Activation-induced Deaminase in B Lymphocyte Maturation and Beyond

Michel Cogné

Activation-induced deaminase (AID), a member of the AID/apolipoprotein B mRNA-editing enzyme-catalytic (APOBEC) family, deaminates DNA cytidines into uridines and is the major trans-acting player of immunoglobulin (Ig) genes' diversification in mature B lymphocytes. It allows multiple antigen-driven Ig modifications through gene conversion and/or somatic hypermutation of variable region genes and also permits to switch from IgM expression to other antibody classes after class switch recombination, or to stop Ig expression after locus suicide recombination. AID is expressed at high levels into germinal center activated B cells with a very stringent temporal and spatial regulation. Despite multiple levels of regulation, off-target effects of AID are quite frequent in the B cell lineage and can affect a number of non-Ig genes, albeit at lower level than Ig genes. Beyond the immune system, AID also contributes to cytosine demethylation in undifferentiated cells by deaminating methylcytosines into thymines which are further processed by thymidine glycosylase. This contributes to the maintenance of pluripotency and to the limitation of genetic imprinting. Since AID attacks on DNA can induce replication errors, base excision repair, or mismatch repair, they are strongly mutagenic and are also driving forces of tumorigenesis, not only in B cell malignancies but also in some non-lymphoid tumors involving ectopic AID expression. *(Biomed J 2013;36:259-268)*

Prof. Michel Cogné

Activation-induced deaminase (AID) is an ancestral member of the AID/APOBEC family of proteins able to deaminate DNA and/or RNA cytidines into uridines; it targets immunoglobulin (Ig) variable (V) regions in all vertebrate species carrying B cells, from cartilaginous fishes to amphibians, birds, and mammals.[1] An AID ortholog is even present in lamprey and is responsible for diversification through gene conversion (GCV) of the variable lymphocyte receptor (VLR) antigen-specific molecules produced by lymphoid cells.[2] Thus, AID supports acquired immune system diversification and acts in antigen-stimulated B cells by allowing antigen-driven Ig diversification through GCV and/or somatic hypermutation (SHM) of V region genes. It also supports the functional diversification of Ig heavy chain (IgH) constant (C_H) regions in those species having later developed class switch recombination (CSR) of C_H

genes, such as frogs, birds, and mammals. In that case, cytosine deamination in repetitive switch (S) regions preceding C_H genes initiates DNA breaks and further recombination. B cell fate can also be more dramatically modified by AID after CSR-like events resulting in complete deletion of the IgH CH gene cluster and thus inducing B cell death through locus suicide recombination (LSR).[3]

AID was first identified as a protein specifically expressed into germinal center (GC) B cells[4] and subsequently shown to be mandatory in mouse B cells for SHM and CSR,[5] while it was absent or mutated in some families of patients with a hyper-IgM immune deficiency.[6] Beyond the immune system where its deaminating activity mostly concerns cytosines, AID is now also considered an important contributor to cytosine demethylation in undifferentiated cells where it can deaminate 5-methyl or 5-hydroxymethyl cytosines into

thymine or 5-hydroxythymines, respectively.[7,8] It was also recently suggested recently that for diversifying Ig genes, the immune system might have hijacked some mechanisms that ancestrally used the AID/APOBEC enzymes into germ cells as actors of the genome evolution.[9]

Biochemical action of AID in the immune system

Targeting of cytosine deamination by AID

In the immune system, both SHM and CSR in activated B cells require transcription of target V and S regions, respectively. Even LSR events joining Sμ with S-like regions located downstream of the IgH locus are concomitant with transcription of the IgH 3′ flanking enhancers, so that all AID on-target (Ig genes-targeted) effects correlate with active transcription.[3] The biochemical activity of AID has now been precisely characterized *in vivo* and *in vitro* [Figure 1], consisting of the deamination of cytosine into uracil in DNA at specific positions of transcribed genes where AID is recruited together with co-factors. AID does not mutate untranscribed Ig genes, and transcription likely allows exposition of short stretches of single-stranded (SS) DNA within transcription bubbles in the V regions, or in repetitive S regions carrying additional R-loops (RNA: DNA hybrids including the template strand). While it was initially shown that AID deaminates preferentially cytosines on the non-template strand of some transcribed DNA substrates *in vitro*, it, by contrast, mutates

both strands of negatively supercoiled DNA *in vitro* and, more importantly, it also mutates both strands of Ig genes *in vivo*.[10] It has now been clarified that AID cannot gain access to the template strand of transcribed Ig gene within a DNA/RNA heteroduplex without help and interaction with the RNA exosome complex, tethered to RNA polymerase II (RNAPII) by the trancription elongation factors Spt5/Spt6.[11]

Within the Ig genes, AID deamination preferentially targets WRC motifs (W = A/T, R = A/G).[12] AID also induces mutations and breaks into numerous non-Ig genes, many of which (but not all) are transcribed in B cells and many of which overlap with sites of translocations/recombinations found in B cell lymphomas or correspond to repetitive DNA sequences.[13,14]

Beside potential constraints about the accessibility of SS DNA, another major link between AID targeting and transcription is that AID is loaded onto Ig genes through its physical interaction with RNAPII under its stalled form. RNAPII stalling during transcription is known to occur immediately downstream of transcription start sites (TSS). It is accompanied with binding by Spt5 which then stimulates AID recruitment to the stalled RNAPII and its transcribed targets.[15] Accordingly, during SHM, the distribution of mutations spans approximately 1 kbp, starting about 120 bp downstream of the TSS and culminating around the V(D) J junction.[16] The RNAPII-associated polymerase-associated factors (PAF) complex also helps recruiting AID.[17]

