BIOCHEMISTRY
including biophysical chemistry & molecular biology

Current Topic

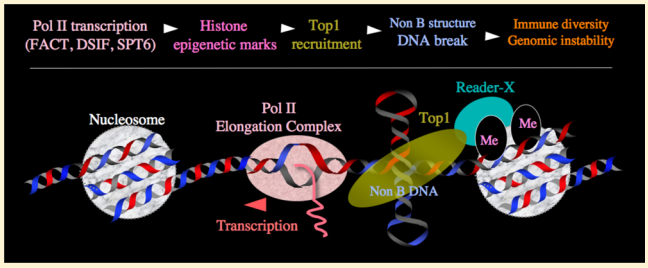
pubs.acs.org/biochemistry

Evolutionary Comparison of the Mechanism of DNA Cleavage with Respect to Immune Diversity and Genomic Instability

Nasim A. Begum and Tasuku Honjo*

Department of Immunology and Genomic Medicine, Graduate School of Medicine, Kyoto University, Yoshida Sakyo-ku, Kyoto 606-8501, Japan

ABSTRACT: It is generally assumed that the genetic mechanism for immune diversity is unique and distinct from that for general genome diversity, in part because of the high efficiency and strict regulation of immune diversity. This expectation was partially met by the discovery of RAG1 and -2, which catalyze V(D)J recombination to generate the immune repertoire of B and T lymphocyte receptors. RAG1 and -2 were later shown to be derived from a transposon. On the other hand, activation-induced cytidine deaminase (AID), which mediates both somatic hypermutation (SHM) and the class-switch recombination (CSR) of the immunoglobulin genes, evolved earlier than RAG1 and -2 in jawless vertebrates. This review compares immune diversity and general genome diversity from an evolutionary perspective, shedding light on the roles of DNA-cleaving enzymes and target recognition markers. This comparison revealed that AID-mediated SHM and CSR share the cleaving enzyme topoisomerase 1 with transcription-associated mutation (TAM) and triplet contraction, which is involved in many genetic diseases. These genome-altering events appear to target DNA with non-B structure, which is induced by the inefficient correction of the excessive supercoiling that is caused by active transcription. Furthermore, an epigenetic modification on chromatin (histone H3K4 trimethylation) is used as a mark for DNA cleavage sites in meiotic recombination, V(D)J recombination, CSR, and SHM. We conclude that acquired immune diversity evolved via the appearance of an AID orthologue that utilized a preexisting mechanism for genomic instability, such as TAM.



The discovery of V(D)J recombination and the subsequent identification of RAG1 and -2 established that a unique sequence-specific recombination is responsible for generating the immune repertoire of T and B cell receptors.[1] Furthermore, transposon insertion sometime early in vertebrate evolution was responsible for introducing RAG1 and -2 and the generation of immunoglobulin (Ig) and T cell receptor genes.[2] These findings led to the assumption that the mechanism for generating acquired immune diversity is different from the mechanisms resulting in general genome diversity.

Subsequently, the antigen stimulation-induced immune diversity in the Ig locus was found to depend on activation-induced cytidine deaminase (AID) and to involve class-switch recombination (CSR), somatic hypermutation (SHM), and gene conversion (GC).[3,4] SHM and GC introduce point mutations into the variable (V) region of Ig genes in a nontemplated and templated manner, respectively. CSR takes place between the donor S region (Sμ) and the acceptor S region (Sγ, Sε, or Sα) of the Ig heavy-chain locus, resulting in a looping-out deletion and replacement of the heavy-chain constant (C_H) region, which is expressed in association with the V_H region. This process changes the isotype from IgM to IgG, IgE, or IgA.

More recently, an AID orthologue was identified in the lamprey *Petromyzon marinus*, the earliest vertebrate ancestor known to exist.[5] *P. marinus* cytidine deaminase PmCAD1 and PmCAD2 carry out GC-type DNA alterations to generate an enormous diversity of non-Ig-type genes called variable lymphocyte receptors (VLR). These findings indicate that the most ancient origin of acquired immune diversity is mediated by an ancestral form of AID.

Most reviews about AID function propose that it directly deaminates C on DNA to create G/U mismatches, which leads to DNA cleavage by either base excision repair or the mismatch repair pathway[6,7] (Figure 1). This DNA deamination hypothesis reinforces the idea that the immune diversity mechanism is unique among genome diversification mechanisms, including mechanisms resulting in genomic instability. In contrast, this review emphasizes the similarities of these mechanisms, based on the RNA editing hypothesis.

General genomic diversification is mediated by the introduction of point mutations and recombinations into somatic and germ cells. The generation of mutations during replication is one of the major causes of genome diversity. However, somatic cells, in which immune diversity is generated, are also equipped with transcription-associated mutagenic mechanisms. Examples are the two- to five-base deletion in transcription-associated mutagenesis (TAM)[8] and triplet contraction or expansion, which is the cause of many genetic

Received: May 8, 2012
Revised: June 12, 2012
Published: June 19, 2012

   dx.doi.org/10.1021/bi3005895 | *Biochemistry* 2012, 51, 5243−5256

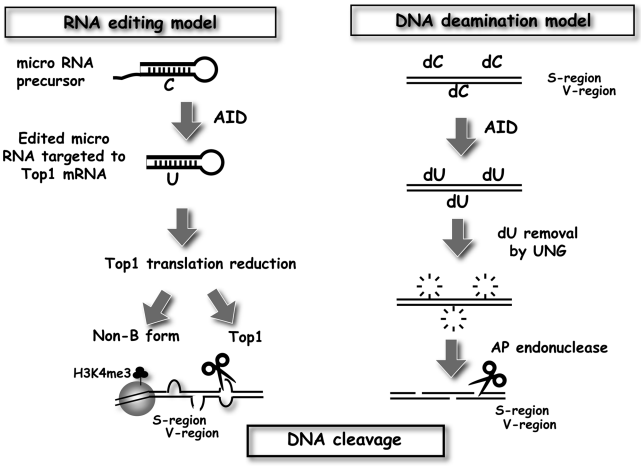


Figure 1. Distinction of two models of AID-induced DNA cleavage of the IgH locus. In the RNA editing model, the cytidine deaminase activity of AID will introduce C to U editing on RNA (a putative microRNA), which in turn will regulate the cellular Top1 to generate a transcription-coupled DNA break at the IgH locus. In the DNA editing model, the cytidine deaminase activity of AID will introduce dC to dU editing on DNA of the IgH locus and the DNA break will be generated at dU during its processing by UNG and AP endonuclease.

diseases.[9] These two mechanisms appear to have several similarities. Both take place at actively transcribed loci, and transcription appears to induce non-B DNA structures and to enhance these genetic alterations.[10−12] Furthermore, topoisomerase 1 (Top1) was recently shown to be responsible for introducing the DNA cleavage in both cases.[13,14] Interestingly, recent studies of the molecular mechanisms for AID-mediated DNA cleavage, based on the RNA editing hypothesis for AID function, revealed that the generation of immune diversity is not completely unique but instead shares features with other types of genome diversification. Two critical features, transcription-associated non-B DNA structure formation and Top1 involvement, are also required for AID-dependent DNA cleavage.[13,14]

The genetic alterations in germ cells are generally mediated by meiotic recombination. The molecular mechanism for this recombination is complex and has been studied most extensively in yeast, using a large number of mutants. The critical step is the introduction of DNA cleavage by a topoisomerase 2-type enzyme, Spo11.[15] It was recently shown that meiotic recombination requires at least two marks to determine where the DNA cleavage takes place. The cis-marking element is a DNA sequence consisting of a loosely conserved 13mer, which is relatively abundant within the genome and usually linked to a promoter. The other mark for cleavage was recently discovered by studies of PRDM9, a histone methyltransferase that recognizes the conserved 13mer DNA sequence by its zinc finger motif and conducts the trimethylation of lysine 4 of histone H3 (H3K4me3) on the surrounding chromatin.[16−18] As PRDM9 is essential for meiotic recombination, the combination of the cis DNA element and trans H3K4me3 histone modification appears to determine the initiation site of the DNA cleavage for meiotic recombination.

An important question in both immune diversity and genomic diversity is how the specific target DNA is chosen and cleaved. The target is not random, because there is clearly biased target specificity in every type of genetic alteration. Surprisingly, the target specificity determination mechanisms for immune diversity and genome diversity, i.e., the combination of cis and trans markers and the enzymes involved, share some common features. In this review, we examine these mechanisms from an evolutionary perspective, which may explain why immune diversity and genome diversity are related at the molecular level.

■ DNA-CLEAVING ENZYMES IN IMMUNE AND GENOME DIVERSITY

Top1 in Genomic Instability. Efficient transcription often induces an unbalanced superhelical structure in the DNA, in which excessively negative supercoiling is created behind the transcription machinery. Top1 is the enzyme that corrects this excessive DNA superhelix (Figure 2). Top1 nicks the DNA,

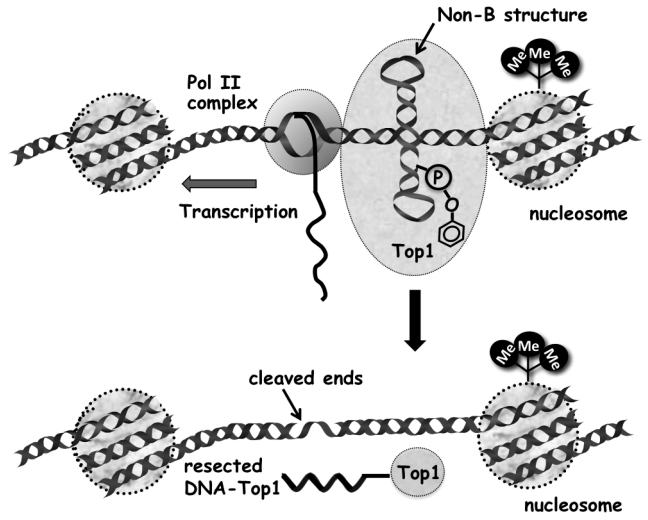


Figure 2. Transcription-coupled DNA cleavage by Top1 during AID-induced CSR and SHM. A negatively supercoiled zone behind the RNA polymerase II complex produces non-B DNA structures, which is enhanced when repetitive sequences or inverted repeats are abundant. Top1 preferentially recognizes non-B DNA structures, introduces a nicking-type cleavage, and forms a covalent intermediate with the DNA. Top1 then normally mediates the directed rotation of the DNA around the axis, to neutralize the negative superhelicity. However, the non-B DNA structure can cause this rotation to fail. As a result, irreversible DNA cleavage takes place, and the Top1−DNA complex remains bound and is removed by end processing repair complexes like MRN or Ctip. The same repair complexes are known to be involved in the resection of Spo11 from meiotic DNA break points. Histone post-translational modification H3K4me3 appears on nucleosomes near the Top1-cleaved non-B structure, indicating that it may serve as a reader−recruiting chromatin mark for AID-induced formation of the DNA cleavage complex.

forms a transient covalent bond at the DNA's 3′ end with one of its tyrosine residues, rotates the DNA around the helix, and then religates the cleaved ends; this set of actions can result in an increased or decreased degree of twisting of the superhelix.[19] Notably, the active transcription of repeat sequences is predicted to form non-B DNA structures.[20,21] When the DNA has aberrant structures like non-B DNA, rotation around the helix may be inhibited, and the cleavage by Top1 can become irreversible and trigger genomic instability (Figure 2).[11,12,20]

TAM has long been known to cause point mutations, microdeletions and -insertions, and recombination.[8,22] Two groups showed by genetic analysis that TAM in yeast is

dx.doi.org/10.1021/bi3005895 | Biochemistry 2012, 51, 5243−5256

Table 1. Genomic Instability Depends on Negative Superhelix Induced by Excessive Transcription

system	biochemical cause	negative superhelix	mutation rate
TAM/triplet D	transcription ↑	excessive, non-B ↑	10^{-6}
SHM/CSR	transcription ↑, Top1 ↓	highly excessive, non-B ↑↑	10^{-3}
human genome	–	normal	10^{-8}

dependent on Top1. Top1-deficient yeast cells are viable but show a drastic reduction in the level of TAM involving two- to five-base deletions.[11,12] Complementation assays clearly showed that the Top1 catalytic activity is essential for TAM. Furthermore, Top1 covalently linked with the target DNA of mutagenesis was demonstrated by ChIP experiments. These results strongly indicated that DNA cleavage by Top1 is the primary cause of TAM.

