Related Mechanisms of Antibody Somatic Hypermutation and Class Switch Recombination

JOYCE K. HWANG,* FREDERICK W. ALT, and LENG-SIEW YEAP*
Howard Hughes Medical Institute, Program in Cellular and Molecular Medicine, Boston Children’s Hospital, and Department of Genetics, Harvard Medical School, Boston, MA 02115

ABSTRACT The primary antibody repertoire is generated by mechanisms involving the assembly of the exons that encode the antigen-binding variable regions of immunoglobulin heavy (IgH) and light (IgL) chains during the early development of B lymphocytes. After antigen-dependent activation, mature B lymphocytes can further alter their IgH and IgL variable region exons by the process of somatic hypermutation (SHM), which allows the selection of B cells in which SHMs resulted in the production of antibodies with increased antigen affinity. In addition, during antigen-dependent activation, B cells can also change the constant region of their IgH chain through a DNA double-strand-break (DSB) dependent process referred to as IgH class switch recombination (CSR), which generates B cell progeny that produce antibodies with different IgH constant region effector functions that are best suited for an elimination of a particular pathogen or in a particular setting. Both the mutations that underlie SHM and the DSBs that underlie CSR are initiated in target genes by activation-induced cytidine deaminase (AID). This review describes in depth the processes of SHM and CSR with a focus on mechanisms that direct AID cytidine deamination in activated B cells and mechanisms that promote the differential outcomes of such cytidine deamination.

OVERVIEW AND INTRODUCTION
Immunoglobulin genes, B cell receptors and antibodies
The B cell receptor (BCR) is expressed on the B lymphocyte cell surface where it serves as a receptor for foreign antigens (1). The BCR is comprised of two immunoglobulin (Ig) heavy (IgH) chains encoded by the IgH heavy chain locus and two Ig light (IgL) chains encoded by, for a given BCR, either the Igκ or Igκ (collectively referred to as IgL) light chain loci (Fig. 1). These three Ig loci lie on different chromosomes in both humans and mice. While there are certain differences in organization, the overall strategies for Ig gene diversification in mice and humans are very much the same (2, 3), so this review will focus mainly on the mouse. The amino-terminal portions of the IgH and IgL chains have a highly variable amino acid sequence from species to species of antibody and are called variable (V) regions. The IgH and IgL variable regions interact to generate the antigen-binding portion of the BCR/antibody. The carboxy-terminal end of IgH and IgL chains have only a few variations in their sequences and thus are called constant (C) regions.

The antigen-independent generation of an extremely large population of B cells in which individual cells express BCRs with unique antigen-binding specificity is of fundamental importance for vertebrates to generate effective humoral adaptive immune responses, as it enables B cells to recognize and respond to an enormous variety of foreign antigens. In this context, IgH and IgL variable region exons are not encoded in the germline,
but rather are assembled during early B cell development prior to antigen exposure in the fetal liver and bone marrow by the V(D)J recombination process (2). V(D)J recombination generates an IgH VH(D)JH variable region exon by assembling different combinations of the numerous IgH variable (VH) segments, diversity (D) segments, and joining (JH) segments that lie within a 1 to 3 Mb region at the 5′ end of the IgH locus. V(D)J recombination assembles an IgL VLJL variable region exon from Igx or Igλ V segments and J segments (2).

V(D)J recombination is initiated by the lymphocyte-specific RAG1 and RAG2 (“RAG”) endonuclease that recognizes conserved recombination signal sequences (RSS) that flank the V, D, and J segments (4). RAG cleaves between the RSSs and the coding sequences of a pair of involved segments, generating a pair of blunt RSS double strand break (DSB) ends that are later joined to each other and a pair of hair-pinned coding DSB ends that are processed and joined to each other (4) by the general cellular classical nonhomologous end-joining (C-NHEJ) DSB repair pathway (5, 6). Coding ends are often further diversified before they are joined, including the de novo additions of N nucleotides by the terminal deoxynucleotidyl transferase (Tdt), another lymphocyte-specific factor involved in V(D)J recombination (2). The combinatorial diversity arising from the numerous V, D, and J segments, as well as the junctional diversity that arises from junctional diversification during joining the segments, generates an enormous repertoire of primary variable region exons (8). Within the IgH and IgL variable regions there are three regions that show “hypervariability” separated by much less variable “framework” regions (FWR). As they are involved in antigen contact, these three hypervariable regions are termed complementarity-determining regions (CDRs) (9). CDR1 and CDR2 are encoded in the different germ-line VH and VL gene segments, and CDR3 is encoded by the VH, D, and JH junctional regions. The heavy and light chain variable regions form the antigen-binding site. The constant region determines the class and effector function of the antibody molecule. doi:10.1128/microbiolspec.MDNA3-0037-2014.f1

FIGURE 1 Antibody structure. The BCR is comprised of two immunoglobulin (Ig) heavy (IgH) chains encoded by the IgH heavy chain locus and two Ig light (IgL) chains. The rectangles represent Ig domains that constitute the structural units of the immunoglobulin heavy and light chains. The variable regions are assembled through V(D)J recombination of VH, D, and JH gene segments on the heavy chain and VL and Jλ gene segments on the light chain. Complementarity-determining regions (CDRs) are indicated as regions in dashed red boxes: CDR 1 and 2 are encoded in the VH or VL gene segments, and CDR 3 is encoded by the VH, D, and JH junctional region or VL and Jλ junctional region. The heavy and light chain variable regions form the antigen-binding site. The constant region determines the class and effector function of the antibody molecule. doi:10.1128/microbiolspec.MDNA3-0037-2014.f1

Translation of fully assembled IgH and IgL chain genes is initiated from the promoter of the V segment used in the V(D)J exon and continues through downstream exons that encode the C regions of IgH and IgL chains (11). The mouse IgH locus contains eight sets of exons that encode different C(H) regions (sometimes termed “C(H) genes”) within the approximately 200 kb region downstream of the JH segment and lying in the order 5′-VDJ-Cμ-Cα-Cγ3-Cγ1-Cγ2b-Cγ2a-Cε-Cα 3′ (12) (Fig. 2a). The set of C(H) exons expressed with the V region exon determines the class of the BCR/antibody (e.g. IgM, IgG, IgE, IgA). Within the IgH locus, there are several B cell specific enhancers, for example iEμ, which lies between the JH segments and the Cμ exons and a 30 kb IgH 3′ regulatory region (“IgH 3′RR”), which is downstream of the Cα exons (11, 13) (Fig. 2a). Initially, the IgH variable region exon is transcribed in association with the immediately downstream Cμ exons, and in some cells, Cα exons. Alternative RNA splicing of these primary IgH transcripts leads to differential expression of Cμ and Cα and also to differential expression
of C₄H sequences that specify whether the IgH chain is expressed as membrane-bound BCR or is secreted as an antibody (14). Thus, prior to antigenic stimulation, resting B cells express IgM (or IgD).