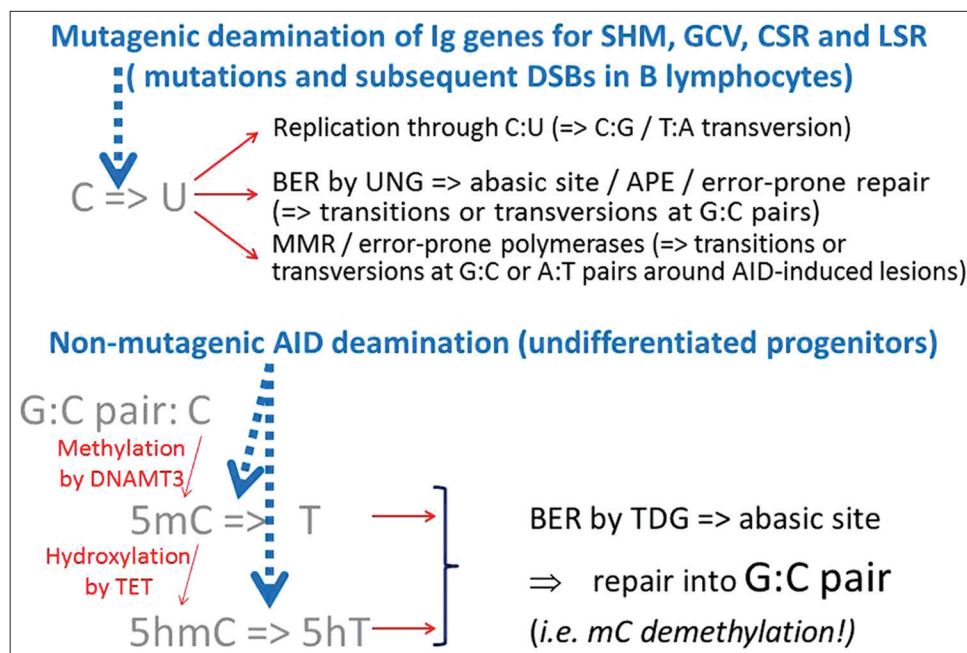RNAPII stalling is likely increased through S regions by the specific repetitive sequences and by secondary structures

Figure 1: AID effects in physiology. AID is predominantly seen as the central enzyme of antibody gene diversification in the immune system, where it can promote somatic hypermutation, gene conversion, class switch recombination, or locus suicide recombination. It is also an important contributor for DNA demethylation in non-lymphoid progenitor cells.

such as R-loops.[18] Indeed, the distribution of AID-induced mutations occurring in S regions parallel to CSR is extended and spans the whole target S region (i.e. up to 10 kbp), although it is again transcription dependent and starts about 150 bp downstream of the I promoter TSS.[19]

In such conditions, AID gains easy access to the non-template DNA strand of transcribed targets; facilitation of its access to the template strand by the RNA exosome might involve removal of the template RNA after binding its 3′ end (eventually released by a backtrack of the stalled RNAPII/Spt5 complex).[11] In fact, RNA exosome components improve AID cytosine deamination and accumulation of mutations on both the template and non-template strands of actively transcribed DNA.

By contrast to V and S regions, C_H regions are protected from AID attacks and this asymmetry is correlated with the presence of paused RNAPII in V and S but not C_H regions of Ig genes. Paused RNAPII is phosphorylated at Ser5 (instead of Ser2 during elongation) on the C-terminal domain (CTD), then modulating the recruitment of histone-modifying enzymes: Pausing often occurs at promoter proximal sites and within S regions, which are enriched for trimethyl histone H3 lysine 4 (H3K4me3) and hyperacetylated (Ac) H3K9, while the repressive marks H3K36me3, H4K20me1, and RNAPII phosphorylated on ser2 feature transcribed Ig C_H regions.[16,20] That H3K4me3 and H3K36me3, respectively, recruit histone acetyltransferases (HATs) and histone deacetylases (HDACs) might

explain why AID targeting focuses on hyperaccessible S regions and not on C_H regions.

By comparison to other genes where H3K4me3 is only found immediately downstream of the promoter, the unusually extended presence of this mark likely involves increased RNAPII stalling in S regions due to their unique repetitive sequences and the frequent occurrence of R-loops impeding RNAPII elongation (as for the extended occurrence of mutations). The role of H3K4me3 marks in target regions of SHM and CSR and the potential participation of topoisomerase I in AID-initiated DNA cleavage were also recently discussed by Honjo,[9] suggesting some parallels with the mechanisms involved in meiotic recombination.

Antisense transcription, which occurs in S but not in C_H regions, might also play a role in recruitment of AID to S regions, together with the presence of multiple alternative TSSs within the S regions.

Transcription and integration of the S region into an intron upstream of a given C_H gene are known prerequisites targeting CSR to that S region. That Spt5 links RNAPII to splicing factors might be related with this need of a splice donor site upstream of an S region for CSR.[16,21] Additional factors connected with splicing were shown to associate with AID and play a role in CSR, including CTNNBL1, a spliceosome-associated factor interacting with AID, and the splicing regulator polypyrimidine tract binding protein 2 (PTBP2) [Figure 2].[22,23] In addition, SRSF1-3, an inhibitory isoform of the SRSF1 splicing factor, was

Figure 2: Molecules interacting with AID. AID expression does not systematically correlate with AID activity on its DNA targets, but is regulated by post-translational changes, subnuclear localization, and complex interactions with multiple partners either preventing, restricting, or promoting the various aspects of AID activity.

recently shown to specifically increase the amount of un-sliced Ig transcripts in the DT40 B cell line and promote the occurrence of on-target SHM on V genes while restricting off-target AID attacks.[24]

Induction of SHM downstream of cytosine deamination by AID

Presence of uracils into DNA has to be repaired and can then eventually induce further DNA mutation and/or recombination. This mutagenic activity needs to be very stringently regulated both spatially and temporally. When affecting the genes encoding Ig V domains, the random untemplated mutagenic activity of AID can alter comple-mentarity determining regions and, thus, the antigen-binding affinity of Igs, in a process where SHM is coordinated with cell competition for optimal cellular interactions. This al-together results in affinity maturation of the B cell receptor and antibody molecules.[25] In some mammals, especially in cattle, AID-mediated SHM can also begin in fetal gut associated lymphoid tissues prior to any contact with ex-ogenous antigens.[26]

Biochemically, the G:U mismatches created through AID deamination can be processed in several ways; but since deamination of cytosines is a frequent event in the whole ge-nome independent of AID targeting and is usually correctly repaired, it is essential to understand how AID-mediated lesions into Ig gene sequences can more frequently lead to mutations rather than repair. In "phase 1" mutations, direct replication across a G U mismatch can generate transitions from G:C to A: T base pairs.