More recently, the molecular mechanism for triplet contraction or expansion, which is widely known to cause a large variety of genetic diseases, was studied in mammalian model cells. Hubert et al.[10] generated FLAH25 human cells with a hypoxanthine guanine phosphoribosyl transferase (HPRT) reporter gene containing a $(CAG)_{95}$ repeat insert that prevents the expression of HPRT. In this system, reduction of the number of CAG repeats to <38 allows the expression of HPRT. Using this system, the authors showed that Top1 is also involved in DNA cleavage of the triplet repeat region. They screened a chemical library to find effectors of triplet contraction and found that inhibition of the Top1 religation step with camptothecin augmented triplet contraction. Similarly, siRNA against Top1 or its tyrosine-hydrolyzing enzyme, Tdp1, accelerated triplet contraction. These data were interpreted as follows. (1) When the level of Top1 is reduced, the negative supercoil is augmented to form non-B DNA at the CAG repeats, which induces the covalent association of Top1 with DNA. (2) Tdp1 removes the Top1 from the DNA, leading to correct repair. (3) Furthermore, the inhibition of transcription-coupled nucleotide excision repair (TC-NER) blocks triplet contraction, suggesting that TC-NER is responsible for the contraction of CAG repeats after Top1 cleaves and binds to the DNA.

Top1 in Immune Diversity. Top1 was first shown to be involved in the cleavage of S-region DNA during CSR, and considerable evidence indicates that Top1 is required for DNA cleavage during AID-induced CSR. Kobayashi et al.[14] found that the expression of AID but not its loss-of-function mutant reduces Top1 and simultaneously causes irreversible DNA cleavage in the S region. The artificial knockdown of Top1 augments CSR as well as DNA cleavage in the S region. The authors showed that CSR can be blocked by brief exposure to a very low concentration of camptothecin (30 nM), which traps the Top1 bound to DNA, probably by inhibiting the subsequent resection step, and thereby exposes the cleaved DNA ends. Although this result appears to be contradictory to those of the triplet contraction experiments described above, the concentration of camptothecin was 300 times lower than that used by Hubert et al.[10] At the higher concentration, camptothecin is known to induce rather nonspecific DNA cleavage, as shown by Pommier's group. In addition, the mechanism for removing the DNA-bound Top1 may differ between CSR and triplet repeat contraction; hydrolysis (Tdp1) and resection (CtIp and MRN) may be involved in CAG repeat contraction and CSR, respectively, because the knockdown of Ctip and MRN but not Tdp1 affects CSR (unpublished data).

More recently, Kobayashi et al.[13] showed that Top1 is also involved in DNA cleavage of the Ig variable (V) region during SHM. They found that a mutant lymphoma cell line expressing a severely reduced amount of Top1 has a higher mutation frequency than the parental cell line. In addition, Top1 overexpression by a transgene suppresses SHM. Furthermore, in Top1 heterozygous animals, in which the amount of Top1 is half that of the wild type, the level of SHM in the Peyer's patch B cells is dramatically increased. These results clearly indicate that Top1 is involved in DNA cleavage in both CSR and SHM. The mutation rate in TAM is on the order of 10^{-6}, whereas that in SHM is 10^{-3}.[13,14] The difference could be due to the efficiency of non-B-DNA formation, because TAM depends only on augmented transcription while SHM utilizes both an elevated level of transcription and the reduction of Top1 (Table 1).

DNA Deamination by AID in Immune Diversity. The DNA deamination hypothesis for AID function proposes that AID directly deaminates DNA, generating U/G mismatches, which are removed by two repair pathways: base excision repair (BER) and mismatch repair (MMR).[23,24] BER is initiated by uracil DNA glycosylase (UNG), which removes U and generates abasic sites whose phosphodiester bonds are cut by apyrimidine/apurine endonuclease (APEX). In the DNA deamination model, instead of faithful repair by polβ in the normal BER reaction, in B cells many error-prone polymerases are recruited to introduce mutations. Similarly, in the MMR pathway, the MMR proteins Msh2 and Msh6 are also proposed to recognize U/G mismatches and to introduce DNA cleavage.[25]

However, experiments in UNG-deficient mice showed that the level of SHM is instead increased, although CSR is severely damaged. The absence of Msh2/6 also does not greatly affect SHM and affects CSR severely. GC is also affected by the absence of UNG. These results clearly indicate that the functions of UNG and Msh2/6 are different in SHM from those in CSR and GC. This could be because both CSR and GC require a recombination step, while SHM does not. Although the mechanism for recombination differs between CSR and GC, both must go through an unstable pairing phase between two cis loci, where UNG and Msh2/6 might be involved.

In general, both BER and MMR are involved in correcting sporadic mutations in the genome, and their most important roles are during replication. Normally, AP endonuclease-cleaved sites are recognized by polβ and correctly repaired. It is not clear why both the BER and MMR pathways are involved in mutagenesis only in B lymphocytes. Importantly, this model proposes a mechanism that is not found in other cells and organisms. The mechanism that causes these proteins to have unique functions in B lymphocytes is unknown. However, evidence indicates that AID can induce mutations in non-lymphocytes such as hepatocytes and epithelial cells of the stomach, and overexpression of AID can generate tumors in T cells, muscle cells, lung cells, and hepatocytes.[26,27] If these mutations are caused by a common mechanism, AID must have

dx.doi.org/10.1021/bi3005895 | *Biochemistry* 2012, 51, 5243−5256

a unique property that converts the mutation correction function of the BER and MMR pathways into a mutagenic function. The mechanism by which repair systems are turned into a mutagenic system is not yet known.

RAG1 and RAG2 in Immune Diversity. V(D)J recombination is catalyzed by RAG1 and RAG2 to generate the antigen receptor repertoire.[28] These enzymes, together with T and B cell receptor genes carrying a recombination signal sequence (RSS), are thought to have been derived from transposon insertion during early vertebrate evolution. RAG1 and RAG2 share many properties with transposase.[2] They recognize relatively conserved sites and have transposase activity. RAG1 is basically a catalytic subunit for DNA cleavage, and RAG2 is a regulatory element. RAG1 can recognize the RSS, but its DNA cleavage activity is inhibited by its interaction with RAG2, unless the histone mark H3K4me3 is recognized by the PHD domain of RAG2.[29,30] RAG1 and RAG2 are rather specifically expressed in T and B lymphocytes, and their contribution to general genomic instability is probably limited to lymphomagenesis. Nonetheless, RAG1 and RAG2 are reported to cause aberrant translocation in addition to the ordinary V(D)J recombination in the immunoglobulin and T cell receptor loci.[31,32]

From an evolutionary point of view, RAG1 and RAG2 are totally different from the rest of the DNA-cleaving enzymes discussed above. Nonetheless, strikingly, they have adopted a regulatory element involving histone modification. Although RAG1 and RAG2 recognize relatively conserved RSSs, the RSS-like sequences are widely distributed in the genome. Therefore, it is important to restrict the target by another mechanism. That is probably the role of the H3K4me3 recognized by the PHD domain of RAG2. In fact, mutations in the PHD domain of the RAG2 protein cause a severe immune deficiency called Omenn syndrome. Mutation in RAG leads to defects in T and B cell development in individuals with Omen syndrome. Patients are extremely susceptible to infection and suffer from various immunological disorders because of abnormal T cells in the circulation and a lack of B cells.

Spo11 in Meiotic Diversity. It is well established that meiotic recombination uses Spo11, a type of topoisomerase 2, as a DNA-cleaving enzyme. Homodimeric Spo11 cleaves DNA on both strands at sites four bases apart.[33,34] Each monomer of Spo11 forms a covalent association at the DNA strands' 5′ ends, and additional enzymes are required to remove Spo11 from the target DNA to expose the cleaved ends. The enzymes involved in this resection process are Ctip and the MRN complex, which is composed of Mre11, Rad50, and Nbs1.[35] After resection, the DNA has a 10−20-base single-stranded region at both sides of the cleavage that was initially made by Spo11. Subsequent single-strand digestion by exonuclease 1 exposes longer single-stranded regions so that strand invasion of the homologous chromosome can occur, to initiate homologous recombination.

TARGET DNA RECOGNITION AND SPECIFICITY: TRANSCRIPTION, REPEAT SEQUENCE, AND DNA STRUCTURE

Non-B-DNA in Genomic Instability. In general, DNA sequences form the canonical right-handed B-DNA structure in the genome. However, DNA can adopt alternative conformations (non-B-DNA structures), especially in repeated sequences. Several types of non-B structures are known, including hairpins and cruciforms, Z-DNA, H-DNA (triplexes), G-quadraplex (tetraplex), sticky DNA, and looped-out slipped structures.[20,36] Non-B-DNA structure-forming sequences are not randomly distributed in the genome; rather, they are enriched at chromosomal breakage or fragile sites and thus are potential contributors to genomic instability and evolution.

The involvement of non-B conformations has been implicated in translocation-related malignant diseases such as myeloma, leukemia, and lymphoma.[37−39] Breakage hot spots often coincide with various forms of repeat sequences such as direct repeats, inverted repeats, and mirror repeats. Non-B structures are detected in the promoter region of the *c-MYC* gene,[40] which overlaps with break points found in various *c-MYC*-induced malignancies. Similarly, the major breakpoint region (Mbr) of the *BCL-2* gene associated with follicular lymphomas adopts a non-B structure, and disruption of the structure markedly reduces the frequency of translocation events.[41,42] Moreover, non-B-DNA structure-induced genomic instability was directly demonstrated in a mouse model utilizing known Z-DNA- and H-DNA-forming sequences.[43]

Long tracts of GAA.TTC.CGG.CCG and CTG.CAG triplet repeats are a potential source of non-B structures and are implicated in the repeat-induced instabilities in various genetic disorders, including Fragile X syndrome, Huntington's disease, Friedreich's ataxia, and myotonic dystrophy.[37,44] It was demonstrated in a bacterial assay system that a high negative supercoil density enhances the mutagenic potential of CCTG.CAGG repeats but not of CAA.TTG repeats,[45] which is incapable of forming non-B-DNA conformations. The same study also revealed that active transcription is the main factor behind the mutagenesis in quasi-stable non-B sequences. Napierala et al.[46] demonstrated that DNA triplet sequences can easily adopt various non-B-DNA structures under the negative superhelical density that is expected to accumulate behind the running RNA polymerase.[47,48] This finding was further supported by a theoretical analysis predicting that CTG.CAG and CGG.CCG repeats begin to writhe at a lower free energy of supercoiling than nonrepeating DNA.[49]

However, it is still not well understood how a non-B structure-forming sequence receives a DNA break or which cellular enzyme cleaves these structures to generate mutagenic genomic rearrangements. Recent reports suggest that Top1 targets triplet-rich repeat sequences and is responsible for cleaving them.[10,11] Transcription-induced non-B structure is irreversibly cleaved by Top1, as described above. The bulky Top1-bound lesions are presumably repaired by several pathways, which may lead to various mutations, including point mutations, deletions, insertions, and inversions, within the repeat tracts. In fact, the repair of non-B structure-induced breaks was found to be facilitated by terminal homologous regions of several base pairs, which suggested the involvement of the error-prone NHEJ and/or AEJ repair pathways in the end processing.[50,51]

Overall, it appears that Top1 is the critical enzyme for controlling the stability of repeat-rich sequences at two phases during transcription: regulating the superhelical density over the repeat tract and breaking repeat-induced non-B structures. Bidirectional convergent transcription over triplet repeats was also recently reported to enhance the non-B structures[52] that are thought to be the potential Top1 cleavage substrates. Top1's involvement was also reported to enhance R-loop formation over repetitive sequences and R-loop-induced genomic instability.[53,54]

The possibility of bidirectional transcription and R-loop formation at the IgH locus has been examined.[55−58] As Top1 is

involved in AID-induced genomic alterations, it is important to understand what feature causes the locus to adopt a non-B structure that eventually becomes the target of Top1 during transcription-associated CSR and SHM. Interestingly, error-prone repair pathways are well-known to be involved in AID-induced genomic instability.[59,60]

Non-B-DNA in Immune Diversity. CSR takes place between a universal donor S region (Sμ) and one of the acceptor switch regions (Sγ, Sα, and Sε) located upstream of each constant region. I promoters are strategically positioned immediately upstream of each S region and drive the active transcription through the two recombining S regions.[61] S regions are 2−12 kb in length and rich in highly repetitive sequences.[62] CSR is abolished if the S regions or their promoters are deleted, suggesting that transcription and repetitive sequences are mandatory for CSR.[58,63−66]

Transcription through long repetitive regions is likely to form various non-B structures. The S region has characteristic repeat units of 25−80 bp that are arranged tandemly and contain motifs such as TGGGG, GGGGT, GGGCT, GAGCT, and AGCT. Among vertebrates, the S regions of mammals and birds are GC-rich and contain tandem repeats with an abundance of such motifs. The *Xenopus* S region is AT-rich but also harbors AGCT motifs, and replacement of the mouse S region with the *Xenopus* S region does not prevent efficient CSR.[67] On the other hand, a change in the orientation of the S region reduces CSR efficiency.[58,68] More recently, a similar replacement study was conducted in a CSR-proficient cell line (CH12 cells) that undergoes IgA switching.[69] The authors demonstrated that, at the acceptor Sα, the transcriptional orientation and the presence of successive AGCT motifs are critical for CSR. Moreover, alteration of sequence motifs such as ACGT can cause dramatic reductions in the level of germline transcription (GLT) and CSR, suggesting that the individual S-region repeats are naturally set for optimal transcription and non-B structure formation for the initiation of DNA cleavage and recombination.

