Diversity in the Cow Ultralong CDR H3 Antibody Repertoire

Jeremy K. Haakenson†, Ruiqi Huang† and Vaughn V. Smider*

Department of Molecular Medicine, The Scripps Research Institute, La Jolla, CA, United States

Typical antibodies found in humans and mice usually have short CDR H3s and generally flat binding surfaces. However, cows possess a subset of antibodies with ultralong CDR H3s that can range up to 70 amino acids and form a unique "stalk and knob" structure, with the knob protruding far out of the antibody surface, where it has the potential to bind antigens with concave epitopes. Activation-induced cytidine deaminase (AID) has a proven role in diversifying antibody repertoires in humoral immunity, and it has been found to induce somatic hypermutation in bovine immunoglobulin genes both before and after contact with antigen. Due to limited use of variable and diversity genes in the V(D)J recombination events that produce ultralong CDR H3 antibodies in cows, the diversity in the bovine ultralong antibody repertoire has been proposed to rely on AID-induced mutations targeted to the IGHD8-2 gene that encodes the entire knob region. In this review, we discuss the genetics, structures, and diversity of bovine ultralong antibodies, as well as the role of AID in creating a diverse antibody repertoire.

Keywords: antibody, ultralong CDRH3, cow immunoglobulin, activation-induced cytidine deaminase, antibody repertoire

ANTIBODY OVERVIEW

Antibodies, also known as immunoglobulins (Ig), are an essential and defining characteristic of the vertebrate immune system. They act by binding and inhibiting foreign substances in the body, such as viruses and bacteria, and then destroying them through complement-dependent cytotoxicity, antibody-dependent cell-mediated cytotoxicity, or antibody-dependent cellular phagocytosis (1, 2). A typical antibody is composed of two heavy chains (HC) and two light chains (LC) bound together through both non-covalent interactions and disulfide bonds. There are five classes of HC, each with a distinct function: IgM, IgD, IgG, IgA, and IgE. In most vertebrates, these HCs can pair with two different classes of LCs: lambda (λ) and kappa (κ). Both HCs and LCs are composed of variable and constant regions. The variable region binds antigen and the constant region dictates which downstream effects will occur, such as activation of complement or the recruitment of macrophages, neutrophils, mast cells, basophils, cytotoxic T cells, or natural killer cells. HCs typically have one variable region and three or four constant regions, while LCs have one variable region and one constant region.