Base excision repair (BER) and removal of uracil by uracil N-glycosylase (UNG) instead generate abasic sites, subsequently undergoing DNA nicking by apurinic/apy-rimidinic endonuclease (APE-1) and finally repaired during replication by error-prone DNA polymerases (leading to both transitions and transversions affecting the attacked G:C pair).

G:U mismatches can also be processed by the mismatch repair (MMR) pathway (involving MSH2/MSH6), with as-sociated error-prone DNA polymerases, and then result in patches of "phase 2" mutations at both G:C and (preferen-tially) A: T base pairs around targeted cytosines.

Providing compelling evidence for this scenario, trans-version mutations at G: C bases are reduced in the absence of UNG, while phase 2 mutations strongly decrease in case of MMR defect. Since all these DNA lesions can contribute to SHM and promote double-strand breaks (DSBs) in S regions, most defects of either AID or UNG or MMR can, in fact, result in simultaneous alterations of both SHM and CSR.

There are several indications of how the processing of AID-mediated DNA lesions can be oriented. Notably, the MMR pathway can apparently choose between ending with either conventional error-free or error-prone repair by dedicated DNA polymerases. This choice of a given MMR signaling cascade downstream of AID occurs in a locus-dependent manner in B cells (and even for Ig loci where repair is preferentially mutagenic, different MMR pathways promote either SHM or CSR).

When an U: G mismatch is recognized by the MSH2/MSH6 complex, ATP-mediated conformational changes promote its recruitment of the proliferating cell nuclear antigen (PCNA) and of the 5′-3′ exonuclease EXO1, which excises the mismatch and leaves a 20-30-nucleotide-long SS patch. Error-free repair of this patch may then be ef-fected by DNA polymerase δ or ε at non-Ig genes, while error-prone (specially for A: T pairs) processing of Ig genes involve DNA polymerase eta (Polη).[27] Mono-ubiq-uitylation of PCNA at lysine 164 has been shown to be important for optimal recruitment of Polη at Ig V and S regions and, thus, for optimal SHM and CSR in activated B cells.[28,29] Rather than competing processes, it was also shown that MMR and BER can occur simultaneously downstream of cytosine deamination of Ig genes and both result in recruitment of Polη.

AID and class switching

The specific features of S region primary sequences have been mentioned above. S region transcription before recruitment of AID is under the control of cytokine-dependent germline promoters preceding the Ig C_H regions and of a se-ries of B cell activation-dependent transcriptional enhancers located in the 3′ regulatory region (3′RR) of the IgH locus.[30-33]

While AID-induced DNA deamination is processed to generate mutations in Ig V regions during SHM, it initiates DNA breaks (DSBs) intermediates in switch (S) regions, thereby promoting large deletions and CSR or LSR.[3,34] Some C-terminal AID mutations disrupting AID interactions with specific partners can compromise CSR without affecting CSR.[35] After their occurrence, DSBs activate the ubiquitous DNA damage response which allows their resolution through classical non-homologous end joining (C-NHEJ) or alterna-tive non-homologous end joining (A-NHEJ). The recruitment of 53BP1 to DNA broken ends (and subsequent formation of γ2AX foci) is required for protection of DNA ends from resection during CSR and during AID-initiated translocations associated with lymphomagenesis; 53BP1 then promotes DNA repair by C-NHEJ rather than A-NHEJ.[36,37]

14-3-3 adaptor proteins are nuclear factors able to bind AGCT repeats present in S regions and to help recruiting and nucleating a complex of AID and its co-factors bound to those regions [Figure 2].[38]

Although incompletely understood, some differences have been demonstrated in the processing of AID lesions occurring in Ig V and S regions. AID activity toward CSR (and off-target recombinations) can be controlled and

down-regulated through AID phosphorylation on serine 3, which is controlled by protein phosphatase 2A (PP2A).[39]

Contrary to V regions, MMR in S regions and in the rest of the genome involves the MutL (MLH2/PMS2) complex. It is also likely that the repetitive sequence of S regions by itself promotes DSBs, either due to pre-existing R-loops or due to repeated and overlapping AID hotspots resulting in a more frequent occurrence of close single-strand gaps on opposite strands ending with staggered DSBs.

It was also recently shown that optimal AID *in vivo* activity for CSR (but not for SHM) required its interaction with a complex including KRAB domain-associated protein 1 (KAP1) and heterochromatin protein 1 (HP1) tethered to H3K9me3 (trimethylated histone H3 at lysine 9) epigenetically marked target DNA, and that this complex bound the donor Sμ region.[40]

S–S regions' synapsis occurring after the occurrence of AID-initiated DSBs is likely favored by the three-dimensional structure and the DNA loops formed by the IgH locus, where physical interactions occur in activated B cells between Eμ and the 3′ RR enhancer elements (presence of the latter being mandatory for CSR).[31,33,32,41]

Regulation of AID activity

AID has been shown to be capable of mutating DNA in almost any living cell with almost no target sequence specificity. A major question is thus to understand how its activity can be regulated and focused on Ig genes in order to avoid dramatic side effects.

Control of AID expression in the B cell lineage at the level of transcription, mRNA stability, and translation

A primary regulation of AID activity in the immune system relies on its strictly controlled tissue-specific and stage-specific transcription in activated B cells due to the simultaneous control of the AID promoter by multiple ubiquitous or lymphoid-specific transcription factors including Pax-5, STAT6, SP1, or C/EBP and due to control of the AID mRNA stability through binding of its 3′ untranslated region by miRNAs, miR155 and miR181b. Altogether, these controls yield high-level AID expression only in the B cell lineage, in cells activated with appropriate cytokine signals as it can occur within GCs upon interaction with follicular dendritic cells and T follicular helper cells.