During B cell development, V(D)J recombination generally occurs first at the IgH locus in progenitor (pro) B cells (2, 6). In this regard, developing B cells generate D to JH rearrangements on both IgH alleles and then append VH segments to pre-existing DJH rearrangements (15). If the first VH to DJH is in frame (productive), the resulting IgH μ heavy chain protein generates a signal that feeds back to prevent VH to DJH joining on the other DJH rearranged allele and to promote development to the precursor (pre) B cell stage. The resulting pre B cells will have a productive V(D)J IgH allele and a “frozen” DJH intermediate allele. If, due to junctional diversification, the first VH to DJH rearrangement is out of frame, the cell can move on and append a VH segment to the second DJH allele which, if productive, will again promote development to the pre B stage with the resulting pre B cells having an in-frame productive V(D)J rearrangement and out-of-frame nonproductive V(D)J rearrangement. Because about two thirds of V to DJH rearrangements are nonproductive, about 40% of normal B cells have two V(D)J rearrangements (one productive and one nonproductive). This “feedback” mechanism for the control of VH to DJH rearrangement is thought to have evolved to ensure mono-specificity of B cell clones in the context of the phenomenon of “allelic exclusion” (see reference 16 for details). Precursor B cells rearrange IgL genes and if they form a productive IgL rearrangement leading to an IgL chain that pairs with μ heavy chain, they then express the complete IgM molecule on their surface as the BCR (2, 6). These newly generated
IgM+ B cells then migrate into the periphery and survey the secondary lymphoid organs, including spleen, lymph nodes and Peyer’s patches, for cognate antigen that binds their BCR.

Overview of SHM, CSR and the role of AID

An encounter with cognate antigen in the secondary lymphoid organs, usually in the context of a T-dependent immune response, can activate mature B cells. Activation can lead to the generation of B cells that secrete their BCR as a secreted antibody. Antigen-dependent B cell activation can also lead to the two somatic processes of genomic rearrangement that enhance the efficacy of the antibody response against specific antigen: SHM further diversifies the variable region exon and alters the affinity of the BCR for antigen (2) (Fig. 2b), while CSR switches the Cμ region exon used and alters the antibody’s antigen elimination function (17) (Fig. 2c). SHM occurs in the germinal centers (GCs) (18), specialized compartments of secondary lymphoid organs, while CSR can occur inside or outside of the GCs (19, 20). Both SHM and CSR are initiated by activation-induced cytidine deaminase (AID) (21, 22). AID is a small (24 kDa) protein that deaminates cytidine residues on single-stranded DNA (ssDNA), usually in the context of preferred sequence substrate motifs (9). Both SHM and CSR require transcription, both to promote specific AID targeting and also to contribute to formation of requisite ssDNA substrates. Both processes also co-opt activities of normal cellular base excision repair (BER) and mismatch repair (MMR) to convert AID cytidine deamination lesions to cellular base excision repair (BER) and mismatch repair (MMR) outcomes. Each of these AID-associated processes will be discussed in depth in following sections.

During SHM, AID deaminates cytosine residues in IgH and IgL V(D)J exons and the deamination products are processed through specific repair pathways into predominantly point mutations, as well as a low frequency of small insertions and deletions (9, 23, 24). SHM produces nucleotide substitutions at all four bases, with a bias towards transitions over transversions such that approximately two thirds of nucleotide substitutions are transitions (25). The mutation frequency over the rearranged variable region exon as a whole is approximately 10⁻³ mutations per base pair per generation, with the highest levels found within complementarity-determining regions (CDRs) (9, 26, 27). In the context of the GC reaction, B cells with SHMs that increase antigen-binding affinity of their BCR are positively selected and those with SHMs that decrease affinity or inactivate the receptor are negatively selected via rounds of SHM, clonal expansion and affinity-based selection; in this manner SHM leads to affinity maturation of the antibody response (18).

For CSR, AID deaminates cytosine residues in long (1 to 10 kb), repetitive, noncoding, switch (S) regions that lie just upstream of each set of Cμ exons (except Cδ exons, which do not undergo traditional CSR) (17, 28). Deamination products at donor and acceptor S regions are processed to DNA DSBs, as well as point mutations (17, 28). CSR is completed when AID-initiated DSBs generated in two participating S regions are fused to delete intervening DNA including the Cμ exons (12, 28). A switch in expression from IgM to different IgH classes such as IgG, IgE, and IgA occurs when Cμ exons are replaced with one of the sets of downstream Cμ exons (e.g. Cγ, Cε, or Cα exons). Each antibody class is specialized for certain pathogen-elimination functions. For example, IgG promotes phagocytosis of antibody-coated particles, IgE triggers mast cell degranulation, and IgA defends against pathogens at mucosal surfaces (29, 30, 31). Thus, CSR alters an antibody’s effector function to one that may be better suited for a given pathogen-elimination response, while maintaining the same variable region exon and thus antibody-binding specificity.

AID FUNCTIONS THROUGH CYTIDINE DEAMINATION OF TARGET DNA

AID was discovered by a subtractive hybridization approach that employed a mouse B cell line stimulated to undergo CSR from IgM to IgA (32). AID knock-out mice were found to be specifically defective for SHM and CSR (21). Likewise, contemporaneous studies of human patients with an autosomal recessive form of hyper-IgM syndrome, characterized by high levels of IgM in the serum and profound defects in IgH CSR and SHM, showed that they had AID mutations (22). These two types of studies showed that AID is required for both SHM and for IgH CSR and, thus, can be considered a master regulator of peripheral antibody diversification. Subsequent studies further showed that AID also is required for the variable region exon DSBs that initiate the gene conversion process that diversifies chicken antibody repertoires (33).

AID target sequences for SHM

AID deaminates cytidines to uridines in ssDNA (24, 35, 36, 37) preferentially deaminating cytidines in the context of “hotspots” described by the consensus motif DGYW (WRCH on the complementary strand, D = A/G/T, Y = C/T, W = A/T, H = T/C/A, R = A/G) (38).
DGYW motifs are very abundant in the tandem repeat units of S regions, with a high density of DGYW motifs a conserved feature among the S regions of species from frogs to mammals (39). V region exons contain a lower density of DGYW motifs than S regions, but their DGYW frequency is still mildly enriched compared to bulk genomic DNA (39). As genomic DNA usually is in duplex form, transcription-based mechanisms have evolved to generate the requisite ssDNA substrates for AID (this is discussed in more detail in subsequent sections).

Mammalian and frog S regions are highly enriched for AGCT motifs, a palindromic variant of the canonical DGYW, which provide AID hotspot motifs on both DNA strands and contribute to DSB generation (see below). DGYW motifs are considered favored mutational hotspots, but not perfect predictors of mutability, since identical DGYW motifs within a given sequence undergo different levels of mutation (40). In addition, DGYW motifs are ubiquitous throughout the genome, but only undergo AID-initiated mutations in a subset of genes and, in most cases, at frequencies orders of magnitude less than at Ig gene targets (38, 41, 42). Thus, additional targeting mechanisms are important, including substrate sequence context beyond the DGYW motif, and higher level mechanisms including transcription (described in later sections). In addition, there is evidence that differential repair of AID cytidine deamination lesions also can influence final mutation and DSB outcome (9, 42).

