Magnac C, et al. 2003. Chronic lymphocytic leukemia B cells expressing AID display dissociation between class switch recombination and somatic hypermutation. *Blood* 101:4029–32

76. McCarthy H, Wierda WG, Barron LL, Cromwell CC, Wang J, et al. 2003. High expression of activation-induced cytidine deaminase (AID) and splice variants is a distinctive feature of poor-prognosis chronic lymphocytic leukemia. *Blood* 101:4903–8
77. Albesiano E, Messmer BT, Damle RN, Allen SL, Rai KR, Chiorazzi N. 2003. Activation-induced cytidine deaminase in chronic lymphocytic leukemia B cells: expression as multiple forms in a dynamic, variably sized fraction of the clone. *Blood* 102:3333–39
78. Jansen JG, Langerak P, Tsaalbi-Shtylik A, Van Den Berk P, Jacobs H, de Wind N. 2006. Strand-biased defect in C/G transversions in hypermutating immunoglobulin genes in Rev1-deficient mice. *J. Exp. Med.* 203:319–23
79. Masuda K, Ouchida R, Hikida M, Kurosaki T, Yokoi M, et al. 2007. DNA polymerases η and θ function in the same genetic pathway to generate mutations at A/T during somatic hypermutation of Ig genes. *J. Biol. Chem.* 282:17387–94
80. Imai K, Slupphaug G, Lee WI, Revy P, Nonoyama S, et al. 2003. Human uracil-DNA glycosylase deficiency associated with profoundly impaired immunoglobulin class-switch recombination. *Nat. Immunol.* 4:1023–28
81. Bardwell PD, Martin A, Wong E, Li Z, Edelmann W, Scharff MD. 2003. Cutting edge: the G-U mismatch glycosylase methyl-CpG binding domain 4 is dispensable for somatic hypermutation and class switch recombination. *J. Immunol.* 170:1620–24
82. Di Noia JM, Rada C, Neuberger MS. 2006. SMUG1 is able to excise uracil from immunoglobulin genes: insight into mutation versus repair. *EMBO J.* 25:585–95
83. Krokan HE, Drablos F, Slupphaug G. 2002. Uracil in DNA–occurrence, consequences and repair. *Oncogene* 21:8935–48
84. Almeida KH, Sobol RW. 2007. A unified view of base excision repair: lesion-dependent protein complexes regulated by post-translational modification. *DNA Repair* 6:695–711
85. Fortini P, Dogliotti E. 2007. Base damage and single-strand break repair: mechanisms and functional significance of short- and long-patch repair subpathways. *DNA Repair* 6:398–409
86. Simpson LJ, Ross AL, Szuts D, Alviani CA, Oestergaard VH, et al. 2006. RAD18-independent ubiquitination of proliferating-cell nuclear antigen in the avian cell line DT40. *EMBO Rep.* 7:927–32
87. Bachl J, Ertongur I, Jungnickel B. 2006. Involvement of Rad18 in somatic hypermutation. *Proc. Natl. Acad. Sci. USA* 103:12081–86
88. Arakawa H, Moldovan GL, Saribasak H, Saribasak NN, Jentsch S, Buerstedde JM. 2006. A role for PCNA ubiquitination in immunoglobulin hypermutation. *PLoS Biol.* 4:e366
89. Langerak P, Nygren AO, Krijger PH, Van Den Berk PC, Jacobs H. 2007. A/T mutagenesis in hypermutated immunoglobulin genes strongly depends on PCNAK164 modification. *J. Exp. Med.* 204:1989–98
90. Rada C, Williams GT, Nilsen H, Barnes DE, Lindahl T, Neuberger MS. 2002. Immunoglobulin isotype switching is inhibited and somatic hypermutation perturbed in UNG-deficient mice. *Curr. Biol.* 12:1748–55
91. Bennett SE, Sung JS, Mosbaugh DW. 2001. Fidelity of uracil-initiated base excision DNA repair in DNA polymerase beta-proficient and -deficient mouse embryonic fibroblast cell extracts. *J. Biol. Chem.* 276:42588–600
92. Larson ED, Cummings WJ, Bednarski DW, Maizels N. 2005. MRE11/RAD50 cleaves DNA in the AID/UNG-dependent pathway of immunoglobulin gene diversification. *Mol. Cell* 20:367–75
93. Yabuki M, Fujii MM, Maizels N. 2005. The MRE11-RAD50-NBS1 complex accelerates somatic hypermutation and gene conversion of immunoglobulin variable regions. *Nat. Immunol.* 6:730–36