The segments of the variable region that bind antigen are termed complementarity determining regions (CDRs), and each HC and LC contains three CDRs. These CDRs usually form loops, and together, the six CDRs (three from the HC and three from the LC) form the antigen binding site of an antibody. The third CDR of the HC (CDR H3) often forms the most significant contact with antigen, is longer than the other CDRs, and usually plays a prominent role in antigen binding. In most species, the CDR H3 forms a simple loop structure, which is typically 8–16 amino acids long in humans. In a significant departure from antibodies of most vertebrates, about 10% of the antibodies found in cows have an ultralong CDR H3 ranging from 40 to 70 amino acids in length (3–9) (**Figure 1A**).
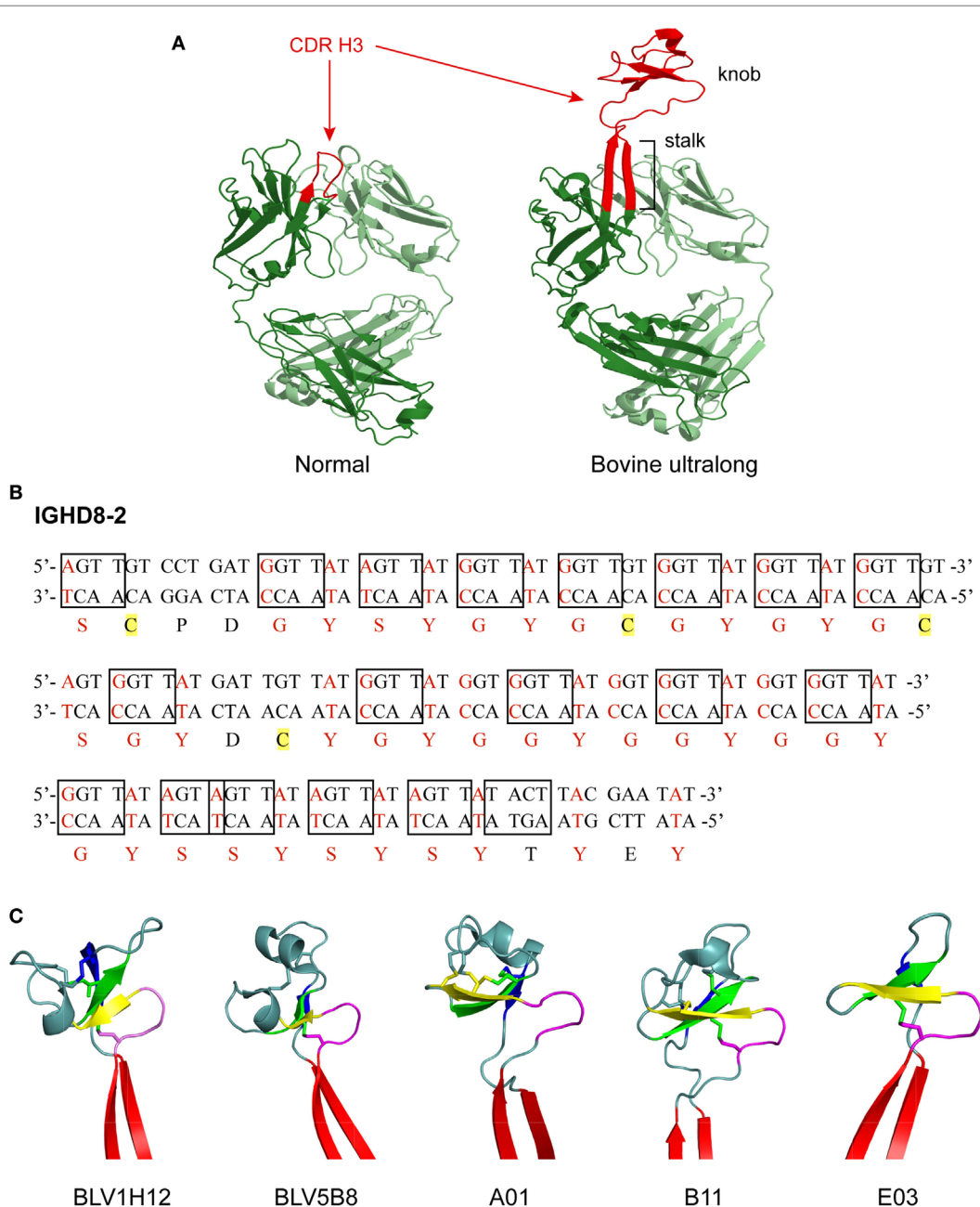
FIGURE 1 | Structure and genetics of ultralong complementarity determining region (CDR) H3 cow antibodies. **(A)** Comparison of normal and ultralong CDR H3 antibody fab fragments. Crystal structures of Yvo, a typical human antibody (left), and B11, an ultralong bovine antibody (right). The CDR H3 of each antibody is highlighted in red. Heavy chains are colored dark green and light chains are colored pale green. Note the two long β strands that make up the stalk of the bovine ultralong CDR H3 and the disulfide-bonded knob domain at the top. PDB IDs of these two fabs are Yvo: 2AGJ; B11: 5IHU. **(B)** Sequence and potential diversity of the ultralong D_H region. Potential AID-induced somatic hypermutation hotspot motifs ("RGYW"/"WRCY") in bovine germline IGHD8-2 are boxed on both strands of the DNA sequence. Nucleotides within a codon that can be mutated to a cysteine-encoding codon with just a single nucleotide change are colored red, and corresponding amino acids encoded by these codons are also colored red. Four cysteine residues encoded by the germline IGHD8-2 are highlighted yellow. **(C)** "Stalk and knob" structures of the ultralong CDR H3s. The β-ribbon stalk is colored red, the type I β-turn is colored magenta, and the three antiparallel β strands are colored yellow, green and blue, respectively. Disulfide bonds formed within the knob region are represented as sticks. PDB IDs of these five fabs are BLV1H12: 4K3D; BLV5B8: 4K3E; A01: 5ILT; B11: 5IHU; and 5IJV.