AID-initiated lesions are processed by normal repair pathways to yield mutations and DSBs

The point mutations and DSBs that occur during SHM and CSR are generated in two steps (9). In the first step, AID deaminates cytidines to uridines (U) in V(D)J exons during SHM, or in S regions during CSR to produce uracil:guanine (U/G) mismatches. The second step involves error-prone resolution of the U/G lesion by co-opted BER and/or MMR pathway activities (Fig. 3). Normally the BER and MMR pathways repair such lesions in an error-free manner. How activities of the BER and MMR pathways that evolved to maintain genome fidelity are coerced into contributing to generating mutations and DSBs downstream of AID lesions is understood only in part. Below we will describe current knowledge of enzymatic processes involved in the two steps leading from AID-generated U/G mismatches to mutations or DSBs, starting with a description of the normal BER and MMR pathways.

BER repairs damaged bases by the following general steps: (1) recognition and excision of a damaged base (e.g. uracil) from the DNA backbone by an initiating DNA glycosylase (e.g. uracil-DNA-glycosylase) to create an abasic site; (2) cleavage of the DNA backbone at the abasic site by an apurinic/apyridimic (AP) endonuclease, generating a ssDNA nick adjacent to the abasic site; (3) processing of the nick to a single-nucleotide gap; (4) filling in of the gap by DNA polymerase β; and finally (5) sealing of the nick by DNA ligase 1 or DNA ligase 3 (43). An alternative to this one-nucleotide short-patch form of BER is long-patch BER. In long-patch BER, after nicking of the DNA by AP endonuclease, DNA polymerase δ/ε displaces and polymerizes an approximately 2 to 10 bp long tract of DNA (43, 44). The displaced strand is removed by activity catalyzed by the flap structure-specific endonuclease 1 (FEN1), and a remaining nick sealed by DNA ligase 1 (43, 44). MMR functions primarily in repair of base-base mismatches by a process involving (1) recognition of the mismatch by the MSH2-MSH6 heterodimer; (2) recruitment of a complex of MLH1 and PMS2 (MutLα); (3) excision of the patch of DNA surrounding the mismatch by exonuclease-1 (Exo1) to generate a gap; (4) gap-filling by DNA polymerase δ bound to PCNA; and (5) ligation by DNA ligase 1 to seal the nick (45, 46).

During SHM, replication over the initiating U/G lesion can produce transition (purine > purine or pyrimidine > pyrimidine) mutations at C/G base pairs (Fig. 3a) (9). In addition, the uracil can also be excised by UNG of the BER pathway which leads to an abasic site. Replication over the abasic site can lead to both transversions (purine > pyrimidine or pyrimidine > purine) as well as transitions at initiating C/G base pairs (Fig. 3b) (9). Thus, UNG deficiency produces mainly transition mutations at C/G base pairs (9, 25). As AID only deaminates C’s, the mutagenic BER processes described above cannot account for SHM at A and T residues. Rather, SHMs at A’s and T’s depends largely on components of the MMR pathway (47, 48). The MSH2-MSH6 heterodimer recognizes the U/G mismatch (49, 50), Exo1 then excises the patch of DNA containing the mismatch (51), and an error-prone polymerase such as polymerase η resynthesizes the patch (52, 53). Thus, mutations are “spread” from the C/G sites of deamination to nearby A/T sites.

The generation of DSBs during CSR also employs the activities of the BER and MMR pathways (Fig. 3b and c) (9, 20). Following the excision of uracil by UNG, AP endonuclease 1 (APE1) may create a nick at the abasic site (20, 44) (Fig. 3b). Adjacent nicks on opposite strands, for example in the context of AGCT motifs, may be sufficient to generate a DSB, particularly if target
motifs are very dense as they are in S regions (54). Overlapping gaps generated during MMR can also lead to DSBs during CSR (55, 56) (Fig. 3c). Accordingly, while BER-deficiency or MMR-deficiency alone reduces CSR, combined BER and MMR deficiencies (e.g. UNG and MSH2 deficiency) abrogate CSR (9, 55). AID-initiated

![Figure 3: Mechanisms of AID cytidine deamination in SHM and CSR. AID deaminates cytidine (C) to uridines (U). The U/G lesion may be repaired with high fidelity (i.e. to C/G) by conventional base excision repair (BER) or mismatch repair (MMR). Mutagenic outcomes during SHM and CSR are generated by the following processes. a. Replication over the U/G lesion produces transition mutations at C/G base pairs. b. Uracil-DNA-Glycosylase (UNG) of the BER pathway excises the U creating an abasic site. Replication over the abasic site generates transition and transversion mutations at C/G base pairs. N indicates any nucleotide A,G,C, or T. AP endonuclease 1 (APE1) may create a nick at the abasic site. Nicks on both DNA strands may lead to DSBs. c. MSH2-MSH6 of the mismatch repair pathway recognize the U/G mismatch. Exo1 excises the patch of DNA containing the mismatch. Error-prone polymerase resynthesizes the patch leading to spreading of mutations to A/T base pairs. Overlapping gaps may lead to DSBs. doi:10.1128/microbiolspec.MDNA3-0037-2014.f3]
SHMs also accompany the DSBs that are generated in CSR-activated B cells, but in this case only mutations at C/G residues are found with little or no spreading to A/T residues. Such differential targeting/outcomes of AID activity during SHM and CSR will be discussed in more depth below. However, it is abating CSR, also eliminates both C/G transversion mutations and spreading of mutations in the context of V region SHM, leaving only C/G transition mutations and also eliminates C/G transversion mutations in S regions during CSR (50, 55, 57). Thus in the absence of BER and MMR, V and S regions exhibit only the footprint of AID deamination, strongly supporting the two-step model in which AID deamination is followed by processing of the resulting lesions by BER and MMR to DSBs, transversion mutations at C and G residues, and transition and transversion mutations at A and T residues.

SOMATIC HYPERMUTATION

SHM occurs during germinal center responses

The process of somatic hypermutation (SHM) introduces point mutations in the assembled V regions of IgH and IgL genes of mature activated B cells (9). SHM takes place in the GC, a specialized structure found in B cell follicles of peripheral lymphoid organs (e.g., lymph nodes and spleen) where rapidly proliferating B cells accumulate after primary immunization (18). Consistent with the role of AID in SHM, GC B cells express high levels of AID (32). In the GC, antigen-activated B cells, usually with the help of T cells, undergo multiple rounds of SHM and Darvinian-like selection for clones with high-affinity antigen binding followed by clonal expansion, leading to the evolution of B cells that express BCRs with increased affinity for the antigen. This process of affinity maturation is fundamental to the production of high-affinity antibodies to particular pathogens (18, 58).

SHM targets in the endogenous IgH and IgL loci

Both productive and nonproductive IgH and IgL variable region exon alleles of B cells undergoing a GC response are subject to SHM (59, 60); however, only mutations in productive alleles affect the BCR and influence the fate of GC B cells that bind the antigen (58). GC B cells with SHMs that decrease BCR affinity, for antigen or lead to auto-reactivity, are not selected. In addition, B cells in which SHMs alter residues necessary for normal BCR functions (e.g., certain framework residues necessary for proper folding) are lost as BCR tonic (e.g., ligand-independent) signaling is required for survival (61). Given the strict selection for and against SHMs on the productive V(D)J alleles, SHMs on the nonproductive allele, which are not selected for or against, are considered a better indicator of intrinsic mutational patterns that are not biased by antigen selection (62, 63).