94. Poltoratsky V, Prasad R, Horton JK, Wilson SH. 2007. Down-regulation of DNA polymerase β accompanies somatic hypermutation in human BL2 cell lines. *DNA Repair* 6:244–53
95. Wu X, Stavnezer J. 2007. DNA polymerase β is able to repair breaks in switch regions and plays an inhibitory role during immunoglobulin class switch recombination. *J. Exp. Med.* 204:1677–89
96. Parlanti E, Locatelli G, Maga G, Dogliotti E. 2007. Human base excision repair complex is physically associated to DNA replication and cell cycle regulatory proteins. *Nucleic Acids Res.* 35:1569–77
97. Iyer RR, Pluciennik A, Burdett V, Modrich PL. 2006. DNA mismatch repair: functions and mechanisms. *Chem. Rev.* 106:302–23
98. Jiricny J. 2006. The multifaceted mismatch-repair system. *Nat. Rev. Mol. Cell Biol.* 7:335–46
99. Kadyrov FA, Dzantiev L, Constantin N, Modrich P. 2006. Endonucleolytic function of MutLα in human mismatch repair. *Cell* 126:297–308
100. Jungnickel B. 2006. False moves for survival: error-prone DNA repair in adaptive immunity. *Cell Cycle* 5:2856–61
101. Woo CJ, Martin A, Scharff MD. 2003. Induction of hypermutation is associated with modifications of variable region chromatin in BL2 cells. *Immunity* 19:479–89
102. Li Z, Zhao C, Iglesias-Ussel MD, Polonskaya Z, Zhuang M, et al. 2006. The mismatch repair protein Msh6 influences the in vivo AID targeting to the Ig locus. *Immunity* 24:393–403
103. Wiesendanger M, Kneitz B, Edelmann W, Scharff MD. 2000. Somatic mutation in MutS homologue (MSH)3-, MSH6-, and MSH3/MSH6-deficient mice reveals a role for the MSH2-MSH6 heterodimer in modulating the base substitution pattern. *J. Exp. Med.* 191:579–84
104. Li Z, Scherer SJ, Ronai D, Iglesias-Ussel MD, Peled JU, et al. 2004. Examination of Msh6- and Msh3-deficient mice in class switching reveals overlapping and distinct roles of MutS homologues in antibody diversification. *J. Exp. Med.* 200:47–59
105. Martomo SA, Yang WW, Gearhart PJ. 2004. A role for Msh6 but not Msh3 in somatic hypermutation and class switch recombination. *J. Exp. Med.* 200:61–68
106. Warren JJ, Pohlhaus TJ, Changela A, Iyer RR, Modrich PL, Beese LS. 2007. Structure of the human MutSα DNA lesion recognition complex. *Mol. Cell* 26:579–92
107. Park K, Kim J, Kim HS, Shin HS. 1998. Isolated human germinal center centroblasts have an intact mismatch repair system. *J. Immunol.* 161:6128–32
108. Li Z, Peled JU, Zhao C, Svetlanov A, Ronai D, et al. 2006. A role for Mlh3 in somatic hypermutation. *DNA Repair* 5:675–82
109. Zeng X, Winter DB, Kasmer C, Kraemer KH, Lehmann AR, Gearhart PJ. 2001. DNA polymerase η is an A-T mutator in somatic hypermutation of immunoglobulin variable genes. *Nat. Immunol.* 2:537–41
110. Delbos F, Aoufouchi S, Faili A, Weill JC, Reynaud CA. 2007. DNA polymerase η is the sole contributor of A/T modifications during immunoglobulin gene hypermutation in the mouse. *J. Exp. Med.* 204:17–23
111. Faili A, Aoufouchi S, Flatter E, Gueranger Q, Reynaud CA, Weill JC. 2002. Induction of somatic hypermutation in immunoglobulin genes is dependent on DNA polymerase iota. *Nature* 419:944–47
112. Zan H, Komori A, Li Z, Cerrutti M, Flajnik MF, et al. 2001. The translesional polymerase zeta plays a major role in Ig and Bcl-6 somatic mutation. *Immunity* 14:643–53