S-Region sequences are all prone to forming non-B-DNA structures whose stability correlates with the CSR efficiency. Indeed, bisulfite probing of C-rich areas in the Sμ region of CH12 cells showed a non-B structure-forming ability[14] that was enhanced by Top1 knockdown, suggesting that a Top1-dependent supercoiling control mechanism is involved in CSR. Studies applying bisulfite techniques in vivo and in vitro also support the idea that repetitive GC-rich sequences can form non-B structures, including R-loops and G-quartets (G4), during transcription.[70,71]

Like CSR, SHM is tightly coupled to transcription, as revealed by the finding that the level of SHM is reduced when the V-region promoter is deleted and enhanced when the promoter is duplicated.[72−74] Several studies also demonstrated that the SHM rate is proportional to transcription efficiency, which decreases with the distance from the promoter.[75,76] Therefore, SHM-associated DNA breakage also occurs in a highly localized and restricted fashion in which the upstream (promoter) and downstream (JH segment) boundaries are defined, and the mutation rate peaks over the rearranged variable region.[76−78]

Does the V-region sequence possess specific sequence elements that induce point mutations? The V region apparently lacks the S-region-type repetitive sequence signature, leading to the question of whether transcription-induced secondary structure occurs. Transcription-induced single-stranded (non-template) DNA is a prerequisite for the formation of various non-B structures. Bisulfite footprinting of isolated genomic DNA or chromatinized DNA indicated that the V region has a large amount of single-stranded DNA that does not conform to R-loop structures.[79,80] Presumably, negative supercoil-induced single-stranded regions form other types of non-B structures, such as stem loops, as predicted by computer modeling.[81,82] Another line of evidence suggests that tripletlike motifs are the targets of AID-induced SHM, such as AGC, TAC, and their inverse repeates (GCT and GTA, respectively). These triplets are not consecutively arranged as a triplet track, but they are abundant and frequent targets of mutation in the V region.[83,84] Moreover, enhanced SHM is observed when E-box-like sequence elements, which may be prone to forming non-B-DNA, are introduced within the V region.[85]

Consistent with these findings, all the SHM substrates in AID-induced SHM studied so far require high transcription activity.[86−88] The most commonly used SHM substrate is the GFP transgene, which undergoes mutation only when it is transcriptionally active, upon activation of an inducible promoter.[86,87] Interestingly, the GFP gene is also GC-rich and contains several AGCT motifs, and invariably, their positions and vicinity are hot spots for mutations. Consistent with this idea, the transcription of certain repetitive sequence elements in a particular orientation leads to the efficient formation of transcription-induced DNA secondary structures, which are the presumed targets of cleavage and subsequent processing.

Recently, Kobayashi et al.[13] investigated the role of Top1 deficiency in SHM in mice carrying a haplodeficiency of *Top1* and in the GFP substrate in a cell line. Astonishingly, they found that the level of AID-induced SHM is elevated several-fold both at the natural IgH locus in the Top1 haploinsufficient mice and at the GFP gene in Top1-reduced cells. Top1 is the only enzyme so far known to play an essential role in balancing transcription-induced supercoiling. Therefore, it is plausible that transcription-induced negative supercoiling favors the formation of non-B structure over the CDR sequences in the V region, which are eventually recognized in conjunction with site-specific chromatin marks and other cofactors, to be discussed below.

Neither a consensus sequence motif, such as RGYW,[76] nor a specific secondary structure, such as the R-loop, can fully explain the specificity of targeting in CSR or SHM. Although specific secondary structures, such as the R-loops generated by polymerase II (pol-II) stalling during elongation, are likely DNA break targets, no direct evidence of an R-loop-induced DNA break has been found for CSR or SHM. Moreover, no mutational preference was seen in a zone containing large amounts of secondary structure.[89] Neither the pol-II pause sites in the *c-Myc* gene nor enhanced pol-II pausing by a DNA-intercalating transcription inhibitor induces SHM.[90,91] Therefore, it is plausible that transcription-coupled supercoiling favors only non-B structures as key target specificity determinants in conjunction with the site-specific chromatin environment.

We recently investigated this assumption by conducting a genome-wide capture of AID-induced break points. We used a novel approach[92] to identify AID-induced break points in a human Burkitt lymphoma line BL2 that expresses a hyper-mutating form of AID. The break-point library, generated by capturing biotin-dUTP-tagged cleaved DNA ends, was analyzed by both whole-genome sequencing and human promoter array

hybridization. This study identified several novel non-Ig AID target loci, including MALAT1, SNHG3, BCL7A, and CUX1.[92] Similar to the IgH locus, the newly identified loci showed enhanced break signals and a characteristic accumulation of mutations surrounding the break zone, suggesting that the DNA cleavage occurred by a similar mechanism. Therefore, we examined the sequence features of the loci in detail and found that all of these loci are highly transcribed and the break regions harbor abundant repetitive sequences, including AGCT elements (Table 2). These findings support the assumption

Table 2. Non-B Structures Are Common Features of AID Target Loci[a]

	repetitive sequences	non-B form
VDJ	direct and inverted repeats	yes
Sμ	GGGCT, GAGCT, GGGGT	yes
MYC	CCGC	yes
SNHG3	GCCTC	yes
MALAT1	GAAG, AGGA	yes
BCL7A	CGCG	yes
CUX1	CCCG, CGGG	yes

[a]From ref 92.

that the non-B-DNA structure induces breaks for both IgH and non-IgH loci during AID-induced genomic instability. More interestingly, this study also revealed a strong correlation between the non-B structure-induced DNA breaks and epigenetic marks on the break-point chromatin, which will be discussed later.

R-Loop in Deamination of DNA by AID. To explain the role of transcription in CSR, Lieber and his team first proposed that an R-loop forms in the S region,[57,93] because of this region's repetitive sequences and G-rich content.[94,95] In this proposed mechanism, the S region's transcripts (GLT) remain bound to the template-strand DNA, and thus, the untranscribed (nontemplate) strand forms a single-stranded loop (R-loop). In fact, it has been shown using in vitro systems that transcription through the S regions adopts R-loops. The existence of R-loops was further demonstrated in vivo in the genomic DNA of primary B cells by a bisulfite footprint assay.[93] No R-loops are detected if the DNA is pretreated with RNaseH, which digests the RNA in the RNA−DNA heteroduplex.

Because AID deaminates single-stranded DNA very efficiently in vitro,[96,97] the R-loop target model gained considerable support to explain how AID can deaminate dC to dU in the top strand of the S region, which eventually leads to DNA breakage and mutations. However, RNaseH-over-expressing B cells are not defective in CSR, and no R-loop has been detected in the V region.[79,80,98] Moreover, the *Xenopus* Sμ, which is AT-rich and thus lacks the ability to form R-loops, can functionally complement mouse Sμ to mediate CSR.[67] To explain these observations, Alt and colleagues[67,99] instead proposed that the common AGCT repeats are targeted by the phosphorylated AID−replication protein A (RPA) complex through a non-R-loop mechanism during transcription.

Alternative modes of R-loop formation by interfering splicing and mRNA transport have also been suggested. Aguilera and colleagues showed that the stability of the R-loop is extended in a THO complex defective strain of yeast.[95] Again, however, the RNAi-based knockdown of THO complex components in B cells does not show any significant alteration in CSR (ref [100] and unpublished data). Because the template strand forms an RNA−DNA hybrid during transcription, the DNA deamination model requires alternative enzymes that will resolve the hybrid structure to allow dC deamination on the template strand.

Alt and colleagues recently proposed that the RNA-degrading exosome complex,[99] which appears to be associated with AID, helps deaminate both the template and nontemplate strands, even without RPA. It remains to be determined whether this in vitro T7 transcription-based study can be recapitulated during Pol II-driven transcription on chromatin in vivo and confirmed in exosome-deficient B cells. Recently, Sen1 and Xrn2, which are involved in RNA polyadenylation and cleavage, were also reported to be involved in R-loop formation in yeast and mice; thus, there are many molecular players, and it is currently unknown which pathway leads to physiological R-loop formation in the S region.[101,102]

The DNA deamination hypothesis also postulates the involvement of specific cofactors for AID to explain AID's targeting in the S and V regions. However, none of the AID-interacting proteins reported so far, which include Pol II, RPA, Spt5, and PKA, has any locus specificity. In fact, extensive searches have revealed a large number of proteins that interact with AID.[103−111] It is difficult to explain how such a large number of proteins can specify or distinguish DNA targets for CSR and SHM.

It is also puzzling how the DNA deamination model can explain why the C-terminal domain of AID is specifically required for CSR.[112] The C-terminal region of AID is strikingly well conserved among vertebrates, suggesting that it has an important function.[113,114] Several functions have been proposed for this region, including stabilization of the AID protein and protection from cell death induced by double-strand breakage.[111,115,116] However, a C-terminal mutation in AID causes a loss of CSR but a less severe defect in SHM, features that have also been observed in hyper-IgM syndrome type 2 (HIGM2) patients.[117] In addition, C-terminally truncated AID mutants that retain the ER domain show a similar loss of CSR with intact or hyper SHM.[118] The deamination of DNA by AID can explain the DNA cleavage activity but not the C-terminally specific activity required for CSR.