SHMs concentrate prominently within the CDRs of the V region exon (9). This accumulation does not appear to be merely due to selection since nonproductive V exons and passenger V transgenes (that provide a transcribed V(D)J exon that is not involved in BCR expression) also show preferential accumulation of SHMs in CDRs (40, 63). Preferential targeting of SHM to CDRs may reflect evolutionary pressure to direct SHM to parts of the V region that bind antigen and away from the intervening FWR, which are important for the Ig’s structural integrity (64). However, the mechanism by which SHMs preferentially accumulate in the CDRs is not known. Differential AID targeting is a likely possibility, with CDRs of VH exons containing somewhat more AID hotspots than other regions of the V exon (40). In this regard, although various codons can encode a given amino acid, CDRs preferentially use those that contain AID hotspot motifs (65, 66). However, it is notable that in nonproductively rearranged V exons, identical AID hotspot motifs (e.g., AGCT) mutate more when located in the CDRs as compared to when located in FWRs (40, 67). Thus, the underlying sequence of CDRs or their flanking regions, beyond AID target motifs, also may have a role in recruiting AID activity for SHM. In addition, the location of the CDRs in terms of their distance from the transcriptional start site (68) and/or other aspects of the overall structure of the V region exon could play a role in directing AID activity. Finally, the possibility of differential repair of AID deamination lesions might also contribute (42). Clearly, there is much to be learned about how AID is targeted within V exons during SHM.

The question of whether or how specific features of V exon sequences contribute to AID targeting must take into account findings that non-Ig sequences, including β-globin, chloramphenicol acetyltransferase and Ig Cκ sequences, are apparently robust substrates when inserted in place of the V exon in transgenic passenger alleles (69, 70, 71). However, the degree of SHM of these transgenes, while sometimes approaching that of the bona fide V exon, is quite dependent on integration site and copy-numbers (69). Thus, it is possible that
features of the integration site or chromatin structural alterations associated with tandem transgene arrays could influence AID targeting in unknown ways. In this regard, targeting of such non-Ig sequences as single copies in place of an endogenous V exon will be required to conclusively address the degree to which these sequences undergo SHM relative to V exon sequences in a physiological endogenous setting. If non-Ig sequences mutate as well as V sequences in the endogenous V location, it would imply that there is nothing absolutely specific to the V exon sequence that targets high levels of SHM. In this case, the question would be why and how SHMs are focused on the CDRs, which may imply evolution of V exon to suppress SHM at FWRs versus CDRs. If non-Ig sequences do not mutate as well as V sequences when expressed in single copies in the normal physiological location, it would support the notion that CDRs evolved to specifically support AID-initiated SHMs.

**Mutational versus deletional outcomes during SHM**

SHM, in contrast to CSR, is generally considered to involve predominantly point mutations and much less frequently DSBs (9). However, deletions have been found at relatively high frequency in some studies of nonproductive V exons (24) and passenger V exons (72). Such deletions generally would result from DSBs, which lead to deletions either through resection or by joining to another DSB in the same V exon (73). Such internal deletions are frequent in S regions in accord with their high DSB frequency (see below). DSBs can also lead to insertions, which have also been found in V exons in association with SHM (24). Together, deletions and insertions are often generically referred to as “indels” (74). DSBs and associated indels must occur at some frequency during SHM but are likely mostly selected against in productive V exon alleles, since they could disrupt reading frame or overall V region structure. The wide variability of the levels of indels found in V exons that have undergone SHM in different experiments (24, 63, 75, 76) could reflect most samples coming from productive alleles, the possibility that different V exon sequences have different propensities to undergo DSBs, limitations of sample size, or other factors. High-throughput sequencing of the Ig variable region exons from HIV-1 infected patients that produce rare broadly neutralizing antibodies have revealed that certain of these broadly neutralizing antibodies are extensively mutated and harbor very frequent indels (77, 78). How these anti-HIV broadly neutralizing antibodies accumulate such high levels of SHM and indels during affinity maturation is still speculative (78). An important question is whether AID-induced SHMs in some unmutated or affinity matured V exons can generate new sequences that further promote or direct DSBs and SHMs.

**Mechanisms that target SHM to specific variable region exon targets**

The mechanisms by which AID is targeted to its substrates is of great interest given the potentially deleterious consequences of AID’s mutagenic activity. Off-target AID activities can activate oncopgenes via mutations or translocations and, thereby, contribute to cancer (79, 80). In this regard, transcription has been shown to be a key factor for targeting AID to V exons (9, 26, 81, 82) as well as to S regions (see below). Correspondingly, deleting the V promoter eliminates SHM (83). In addition, non-Ig promoters can support SHM at least to some degree (83, 84, 85), suggesting that the V promoter per se may not direct AID targeting; but rather that such targeting is provided by transcription in general. Consistent with a key role for transcription, the spatial distribution of SHMs in a V exon is influenced by distance from the transcription start site (TSS), with the TSS defining the 5’ boundary of SHM (9, 76) and SHM frequency decreasing with distance from the TSS (68).

Ig enhancers, which are known transcriptional regulators (13, 82), promote SHM within transgenic V(D)J substrates (84, 86, 87). However, such enhancers, including the IgH intronic enhancer (iEμ), intronic Igκ enhancer (iEx) and 3′ Igκ enhancer were deleted in mice and found to not be required for SHM (87, 88, 89). The difference between the transgene and endogenous findings may reflect redundancy of tested enhancer elements with other enhancers or other types of elements in the endogenous setting (82). The 30 kb IgH 3′RR contains a number of different enhancers and deletion of several of them in the endogenous locus can abolish germline Cμ transcription and CSR to most Cμs (91; see below), without affecting V(D)J transcription or SHM (92). However, recent studies showed that complete deletion of the 30 kb IgH 3′RR in the endogenous locus completely abolishes germline Cμ transcription and CSR (93) and also severely impairs SHM (94). However, the impairment of SHM in the absence of the 30 kb IgH 3′ RR was accompanied by only marginal reduction in transcription (94). These studies imply that the full IgH 3′RR contains elements that may impact AID targeting during SHM via mechanisms beyond transcription (94), as has also been suggested by mutation targeting studies in chicken Igκ (95, see below). Another possibility is
that the type of transcription (e.g. sense versus anti-
sense) is important.

AID must be directed to its intended Ig gene substrate
sequences versus other sequences to maintain specificity
and reduce its potential off-target mutational activity.
In addition, once at a target sequence, AID must gain
access to a requisite ssDNA substrate. Transcription
and transcription-related mechanisms have been implicated
in facilitating both of these steps. In this context, a
number of AID-associated proteins have been described
that may contribute to these activities (96, 97). Tran-
scriptional stalling has been implicated in directing AID
to its targets (71, 81). The transcription associated fac-
tor Suppressor of Ty5 homolog (Spt5) associates with
both AID (98) and RNA pol II (Pol II) (99, 100).
Colocalization of Pol II, AID and Spt5 on genomic sites
is predictive of AID-induced mutated sites, suggesting that
Spt5 may target AID to genomic loci by recruiting AID
to sites of stalled Pol II (98). The factors that lead to Pol
II stalling in AID targets are unknown.