113. Diaz M, Verkoczy LK, Flajnik MF, Klinman NR. 2001. Decreased frequency of somatic hypermutation and impaired affinity maturation but intact germinal center formation in mice expressing antisense RNA to DNA polymerase zeta. *J. Immunol.* 167:327–36
114. Garg P, Burgers PM. 2005. Ubiquitinated proliferating cell nuclear antigen activates translesion DNA polymerases η and REV1. *Proc. Natl. Acad. Sci. USA* 102:18361–66
115. Ulrich HD. 2006. Deubiquitinating PCNA: a downside to DNA damage tolerance. *Nat. Cell Biol.* 8:303–5
116. Friedberg EC. 2006. Reversible monoubiquitination of PCNA: a novel slant on regulating translesion DNA synthesis. *Mol. Cell* 22:150–52
117. Goodman MF. 2002. Error-prone repair DNA polymerases in prokaryotes and eukaryotes. *Annu. Rev. Biochem.* 71:17–50
118. Moldovan GL, Pfander B, Jentsch S. 2007. PCNA, the maestro of the replication fork. *Cell* 129:665–79
119. Mao C, Jiang L, Melo-Jorge M, Puthenveetil M, Zhang X, et al. 2004. T cell-independent somatic hypermutation in murine B cells with an immature phenotype. *Immunity* 20:133–44
120. Han JH, Akira S, Calame K, Beutler B, Selsing E, Imanishi-Kari T. 2007. Class switch recombination and somatic hypermutation in early mouse B cells are mediated by B cell and Toll-like receptors. *Immunity* 27:64–75
121. Cattoretti G, Buttner M, Shaknovich R, Kremmer E, Alobeid B, Niedobitek G. 2006. Nuclear and cytoplasmic AID in extrafollicular and germinal center B cells. *Blood* 107:3967–75
122. Crouch EE, Li Z, Takizawa M, Fichtner-Feigl S, Gourzi P, et al. 2007. Regulation of AID expression in the immune response. *J. Exp. Med.* 204:1145–56
123. Martin A, Bardwell PD, Woo CJ, Fan M, Shulman MJ, Scharff MD. 2002. Activation-induced cytidine deaminase turns on somatic hypermutation in hybridomas. *Nature* 415:802–6
124. Martin A, Scharff MD. 2002. Somatic hypermutation of the AID transgene in B and non-B cells. *Proc. Natl. Acad. Sci. USA* 99:12304–8
125. Okazaki IM, Hiai H, Kakazu N, Yamada S, Muramatsu M, et al. 2003. Constitutive expression of AID leads to tumorigenesis. *J. Exp. Med.* 197:1173–81
126. Ito S, Nagaoka H, Shinkura R, Begum N, Muramatsu M, et al. 2004. Activation-induced cytidine deaminase shuttles between nucleus and cytoplasm like apolipoprotein B mRNA editing catalytic polypeptide 1. *Proc. Natl. Acad. Sci. USA* 101:1975–80
127. McBride KM, Barreto V, Ramiro AR, Stavropoulos P, Nussenzweig MC. 2004. Somatic hypermutation is limited by CRM1-dependent nuclear export of activation-induced deaminase. *J. Exp. Med.* 199:1235–44
128. Gordon MS, Kanegai CM, Doerr JR, Wall R. 2003. Somatic hypermutation of the B cell receptor genes *B29* (*Igβ*, CD79b) and *mb1* (*Igα*, CD79a). *Proc. Natl. Acad. Sci. USA* 100:4126–31
129. Shen HM, Peters A, Baron B, Zhu X, Storb U. 1998. Mutation of BCL-6 gene in normal B cells by the process of somatic hypermutation of Ig genes. *Science* 280:1750–52
130. Muschen M, Re D, Jungnickel B, Diehl V, Rajewsky K, Kuppers R. 2000. Somatic mutation of the CD95 gene in human B cells as a side-effect of the germinal center reaction. *J. Exp. Med.* 192:1833–40
131. Pasqualucci L, Neumeister P, Goossens T, Nanjangud G, Chaganti RS, et al. 2001. Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. *Nature* 412:341–46