In addition to activated B cell-specific expression, it is noteworthy that AID is expressed at low levels in some bone marrow developing B cells, especially after stimulation of toll-like receptors.[42,43]

The ancestral hormone-dependent regulation of AID/APOBEC family proteins is also maintained in mammalian B cells and non B cells.[44,45] Beside Ig genes' diversification, another physiologic activity of AID is devoted to its participation in DNA demethylation in various progenitor cells, and will be developed subsequently. This activity is correlated with low-level AID expression into various tissues and notably into germ cells. Finally, we will also precisely explain below how ectopic AID expression can be induced in pathology and plays a role in non-lymphoid tumors.

Control of AID activity by its subcellular and subnuclear and localization

AID shuttles between the nucleus (where it can be actively imported thanks to a specific nuclear localization signal) and the cytoplasm where it is mostly retained at steady state through its C-terminal region (mediating both cytoplasmic retention and exportin-1–dependent nuclear export). The cytoplasm retention signal of AID allows its stoichiometric interaction with the elongation factor eEF1A, within a > 300 kDa complex; this retention prevents AID to reach its DNA substrate into the nucleus. How this complex is dissociated and regulated to allow AID nuclear import is still unclear, but might involve AID phosphorylation.[46] The cytoplasmic form of AID might also interact with LINE-1 retroelements mRNA and inhibit their integration into the genome, a known function of other APOBEC family members.[47]

After its nuclear import, AID is transiently found in the nucleolus where it interacts with CTNNBL1, nucleolin, nucleophosmin, and probably other factors with nucleolar localization such as the RNA exosome and where it could thus pre-assemble into various types of active complexes.[48] Another factor, germinal center-associated nuclear protein (GANP) associates with AID and increases its nuclear import together with its loading on V regions genes targeted by SHM.[49]

Control of AID stability and nuclear degradation

After translation, AID is stabilized by interactions with chaperones, with a special role of DnaJA1.[50] Availabililty of nuclear AID in its active form is also regulated by ubiquitylation and further proteosomal degradation.[51,52] Nuclear AID can also be targeted to degradation independent of ubiquitin through interaction with REG-γ.[53] Altogether, the control of degradation obviously impacts the level of AID nuclear availability.

Proteins interacting with AID

AID is a sticky protein with multiple interactions. It can even interact with agarose beads and yield false-positive results in immunoprecipitation experiments.[54] While multiple factors physically interacting or co-precipitating with AID have been described, functional roles were only demonstrated for some of these interactions.

These include the above-mentioned regulation of AID cytoplasmic retention by eEF1A, AID subnuclear distribu-

tion by nucleolin, nucleophosmin, and CTNNBL1, AID nuclear degradation after ubiquitylation by RNF126 or interaction with REG-γ, AID target DNA recruitment by the Spt5/RNAPII/PAF complex, 14-3-3, and GANP, AID increased activity by interaction with the splicing regulator PTBP2, the nascent RNA degrading exosome or with the heterochromatin factors Kap/HP1, AID attacks outcome by RPA bound to pSer38-AID [Figure 2], etc.

Control of AID activity by phosphorylation

AID can be regulated through phosphorylation at various sites, with either no (serine 41 or 43, tyrosine 124), inhibitory (serine 3) or stimulatory (serine 38, threonine 27 or 140) effect.[34] AID activity promoting CSR and recombinations is inhibited by AID phosphorylation on serine 3.[39] The role of serine 38 phosphorylation has been extensively studied. It is mediated *in vivo* by protein kinase A (PKA) and then allows interaction between pSer38-AID and RPA, likely stabilizing SS DNA within transcription bubbles and favoring AID action on V or S regions in the absence of R-loops. S38 phosphorylation and RPA interaction are necessary for both optimal SHM and CSR.[55] It was recently precised that AID deamination initially occurs independent of S38 phosphorylation/RPA binding in G1 phase where most of the subsequent DNA breaks are bound by 53BP1 and γH2AX and repaired in G1 by classical NHEJ (C-NHEJ); RPA associates at low levels to unrepaired ends of SS DNA during the G1 phase in a S38-pAID- and ATM-dependent manner (ATM standing as an essential mediator of the G1–S phase checkpoint), and it then promotes repair by micro-homology–dependent alternate NHEJ (A-NHEJ); finally, resected ends of DNA beaks persisting in the S/G2 phase recruit higher amounts of RPA together with Rad51 to be preferentially repaired by error-free homologous recombination.[56]

Regulation of target Ig gene accessibility by cis-regulatory elements

Several elements acting in *cis* and located within Ig loci have been shown to contribute to the recruitment of AID activity onto target V regions. This was the case for a matrix anchoring region associated with the Eκ intronic enhancer within the Igκ locus.[57] The group of U. Storb also showed that the E-motif CAGGTG increased mutability of an associated sequence without influencing its transcription, while AID SHM was shown in parallel to increase with G1 expression of the E12 and E47 E2A factors binding this motif.[58] In the mouse IgH locus, several *cis*-acting sequences were shown to play a role in targeting AID: Multiple alterations of the IgH 3′RR region compromised CSR in a way that notably relied onto germline transcription of the target S regions.[30-33] This effect was unlikely

to be entirely mediated by transcription since in several instances, partial transcription blockades resulting from 3′RR alterations showed association with complete CSR blockades, suggesting that the 3′RR not only promotes germline transcription of S regions but also promotes by other means AID activity, either by inducing epigenetic changes or by attracting and recruiting AID or by promoting AID activation and/or by recruiting transcription factors that are AID partners. While CSR alterations were associated with all major disruptions of the 3′RR, SHM was preserved by a deletion of the two more downstream 3′RR enhancers, hs3b and hs4.[59] By contrast, BAC transgenes lacking the 3′RR failed to undergo SHM.[60] We also recently observed in the endogenous IgH locus that the complete 3′RR deletion resulted in an almost complete loss of SHM while V region transcription was only minimally affected.[61]