How AID gains access to the ssDNA
substrate following targeting
Following recruitment to targets, AID must gain access
to the ssDNA template. Purified AID deaminates the
nontemplate strand of mammalian S regions transcribed
in vitro because mammalian S regions loop out the
ssDNA nontemplate strand in the form of R loops upon
transcription (35; see below). Purified AID does not
deaminate T7-transcribed V exon substrates, which do
not form R loops (101). However, serine 38 (S38)-
phosphorylated AID in association with replication
protein A (RPA) deaminates V exon substrates in vitro
but only on the nontemplate strand, suggesting that
RPA may assist in stabilizing ssDNA template for AID
deamination and/or enhance further AID recruitment
(101). Consistent with this model, disrupting the S38
phosphorylation of AID dramatically reduces SHM in
GC B cells and CSR in activated B cells in vivo (102, 103).
Both strands of DNA are targeted for SHM. Thus, while
transcriptional stalling may target AID, and co-factors
such as RPA may stabilize ssDNA substrates, a mecha-
nism must exist to provide AID with access to the tem-
plate strand, which may be masked by nascent RNA
transcripts. The RNA exosome has been implicated in
this role as it allows AID to robustly deaminate both
strands of T7 transcribed substrates in vitro and is re-
quired for normal levels of CSR in vivo (104). The RNA
exosome is an evolutionarily conserved exonuclease
that processes nuclear RNA precursors and degrades
RNA in the nucleus and cytoplasm (105). A working
model suggests that once AID is brought to a target via
stalled Pol II and Spt5, the RNA exosome displaces or
degrades the nascent RNA, thus making the template
strand available for deamination, which may in vivo be
further augmented by RPA association (104) (Fig. 4b).
Negative supercoiling is an additional mechanism that
has been proposed to make both DNA strands available
as ssDNA substrates (106, 107).

Factors that promote AID activity
at “on-” and “off-target” genes
As transcription occurs in a large number of genes in
activated B cells, transcription alone cannot explain
occurrence of SHM (and CSR) specifically within Ig loci
and also within just a limited number of “off-target”
genomes, including potential B cell oncogenes, that undergo
SHM at lower levels (42, 108, 109, 110). Thus many
transcribed genes in B cells do not appear to undergo
SHM. In this regard, unknown aspects of sequence
context of AID targeting motifs in V(D)J exon, type of
transcription (see above), aspects of specific chromatin
structure, or cis elements beyond those involved in
transcription may contribute to focusing SHM to Ig
substrates and also contribute to AID mutation to par-
ticular off-targets. Recent studies of the DT40 chicken
B cell line identified within the chicken Igκ gene a di-
versification activator (DIVAC) that targets SHM to the
V exon (95, 111, 112). The DIVAC contains multiple,
redundant transcription factor binding motifs yet en-
hances SHM without stimulating a major increase in
transcription (at least as measured by steady-state tran-
script levels), suggesting that it may target SHM by a
transcription-independent mechanism (95). Notably,
variable mouse and human Ig enhancers also served as
very strong DIVACs in this system, suggesting possible
functional redundancy between enhancers and DIVACs
in promoting SHM in the chicken Igκ locus, by mecha-
nisms that do not involve influences on transcription
per se (95). If so, however, then there still must be ad-
tional elements that function in this context in mam-
alian cells since mouse iEκ and mouse iEx functioned as
DIVACs in the DT40 assay (95), but their deletion
does not markedly impair SHM in mouse GC B cells (88,
90).

IgH CLASS SWITCH RECOMBINATION
Overview
During class switch recombination (CSR), AID intro-
duces cytidine deamination C to U mismatch lesions
into large repetitive S regions that flank the upstream portion of C_H genes. These AID-initiated lesions are then converted into DSBs by co-opted BER and MMR activities (Fig. 3b and c). Deletional end-joining, usually via C-NHEJ of the upstream end of a DSB in Sμ to the downstream end of a DSB in a downstream S region, juxtaposes the V(D)J exon and downstream CH gene to effect CSR. Whether there are mechanisms that promote such deletional joining versus the inversional joining alternative is unknown (6). The primary IgH class switching event involves switching from IgM to the various IgGs, IgE, and IgA, in which DSBs in the "donor" Sμ is joined to "acceptor" Sγ, Sε, or Sa. CSR also can in some cases occur successively. For example, an initial CSR event from Sμ to Sγ1 can generate a hybrid Sμ/Sγ1 donor S region that can subsequently successively undergo CSR with Sh to generate an IgE-producing B cell (30). Downstream "CSR" events between Sγ and Se can also occur in cells in which Sμ has been truncated (113); in theory such a pathway could also contribute to successive CSR but the physiological significance of such downstream CSR events in normal B cells remains to be determined. The targeting of AID to S-region sequences shares at least some common mechanistic aspects with AID targeting during SHM. A primary mechanistic feature of AID targeting for both SHM and CSR is the requirement for transcription, which again both contributes to directing AID to its target S region sequence and which contributes to providing AID with access to ssDNA template.

**Targeting of specific IgH CSR events via differential activation of transcription in various C_H genes**

Each set of C_H exons is part of a “germline” transcription unit referred to as a C_H gene, even though its transcribed RNA does not encode a known protein product (17) (Fig. 2). In these germline C_H genes, transcription is initiated from a cytokine activation-specific promoter upstream of a noncoding “I” (for “intervening”) exon, continues through the associated S region, and terminates downstream of the C_H exons (17, 114). Different cytokines secreted by T cells and other immune cells can stimulate transcription from different I region promoters and, thereby, transcriptionally direct CSR to the C_H region most appropriate for a given type of pathogen or setting of pathogen infection (17, 20). As an example of such regulation, B cells activated through the CD40/CD40L pathway of B and T cell interaction that are also exposed to interleukin-4 (IL-4) (a cytokine secreted by T helper cells) will activate germline transcription from the IL-4 inducible Iγ1 and Iε promoters, and thereby direct CSR to IgG1 and IgE (114, 115, 116).

Induction of such germline C_H transcription occurs on both productive and nonproductive alleles (117),

**FIGURE 4** Transcriptional targeting of AID. 

(a) R-loop structure. An R loop forms from G-rich RNA transcribed from the C-rich template strand forming a stable RNA-DNA hybrid with the C-rich template strand and looping out the G-rich nontemplate strand as ssDNA. 