132. Pasqualucci L, Migliazza A, Fracchiolla N, William C, Neri A, et al. 1998. BCL-6 Mutations in normal germinal center B cells: evidence of somatic hypermutation acting outside Ig loci. *Proc. Natl. Acad. Sci. USA* 95:11816–21
133. Yoshikawa K, Okazaki IM, Eto T, Kinoshita K, Muramatsu M, et al. 2002. AID enzyme-induced hypermutation in an actively transcribed gene in fibroblasts. *Science* 296:2033–36
134. Parsa JY, Basit W, Wang CL, Gommerman JL, Carlyle JR, Martin A. 2007. AID mutates a nonimmunoglobulin transgene independent of chromosomal position. *Mol. Immunol.* 44:567–75
135. Wang CL, Harper RA, Wabl M. 2004. Genome-wide somatic hypermutation. *Proc. Natl. Acad. Sci. USA* 101:7352–56
136. Petersen S, Casellas R, Reina-San-Martin B, Chen HT, Difilippantonio MJ, et al. 2001. AID is required to initiate Nbs1/gamma-H2AX focus formation and mutations at sites of class switching. *Nature* 414:660–65
137. Okazaki IM, Kotani A, Honjo T. 2007. Role of AID in tumorigenesis. *Adv. Immunol.* 94:245–73
138. Yang SY, Schatz DG. 2007. Targeting of AID-mediated sequence diversification by cis-acting determinants. *Adv. Immunol.* 94:109–25
139. Li Z, Luo Z, Ronai D, Kuang FL, Peled JU, et al. 2007. Targeting AID to the Ig genes. *Adv. Exp. Med. Biol.* 596:93–109
140. Fraenkel S, Mostoslavsky R, Novobrantseva TI, Pelanda R, Chaudhuri J, et al. 2007. Allelic 'choice' governs somatic hypermutation in vivo at the immunoglobulin kappa-chain locus. *Nat. Immunol.* 8:715–22
141. Odegard VH, Kim ST, Anderson SM, Shlomchik MJ, Schatz DG. 2005. Histone modifications associated with somatic hypermutation. *Immunity* 23:101–10
142. Larijani M, Frieder D, Sonbuchner TM, Bransteitter R, Goodman MF, et al. 2005. Methylation protects cytidines from AID-mediated deamination. *Mol. Immunol.* 42:599–604
143. Li Z, Luo Z, Scharff MD. 2004. Differential regulation of histone acetylation and generation of mutations in switch regions is associated with Ig class switching. *Proc. Natl. Acad. Sci. USA* 101:15428–33
144. Wang L, Whang N, Wuerffel R, Kenter AL. 2006. AID-dependent histone acetylation is detected in immunoglobulin S regions. *J. Exp. Med.* 203:215–26
145. Odegard VH, Schatz DG. 2006. Targeting of somatic hypermutation. *Nat. Rev. Immunol.* 6:573–83
146. Bachl J, Wabl M. 1996. Enhancers of hypermutation. *Immunogenetics* 45:59–64
147. Lin MM, Green NS, Zhang W, Scharff MD. 1998. The effects of Em, 3′α (hs 1,2) and 3′κ enhancers on mutation of an Ig-VDJ-Cγ2a immunoglobulin heavy gene in cultured B cells. *Int. Immunol.* 10:1121–29
148. Yang SY, Fugmann SD, Schatz DG. 2006. Control of gene conversion and somatic hypermutation by immunoglobulin promoter and enhancer sequences. *J. Exp. Med.* 203:2919–28
149. Perlot T, Alt FW, Bassing CH, Suh H, Pinaud E. 2005. Elucidation of IgH intronic enhancer functions via germ-line deletion. *Proc. Natl. Acad. Sci USA* 102:14362–67
150. Morvan CL, Pinaud E, Decourt C, Cuvillier A, Cogne M. 2003. The immunoglobulin heavy-chain locus hs3b and hs4 3′ enhancers are dispensable for VDJ assembly and somatic hypermutation. *Blood* 102:1421–27
151. Garrett FE, Emelyanov AV, Sepulveda MA, Flanagan P, Volpi S, et al. 2005. Chromatin architecture near a potential 3′ end of the IgH locus involves modular regulation of