For these unusual ultralong CDR H3 bovine antibodies, it is likely that only CDR H3 binds antigen, while the other CDRs play a merely structural role (10, 11).

Some vertebrate species other than cows also contain unique antibody structures. For example, camels and llamas contain antibodies composed of HC homodimers that do not have an LC (12, 13). Interestingly, the CDR H3s of these antibodies often contain a free cysteine which can form a disulfide bond with a cysteine in CDR H1 (12). Sharks also generate antibodies composed of HC homodimers with no LCs called immunoglobulin

new antigen receptors, which consist of one variable domain and five constant domains (14–16).

The variable region of an antibody is encoded by variable (V), diversity (D), and junction (J) gene segments (V_H, D_H, and J_H for the HC genes, V_L or J_L for the light chain genes which can be specifically denoted V_λ, V_κ or J_λ, and J_κ respectively). In humans and mice, antibody repertoire diversity is achieved through V(D)J recombination, which occurs prior to antigen exposure, and somatic hypermutation (SHM), which occurs after antigen exposure. In V(D)J recombination, typically any V_H to D_H, or D_H to J_H recombination event can occur to form the full length gene encoding a HC. Since LCs do not contain D genes, any V_L gene can combine with any J_L gene to form the LC variable region gene. D–J recombination occurs first at the HC locus, followed by V–DJ joining. Further diversity is created during V(D)J recombination through the addition and/or deletion of nucleotides at the junctions between V and D, D and J, or V and J gene segments. Two different types of nucleotide additions can contribute to this junctional diversity: palindromic P nucleotides can be added first, followed by random N nucleotide insertions by terminal deoxynucleotidyl transferase (TdT) (17). Once an organism has been exposed to antigen, SHM further contributes to antibody diversity during the affinity maturation process. This will be described in depth in later sections of this review as it appears critically important to diversity generation in the cow. The sequential nature of V(D)J recombination to form the primary repertoire followed by SHM after exposure to antigen is treated as dogma in molecular immunology, likely due to the murine and human model systems that predominate in immunology research. However, it is clear that some vertebrates, including cows, utilize SHM to diversify the primary repertoire (18–20).

Antibody gene transformations occur in parallel with development of the B cells that produce them. In the bone marrow, hematopoietic stem cells differentiate into pro-B cells, in which V(D)J recombination occurs. These pro-B cells further differentiate into pre-B cells, which make rearranged HCs paired with a LC proxy termed the surrogate light chain. Pre-B cells further differentiate into immature naïve B cells, which produce cell surface IgM paired with a conventional LC. These immature naïve B cells then differentiate into mature naïve B cells that co-express membrane IgM and IgD and can be found throughout the body, especially in the lymphatic system and spleen. Finally, after exposure to antigen, often with the help of T cells, mature naïve B cells proliferate and differentiate into antibody-secreting effector plasma cells or into memory B cells, which produce affinity-matured antibodies of all of the Ig isotypes. It is unclear whether the unusual cow antibodies described below have equivalent distributions in all isotypes and tissue compartments. This is an area of active research.