In chicken B cells, it was also shown that GCV was modulated by the expression of the PU.1 and IRF4 factors, which acted both by binding sites close to the Ig V regions targeted by GCV and sites close to the AID promoter, thereby enhancing both AID recruitment and expression.[62] NF-κB factors might also play a role in the recruitment of AID at target regions for GCV in chicken B cells, while several *cis*-acting sequences enhancing SHM and/or GCV but not transcription (and binding still poorly defined factors) have recently been described in the chicken Ig λ locus; these sequences did not appear to promote epigenetic changes in the associated genes but might increase the presence of the serine-5 phosphorylated stalled RNAPII known to recruit AID activity.[63,64]

AID off-target effects in B cell malignancies

Studies of B cell malignancies since the early 1980s allowed to identify a number of translocations linking oncogenes such as c-MYC, c-MAF, CYCLIN D1, BCL2, etc., to breakpoints located within Ig loci. Some of these events were recurrently identified as the clonal driving forces of lymphomagenesis. This pointed to a major role of errors in the processes of antibody gene diversification in tumorigenesis. Recently, it was also shown that breakpoints located outside of Ig loci constitute frequent side targets of AID attacks in activated B cells [Figure 3]. It is thus now clear that in B cell malignancies as in normal B cells, AID-induced mutations accumulate on multiple loci including those of Ig (with the higher level of error-prone repair after AID deamination), bcl6, and a large list of oncogenes and tumor suppressors.[13] Precise characterization of translocation breakpoints and the development of whole genome sequencing showed that mutations and recombinations detected in activated B cells and those seen in tumors with mature B cell phenotypes all carry
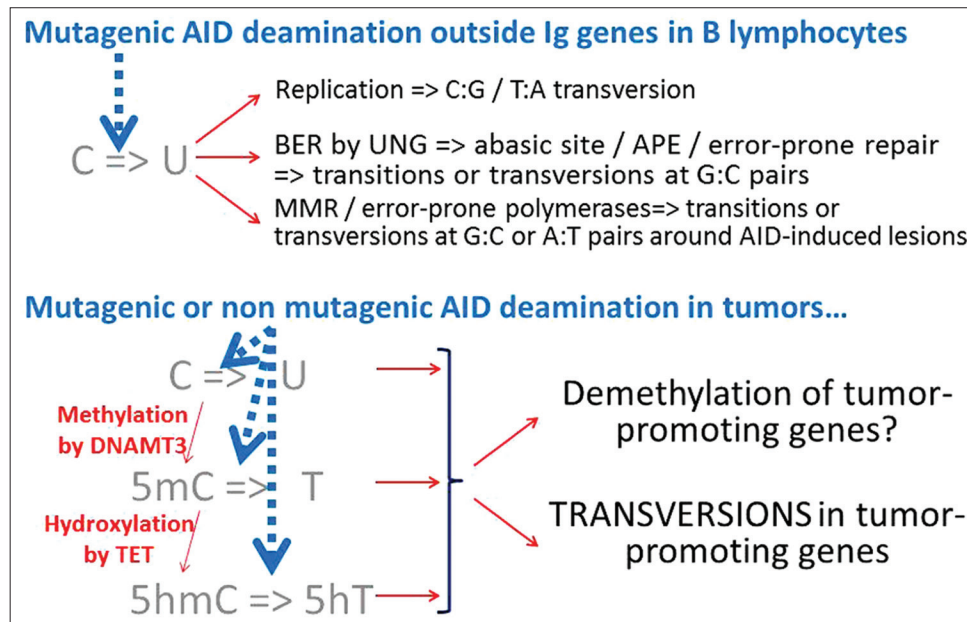
Figure 3: AID off-target effects in cancer. AID-dependent modifications of immunoglobulin genes often lead to aberrant repair and off-target lesions on non-Ig genes. These anomalies play a major role in the process of lymphomagenesis. In addition in non-lymphoid cells, ectopic expression of AID and/or other members of the AID/APOBEC family is increasingly considered as a driving force of tumorigenesis in various inflamation-initiated sarcomas.

hallmarks of WRC sequence-specific AID on-target and off-target attacks.[65,66] In addition, more than half of the translocations seen in immature B cell malignancies also show breakpoints at sites of AID deamination corresponding to mCG dinucleotides and might imply the combined action of AID, RAG, and ARTEMIS.[67,68] However, this issue can still be debated and it is noteworthy that lymphoma breakpoints targeting CG dinucleotides were also recently identified by the group of A. Nussenzweig as early replicating fragile sites susceptible to AID-independent genomic instability.[69]

AID effects in non-lymphoid tissues and non-lymphoid tumors

An ancestral upregulation of AID/APOBEC family proteins by estrogens (and down-regulation by progesterone) is maintained for AID in mammalian B cells and accounts as well for a low-level AID expression in some hormone-dependent tissues or tumors.[44,45]

In addition, AID has been recently recognized as a factor expressed at a low level and acting in mammalian progenitor cells in order to maintain pluripotency and restrict the inheritance of epimutations (with the ability to participate into epigenetic reprogramming) through its ability to deaminate 5-methylcytosine and hydroxymethyl cytosine into thymine and 5-hydroxythymine, respectively [Figure 3]. AID is thus widely involved in the demethylation of cytosines into progenitor cells and this function is ensured by a complex

between AID, thymidine glycosylase (TDG), and the DNA damage protein, GADD45a.[7,8] TDG is known as necessary for protection of CpG islands from hypermethylation and for demethylating and/or recruiting histone acetyltransferase to tissue-specific developmentally, hormonally, and retinoic acid-regulated promoters and enhancers.[8] Interestingly, this demethylation action of AID does not seem to require prior transcription of target sequences and/or interaction with RNAPII.