(b) A working model suggests that once AID is brought to a target via stalled Pol II and Spt5, the RNA exosome displaces or degrades the nascent RNA, thus making the template strand available for deamination, which may in vivo be further augmented by RPA association. Figure adapted from reference 104. doi:10.1128/MicrobiolSpectrum.MDNA3-0037-2014.f4

Hwang et al.
frequently leading to CSR to the same C_H gene on both alleles (118, 119), showing that CSR is a directed recombination process. Deletion of the I exon abolishes germline C_H transcription from the associated I promoter and, thus, eliminates CSR to the corresponding S regions (120, 121). Moreover, replacement of I promoters with heterologous promoters ectopically targets CSR under B cell activation conditions in which it would not normally occur, implicating a direct role for transcription in directing CSR (122, 123, 124, 125). In addition to the cytokine and/or activation of specific I region promoters, differential regulation of transcription from these promoters is modulated by the IgH 3RR located just downstream of the IgH locus (126).

The 30 kb region of IgH 3′RR contains multiple DNase I hypersensitive sites that correspond to enhancers (13). Combined deletion of one subset of these enhancers severely impairs germline C_H transcription from all I region promoters, except the IgY1 promoter, and, correspondingly, severely impairs CSR to the transcriptionally inhibited C_H genes (91, 126). Deletion of the entire 30 kb region deletes additional enhancers and abrogates CSR to all C_H genes, including IgY1, confirming that the IgH 3′RR is a master regulator of CSR, and implicating differential activity of elements within it in controlling CSR (93) and, as described earlier SHM (94). The IgH 3′RR has been implicated in differentially regulating CSR to different C_H genes by a “promoter competition” mechanism, by which certain activated I region promoters compete with and exclude interaction of other I promoters with the IgH 3′RR (126, 127). For example, LPS treatment of B cells induces germline transcription of Cγ2b and Cγ3 and CSR to these C_H genes. However, including IL-4 along with LPS in the treatment activates the IgY1 and Igε promoters which are proposed to inhibit germline transcription from IgY2b and IgY3 promoters via competition for the IgH 3′RR, resulting in inhibition of CSR to Cγ2b and Cγ3 and activation of CSR to Cγ1 and Cε (127). Such regulation is consistent with the finding that the IgH 3′RR forms chromosomal loops with activated I region promoters (128, 129).

**Targeting AID activity within S regions**

Once AID is targeted to S regions via transcriptionally related mechanisms, other features of targeting may overlap at least in part with those discussed for SHM, but there also are notable differences that are discussed below. Mammalian S regions are 1 to 10 kb long sequences composed primarily of tandem repetitive units (39). The deletion of S regions and their replacement with a sequence that lacks S-region features severely impairs CSR to the associated C_H exon, demonstrating that the S region plays a specific role in CSR (117, 130, 131). Mouse and human Sμ, Sα, and Sε are comprised of pentameric repeats, while the Sy regions are comprised of 49 to 52 bp long repeats, that are enriched in smaller repeats including the AGCT and other DGYW targeting motifs (3, 39). In addition, mammalian S regions are C/G-rich and G-rich on the non-template strand (39). In contrast to mammalian S regions, the Sμ of *Xenopus* (frog) is A/T-rich and contains a high density of DGYW sequences, in particular AGCT motifs (39, 132). As mentioned earlier, the palindromic nature of the AGCT motif may make it an optimal substrate for generating DSB breaks by providing AID hotspot motifs on both DNA strands (130, 131), with the high density of such motifs in S regions thereby promoting DSBs.

Transcription through mammalian S regions generates stable R loops in vivo (133). Such R loops result from G-rich RNA transcribed from the C-rich non-template S-region strand forming a stable RNA-DNA hybrid with the C-rich DNA template strand and looping out the G-rich non-template strand as ssDNA (134, 135, 136, 137, 138) (Fig. 4a). In biochemical experiments, AID robustly deaminates the non-template strand of a T7 transcribed S-region substrate that forms an R loop but not a C-rich substrate that does not form an R loop (35). R-loop formation during CSR in B cells is also abolished by inversion of the S-region sequence, which decreases, but does not eliminate, CSR (117). Based on such findings, the formation of R loops in mammalian S regions has been suggested to have evolved to enhance AID access to S regions (35, 138). However, as mentioned, inversion experiments show that R-loop formation is not absolutely required for S-region function in CSR (117). Likewise, *Xenopus* Sμ regions, which are A/T-rich and do not form R loops support substantial CSR when substituted for mouse S regions in activated B cells (131). We note, though, that it remains possible that R-loop formation in mammalian S regions evolved to enhance some other CSR function, for example by playing a role in S-region synopsis (see below).

As for SHM, AID targeting in CSR also appears to involve Pol II stalling, as revealed by accumulation of Pol II in transcribed S regions (139, 140). R loops generated in the transcribed S region may enhance stalling (139, 140). Spt5 also has been implicated in recruiting AID in the context of stalled Pol II in S regions during CSR (98). In vitro, RPA facilitates AID access to the non-template strand of T7 transcribed *Xenopus* Sμ by association with S38-phosphorylated AID (101) and also binds R loop forming S regions in a S38-phosphorylated-AID-
dependent manner (101, 141, 142). Thus, the prevention of AID S38-phosphorylation by an S38A mutation severely impairs RPA association and CSR in mouse B cells without affecting AID catalytic activity, supporting the notion that RPA interaction with S38-phosphorylated AID is important for CSR in vivo (102, 103, 143). The S38-phosphorylation of AID by PKA (143) may be important for CSR in a feedback loop mechanism that involves RPA and downstream repair factors (141, 144). Within transcribed mammalian S regions, template DNA is likely shielded from AID activity by stably hybridized RNA in the form of R loops (133). In this regard, the RNA exosome could function to displace/degrade the nascent RNA, facilitating targeting to the template strand (104) (Fig. 4b). In vitro, at least, the RNA exosome facilitates AID access to both DNA strands of non-R-loop-forming transcribed substrates (104, 145). Whether it plays such a role in vivo is unknown (104). Finally, RPA may stabilize the ssDNA substrates and augment AID activity in a phosphorylation-dependent manner (101, 143).

Evolution of CSR from SHM

Evolutionarily, SHM precedes CSR, with SHM emerging in early jawed vertebrates and CSR emerging in amphibians (146, 147). It has been proposed that Xenopus Sμ evolved to employ mechanisms utilized for AID targeting during SHM of variable region exons, which are not C/G-rich and which do not form R loops when transcribed (131). In this context, the AGCT-rich Xenopus Sμ region may have evolved via duplication of AGCT motif-dense CDR regions of V exons and in that context would target AID by SHM-like mechanisms. The novel features of mammalian S regions (C/G-richness, R-loop formation, much higher AGCT content) may have evolved to further enhance AID targeting or other aspects of CSR (e.g. synopsis, see below; 148). Further understanding of how S-region structure contributes to CSR may be illuminated by studying the divergent S regions of other species which may have evolved alternative or additional solutions for optimizing S-region substrates for CSR. For example, with respect to base composition, duck Sμ is C/G-rich but G:C content is equal between the two strands, while the putative duck Sα has an almost even distribution of the four bases on both strands with only a minor enrichment of G on one strand (29, 149). Assays of the abilities of these divergent S regions to support AID targeting and/or CSR in mammalian B cells could provide new insights into the S-region elements and the types of functions they support.