Recombination Signal Sequence (RSS) in V(D)J Recombination. Transcription is also required for the RAG-induced recombination of B and T lymphocyte antigen receptor genes, and the importance of promoters and enhancers for V(D)J recombination is well-established.[119,120] Compared to AID-induced CSR, which is a region-specific recombination, V(D)J recombination is sequence-specific.[1,121] The RAG1−RAG2 complex recognizes RSSs consisting of a palindromic heptamer (5′-CACAGTG) and an A-rich nonamer (5′-ACAAAAACC) with a pair of intervening spacers of 12 or 23 bp. Recombination takes place between a pair of RSSs with different spacer lengths, thereby specifying the 12/23 rule.[122] The V, D, and J gene segments are flanked by RSS elements; they are located at the 3′ end of each V segment, the 5′ and 3′ ends of each D segment, and the 5′ end of each J segment. Although RSSs are well-conserved DNA elements, they are numerous, and most endogenous RSSs deviate from the consensus sequence.[123,124] The nonamer motif is recognized by the nonamer-binding domain of RAG1 and functions to anchor the RAG proteins to the DNA. The heptamer enhances binding of the RAG recombinase to the RSS and specifies the site of DNA cleavage. The first three nucleotides of the heptamer (5′-CAC) are essentially invariant in functional RSSs. In contrast to the nonamer and heptamer sequences, the spacer sequences are

much more variable, but the length of the spacer is highly conserved and helps align the heptamer and nonamer. The sequence of the spacer can also influence the efficiency of V(D)J recombination, although to a small extent.[125,126]

It appears that RSS, RAG1, and RAG2 were introduced in early vertebrates by transposon insertion. The catalytic core sequence of RAG1 shows significant similarity with the transposase of the *Transib* superfamily, including the conserved DDE catalytic motif.[127] It has been demonstrated that RSSs were derived from terminal inverted repeats of an ancient *Transib* transposon. Several studies have suggested that RAG-mediated transposition can lead to translocations, and cryptic RSS sites can be detected in some proto-oncogenes in T cell leukemia.[124,128] Interestingly, several in vitro studies indicate that RAG can cleave specific DNA structures, including non-B structures such as bubbles, flaps, cruciforms, and G4.[71,129] The best-studied example is the Mbr region of the BCL2 gene, which undergoes a t(14;18) translocation; the Mbr region was shown biochemically to form a non-B-DNA structure and to be cleaved by the RAG complex.[42]

The RAG-induced cleavage of non-B structures occurs in the context of a certain sequence and/or base specificity. Theoretical analysis of chromosomal breakage points showed statistically significant CpG enrichment; a mutation in C produces a mismatched bubble, which can be cleaved by RAG alone or in cooperation with other nucleases to introduce DNA breaks.[130,131] Therefore, a bona fide recombination-specific recombinase not only recognizes sequence-specific targets but also can attack non-B structure-forming sites, directly or in cooperation with other enzymes, such as topoisomerases.

Conserved 13mer Sequences in Meiotic Recombination. How is the target specificity of Spo11 determined in germ cells? Spo11-induced double-stranded breaks (DSBs) occur at discrete genomic loci known as recombination hot spots.[132] High-resolution mapping of the recombination crossover sites identified more than 30000 recombination hot spots in the human genome.[133] Remarkably, 40% of them have a 13 bp consensus sequence motif (5′-CCNCCNTNNCCNC-3′), which is recognized by the zinc finger domain of Prdm9, a germ cell-specific histone methyltransferase, in both humans and mice.[16,17,134] Hot spot-specific recombination is severely impaired by an introduced mutation or polymorphism that disrupts this consensus motif. A recent study also found a significant association between DSB hot spots and GC content, short interspersed nuclear elements (SINEs), long interspersed nuclear elements (LINEs), long terminal repeats (LTRs), and additional features like CCTCCCT and CCCCACCCC elements.[133]

Similarly, five classes of recombination hot spot elements were identified in the yeast genome, called motif M26 (5′-ATGACTG-3′), motif 5′-CCAAT-3′, motif-oligo-C (5′-CCCCCGCA-3′), motif 4095 (5′-GGTCTRGAC-3′), and motif 8-6 (5′-WTCGGCCGA-3′).[135,136] Interestingly, the yeast genome lacks the 13 bp motif recognized by the zinc finger domain of human Prdm9. However, recombination hot spots in yeast are often found at promoters or in their vicinity and thus appear to be associated with active transcription. A large body of work in the yeast system suggests that a single site is not sufficient for the specificity but, rather, that the recombination hot spots are context-dependent;[135,137] therefore, sequence-specific binding proteins and the local chromatin environment may both contribute to the selection of a hot spot motif.

TARGET CHROMATIN RECOGNITION: SPECIFICITY BY CHROMATIN MARKS

H3K4me3 in CSR and SHM. Stanlie et al.[138] revealed a requirement for H3K4me3 in the AID-induced DNA break in CSR during an investigation of the role of SSRP1, a subunit of the FACT complex histone chaperone. Furtheremore, depletion of the histone chaperone FACT or of Set1 complex core components does not perturb S-region transcription but inhibits DNA breaks and CSR.[138] They showed that the knockdown of FACT components decreases the amount of H3K4me3 in the acceptor and donor S regions and concomitantly blocks DSB formation in both the S regions. To compare these results with the situation in V(D)J and meiotic recombination, they examined whether knockdown of the core components of the H3K4 methyltransferase complex, ASH2, WDR5, or both, leads to a CSR defect. The depletion of core components of the Set1 H3K4 methyltransferase complex strongly inhibited CSR, which was reminiscent of the decrease in the level of V(D)J recombination by WDR5 depletion or by Set1 H3K4 demethylase overexpression.[139] Notably, depletion of the H3K4 methyltransferase MLL1 did not cause any defect in CSR or in H3K4me3 deposition in the S region, although ASH2 and WDR5 are the core components of *MLL1*.[140]

Because both transcription- and meiotic recombination-associated H3K4me3 marks are mediated by the Set1 H3K4 methyltransferase in yeast,[137,141] the authors investigated the role of Set1 in CSR instead by knocking down Wdr82 and CxxC1, the specific cofactors of the mammalian Set1 enzyme. Indeed, the knockdown of Set1a and Set1b affected CSR more severely compared to the knockdown of other MLLs (1−5) (our unpublished data). Furthermore, meiotic recombination is defective in both yeast and mice if the H2B ubiquitination is blocked, which is a prerequisite for H3K4 formation. The Bre1 ubiquitinating enzyme and Paf1 complex play critical roles in H2B ubiquitination,[142] and Bre1 or Paf1 knockdown also leads to a CSR defect,[138] pointing to a tight link between the chromatin modification cascade and CSR.

Moreover, the trans-histone chromatin modification cascade is intimately associated with transcription elongation involving various accessory elongation complexes such as DSIF,[143,144] which is composed of two subunits, Spt5 and Spt4. Our recent work[145] revealed that Spt5, the larger subunit of the DSIF complex, is involved in H3K4me3 regulation specifically in the Sα region, presumably through its known ability to promote H2B ubiquitination.[143,144,146]

Consistent with the S-region-specific H3K4me3 regulation by Spt5, the levels of CSR and CSR-associated DNA breaks are decreased upon the siRNA-mediated knockdown of Spt5 in B cells. Interestingly, Stanlie et al.[150] observed that this function of the Spt5 subunit of DSIF is independent of the smaller subunit, Spt4, suggesting that the two subunits of this elongation complex have different regulatory roles in CSR. Consistent with this notion, these authors also revealed a novel transcription-coupled DNA repair feature of this complex. Thus, it appears that elongation-associated auxiliary factors that are directly or indirectly involved in the regulation of histone marks like H3K4me3 are critical regulators of CSR.

Accordingly, a second key histone chaperone, Spt6, was found to regulate CSR very similarly. It appears that while FACT and Spt6 are both involved in core histone maintenance in Sμ, they can regulate H3K4me3 in both the acceptor and donor S regions.[145,152] It was initially puzzling that CSR but not

dx.doi.org/10.1021/bi3005895 | *Biochemistry* 2012, 51, 5243−5256

SHM was affected by Spt6 depletion,[108] but it was later discovered that the GFP transgene, which is often used as an SHM substrate, is insensitive to Spt6-mediated H3K4me3 regulation.

On the other hand, SHM in the V region and other non-Ig target loci of AID, such as MALAT1 and SNHG3, show a tight correlation with the formation of Spt6-dependent H3K4me3.[92,147] Remarkably, CDR regions in the V region, which are hot spots for SHM, are strongly correlated with H3K4me3 enrichment (Figure 3). These observations further strengthen the idea that chromatin marks at a particular locus can influence AID-induced genomic alterations. Consistent with this idea, a recent study of AID-induced translocations and chromatin marks at breakpoints indicated that many of the AID target loci are indeed enriched with H3K4me3 and other active histone marks, like H3Kac and H3K36me3.[148]

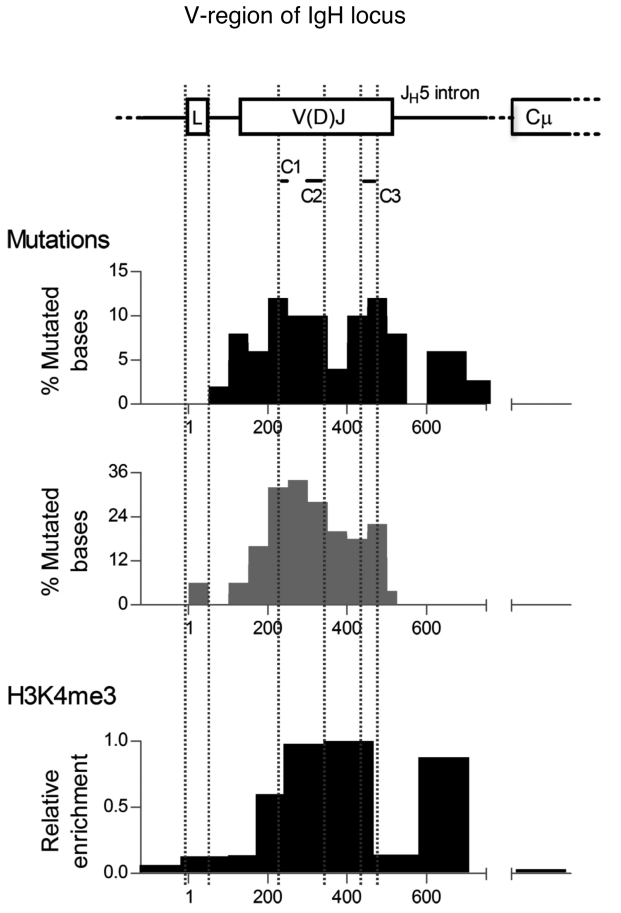


Figure 3. Mutation frequency is correlated with H3K4me3 enrichment in the variable region of the IgH locus. Mutation analysis was performed in a hypermutating human B cell line, BL2, and the mutation rate was calculated as the percentage of mutated bases per 50 bp length. The mutation frequency plots colored black and gray represent data from two independent studies.[92] Three CDRs (C1, C2, and C3) that are hot spots for SHM showed H3K4me3 peaks, but the leader region (L) and its 3′ region were devoid of both H3K4me3 and mutations.

It will be interesting to determine whether H3K4me3, alone or in combination with other marks, determines target specificity, because actively transcribed S regions are enriched with various active histone PTMs.[138,149,150] In particular, the level of histone acetylation is higher in gene segments that are poised to recombine and in meiotic recombination sites. Conversely, histones H3K9 and H3K27, markers of inactive chromatin, are correlated with recombinational inactivity and are absent from meiotic hot spots,[151] although H3K9 has been found only at the donor Sμ, and a few reports suggest that it might be associated with active loci, as well.[105,152,153]

Interestingly, deficiency of a DNA repair factor PTIP, which also acts as a cofactor of MLL2 and MLL3, selectively decreases the level of H3K4me3 at the Iγ2b and Iγ3 promoters, leading to transcriptional inhibition from the respective loci.[154] As both repair and transcription are affected in the PTIP-deficient B cells, the biological significance of H3K4me3 per se in the DNA cleavage step of CSR was not defined by this study. It will be interesting to determine whether a MLL2 or MLL3 deficiency causes an Ig isotype-specific CSR defect in mouse models.

H3K4me3 in V(D)J Recombination. RAG's accessibility to the RSS site is also known to depend on chromatin structure, and earlier studies pointed to the importance of modifications in histone acetylation,[121,155,156] a common characteristic of transcriptionally active domains or open chromatin conformations. Another line of evidence suggests that the ATP-dependent chromatin remodeling complex SWI/SNF plays a role in the accessibility by a nucleosome sliding activity that exposes the recombination site to the RAG complex.[157,158] However, the most compelling discovery emerged recently, which was that RAG2 itself has the ability to recognize H3K4me3.[139,159] While the N-terminus of RAG2 interacts with RAG1 through RAG2's Kelech motif, the C-terminus of RAG2 interacts with chromatin through its PHD domain,[139,159,160] suggesting a novel pathway for target site selection by V(D)J recombinase. An in vitro interaction study indicated that the RAG2 PHD domain has a higher binding affinity for H3K4me3 ($K_d \approx 4\ \mu M$) than for H3K4me2 ($K_d \approx 60\ \mu M$), H3K4me1, or H3K4me0.