histone modifications during B-cell development and in vivo occupancy of CTCF sites. *Mol. Cell. Biol.* 25:1511–25
152. Ronai D, Iglesias-Ussel MD, Fan M, Shulman MJ, Scharff MD. 2005. Complex regulation of somatic hypermutation by cis-acting sequences in the endogenous IgH gene in hybridoma cells. *Proc. Natl. Acad. Sci. USA* 102:11829–34
153. Komori A, Xu Z, Wu X, Zan H, Casali P. 2006. Biased dA/dT somatic hypermutation as regulated by the heavy chain intronic iEμ enhancer and 3′Eα enhancers in human lymphoblastoid B cells. *Mol. Immunol.* 43:1817–26
154. Kodama M, Hayashi R, Nishizumi H, Nagawa F, Takemori T, Sakano H. 2001. The PU.1 and NF-EM5 binding motifs in the Igκ 3′ enhancer are responsible for directing somatic hypermutations to the intrinsic hotspots in the transgenic Vκ gene. *Int. Immunol.* 13:1415–22
155. Duquette ML, Pham P, Goodman MF, Maizels N. 2005. AID binds to transcription-induced structures in c-MYC that map to regions associated with translocation and hypermutation. *Oncogene* 24:5791–98
156. Yu K, Lieber MR. 2003. Nucleic acid structures and enzymes in the immunoglobulin class switch recombination mechanism. *DNA Repair* 2:1163–74
157. Kinoshita K, Honjo T. 2001. Linking class-switch recombination with somatic hypermutation. *Nat. Rev. Mol. Cell Biol.* 2:493–503
158. Wright BE, Schmidt KH, Minnick MF. 2004. Mechanisms by which transcription can regulate somatic hypermutation. *Genes Immun.* 5:176–82
159. Wu X, Geraldes P, Platt JL, Cascalho M. 2005. The double-edged sword of activation-induced cytidine deaminase. *J. Immunol.* 174:934–41
160. MacDuff DA, Neuberger MS, Harris RS. 2006. MDM2 can interact with the C-terminus of AID but it is inessential for antibody diversification in DT40 B cells. *Mol. Immunol.* 43:1099–108
161. Faili A, Aoufouchi S, Gueranger Q, Zober C, Leon A, et al. 2002. AID-dependent somatic hypermutation occurs as a DNA single-strand event in the BL2 cell line. *Nat. Immunol.* 3:815–21
162. Yang G, Obiakor H, Sinha RK, Newman BA, Hood BL, et al. 2005. Activation-induced deaminase cloning, localization, and protein extraction from young VH-mutant rabbit appendix. *Proc. Natl. Acad. Sci. USA* 102:17083–88
163. Kosak ST, Skok JA, Medina KL, Riblet R, Le Beau MM, et al. 2002. Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development. *Science* 296:158–62
164. Reina-San-Martin B, Difilippantonio S, Hanitsch L, Masilamani RF, Nussenzweig A, Nussenzweig MC. 2003. H2AX is required for recombination between immunoglobulin switch regions but not for intra-switch region recombination or somatic hypermutation. *J. Exp. Med.* 197:1767–78
165. Osborne CS, Chakalova L, Mitchell JA, Horton A, Wood AL, et al. 2007. *Myc* dynamically and preferentially relocates to a transcription factory occupied by *Igh*. *PLoS Biol.* 5:e192
166. Neuberger MS, Rada C. 2007. Somatic hypermutation: activation-induced deaminase for C/G followed by polymerase η for A/T. *J. Exp. Med.* 204:7–10
167. Monti S, Savage KJ, Kutok JL, Feuerhake F, Kurtin P, et al. 2005. Molecular profiling of diffuse large B-cell lymphoma identifies robust subtypes including one characterized by host inflammatory response. *Blood* 105:1851–61
168. Shaffer AL, Wright G, Yang L, Powell J, Ngo V, et al. 2006. A library of gene expression signatures to illuminate normal and pathological lymphoid biology. *Immunol. Rev.* 210:67–85

169. Ranuncolo SM, Polo JM, Dierov J, Singer M, Kuo T, et al. 2007. Bcl-6 mediates the germinal center B cell phenotype and lymphomagenesis through transcriptional repression of the DNA-damage sensor ATR. *Nat. Immunol.* 8:705–14
170. Phan RT, Dalla-Favera R. 2004. The BCL6 proto-oncogene suppresses p53 expression in germinal-centre B cells. *Nature* 432:635–39
171. Reynaud CA, Aoufouchi S, Faili A, Weill JC. 2003. What role for AID: mutator, or assembler of the immunoglobulin mutasome? *Nat. Immunol.* 4:631–38
172. Kavli B, Otterlei M, Slupphaug G, Krokan HE. 2007. Uracil in DNA—general mutagen, but normal intermediate in acquired immunity. *DNA Repair* 6:505–16

RELATED RESOURCES

Stavnezer, J, Guikema JEJ, Schrader CE, 2008. Mechanism and regulation of class switch recombination. *Annu. Rev. Immunol.* 26:261–92

Database of HIGM Type II causing mutations: **http://bioinf.uta.fi/AICDAbase**

Annual Review of Immunology
Volume 26, 2008

Contents