THE BOVINE ANTIBODY REPERTOIRE

The germline genetic components of the bovine antibody repertoire are limited compared to humans, mice, and other vertebrates. At the HC locus, cows have 12 functional V_H, 23 D_H, and 4 J_H gene segments, while humans have 36–49 V_H, 23 D_H, and 6 J_H gene segments (10). This means that, in theory, there are up to 6,762

possible VDJ combinations in humans compared to only 920 in cows, allowing for much less combinatorial diversity in the cow.

At the LC lambda locus, cows have 25 V_λ and 3 J_λ gene segments compared to 73–74 V_λ and 7–11 J_λ gene segments in humans. The bovine κ locus contains 8 V_κ and 3 J_κ gene segments compared to 31–36 V_κ and 5 J_κ gene segments in humans (21). It has been observed that cows preferentially use λ LCs over κ (22), perhaps due to the increased potential for diversity (21).

Bovine antibodies with ultralong CDR H3 occupy at least 10% of the entire bovine antibody repertoire (5, 9), and they are found in all bovine Ig isotypes (23). However, the length distribution of ultralong CDR H3 in bovine Ig isotypes is different (9). It has been established that bovine ultralong antibodies almost always use the same HC V gene segment (IGHV1-7) and the same HC D gene segment (IGHD8-2), which is longer than any other D gene segment known to exist in nature (6, 8–10, 24). In addition, ultralong HCs appear to preferentially pair with λ LCs containing the V30 gene segment which contains residues that bind and may stabilize the ultralong CDR H3 of the HC (9, 25). This restricted use of V and D gene segments suggests that the diversity seen in bovine ultralong antibodies may not primarily result from V(D)J recombination.

GENETICS OF BOVINE ULTRALONG CDR H3 ANTIBODIES

Distinct features of germline IGHV1-7 and IGHD8-2 genes facilitate the formation of ultralong CDR H3 in cow antibodies. Nucleotide alignment of IGHV1-7 with other functional IGHV genes reveals a eight-nucleotide duplication, "TACTACTG" at its 3′ end (10). This duplication changes the reading frame at its 3′ end, and results in a longer V-region (encoding "TTVHQ" instead of "AR/K"), which contributes to the initiation of the ascending β-strand of the CDR H3 "stalk" (see the next section). Moreover, compared to other IGHV genes used in cow, IGHV1-7 features low amino acid variability in mature CDR H1 and H2 sequences, which are usually involved in antigen binding in antibodies with short CDR H3 in other species (10). Therefore, the diversity of the cow ultralong antibody repertoire is heavily concentrated in the CDR H3 which is mainly encoded by the IGHD8-2 gene.

IGHD8-2 encodes the majority of the ultralong CDR H3 after the "TTVHQ" motif. It encodes 48 amino acid residues and features 19 activation-induced cytidine deaminase (AID) induced SHM hotspots ("RGYW or WRCY" discussed later) (9, 10) (**Figure 1B**). In addition, IGHD8-2 contains 4 cysteine codons and another 38 codons that can readily be mutated to cysteine with just one nucleotide change (9) (**Figure 1B**). These existing and potential cysteine codons would enable disulfide bond formation within the CDR H3, and thus diversify the bovine ultralong antibodies at the structural level (9).

STRUCTURAL FEATURES OF BOVINE ULTRALONG CDR H3 ANTIBODIES

Crystal structures of five bovine antibody Fab fragments with ultralong CDR H3 regions have been published (9, 25). Despite

significant amino acid differences, as well as different numbers of amino acids in the CDR H3 region of these five antibodies, all of them adopt the same general CDR H3 structure, where a β-ribbon stalk supports a knob region that contains one type I β-turn, one conserved disulfide bond, and three antiparallel β strands (**Figure 1C**). The ascending β-strand in the stalk is initiated by the extended C-terminus encoded by IGHV1-7, and the descending β-strand is contributed by the C-terminus encoded by IGHD8-2, while the knob region is encoded solely by IGHD8-2. The supported knob region extends far out from the traditional antigen binding surface generated by the other five CDRs and has the potential to bind antigens with concave epitopes (9, 11, 25).