RAG2's interaction with H3K4me3 was also confirmed in vivo by genome-wide ChIP sequencing. Furthermore, residues Y415, M443, and W453 of the RAG2 PHD domain form an aromatic binding channel with the K4me3 side chain of histone 3.[139,160] Mutation of any one of the key residues that destroys RAG2's interaction with H3K4me3 severely reduces V(D)J recombination activity. Consistent with this observation, several PHD domain mutations, including W453, have been identified in a rare immunodeficiency syndrome (Omenn), which shows a complete lack of peripheral T and B cells due to the defect in V(D)J recombination.[161] Findings also suggest that the interaction between the PHD domain and H3K4me3 enhances not only RSS recognition but also the catalytic activity of the RAG complex.[29] Another line of investigation showed that RAG2 may remain repressed by a self-inhibitory conformation between its C-terminal end and the PHD domain. Therefore, the interaction of RAG2's PHD domain with H3K4me3 stimulates RAG1 and RAG2 activity by relieving the repressive conformation.[29,30] Overall, the evidence indicates that the structural integrity of the RAG2 PHD domain in binding H3K4me3 is critical for target chromatin recognition and efficient V(D)J recombination.

H3K4me3 in Meiotic Recombination. The H3K4me3 chromatin mark is critical for Spo11-induced meiotic recombination. The chromatin at recombination hot spots is enriched with the H3K4me3 mark, the level of which strongly correlates with the recombination efficiency in yeast, mice, and humans;[18,137,162,163] Spo11-induced meiotic breaks predominantly occur near promoters in yeast, and the promoter areas

dx.doi.org/10.1021/bi3005895 | *Biochemistry* 2012, 51, 5243−5256

of transcribed loci are invariably enriched with the H3K4me3 mark.[135,151] Set1 is the budding yeast H3K4 methyl transferase, which associates with pol-II and is involved in promoter-proximal H3K4me3 formation.[164] The number of meiotic DSBs is dramatically reduced in *set1* deletion mutants, and a direct correlation between the local H3K4me3 level and DSB frequency was demonstrated.[137,141]

In humans and mice, a meiosis-specific methyltransferase, PRDM9, is responsible for the trimethylation of H3K4 at recombination hot spots or crossover (CO) sites.[16,17,134,163] Strikingly, the DSB/CO sites are not exclusively localized to promoters, as observed in yeast, but are nonetheless enriched with H3K4me3.[133] This phenomenon can be explained by the properties of PRDM9, which has a characteristic SET domain and zinc finger module that recognizes recombination hot spot-specific consensus sequences in mammals. Because of the highly polymorphic nature of this zinc finger, PRDM9 recognizes a huge variety of CO hot spots within the genome.[165,166] However, approximately 40% of human recombination hot spots possess a consensus motif that matches the predicted binding site of human PRDM9. Examination of the nucleosome forming ability of the hot spot sequence also revealed that the hot spot consensus for Prdm9 has an intrinsic ability to assemble a nucleosome.[133,167,168] Intriguingly, the level of Prdm9-mediated H3K4me3 peaks at hot spots just before Spo11-induced DSB during meiotic division.

Comparative H3K4me3 profiling between germ (testis) and somatic (liver) tissues revealed that 94% of the hot spots with H3K4me3 overlap, with the majority overlapping the testis-specific H3K4me3 marks (87%).[133] Notably, only a small fraction (16.7%) of the total H3K4me3 marks in the testis correspond with hot spots. The authors suggested that recruitment of the DSB machinery through the PRDM9 complex or H3K4me3 might introduce different binding specificities into its alleles. It is possible that the rest of the testis-specific or Prdm9-induced H3K4me3 has another meiosis-specific function.

Alternatively, other histone epigenetic marks may contribute to meiotic DSB formation in addition to H3K4me3. In fact, other active histone marks like H3K9ac but not the silent histone post-translational modifications (PTMs) (H3K9me3 or H3K27me3) are detected at hot spots.[151] Furthermore, not only active chromatin marks but also various transcription factors are known to contribute to yeast meiotic recombination, such as binding of the Atf1−Pcr1 transcription complex to the ade6-M26 hot spot motif in *Schizosaccharomyces pombe*.[135] Thus, it appears that hot spot consensus sequences are not the sole determinant of target-specific DNA cleavage by Spo11; chromatin marks like H3K4me3 and presumably others play an essential role in determining target site specificity.

CONCLUSIONS

We have compared the DNA cleavage mechanisms involved in genomic instability and immune diversity, focusing on the actual DNA-cleaving enzymes and target specificity determination. The RAG1- and RAG2-mediated immune diversity mechanism is well-established in site-specific recombination and was introduced by transposon insertion. By contrast, AID-mediated Ig diversification evolved earlier in jawless vertebrates, and its mechanism is still being debated. We found that Top1 is used as a cleaving enzyme in CSR and SHM in a manner similar to its function in TAM and triplet diseases. In both immune diversity and genomic instability, the formation of non-B DNA structures by the excessive transcription of repetitive sequences appears to cause irreversible cleavage by Top1. AID further enhances non-B-DNA structure formation by reducing the level of Top1, which we hypothesize to be caused by miRNA editing by an RNA editing activity of AID. Because of the highly efficient formation of non-B structure, the mutation rate in SHM is 1000 times higher than that in TAM.

In all the programmed recombination mechanisms, locus-specific unique cleavage complexes assemble during transcription by recognizing combinatorial histone marks, including H3K4me3. We hypothesize that locus-specific cis elements (non-B structures), combinatorial histone codes, and their associated reader proteins determine the locus-specific DNA breaks in CSR and SHM. According to this model, aberrant AID-induced targeting occurs in many transcriptionally active loci, because these loci may form a similar cis−trans complex that facilitates non-IgH locus DNA cleavage by Top1 (Table 3).

Table 3. Target Specificity Determination by Cis and Trans Marks

cleaving enzyme	cis (DNA)	trans (histone)
RAG1 and -2 (VDJ Rb)	RSS (12/23)	H3K4me3
Spo11 (meiotic Rb)	13mer	H3K4me3
Top1 (CSR, SHM)	non-B	H3K4me3
(TAM, triplet D)	non-B	?

AUTHOR INFORMATION

Corresponding Author

*Telephone: +81-75-753-4371. Fax: +81-75-753-4388. E-mail: honjo@mfour.med.kyoto-u.ac.jp.

Funding

This work was supported by a Grant-in-Aid for Specially Promoted Research (17002015) and a Grant-in-Aid for Scientific Research (24590352) from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We are thankful to Ms. Y. Shiraki for preparing the manuscript.

REFERENCES

(1) Schatz, D. G., and Swanson, P. C. (2011) V(D)J recombination: Mechanisms of initiation. *Annu. Rev. Genet. 45*, 167−202.

(2) Thompson, C. B. (1995) New insights into V(D)J recombination and its role in the evolution of the immune system. *Immunity 3*, 531−539.

(3) Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y., and Honjo, T. (2000) Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell 102*, 553−563.

(4) Muramatsu, M., Nagaoka, H., Shinkura, R., Begum, N. A., and Honjo, T. (2007) Discovery of activation-induced cytidine deaminase, the engraver of antibody memory. *Adv. Immunol. 94*, 1−36.

(5) Rogozin, I. B., Iyer, L. M., Liang, L., Glazko, G. V., Liston, V. G., Pavlov, Y. I., Aravind, L., and Pancer, Z. (2007) Evolution and diversification of lamprey antigen receptors: Evidence for involvement of an AID-APOBEC family cytosine deaminase. *Nat. Immunol. 8*, 647−656.