Joining AID-initiated S-region DSBs to complete CSR: overview

Productive CSR requires the upstream end of a DSB in Sμ to join to the downstream end of a DSB in a downstream S region that lies between 60 and 160 kb away, depending on the targeted S region. AID deaminates multiple cytosines within a given targeted S region (57); this may lead to multiple DSBs, with Sμ thought to be a particularly strong target. In addition to joining to DSBs in other S regions, S-region DSBs may be directly joined back together, be joined back together following end resection, or be joined to another DSB within the same S region (Fig. 5). The latter two outcomes result in internal S-region deletions (ISDs). In addition, DSBs generated in an S region may participate in chromosomal translocations by joining to other non-S-region DSBs on the cis chromosome or to DSBs on other chromosomes (80). Yet, in activated B cell populations in vitro, CSR joins can occur in up to 50% or more cells over a 4-day period (102), raising the question of the nature of the mechanisms that promote CSR events over substantial distances within the IgH locus.

Factors that promote S-region synopsis during CSR

For two DSBs in different genomic locations to be joined, they must be simultaneously broken and physically juxtaposed (synapsed). Thus, joining of two separate DSBs will be influenced by the frequency of DSBs at each site (which reflects both frequency of generation and time of persistence) and by the frequency with which the DSB target sequences are synapsed (6, 150). During V(D)J recombination, appropriate pairs of RSSs are likely synapsed by stochastic mechanisms that are enhanced by locus contraction. Then, RAG1/2 binds to and cleaves the synapsed pairs of RSS ends and subsequently holds them in a post-cleavage complex in which the appropriate ends are joined (e.g. coding end to coding end and RSS end to RSS end) exclusively by C-NHEJ (4, 5, 6). In contrast, AID can clearly act on S regions in the absence of their synapsis, as evidenced by frequent ISDs and other experiments (151). Thus, the question arises as to how joining between DSBs in two separate S regions occurs at sufficient levels to yield physiologic levels of CSR, as opposed to just being rejoined or joined as ISDs.

Studies of recombinational IgH class switching in the absence of S regions or AID have provided insight. Specifically, B cells in which Sμ and Sγ1 are replaced with yeast I-SceI meganuclease target sites can undergo
I-SceI-dependent recombinational class switching to IgG1 at levels approaching those of the wild type (WT) (152, 153). High-throughput analyses of joining frequencies between these widely separated IgH locus I-SceI-generated DSBs indicated that high level joining frequencies also occur in other cell types (e.g. T cells); in addition, similarly separated I-SceI or Cas-9/gRNA separated DSBs joined at high frequency in other chromosomal locations (e.g. around the c-myc gene) (153). Indeed, high-throughput DSB joining assays showed that DSBs within so-called megabase or submegabase topological “domains” (154, 155, 156, 157) are joined at surprisingly high frequency in several tested sites across the genome (6, 150).

The occurrence of sequences, such as of different S regions, within such megabase domains would generate a much greater chance of them being transiently synapsed by diffusion or related mechanisms (153). Thus, if two sequences in such a domain also were frequently broken, they would have a higher probability of being joined when broken while synapsed, thereby driving frequent joining (6, 153) (Fig. 6). CSR has been speculated to have evolved to exploit this general property of gene organization in chromatin (6, 17, 153). In addition, IgH-locus-specific mechanisms may further promote S-region synapsis in activated B cells; such mechanisms might involve the organization of the locus into specific loops via interactions of activated I-region promoters with enhancers and or other types of intra-locus interactions (158). While AID has no known properties of stabilizing the synapsis of two S-region DSBs similar to RAG holding the DSBs in a post-cleavage complex, other factors including those of the DNA damage response regulated by the ataxia telangiectasia mutated kinase (ATM), as discussed below, may contribute such stabilization activities that increase the duration of S-region DSB synapsis to further promote CSR (97, 159, 160) (Fig. 6).
Contribution of the ATM-dependent DNA damage response to both synapsis and joining during CSR

DSBs, including AID-dependent S-region DSBs during CSR, activate the ATM-dependent DNA damage response. In the ATM-dependent DNA damage response, ATM phosphorylates substrates that mediate cell cycle checkpoints, such as p53, and DNA repair factors, including the proteins histone H2AX, mediator of DNA damage checkpoint protein 1 (MDC1), the nibrin (Nbs1) component of the Mre11-Rad50-Nbs1 complex (MRN), and tumor suppressor p53 binding protein 1 (53BP1) (160). ATM-dependent double-strand-break response factors (DSBR factors) form foci that extend megabases along chromatin flanking DSB sites and provide docking sites for protein complexes that bind and tether broken DNA ends (159, 161). DSBs in two different S regions are well within the range of overlapping DSBR foci, which might contribute to stabilizing the synapsis of broken S regions (159). Thus, deficiency for ATM or for several of its substrates, including H2AX, MDC1, and 53BP1 reduce CSR while increasing AID-dependent IgH breaks and translocations (162, 163).

Of the ATM substrates implicated in CSR, 53BP1 appears to have a specialized and especially critical role in CSR. In this regard, CSR in 53BP1-deficient B cells is profoundly reduced (164, 165). In addition, while overall levels of genome instability in CSR-activated 53BP1-deficient B cells are similar to those of ATM- or H2AX-deficient B cells, most of the instability arising in the context of 53BP1 deficiency occurs at the IgH locus, in contrast to the more widespread genomic instability found in the context of ATM or H2AX deficiency (162, 163). Proposed roles for 53BP1 include preventing resection of DSB ends (166, 167), influencing repair pathway choice between C-NHEJ and alternative end-joining (A-EJ) (166), and promoting long-range joining (168). In the latter context, one potential function of 53BP1 would be to stabilize synapsed S regions once they are broken. ATM-dependent phosphorylation of 53BP1 recruits Rif1 to DSBs where it counteracts resection of DNA breaks that would inhibit CSR (169, 170, 171). The ATM-dependent DSB response also contributes directly to the joining of CSR DSBs via the C-NHEJ pathway discussed next (172).
End-joining of synapsed DSB ends to complete IgH CSR

DSBs in two synapsed S regions are end-joined to complete CSR. Most CSR DSBs appear to be joined by C-NHEJ, one of the two major known DSB joining pathways. C-NHEJ DSB end-joining may generate “direct” (or “blunt”) DSB joins, which involves ligating blunt DNA ends (either generated as blunt ends or blunted by end-processing) or, alternatively, C-NHEJ can employ base pairing interactions between short (usually 1 to 4 bp) stretches of homology (“micro-homologies”, MH) present in single-strand overhangs. Consistent with C-NHEJ, most CSR junctions are direct or use short MHs. C-NHEJ and a number of C-NHEJ factors were originally discovered based on the exclusive role of C-NHEJ for joining during V(D)J recombination. During C-NHEJ, DSBs are bound by the Ku70 and Ku80 end-binding complex (Ku), which promotes end ligation carried out by the XRCC4/ligase 4 (Lig4) complex. These factors Ku70, Ku80, XRCC4, and Lig4 are considered the “core” C-NHEJ factors as they are required for joining of all types of broken ends during C-NHEJ. Ku also recruits DNA-PKcs which appears to function both in end synopsis and also by recruiting other factors, most notably the Artemis endonuclease, that process certain types of ends before they can be joined by core C-NHEJ factors.