Although these ultralong CDR H3 antibodies share the general "stalk & knob" scaffold, each antibody also possesses distinct structural variations in the CDR H3 (**Figure 1C**). These structural variations are reflected in differences in the length of the stalk and the orientation of the knob relative to the rest of the antibody structure (25). For example, the stalk length of BLV1H12 is the longest among these five antibodies, while the stalk length of A1 is the shortest (25). When the five Fab structures are superimposed by the shared type I β-turn and the three antiparallel β strands in the knob region, obvious differences in stalk positions are observed, reflecting different knob orientations that are supported by the stalks (25). Furthermore, as the number and positions of cysteine residues in the knob region differ significantly among these cow ultralong antibodies, each of them possesses a knob region with unique disulfide bond patterns (25). Therefore, different stalk lengths, knob orientations, and disulfide bond patterns, in addition to diverse amino acid content within the knob regions of these antibodies provide remarkable structural diversity generated in the bovine ultralong CDR H3 antibody repertoire.

DIVERSITY AND DEVELOPMENT OF THE BOVINE ULTRALONG CDR H3 ANTIBODY REPERTOIRE

Although combinatorial antibody diversity generated through V(D)J recombination in the cow is limited, some diversity of these ultralong CDR H3 antibodies can arise from junctional diversity; P-nucleotide insertions, and TdT-catalyzed N additions and deletions can potentially occur at the V–D and D–J junctions (8, 26). It has been reported that some ultralong CDR H3 antibody genes can gain additional diversity through conserved short nucleotide sequence insertions (up to 6 codons) at the V–D junction (8). Importantly, the diversity of the ultralong CDR H3 antibody repertoire may be primarily generated through extensive SHM mediated by AID, which can occur either before or after antigen contact (**Figure 2**). For example, AID is expressed in naïve B cells in fetal bovine ileal Peyer's patches (lymphoid follicles in the