(6) Longerich, S., and Storb, U. (2005) The contested role of uracil DNA glycosylase in immunoglobulin gene diversification. *Trends Genet. 21*, 253−256.
(7) Storb, U., and Stavnezer, J. (2002) Immunoglobulin genes: Generating diversity with AID and UNG. *Curr. Biol. 12*, R725−R727.
(8) Aguilera, A. (2002) The connection between transcription and genomic instability. *EMBO J. 21*, 195−201.
(9) La Spada, A. R., and Taylor, J. P. (2010) Repeat expansion disease: Progress and puzzles in disease pathogenesis. *Nat. Rev. Genet. 11*, 247−258.
(10) Hubert, L., Jr., Lin, Y., Dion, V., and Wilson, J. H. (2011) Topoisomerase 1 and Single-Strand Break Repair Modulate Transcription-Induced CAG Repeat Contraction in Human Cells. *Mol. Cell. Biol. 31*, 3105−3112.
(11) Lippert, M. J., Kim, N., Cho, J. E., Larson, R. P., Schoenly, N. E., O'Shea, S. H., and Jinks-Robertson, S. (2011) Role for topoisomerase 1 in transcription-associated mutagenesis in yeast. *Proc. Natl. Acad. Sci. U.S.A. 108*, 698−703.
(12) Takahashi, T., Burguiere-Slezak, G., Van der Kemp, P. A., and Boiteux, S. (2011) Topoisomerase 1 provokes the formation of short deletions in repeated sequences upon high transcription in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U.S.A. 108*, 692−697.
(13) Kobayashi, M., Sabouri, Z., Sabouri, S., Kitawaki, Y., Pommier, Y., Abe, T., Kiyonari, H., and Honjo, T. (2011) Decrease in topoisomerase I is responsible for activation-induced cytidine deaminase (AID)-dependent somatic hypermutation. *Proc. Natl. Acad. Sci. U.S.A. 108*, 19305−19310.
(14) Kobayashi, M., Aida, M., Nagaoka, H., Begum, N. A., Kitawaki, Y., Nakata, M., Stanlie, A., Doi, T., Kato, L., Okazaki, I. M., Shinkura, R., Muramatsu, M., Kinoshita, K., and Honjo, T. (2009) AID-induced decrease in topoisomerase 1 induces DNA structural alteration and DNA cleavage for class switch recombination. *Proc. Natl. Acad. Sci. U.S.A. 106*, 22375−22380.
(15) Gerton, J. L., and Hawley, R. S. (2005) Homologous chromosome interactions in meiosis: Diversity amidst conservation. *Nat. Rev. Genet. 6*, 477−487.
(16) Baudat, F., Buard, J., Grey, C., Fledel-Alon, A., Ober, C., Przeworski, M., Coop, G., and de Massy, B. (2010) PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. *Science 327*, 836−840.
(17) Myers, S., Bowden, R., Tumian, A., Bontrop, R. E., Freeman, C., MacFie, T. S., McVean, G., and Donnelly, P. (2010) Drive against hotspot motifs in primates implicates the PRDM9 gene in meiotic recombination. *Science 327*, 876−879.
(18) Parvanov, E. D., Petkov, P. M., and Paigen, K. (2010) Prdm9 controls activation of mammalian recombination hotspots. *Science 327*, 835.
(19) Pommier, Y. (2006) Topoisomerase I inhibitors: Camptothecins and beyond. *Nat. Rev. Cancer 6*, 789−802.
(20) Zhao, J., Bacolla, A., Wang, G., and Vasquez, K. M. (2010) Non-B DNA structure-induced genetic instability and evolution. *Cell. Mol. Life Sci. 67*, 43−62.
(21) Chuzhanova, N., Chen, J. M., Bacolla, A., Patrinos, G. P., Ferec, C., Wells, R. D., and Cooper, D. N. (2009) Gene conversion causing human inherited disease: Evidence for involvement of non-B-DNA-forming sequences and recombination-promoting motifs in DNA breakage and repair. *Hum. Mutat. 30*, 1189−1198.
(22) Kim, N., and Jinks-Robertson, S. (2012) Transcription as a source of genome instability. *Nat. Rev. Genet. 13*, 204−214.
(23) Petersen-Mahrt, S. K., Harris, R. S., and Neuberger, M. S. (2002) AID mutates *E. coli* suggesting a DNA deamination mechanism for antibody diversification. *Nature 418*, 99−103.
(24) Chaudhuri, J., Tian, M., Khuong, C., Chua, K., Pinaud, E., and Alt, F. W. (2003) Transcription-targeted DNA deamination by the AID antibody diversification enzyme. *Nature 422*, 726−730.
(25) Larson, E. D., Cummings, W. J., Bednarski, D. W., and Maizels, N. (2005) MRE11/RAD50 cleaves DNA in the AID/UNG-dependent pathway of immunoglobulin gene diversification. *Mol. Cell 20*, 367−375.
(26) Okazaki, I. M., Hiai, H., Kakazu, N., Yamada, S., Muramatsu, M., Kinoshita, K., and Honjo, T. (2003) Constitutive expression of AID leads to tumorigenesis. *J. Exp. Med. 197*, 1173−1181.
(27) Okazaki, I. M., Kotani, A., and Honjo, T. (2007) Role of AID in tumorigenesis. *Adv. Immunol. 94*, 245−273.
(28) Schlissel, M. S. (2003) Regulating antigen-receptor gene assembly. *Nat. Rev. Immunol. 3*, 890−899.
(29) Shimazaki, N., Tsai, A. G., and Lieber, M. R. (2009) H3K4me3 stimulates the V(D)J RAG complex for both nicking and hairpinning in trans in addition to tethering in cis: Implications for translocations. *Mol. Cell 34*, 535−544.
(30) Grundy, G. J., Yang, W., and Gellert, M. (2010) Autoinhibition of DNA cleavage mediated by RAG1 and RAG2 is overcome by an epigenetic signal in V(D)J recombination. *Proc. Natl. Acad. Sci. U.S.A. 107*, 22487−22492.
(31) Curry, J. D., Schulz, D., Guidos, C. J., Danska, J. S., Nutter, L., Nussenzweig, A., and Schlissel, M. S. (2007) Chromosomal reinsertion of broken RSS ends during T cell development. *J. Exp. Med. 204*, 2293−2303.
(32) Marculescu, R., Vanura, K., Montpellier, B., Roulland, S., Le, T., Navarro, J. M., Jager, U., McBlane, F., and Nadel, B. (2006) Recombinase, chromosomal translocations and lymphoid neoplasia: Targeting mistakes and repair failures. *DNA Repair 5*, 1246−1258.
(33) Keeney, S., Giroux, C. N., and Kleckner, N. (1997) Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. *Cell 88*, 375−384.
(34) Keeney, S. (2008) Spo11 and the Formation of DNA Double-Strand Breaks in Meiosis. *Genome Dynamics and Stability 2*, 81−123.
(35) Hartsuiker, E., Mizuno, K., Molnar, M., Kohli, J., Ohta, K., and Carr, A. M. (2009) Ctp1CtIP and Rad32Mre11 nuclease activity are required for Rec12Spo11 removal, but Rec12Spo11 removal is dispensable for other MRN-dependent meiotic functions. *Mol. Cell. Biol. 29*, 1671−1681.
(36) Wang, A. J., Quigley, G. J., Kolpak, F. J., van der Marel, G., van Boom, J. H., and Rich, A. (1981) Left-handed double helical DNA: Variations in the backbone conformation. *Science 211*, 171−176.
(37) Wang, G., and Vasquez, K. M. (2006) Non-B DNA structure-induced genetic instability. *Mutat. Res. 598*, 103−119.
(38) Nambiar, M., Kari, V., and Raghavan, S. C. (2008) Chromosomal translocations in cancer. *Biochim. Biophys. Acta 1786*, 139−152.
(39) Nambiar, M., and Raghavan, S. C. (2011) How does DNA break during chromosomal translocations? *Nucleic Acids Res. 39*, 5813−5825.
(40) Michelotti, G. A., Michelotti, E. F., Pullner, A., Duncan, R. C., Eick, D., and Levens, D. (1996) Multiple single-stranded cis elements are associated with activated chromatin of the human c-myc gene in vivo. *Mol. Cell. Biol. 16*, 2656−2669.
(41) Raghavan, S. C., Chastain, P., Lee, J. S., Hegde, B. G., Houston, S., Langen, R., Hsieh, C. L., Haworth, I. S., and Lieber, M. R. (2005) Evidence for a triplex DNA conformation at the bcl-2 major breakpoint region of the t(14;18) translocation. *J. Biol. Chem. 280*, 22749−22760.
(42) Raghavan, S. C., Swanson, P. C., Wu, X., Hsieh, C. L., and Lieber, M. R. (2004) A non-B-DNA structure at the Bcl-2 major breakpoint region is cleaved by the RAG complex. *Nature 428*, 88−93.
(43) Wang, G., Carbajal, S., Vijg, J., DiGiovanni, J., and Vasquez, K. M. (2008) DNA structure-induced genomic instability in vivo. *J. Natl. Cancer Inst. 100*, 1815−1817.
(44) Wells, R. D. (1996) Molecular basis of genetic instability of triplet repeats. *J. Biol. Chem. 271*, 2875−2878.
(45) Wojciechowska, M., Napierala, M., Larson, J. E., and Wells, R. D. (2006) Non-B DNA conformations formed by long repeating tracts of myotonic dystrophy type 1, myotonic dystrophy type 2, and Friedreich's ataxia genes, not the sequences per se, promote mutagenesis in flanking regions. *J. Biol. Chem. 281*, 24531−24543.
(46) Napierala, M., Bacolla, A., and Wells, R. D. (2005) Increased negative superhelical density in vivo enhances the genetic instability of triplet repeat sequences. *J. Biol. Chem. 280*, 37366−37376.

(47) Tsao, Y. P., Wu, H. Y., and Liu, L. F. (1989) Transcription-driven supercoiling of DNA: Direct biochemical evidence from in vitro studies. *Cell 56*, 111−118.
(48) Albert, A. C., Spirito, F., Figueroa-Bossi, N., Bossi, L., and Rahmouni, A. R. (1996) Hyper-negative template DNA supercoiling during transcription of the tetracycline-resistance gene in topA mutants is largely constrained in vivo. *Nucleic Acids Res. 24*, 3093−3099.
(49) Bacolla, A., Gellibolian, R., Shimizu, M., Amirhaeri, S., Kang, S., Ohshima, K., Larson, J. E., Harvey, S. C., Stollar, B. D., and Wells, R. D. (1997) Flexible DNA: Genetically unstable CTG.CAG and CGG.CCG from human hereditary neuromuscular disease genes. *J. Biol. Chem. 272*, 16783−16792.
(50) Kozlowski, P., Sobczak, K., and Krzyzosiak, W. J. (2010) Trinucleotide repeats: Triggers for genomic disorders? *Genome Med. 2*, 29.
(51) Crespan, E., Czabany, T., Maga, G., and Hubscher, U. (2012) Microhomology-mediated DNA strand annealing and elongation by human DNA polymerases λ and β on normal and repetitive DNA sequences. *Nucleic Acids Res.*, DOI: 10.1093/nar/gks186.
(52) Nakamori, M., Pearson, C. E., and Thornton, C. A. (2011) Bidirectional transcription stimulates expansion and contraction of expanded (CTG)*(CAG) repeats. *Hum. Mol. Genet. 20*, 580−588.
(53) El Hage, A., French, S. L., Beyer, A. L., and Tollervey, D. (2010) Loss of topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis. *Genes Dev. 24*, 1546−1558.
(54) Lin, Y., Dent, S. Y., Wilson, J. H., Wells, R. D., and Napierala, M. (2010) R loops stimulate genetic instability of CTG.CAG repeats. *Proc. Natl. Acad. Sci. U.S.A. 107*, 692−697.
(55) Tian, M., and Alt, F. W. (2000) Transcription-induced cleavage of immunoglobulin switch regions by nucleotide excision repair nucleases in vitro. *J. Biol. Chem. 275*, 24163−24172.
(56) Perlot, T., Li, G., and Alt, F. W. (2008) Antisense transcripts from immunoglobulin heavy-chain locus V(D)J and switch regions. *Proc. Natl. Acad. Sci. U.S.A. 105*, 3843−3848.
(57) Huang, F. T., Yu, K., Balter, B. B., Selsing, E., Oruc, Z., Khamlichi, A. A., Hsieh, C. L., and Lieber, M. R. (2007) Sequence dependence of chromosomal R-loops at the immunoglobulin heavy-chain Smu class switch region. *Mol. Cell. Biol. 27*, 5921−5932.
(58) Haddad, D., Oruc, Z., Puget, N., Laviolette-Malirat, N., Philippe, M., Carrion, C., Le Bert, M., and Khamlichi, A. A. (2011) Sense transcription through the S region is essential for immunoglobulin class switch recombination. *EMBO J. 30*, 1608−1620.
(59) Yan, C. T., Boboila, C., Souza, E. K., Franco, S., Hickernell, T. R., Murphy, M., Gumaste, S., Geyer, M., Zarrin, A. A., Manis, J. P., Rajewsky, K., and Alt, F. W. (2007) IgH class switching and translocations use a robust non-classical end-joining pathway. *Nature 449*, 478−482.
(60) Rooney, S., Chaudhuri, J., and Alt, F. W. (2004) The role of the non-homologous end-joining pathway in lymphocyte development. *Immunol. Rev. 200*, 115−131.
(61) Stavnezer, J., Kinoshita, K., Muramatsu, M., and Honjo, T. (2004) Chapter 20: Molecular mechanism of class switch recombination. In *Molecular Biology of B Cells* (Honjo, T., Alt, F. W., and Neuberger, M. S., Eds.) pp 307−326, Elsevier Ltd., Amsterdam.
(62) Dunnick, W., Hertz, G. Z., Scappino, L., and Gritzmacher, C. (1993) DNA sequences at immunoglobulin switch region recombination sites. *Nucleic Acids Res. 21*, 365−372.
(63) Luby, T. M., Schrader, C. E., Stavnezer, J., and Selsing, E. (2001) The mu switch region tandem repeats are important, but not required, for antibody class switch recombination. *J. Exp. Med. 193*, 159−168.
(64) Harriman, G. R., Bradley, A., Das, S., Rogers-Fani, P., and Davis, A. C. (1996) IgA class switch in Iα exon-deficient mice. Role of germline transcription in class switch recombination. *J. Clin. Invest. 97*, 477−485.
(65) Xu, L., Gorham, B., Li, S. C., Bottaro, A., Alt, F. W., and Rothman, P. (1993) Replacement of germ-line ε promoter by gene targeting alters control of immunoglobulin heavy chain class switching. *Proc. Natl. Acad. Sci. U.S.A. 90*, 3705−3709.
(66) Oruc, Z., Boumediene, A., Le Bert, M., and Khamlichi, A. A. (2007) Replacement of Iγ3 germ-line promoter by Iγ1 inhibits class-switch recombination to IgG3. *Proc. Natl. Acad. Sci. U.S.A. 104*, 20484−20489.
(67) Zarrin, A. A., Alt, F. W., Chaudhuri, J., Stokes, N., Kaushal, D., Du Pasquier, L., and Tian, M. (2004) An evolutionarily conserved target motif for immunoglobulin class-switch recombination. *Nat. Immunol. 5*, 1275−1281.
(68) Shinkura, R., Tian, M., Smith, M., Chua, K., Fujiwara, Y., and Alt, F. W. (2003) The influence of transcriptional orientation on endogenous switch region function. *Nat. Immunol. 4*, 435−441.
(69) Han, L., Masani, S., and Yu, K. (2011) Overlapping activation-induced cytidine deaminase hotspot motifs in Ig class-switch recombination. *Proc. Natl. Acad. Sci. U.S.A. 108*, 11584−11589.
(70) Raghavan, S. C., Tsai, A., Hsieh, C. L., and Lieber, M. R. (2006) Analysis of non-B DNA structure at chromosomal sites in the mammalian genome. *Methods Enzymol. 409*, 301−316.
(71) Nambiar, M., Goldsmith, G., Moorthy, B. T., Lieber, M. R., Joshi, M. V., Choudhary, B., Hosur, R. V., and Raghavan, S. C. (2011) Formation of a G-quadruplex at the BCL2 major breakpoint region of the t(14;18) translocation in follicular lymphoma. *Nucleic Acids Res. 39*, 936−948.
(72) Fukita, Y., Jacobs, H., and Rajewsky, K. (1998) Somatic hypermutation in the heavy chain locus correlates with transcription. *Immunity 9*, 105−114.
(73) Storb, U., Peters, A., Klotz, E., Rogerson, B., and Hackett, J., Jr. (1996) The mechanism of somatic hypermutation studied with transgenic and transfected target genes. *Semin. Immunol. 8*, 131−140.
(74) Peters, A., and Storb, U. (1996) Somatic hypermutation of immunoglobulin genes is linked to transcription initiation. *Immunity 4*, 57−65.
(75) Bachl, J., Carlson, C., Gray-Schopfer, V., Dessing, M., and Olsson, C. (2001) Increased transcription levels induce higher mutation rates in a hypermutating cell line. *J. Immunol. 166*, 5051−5057.
(76) Peled, J. U., Kuang, F. L., Iglesias-Ussel, M. D., Roa, S., Kalis, S. L., Goodman, M. F., and Scharff, M. D. (2008) The biochemistry of somatic hypermutation. *Annu. Rev. Immunol. 26*, 481−511.
(77) Winter, D. B., and Gearhart, P. J. (1998) Dual enigma of somatic hypermutation of immunoglobulin variable genes: Targeting and mechanism. *Immunol. Rev. 162*, 89−96.
(78) Lebecque, S. G., and Gearhart, P. J. (1990) Boundaries of somatic mutation in rearranged immunoglobulin genes: 5′ boundary is near the promoter, and 3′ boundary is approximately 1 kb from V(D)J gene. *J. Exp. Med. 172*, 1717−1727.
(79) Ronai, D., Iglesias-Ussel, M. D., Fan, M., Li, Z., Martin, A., and Scharff, M. D. (2007) Detection of chromatin-associated single-stranded DNA in regions targeted for somatic hypermutation. *J. Exp. Med. 204*, 181−190.
(80) Parsa, J. Y., Ramachandran, S., Zaheen, A., Nepal, R. M., Kapelnikov, A., Belcheva, A., Berru, M., Ronai, D., and Martin, A. (2012) Negative Supercoiling Creates Single-Stranded Patches of DNA That Are Substrates for AID-Mediated Mutagenesis. *PLoS Genet. 8*, e1002518.
(81) Wright, B. E., Schmidt, K. H., Davis, N., Hunt, A. T., and Minnick, M. F. (2008) II. Correlations between secondary structure stability and mutation frequency during somatic hypermutation. *Mol. Immunol. 45*, 3600−3608.
(82) Wright, B. E., Schmidt, K. H., Minnick, M. F., and Davis, N. (2008) I. VH gene transcription creates stabilized secondary structures for coordinated mutagenesis during somatic hypermutation. *Mol. Immunol. 45*, 3589−3599.
(83) Shapiro, G. S., Aviszus, K., Ikle, D., and Wysocki, L. J. (1999) Predicting regional mutability in antibody V genes based solely on di- and trinucleotide sequence composition. *J. Immunol. 163*, 259−268.
(84) Smith, D. S., Creadon, G., Jena, P. K., Portanova, J. P., Kotzin, B. L., and Wysocki, L. J. (1996) Di- and trinucleotide target preferences of somatic mutagenesis in normal and autoreactive B cells. *J. Immunol. 156*, 2642−2652.