Non-lymphoid tumors also frequently show mutations of oncogenes or anti-oncogenes and a number of recent studies have analyzed mutation patterns looking for involvement of the AID/APOBEC family enzymes. Since AID mostly targets cytosines included into WRC motifs or mCG dinucleotides (while APOBEC1 prefers the final C of TC dinucleotides and APOBEC3 prefers the final C of CC dinucleotides), statistical analysis of the observed mutations can allow to deduce which member of the AID/APOBEC family has been involved in the mutagenesis of a given tumor type.

In pathology, ectopic expression of AID has been suggested in various instances. For example, the APC tumor suppressor mutated in the majority of colorectal cancers might be targeted by AID. In agreement with this hypothesis, colonic cancer development upon bowel inflammation is strongly reduced in AID-deficient mice.[70] In gastric epithelial cells infected with *Helicobacter pylori*, AID expression is also triggered and likely result in mutations accumulating in the *TP53* gene.[71]

Conclusion

AID appears as a major and ancestral enzyme involved in DNA demethylation in many tissues and notably in stem cells where it contributes to the maintenance of pluripotency and to the limitation of genetic imprinting. By its ability to alter cytosines and induce replication errors or to recruit DNA glycosylases or MMR, AID is also strongly mutagenic and has evolved as the master player of Ig gene secondary diversification. AID prevalent role in mammalian physiology, thus, relates to its specific high expression in activated B cells and its connection with downstream pathways inducing mutations, DSBs, or recombinations within Ig genes. This obviously needs a very stringent regulation, and AID activity appears to to be controlled at multiple levels (expression, nuclear importation, post-translational modifications, recruitment to targets, degradation, etc.). It is still currently a major issue to understand how the various protein factors interacting with AID, the various chromosomal marks, or the various *cis*-regulatory sequences controlling AID accessibility within Ig heavy or light chain loci can altogether mediate this regulation. Finally, frequent errors in these processes make AID an important player in B cell malignancies (and potentially also in sarcomas from tissues where inflammation promotes AID ectopic expression).

Acknowledgments

REFERENCES

1. Conticello SG, Thomas CJ, Petersen-Mahrt SK, Neuberger MS. Evolution of the AID/APOBEC family of polynucleotide (deoxy) cytidine deaminases. Mol Biol Evol 2005;22:367-77.

2. Boehm T, McCurley N, Sutoh Y, Schorpp M, Kasahara M, Cooper MD. VLR-based adaptive immunity. Annu Rev Immunol 2012;30:203-20.

3. Peron S, Laffleur B, Denis-Lagache N, Cook-Moreau J, Tinguely A, Delpy L, *et al*. AID-driven deletion causes immunoglobulin heavy chain locus suicide recombination in B cells. Science 2012;336:931-4.

4. Muramatsu M, Sankaranand VS, Anant S, Sugai M, Kinoshita K, Davidson NO, *et al*. Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. J Biol Chem 1999;274:18470-6.

5. Muramatsu M, Kinoshita K, Fagarasan S, Yamada S, Shinkai Y, Honjo T. Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. Cell 2000;102:553-63.

6. Revy P, Muto T, Levy Y, Geissmann F, Plebani A, Sanal O, *et al*. Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2). Cell 2000;102:565-75.

7. Popp C, Dean W, Feng S, Cokus SJ, Andrews S, Pellegrini M, *et al*. Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. Nature 2010;463:1101-5.

8. Cortellino S, Xu J, Sannai M, Moore R, Caretti E, Cigliano A, *et al*. Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. Cell 2011;146:67-79.

9. Kato L, Stanlie A, Begum NA, Kobayashi M, Aida M, Honjo T. An evolutionary view of the mechanism for immune and genome diversity. J Immunol 2012;188:3559-66.

10. Shen HM, Storb U. Activation-induced cytidine deaminase (AID) can target both DNA strands when the DNA is supercoiled. Proc Natl Acad Sci U S A 2004;101:12997-3002.

11. Basu U, Meng FL, Keim C, Grinstein V, Pefanis E, Eccleston J, *et al*. The RNA exosome targets the AID cytidine deaminase to both strands of transcribed duplex DNA substrates. Cell 2011;144:353-63.

12. Maul RW, Gearhart PJ. AID and somatic hypermutation. Adv Immunol 2010;105:159-91.

13. Liu M, Duke JL, Richter DJ, Vinuesa CG, Goodnow CC, Kleinstein SH, *et al*. Two levels of protection for the B cell genome during somatic hypermutation. Nature 2008;451:841-5.

14. Staszewski O, Baker RE, Ucher AJ, Martier R, Stavnezer J, Guikema JE. Activation-induced cytidine deaminase induces reproducible DNA breaks at many non-Ig Loci in activated B cells. Mol Cell 2011;41:232-42.

15. Pavri R, Gazumyan A, Jankovic M, Di Virgilio M, Klein I, Ansarah-Sobrinho C, *et al*. Activation-induced cytidine deaminase targets DNA at sites of RNA polymerase II stalling by interaction with Spt5. Cell 2010;143:122-33.

16. Kenter AL. AID targeting is dependent on RNA polymerase II pausing. Semin Immunol 2012;24:281-6.

17. Willmann KL, Milosevic S, Pauklin S, Schmitz KM, Rangam G, Simon MT, *et al*. A role for the RNA pol II-associated PAF complex in AID-induced immune diversification. J Exp Med 2012;209:2099-111.

18. Maul RW, Gearhart PJ. Controlling somatic hypermutation in immunoglobulin variable and switch regions. Immunol Res 2010;47:113-22.

19. Xue K, Rada C, Neuberger MS. The *in vivo* pattern of AID targeting to immunoglobulin switch regions deduced from mutation spectra in msh2-/- ung-/- mice. J Exp Med 2006;203:2085-94.

20. Wang L, Wuerffel R, Feldman S, Khamlichi AA, Kenter AL. S region sequence, RNA polymerase II, and histone modifications create chromatin accessibility during class switch recombination. J Exp Med 2009;206:1817-30.

21. Lorenz M, Jung S, Radbruch A. Switch transcripts in immunoglobulin class switching. Science 1995;267:1825-8.

22. Conticello SG, Ganesh K, Xue K, Lu M, Rada C, Neuberger MS. Interaction between antibody-diversification enzyme AID and spliceosome-associated factor CTNNBL1. Mol Cell 2008;31:474-84.