Although V(D)J recombination and, thus, mature B cell development is strictly dependent on C-NHEJ, this developmental requirement can be circumvented by providing developing B cells with preassembled IgH and IgL variable region exons in their germline (“HL mice”). Germline inactivation of Ku, XRCC4 or Lig4 in HL mice, or conditional inactivation of the latter two factors in mature B cells, demonstrated that CSR can occur at up to 40% of normal levels in the absence of core C-NHEJ factors. In such C-NHEJ deficient B cells, unjoined AID-initiated DSBs progress into AID-dependent IgH locus chromosomal breaks and translocations. DNA-PKcs and Artemis deficiencies generally have a milder effect on CSR levels, consistent at least in part with major roles for these factors in processing certain types of ends but not others for C-NHEJ. In addition, Artemis-independent functions of DNA-PKcs, for example DSB end synopsis, can be provided by functionally redundant activities of the ATM kinase. Despite their more modest contribution to CSR joining than core C-NHEJ factors, the roles of DNA-PKcs and Artemis in CSR are clear since in the absence of either factor, activated B cells accumulate significantly increased levels of IgH locus chromosomal breaks and translocations.

CSR in the absence of C-NHEJ is carried out by one or more alternative end-joining pathways. A-EJ has been described in many ways, including being considered MH-dependent. Yet, depending on context, A-EJ in mammalian cells is not necessarily MH-dependent. Thus, perhaps the best working definition is any end-joining that occurs in the absence of the core C-NHEJ factors. In the absence of XRCC4 or Lig4, nearly all CSR junctions have MHs at their junctions in contrast to only 50% of CSR junctions having MHs in WT B cells. Notably, a fraction of these MHs in CSR junctions from C-NHEJ deficient cells are longer than those typically associated with C-NHEJ. The use of long MH may be influenced by sequence context, with ends from more related S regions having a greater chance to produce long MHs to support joining by A-EJ. Currently, however, the degree to which A-EJ contributes to normal CSR in the presence of intact C-NHEJ remains to be determined.

A-EJ during CSR likely represents more than one pathway. These pathways include a Ku-independent pathway, as Ku-deficient cells also have reduced yet substantial CSR that is somewhat less MH-dependent than that of XRCC4 or Lig4 deficient B cells, and a Lig4-independent pathway that uses a different ligase downstream of Ku and other C-NHEJ components, as B cells lacking both Ku and Lig4 undergo CSR similarly to B cells lacking Ku. Factors thus far implicated in A-EJ during CSR include components of other known DNA repair pathways, including poly (ADP-ribose) polymerase 1 (PARP-1), which may provide a DSB end recognition function, Mre11 and the C-terminal-binding protein 1-interacting protein (CtIP), which may be involved in end-processing, and X-ray repair cross-complementing protein 1 (XRCC1), ligase 1, and ligase 3, which may be involved in end-joining. A-EJ, including the potential role of this pathway in CSR and translocations, has been recently reviewed in depth.

Perspective

Elucidating mechanisms that promote differential targeting and outcome of AID activity in SHM and CSR is a major ongoing question. While many advances have been made, the question of how AID cytidine deamination activity is specifically targeted during SHM and...
CSR remains in substantial part unanswered. As outlined above, transcription of target sequences is required for CSR and SHM. Yet transcription per se is not sufficient to explain the specificity of AID targeting since most transcribed genes in CSR-activated or GC B cells are not subject to detectable AID deamination. In this context, the non-Ig loci that are recurrent targets of lower level AID targeting in CSR-activated or GC B cells also show no readily apparent similarity with S regions and V(D)J exons at the sequence level (193). A longstanding mystery involves the differential targeting of AID to V exons during the GC SHM response versus to S regions during CSR. Thus, AID acts on the S regions in CSR-activated B cells, but not (detectably) on the adjacent V(D)J exons, even though they are actively transcribed (194, 195). Likewise, some GC B cells have robust SHM within their IgH V exons in the absence of CSR (196, 197, 198). How AID achieves such specificity with respect to physiological targets is unknown. Another unanswered question is how the mutational outcome of AID activity is targeted to the three CDRs within variable region exons and the relative contributions of cellular selection versus actual targeting. Clearly, in all of these contexts, there are likely, yet to be defined, sequence motifs that may couple with unique transcription features of particular genes and contribute to make them AID targets. The chicken DIVAC elements, which function to enhance SHM through unknown mechanisms, are one potential example (95, 111, 112).

Once AID cytidine deamination is targeted, various mechanisms may lead to differential outcomes of this potently mutagenic activity during SHM and CSR. In this regard, the level of AID targeting to a particular sequence is usually estimated based on the accumulation of mutations or DSBs and rearrangements/translocations. However, it remains possible that, for at least some sequences, the level at which such genomic alterations are found at particular sequences may be influenced by their predisposition to undergo high fidelity versus mutagenic repair outcomes of the AID-generated cytidine deamination lesions (42). It also is generally considered that AID targeting can preferentially lead to point mutational versus DSB outcomes, respectively, during SHM and CSR. However, more rigorous analyses need to be performed to determine the extent to which these generalizations actually apply, since AID activity at S regions clearly can lead to point mutations and AID activity on V regions can lead to DSBs. Beyond this potential caveat, sequences likely play an important role in the DSB outcome of AID activity in S regions, with a prime example being the abundance in S regions of the palindromic AGCT motif that could promote DSBs by leading to AID targeting on both DNA strands (6, 9, 26). In addition, other aspects of target sequences, differential repair pathways, or SHM versus CSR specific co-factors that favor the generation of point mutation versus DSBs may also play a role.

Another question that is not fully answered is how AID activity leads to C/G mutations in cell lines (199, 200, 201, 202, 203) or in CSR-activated B cells (194) versus spreading of the C/G mutations to A/T base pairs in GC B cells (204, 205). Error-prone DNA polymerases during BER and MMR in GC B cells have been implicated in the spreading process that generates mutations at A/T base pairs (206, 207), raising the possibility that differential expression of these enzymes or other related factors in GC B cells could contribute to SHM spreading (208). However, since V(D)J exons are not AID targets in B cells activated in culture for CSR, this notion has not yet been tested directly.

ACKNOWLEDGMENTS

We thank Feilong Meng, Ming Tian, and Vipul Kumar for helpful comments. This work was supported by NIH R01 AI077595 (to FA) and a fellowship from the Cancer Research Institute (to LSY). FWA is an Investigator of the Howard Hughes Medical Institute.

REFERENCES


Antibody Somatic Hypermutation and Class Switch Recombination


125. Qiu G, Harriman GR, Stavnezer J. 1999. Ilaphe exon-replacement mice synthesize a spliced HPRT(α/-)alpha transcript which may explain their ability to switch to IgA. Inhibition of switching to IgA in these mice. *Int Immunol* **11**(1):37–46.