(85) Michael, N., Shen, H. M., Longerich, S., Kim, N., Longacre, A., and Storb, U. (2003) The E box motif CAGGTG enhances somatic hypermutation without enhancing transcription. *Immunity 19*, 235−242.
(86) Bachl, J., and Olsson, C. (1999) Hypermutation targets a green fluorescent protein-encoding transgene in the presence of immunoglobulin enhancers. *Eur. J. Immunol. 29*, 1383−1389.
(87) Yoshikawa, K., Okazaki, I. M., Eto, T., Kinoshita, K., Muramatsu, M., Nagaoka, H., and Honjo, T. (2002) AID enzyme-induced hypermutation in an actively transcribed gene in fibroblasts. *Science 296*, 2033−2036.
(88) Pham, P., Bransteitter, R., Petruska, J., and Goodman, M. F. (2003) Processive AID-catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation. *Nature 424*, 103−107.
(89) Michael, N., Martin, T. E., Nicolae, D., Kim, N., Padjen, K., Zhan, P., Nguyen, H., Pinkert, C., and Storb, U. (2002) Effects of sequence and structure on the hypermutability of immunoglobulin genes. *Immunity 16*, 123−134.
(90) Faili, A., Aoufouchi, S., Gueranger, Q., Zober, C., Leon, A., Bertocci, B., Weill, J. C., and Reynaud, C. A. (2002) AID-dependent somatic hypermutation occurs as a DNA single-strand event in the BL2 cell line. *Nat. Immunol. 3*, 815−821.
(91) Bemark, M., and Neuberger, M. S. (2000) The c-MYC allele that is translocated into the IgH locus undergoes constitutive hypermutation in a Burkitt's lymphoma line. *Oncogene 19*, 3404−3410.
(92) Kato, L., Begum, N. A., Burroughs, A. M., Doi, T., Kawai, J., Daub, C. O., Kawaguchi, T., Matsuda, F., Hayashizaki, Y., and Honjo, T. (2012) Nonimmunoglobulin target loci of activation-induced cytidine deaminase (AID) share unique features with immunoglobulin genes. *Proc. Natl. Acad. Sci. U.S.A. 109*, 2479−2484.
(93) Yu, K., Chedin, F., Hsieh, C. L., Wilson, T. E., and Lieber, M. R. (2003) R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. *Nat. Immunol. 4*, 442−451.
(94) Kim, N., and Jinks-Robertson, S. (2011) Guanine repeat-containing sequences confer transcription-dependent instability in an orientation-specific manner in yeast. *DNA Repair 10*, 953−960.
(95) Dominguez Sanchez, M. S., Barroso, S., Gomez-Gonzalez, B., Luna, R., and Aguilera, A. (2011) Genome instability and transcription elongation impairment in human cells depleted of THO/TREX. *PLoS Genet. 7*, e1002386.
(96) Bransteitter, R., Pham, P., Scharff, M. D., and Goodman, M. F. (2003) Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. *Proc. Natl. Acad. Sci. U.S.A. 100*, 4102−4107.
(97) Yu, K., Roy, D., Bayramyan, M., Haworth, I. S., and Lieber, M. R. (2005) Fine-structure analysis of activation-induced deaminase accessibility to class switch region R-loops. *Mol. Cell. Biol. 25*, 1730−1736.
(98) Lee, C. G., Kinoshita, K., Arudchandran, A., Cerritelli, S. M., Crouch, R. J., and Honjo, T. (2001) Quantitative regulation of class switch recombination by switch region transcription. *J. Exp. Med. 194*, 365−374.
(99) Basu, U., Meng, F. L., Keim, C., Grinstein, V., Pefanis, E., Eccleston, J., Zhang, T., Myers, D., Wasserman, C. R., Wesemann, D. R., Januszyk, K., Gregory, R. I., Deng, H., Lima, C. D., and Alt, F. W. (2011) The RNA exosome targets the AID cytidine deaminase to both strands of transcribed duplex DNA substrates. *Cell 144*, 353−363.
(100) Dominguez-Sanchez, M. S., Barroso, S., Gomez-Gonzalez, B., Luna, R., and Aguilera, A. (2011) Genome instability and transcription elongation impairment in human cells depleted of THO/TREX. *PLoS Genet. 7*, e1002386.
(101) Mischo, H. E., Gomez-Gonzalez, B., Grzechnik, P., Rondon, A. G., Wei, W., Steinmetz, L., Aguilera, A., and Proudfoot, N. J. (2011) Yeast Sen1 helicase protects the genome from transcription-associated instability. *Mol. Cell 41*, 21−32.
(102) Skourti-Stathaki, K., Proudfoot, N. J., and Gromak, N. (2011) Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Mol. Cell 42*, 794−805.
(103) Basu, U., Chaudhuri, J., Alpert, C., Dutt, S., Ranganath, S., Li, G., Schrum, J. P., Manis, J. P., and Alt, F. W. (2005) The AID antibody diversification enzyme is regulated by protein kinase A phosphorylation. *Nature 438*, 508−511.
(104) Chaudhuri, J., Khuong, C., and Alt, F. W. (2004) Replication protein A interacts with AID to promote deamination of somatic hypermutation targets. *Nature 430*, 992−998.
(105) Jeevan-Raj, B. P., Robert, I., Heyer, V., Page, A., Wang, J. H., Cammas, F., Alt, F. W., Losson, R., and Reina-San-Martin, B. (2011) Epigenetic tethering of AID to the donor switch region during immunoglobulin class switch recombination. *J. Exp. Med. 208*, 1649−1660.
(106) MacDuff, D. A., Neuberger, M. S., and Harris, R. S. (2006) MDM2 can interact with the C-terminus of AID but it is inessential for antibody diversification in DT40 B cells. *Mol. Immunol. 43*, 1099−1108.
(107) Nambu, Y., Sugai, M., Gonda, H., Lee, C. G., Katakai, T., Agata, Y., Yokota, Y., and Shimizu, A. (2003) Transcription-coupled events associating with immunoglobulin switch region chromatin. *Science 302*, 2137−2140.
(108) Okazaki, I. M., Okawa, K., Kobayashi, M., Yoshikawa, K., Kawamoto, S., Nagaoka, H., Shinkura, R., Kitawaki, Y., Taniguchi, H., Natsume, T., Iemura, S., and Honjo, T. (2011) Histone chaperone Spt6 is required for class switch recombination but not somatic hypermutation. *Proc. Natl. Acad. Sci. U.S.A. 108*, 7920−7925.
(109) Pasqualucci, L., Kitaura, Y., Gu, H., and Dalla-Favera, R. (2006) PKA-mediated phosphorylation regulates the function of activation-induced deaminase (AID) in B cells. *Proc. Natl. Acad. Sci. U.S.A. 103*, 395−400.
(110) Pavri, R., Gazumyan, A., Jankovic, M., Di Virgilio, M., Klein, I., Ansarah-Sobrinho, C., Resch, W., Yamane, A., Reina San-Martin, B., Barreto, V., Nieland, T. J., Root, D. E., Casellas, R., and Nussenzweig, M. C. (2010) Activation-induced cytidine deaminase targets DNA at sites of RNA polymerase II stalling by interaction with Spt5. *Cell 143*, 122−133.
(111) Wu, X., Geraldes, P., Platt, J. L., and Cascalho, M. (2005) The double-edged sword of activation-induced cytidine deaminase. *J. Immunol. 174*, 934−941.
(112) Ta, V. T., Nagaoka, H., Catalan, N., Durandy, A., Fischer, A., Imai, K., Nonoyama, S., Tashiro, J., Ikegawa, M., Ito, S., Kinoshita, K., Muramatsu, M., and Honjo, T. (2003) AID mutant analyses indicate requirement for class-switch-specific cofactors. *Nat. Immunol. 4*, 843−848.
(113) Barreto, V. M., and Magor, B. G. (2011) Activation-induced cytidine deaminase structure and functions: A species comparative view. *Dev. Comp. Immunol. 35*, 991−1007.
(114) Wakae, K., Magor, B. G., Saunders, H., Nagaoka, H., Kawamura, A., Kinoshita, K., Honjo, T., and Muramatsu, M. (2006) Evolution of class switch recombination function in fish activation-induced cytidine deaminase, AID. *Int. Immunol. 18*, 41−47.
(115) Chaudhuri, J., and Alt, F. W. (2004) Class-switch recombination: Interplay of transcription, DNA deamination and DNA repair. *Nat. Rev. Immunol. 4*, 541−552.
(116) Hasler, J., Rada, C., and Neuberger, M. S. (2011) Cytoplasmic activation-induced cytidine deaminase (AID) exists in stoichiometric complex with translation elongation factor 1α (eEF1A). *Proc. Natl. Acad. Sci. U.S.A. 108*, 18366−18371.
(117) Durandy, A., Peron, S., Taubenheim, N., and Fischer, A. (2006) Activation-induced cytidine deaminase: Structure-function relationship as based on the study of mutants. *Hum. Mutat. 27*, 1185−1191.
(118) Doi, T., Kato, L., Ito, S., Shinkura, R., Wei, M., Nagaoka, H., Wang, J., and Honjo, T. (2009) The C-terminal region of activation-induced cytidine deaminase is responsible for a recombination function other than DNA cleavage in class switch recombination. *Proc. Natl. Acad. Sci. U.S.A. 106*, 2758−2763.