23. Nowak U, Matthews AJ, Zheng S, Chaudhuri J. The splicing regulator PTBP2 interacts with the cytidine deaminase AID and promotes binding of AID to switch-region DNA. Nat Immunol 2011;12:160-6.

24. Kanehiro Y, Todo K, Negishi M, Fukuoka J, Gan W, Hikasa T, *et al*.

Activation-induced cytidine deaminase (AID)-dependent somatic hypermutation requires a splice isoform of the serine/arginine-rich (SR) protein SRSF1. Proc Natl Acad Sci 2012;109:1216-21.

25. Pavri R, Nussenzweig MC. AID targeting in antibody diversity. Adv Immunol 2011;110:1-26.

26. Liljavirta J, Ekman A, Knight JS, Pernthaner A, Iivanainen A, Niku M. Activation-induced cytidine deaminase (AID) is strongly expressed in the fetal bovine ileal Peyer's patch and spleen and is associated with expansion of the primary antibody repertoire in the absence of exogenous antigens. Mucosal Immunol 2013;6:942-9.

27. Chahwan R, Edelmann W, Scharff MD, Roa S. AIDing antibody diversity by error-prone mismatch repair. Semin Immunol 2012;24:293-300.

28. Krijger PH, Langerak P, Van den Berk PC, Jacobs H. Dependence of nucleotide substitutions on Ung2, Msh2, and PCNA-Ub during somatic hypermutation. J Exp Med 2009;206:2603-11.

29. Roa S, Avdievich E, Peled JU, Maccarthy T, Werling U, Kuang FL, et al. Ubiquitylated PCNA plays a role in somatic hypermutation and class-switch recombination and is required for meiotic progression. Proc Natl Acad Sci U S A 2008;105:16248-53.

30. Cogné M, Lansford R, Bottaro A, Zhang J, Gorman J, Young F, et al. A class switch control region at the 3' end of the immunoglobulin heavy chain locus. Cell 1994;77:737-47.

31. Pinaud E, Khamlichi AA, Le Morvan C, Drouet M, Nalesso V, Le Bert M, et al. Localization of the 3' IgH locus elements that effect long-distance regulation of class switch recombination. Immunity 2001;15:187-99.

32. Pinaud E, Marquet M, Fiancette R, Péron S, Vincent-Fabert C, Denizot Y, et al. The IgH locus 3' regulatory region: Pulling the strings from behind. Adv Immunol 2011;110:27-70.

33. Vincent-Fabert C, Fiancette R, Pinaud E, Truffinet V, Cogné N, Cogné M, et al. Genomic deletion of the whole IgH 3' regulatory region (hs3a, hs1,2, hs3b, and hs4) dramatically affects class switch recombination and Ig secretion to all isotypes. Blood 2010;116:1895-8.

34. Vuong BQ, Chaudhuri J. Combinatorial mechanisms regulating AID-dependent DNA deamination: Interacting proteins and post-translational modifications. Semin Immunol 2012;24:264-72.

35. Durandy A, Peron S, Taubenheim N, Fischer A. Activation-induced cytidine deaminase: Structure-function relationship as based on the study of mutants. Hum Mutat 2006;27:1185-91.

36. Jankovic M, Feldhahn N, Oliveira TY, Silva IT, Kieffer-Kwon KR, Yamane A, et al. 53BP1 Alters the Landscape of DNA Rearrangements and Suppresses AID-Induced B Cell Lymphoma. Molecular Cell. Available from: http://www.sciencedirect.com/science/article/pii/S1097276512009884. [Last accessed on 2013 Jan 13].

37. Bothmer A, Robbiani DF, Di Virgilio M, Bunting SF, Klein IA, Feldhahn N, et al. Regulation of DNA end joining, resection, and immunoglobulin class switch recombination by 53BP1. Mol Cell 2011;42:319-29.

38. Xu Z, Fulop Z, Wu G, Pone EJ, Zhang J, Mai T, et al. 14-3-3 adaptor proteins recruit AID to 5′-AGCT-3′–rich switch regions for class switch recombination. Nat Struct Mol Biol 2010;17:1124-35.

39. Gazumyan A, Timachova K, Yuen G, Siden E, Di Virgilio M, Woo EM, et al. Amino-terminal phosphorylation of activation-induced cytidine deaminase suppresses c-myc/IgH translocation. Mol Cell Biol 2011;31:442-9.

40. Jeevan-Raj BP, Robert I, Heyer V, Page A, Wang JH, Cammas F, et al. Epigenetic tethering of AID to the donor switch region during immunoglobulin class switch recombination. J Exp Med 2011;208:1649-60.

41. Wuerffel R, Wang L, Grigera F, Manis J, Selsing E, Perlot T, et al. S-S synapsis during class switch recombination is promoted by distantly located transcriptional elements and activation-induced deaminase. Immunity 2007;27:711-22.

42. Han JH, Akira S, Calame K, Beutler B, Selsing E, Imanishi-Kari T. Class switch recombination and somatic hypermutation in early mouse B cells are mediated by B cell and Toll-like receptors. Immunity 2007;27:64-75.

43. Kuraoka M, Holl TM, Liao D, Womble M, Cain DW, Reynolds AE, et al. Activation-induced cytidine deaminase mediates central tolerance in B cells. Proc Natl Acad Sci 2011;108:11560-5.

44. Pauklin S, Sernández IV, Bachmann G, Ramiro AR, Petersen-Mahrt SK. Estrogen directly activates AID transcription and function. J Exp Med 2009;206:99-111.

45. Pauklin S, Petersen-Mahrt SK. Progesterone inhibits activation-induced deaminase by binding to the promoter. J Immunol 2009;183:1238-44.

46. Hasler J, Rada C, Neuberger MS. Cytoplasmic activation-induced cytidine deaminase (AID) exists in stoichiometric complex with translation elongation factor 1 (eEF1A). Proc Natl Acad Sci 2011;108:18366-71.