(119) Abarrategui, I., and Krangel, M. S. (2006) Regulation of T cell receptor-α gene recombination by transcription. *Nat. Immunol. 7*, 1109−1115.
(120) Jung, D., Giallourakis, C., Mostoslavsky, R., and Alt, F. W. (2006) Mechanism and control of V(D)J recombination at the immunoglobulin heavy chain locus. *Annu. Rev. Immunol. 24*, 541−570.
(121) Bassing, C. H., Swat, W., and Alt, F. W. (2002) The mechanism and regulation of chromosomal V(D)J recombination. *Cell 109* (Suppl.), S45−S55.
(122) Sakano, H., Huppi, K., Heinrich, G., and Tonegawa, S. (1979) Sequences at the somatic recombination sites of immunoglobulin light-chain genes. *Nature 280*, 288−294.
(123) Rahman, N. S., Godderz, L. J., Stray, S. J., Capra, J. D., and Rodgers, K. K. (2006) DNA cleavage of a cryptic recombination signal sequence by RAG1 and RAG2. Implications for partial V(H) gene replacement. *J. Biol. Chem. 281*, 12370−12380.
(124) Zhang, M., and Swanson, P. C. (2008) V(D)J recombinase binding and cleavage of cryptic recombination signal sequences identified from lymphoid malignancies. *J. Biol. Chem. 283*, 6717−6727.
(125) Lee, A. I., Fugmann, S. D., Cowell, L. G., Ptaszek, L. M., Kelsoe, G., and Schatz, D. G. (2003) A functional analysis of the spacer of V(D)J recombination signal sequences. *PLoS Biol. 1*, E1.
(126) Schatz, D. G., and Ji, Y. (2011) Recombination centres and the orchestration of V(D)J recombination. *Nat. Rev. Immunol. 11*, 251−263.
(127) Kapitonov, V. V., and Jurka, J. (2005) RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons. *PLoS Biol. 3*, e181.
(128) Marculescu, R., Le, T., Simon, P., Jaeger, U., and Nadel, B. (2002) V(D)J-mediated translocations in lymphoid neoplasms: A functional assessment of genomic instability by cryptic sites. *J. Exp. Med. 195*, 85−98.
(129) Nambiar, M., and Raghavan, S. C. (2012) Mechanism of Fragility at BCL2 Gene Minor Breakpoint Cluster Region during t(14;18) Chromosomal Translocation. *J. Biol. Chem. 287*, 8688−8701.
(130) Mahowald, G. K., Baron, J. M., and Sleckman, B. P. (2008) Collateral damage from antigen receptor gene diversification. *Cell 135*, 1009−1012.
(131) Tsai, A. G., Lu, H., Raghavan, S. C., Muschen, M., Hsieh, C. L., and Lieber, M. R. (2008) Human chromosomal translocations at CpG sites and a theoretical basis for their lineage and stage specificity. *Cell 135*, 1130−1142.
(132) Arnheim, N., Calabrese, P., and Tiemann-Boege, I. (2007) Mammalian meiotic recombination hot spots. *Annu. Rev. Genet. 41*, 369−399.
(133) Smagulova, F., Gregoretti, I. V., Brick, K., Khil, P., Camerini-Otero, R. D., and Petukhova, G. V. (2011) Genome-wide analysis reveals novel molecular features of mouse recombination hotspots. *Nature 472*, 375−378.
(134) Mihola, O., Trachtulec, Z., Vlcek, C., Schimenti, J. C., and Forejt, J. (2009) A mouse speciation gene encodes a meiotic histone H3 methyltransferase. *Science 323*, 373−375.
(135) Wahls, W. P., and Davidson, M. K. (2010) Discrete DNA sites regulate global distribution of meiotic recombination. *Trends Genet. 26*, 202−208.
(136) Wahls, W. P., and Davidson, M. K. (2011) DNA sequence-mediated, evolutionarily rapid redistribution of meiotic recombination hotspots. *Genetics 189*, 685−694.
(137) Borde, V., Robine, N., Lin, W., Bonfils, S., Geli, V., and Nicolas, A. (2009) Histone H3 lysine 4 trimethylation marks meiotic recombination initiation sites. *EMBO J. 28*, 99−111.
(138) Stanlie, A., Aida, M., Muramatsu, M., Honjo, T., and Begum, N. A. (2010) Histone3 lysine4 trimethylation regulated by the facilitates chromatin transcription complex is critical for DNA cleavage in class switch recombination. *Proc. Natl. Acad. Sci. U.S.A. 107*, 22190−22195.
(139) Matthews, A. G., Kuo, A. J., Ramon-Maiques, S., Han, S., Champagne, K. S., Ivanov, D., Gallardo, M., Carney, D., Cheung, P., Ciccone, D. N., Walter, K. L., Utz, P. J., Shi, Y., Kutateladze, T. G., Yang, W., Gozani, O., and Oettinger, M. A. (2007) RAG2 PHD finger couples histone H3 lysine 4 trimethylation with V(D)J recombination. *Nature 450*, 1106−1110.
(140) Trievel, R. C., and Shilatifard, A. (2009) WDR5, a complexed protein. *Nat. Struct. Mol. Biol. 16*, 678−680.
(141) Sollier, J., Lin, W., Soustelle, C., Suhre, K., Nicolas, A., Geli, V., and de La Roche Saint-Andre, C. (2004) Set1 is required for meiotic S-phase onset, double-strand break formation and middle gene expression. *EMBO J. 23*, 1957−1967.
(142) Weake, V. M., and Workman, J. L. (2008) Histone ubiquitination: Triggering gene activity. *Mol. Cell 29*, 653−663.
(143) Chen, Y., Yamaguchi, Y., Tsugeno, Y., Yamamoto, J., Yamada, T., Nakamura, M., Hisatake, K., and Handa, H. (2009) DSIF, the Paf1 complex, and Tat-SF1 have nonredundant, cooperative roles in RNA polymerase II elongation. *Genes Dev. 23*, 2765−2777.
(144) Chandrasekharan, M. B., Huang, F., and Sun, Z. W. (2010) Histone H2B ubiquitination and beyond: Regulation of nucleosome stability, chromatin dynamics and the trans-histone H3 methylation. *Epigenetics 5*, 460−468.
(145) Stanlie, A., Begum, N. A., Akiyama, H., and Honjo, T. (2012) The DSIF Subunits Spt4 and Spt5 Have Distinct Roles at Various Phases of Immunoglobulin Class Switch Recombination. *PLoS Genet. 8*, e1002675.
(146) Zhou, K., Kuo, W. H., Fillingham, J., and Greenblatt, J. F. (2009) Control of transcriptional elongation and cotranscriptional histone modification by the yeast BUR kinase substrate Spt5. *Proc. Natl. Acad. Sci. U.S.A. 106*, 6956−6961.
(147) Begum, N. A., Stanlie, A. S., Akiyama, H., and Honjo, T. (2012) The histone chaperone SPT6 is required for AID Target determination through H3K4me3 regulation, submitted for publication.
(148) Klein, I. A., Resch, W., Jankovic, M., Oliveira, T., Yamane, A., Nakahashi, H., Di Virgilio, M., Bothmer, A., Nussenzweig, A., Robbiani, D. F., Casellas, R., and Nussenzweig, M. C. (2011) Translocation-capture sequencing reveals the extent and nature of chromosomal rearrangements in B lymphocytes. *Cell 147*, 95−106.
(149) Wang, L., Wuerffel, R., Feldman, S., Khamlichi, A. A., and Kenter, A. L. (2009) S region sequence, RNA polymerase II, and histone modifications create chromatin accessibility during class switch recombination. *J. Exp. Med. 206*, 1817−1830.
(150) Yamane, A., Resch, W., Kuo, N., Kuchen, S., Li, Z., Sun, H. W., Robbiani, D. F., McBride, K., Nussenzweig, M. C., and Casellas, R. (2011) Deep-sequencing identification of the genomic targets of the cytidine deaminase AID and its cofactor RPA in B lymphocytes. *Nat. Immunol. 12*, 62−69.
(151) Buard, J., Barthes, P., Grey, C., and de Massy, B. (2009) Distinct histone modifications define initiation and repair of meiotic recombination in the mouse. *EMBO J. 28*, 2616−2624.
(152) Brinkman, A. B., Roelofsen, T., Pennings, S. W., Martens, J. H., Jenuwein, T., and Stunnenberg, H. G. (2006) Histone modification patterns associated with the human X chromosome. *EMBO Rep. 7*, 628−634.
(153) Vakoc, C. R., Mandat, S. A., Olenchock, B. A., and Blobel, G. A. (2005) Histone H3 lysine 9 methylation and HP1γ are associated with transcription elongation through mammalian chromatin. *Mol. Cell 19*, 381−391.
(154) Daniel, J. A., Santos, M. A., Wang, Z., Zang, C., Schwab, K. R., Jankovic, M., Filsuf, D., Chen, H. T., Gazumyan, A., Yamane, A., Cho, Y. W., Sun, H. W., Ge, K., Peng, W., Nussenzweig, M. C., Casellas, R., Dressler, G. R., Zhao, K., and Nussenzweig, A. (2010) PTIP promotes chromatin changes critical for immunoglobulin class switch recombination. *Science 329*, 917−923.
(155) McBlane, F., and Boyes, J. (2000) Stimulation of V(D)J recombination by histone acetylation. *Curr. Biol. 10*, 483−486.
(156) McMurry, M. T., and Krangel, M. S. (2000) A role for histone acetylation in the developmental regulation of VDJ recombination. *Science 287*, 495−498.

(157) Roth, D. B., and Roth, S. Y. (2000) Unequal access: Regulating V(D)J recombination through chromatin remodeling. *Cell 103*, 699–702.
(158) Du, H., Ishii, H., Pazin, M. J., and Sen, R. (2008) Activation of 12/23-RSS-dependent RAG cleavage by hSWI/SNF complex in the absence of transcription. *Mol. Cell 31*, 641–649.
(159) Liu, Y., Subrahmanyam, R., Chakraborty, T., Sen, R., and Desiderio, S. (2007) A plant homeodomain in RAG-2 that binds hypermethylated lysine 4 of histone H3 is necessary for efficient antigen-receptor-gene rearrangement. *Immunity 27*, 561–571.
(160) Ramon-Maiques, S., Kuo, A. J., Carney, D., Matthews, A. G., Oettinger, M. A., Gozani, O., and Yang, W. (2007) The plant homeodomain finger of RAG2 recognizes histone H3 methylated at both lysine-4 and arginine-2. *Proc. Natl. Acad. Sci. U.S.A. 104*, 18993–18998.
(161) Villa, A., Sobacchi, C., and Vezzoni, P. (2002) Omenn syndrome in the context of other B cell-negative severe combined immunodeficiencies. *Isr. Med. Assoc. J. 4*, 218–221.
(162) Kniewel, R., and Keeney, S. (2009) Histone methylation sets the stage for meiotic DNA breaks. *EMBO J. 28*, 81–83.
(163) Grey, C., Barthes, P., Chauveau-Le Friec, G., Langa, F., Baudat, F., and de Massy, B. (2011) Mouse PRDM9 DNA-binding specificity determines sites of histone H3 lysine 4 trimethylation for initiation of meiotic recombination. *PLoS Biol. 9*, e1001176.
(164) Ng, H. H., Robert, F., Young, R. A., and Struhl, K. (2003) Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. *Mol. Cell 11*, 709–719.
(165) Ponting, C. P. (2011) What are the genomic drivers of the rapid evolution of PRDM9? *Trends Genet. 27*, 165–171.
(166) Segurel, L., Leffler, E. M., and Przeworski, M. (2011) The case of the fickle fingers: how the PRDM9 zinc finger protein specifies meiotic recombination hotspots in humans. *PLoS Biol. 9*, e1001211.
(167) Getun, I. V., Wu, Z. K., Khalil, A. M., and Bois, P. R. (2010) Nucleosome occupancy landscape and dynamics at mouse recombination hotspots. *EMBO Rep. 11*, 555–560.
(168) Wu, Z. K., Getun, I. V., and Bois, P. R. (2010) Anatomy of mouse recombination hot spots. *Nucleic Acids Res. 38*, 2346–2354.