47. Metzner M, Jäck HM, Wabl M. LINE-1 retroelements complexed and inhibited by activation induced cytidine deaminase. PLoS ONE 2012;7:e49358.

48. Hu Y, Ericsson I, Torseth K, Methot SP, Sundheim O, Liabakk NB, et al. A combined nuclear and nucleolar localization motif in Activation-Induced Cytidine Deaminase (AID) controls immunoglobulin class switching. J Mol Biol 2013;425:424-43.

49. Maeda K, Singh SK, Eda K, Kitabatake M, Pham P, Goodman MF, et al. GANP-mediated recruitment of activation-induced cytidine deaminase to cell nuclei and to immunoglobulin variable region DNA. J Biol Chem 2010;285:23945-53.

50. Orthwein A, Zahn A, Methot SP, Godin D, Conticello SG, Terada K, et al. Optimal functional levels of activation-induced deaminase specifically require the Hsp40 DnaJa1. EMBO J 2012;31:679-91.

51. Aoufouchi S, Faili A, Zober C, D'Orlando O, Weller S, Weill JC, et al. Proteasomal degradation restricts the nuclear lifespan of AID. J Exp Med 2008;205:1357-68.

52. Delker RK, Zhou Y, Strikoudis A, Stebbins CE, Papavasiliou FN. Solubility-based genetic screen identifies RING finger protein 126 as an E3 ligase for activation-induced cytidine deaminase. PNAS. 2012. Available from: http://www.pnas.org/content/early/2012/12/27/1214538110. [Last accessed on 2013 Jan 13].

53. Uchimura Y, Barton LF, Rada C, Neuberger MS. REG-γ associates with and modulates the abundance of nuclear activation-induced deaminase. J Exp Med 2011;208:2385-91.

54. Metzner M, Schuh W, Roth E, Jäck HM, Wabl M. Two forms of activation-induced cytidine deaminase differing in their ability to bind agarose. PLoS ONE 2010;5:e8883.

55. Cheng HL, Vuong BQ, Basu U, Franklin A, Schwer B, Astarita J, et al. Integrity of the AID serine-38 phosphorylation site is critical for class switch recombination and somatic hypermutation in mice. PNAS 2009;106:2717-22.

56. Yamane A, Robbiani DF, Resch W, Bothmer A, Nakahashi H, Oliveira T, *et al*. RPA accumulation during class switch recombination represents 5'-3' DNA-end resection during the S-G2/M phase of the cell cycle. Cell Rep 2013;3:138-47.

57. Betz AG, Milstein C, González-Fernández A, Pannell R, Larson T, Neuberger MS. Elements regulating somatic hypermutation of an immunoglobulin κ gene: Critical role for the intron enhancer/matrix attachment region. Cell 1994;77:239-48.

58. Yabuki M, Ordinario EC, Cummings WJ, Fujii MM, Maizels N. E2A Acts in cis in G1 phase of cell cycle to promote Ig gene diversification. J Immunol 2009;182:408-15.

59. Morvan CL, Pinaud E, Decourt C, Cuvillier A, Cogné M. The immunoglobulin heavy-chain locus hs3b and hs4 3' enhancers are dispensable for VDJ assembly and somatic hypermutation. Blood 2003;102:1421-7.

60. Dunnick WA, Collins JT, Shi J, Westfield G, Fontaine C, Hakimpour P, *et al*. Switch recombination and somatic hypermutation are controlled by the heavy chain 3' enhancer region. J Exp Med 2009;206:2613-23.

61. Rouaud P, Vincent-Fabert C, Saintamand A, Fiancette R, Marquet M, Robert I, *et al*. The IgH 3' regulatory region controls somatic hypermutation in germinal centre B-cells. J Exp Med 2013;210:1501-7.

62. Luo H, Tian M. Transcription factors PU.1 and IRF4 regulate activation induced cytidine deaminase in chicken B cells. Mol Immunol 2010;47:1383-95.

63. Kim Y, Tian M. NF-kappaB family of transcription factor facilitates gene conversion in chicken B cells. Mol Immunol 2009;46:3283-91.

64. Kohler KM, McDonald JJ, Duke JL, Arakawa H, Tan S, Kleinstein SH, *et al*. Identification of core DNA elements that target somatic hypermutation. J Immunol 2012;189:5314-26.

65. Chiarle R, Zhang Y, Frock RL, Lewis SM, Molinie B, Ho YJ, *et al*. Genome-wide translocation sequencing reveals mechanisms of chromosome breaks and rearrangements in B cells. Cell 2011;147:107-19.

66. Gu X, Shivarov V, Strout MP. The role of activation-induced cytidine deaminase in lymphomagenesis. Curr Opin Hematol 2012;19:292-8.

67. Tsai AG, Lu H, Raghavan SC, Muschen M, Hsieh CL, Lieber MR. Human chromosomal translocations at CpG sites and a theoretical basis for their lineage and stage specificity. Cell 2008;135:1130-42.

68. Cui X, Lu Z, Kurosawa A, Klemm L, Bagshaw A, Tsai AG, *et al*. Both CpG methylation and AID are required for the fragility of the human Bcl-2 major breakpoint region: Implications for the timing of the breaks in the t (14;18). Mol Cell Biol 2013;33:947-57.

69. Barlow JH, Faryabi RB, Callén E, Wong N, Malhowski A, Chen HT, *et al*. Identification of early replicating fragile sites that contribute to genome instability. Cell 2013;152:620-32.

70. Takai A, Marusawa H, Minaki Y, Watanabe T, Nakase H, Kinoshita K, *et al*. Targeting activation-induced cytidine deaminase prevents colon cancer development despite persistent colonic inflammation. Oncogene 2012;31:1733-42.

71. Matsumoto Y, Marusawa H, Kinoshita K, Endo Y, Kou T, Morisawa T, *et al*. *Helicobacter pylori* infection triggers aberrant expression of activation-induced cytidine deaminase in gastric epithelium. Nat Med 2007;13:470-6.