small intestine) and spleen, where it induces SHM by mutating the CDR regions more often than the framework (FR) regions, a pattern similar to AID-induced SHM upon antigen contact in adult B cells (20). It has also been proposed that AID may mediate a novel diversification mechanism in cow either by deleting interior codons in the IGHD8-2 region without disrupting the knob and stalk structure, or by varying the number
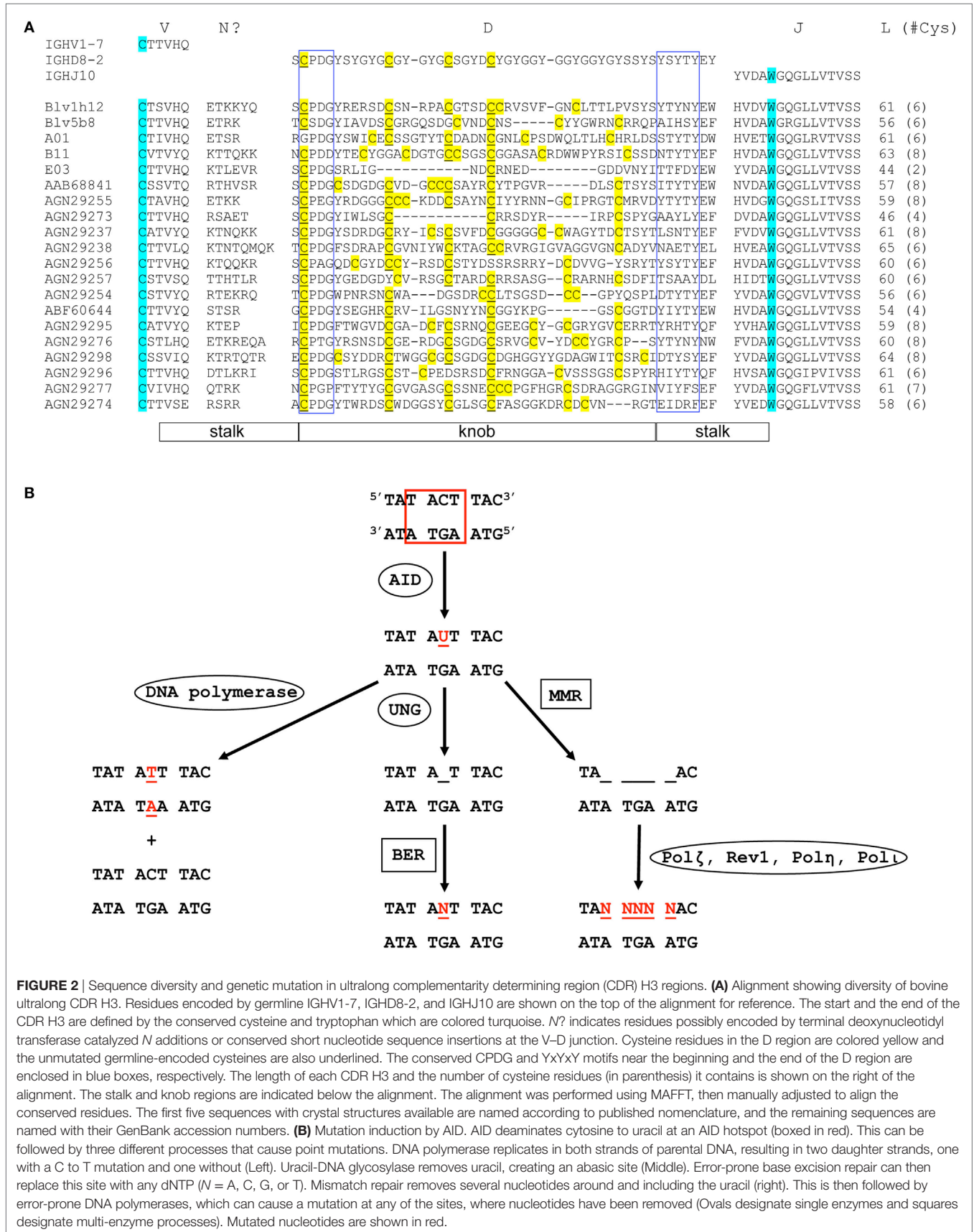
of cysteine residues in the knob, thus creating different disulfide bond patterns (**Figure 2A**) (9, 10, 25). Given that relatively limited diversity is generated by recombinational mechanisms in the bovine ultralong CDR H3 antibody repertoire, AID may be the primary driver for diversity development in the bovine ultralong CDR H3 antibody system.

EVOLUTIONARY ORIGIN OF AID

AID, which drives antibody class switch recombination (CSR) (27), SHM (27), and gene conversion (GC) (28), plays an important role in vertebrate immune responses and antibody diversification, and may play a particularly important role in diversifying cattle antibody genes. Given the potential unique importance of AID in the cow antibody repertoire, here we will review AID, whose activity and characteristics have been mostly studied in mouse and human systems, but whose properties are likely also applicable to bovines. AID belongs to the AID/APOBEC family which includes AID and APOBEC1 to APOBEC4. Although *in vivo* functions of APOBEC2 and APOBEC4 are still unclear, the other three members are zinc-dependent cytidine deaminases that catalyze the deamination of cytidine (C) to uracil (U) in either DNA (AID and APOBEC3) and/or RNA substrates (APOBEC1) (29). The first member characterized in this family was APOBEC1, which is an RNA editing enzyme that deaminates C6666 to U in the apolipoprotein B pre-mRNA in the small intestine, resulting in an expressed truncated apolipoprotein B, in contrast to the longer form which is synthesized in the liver (30, 31). AID, as a close relative of APOBEC1, is a 24 kDa enzyme that shares a similar structural and catalytic scaffold to APOBEC1, but only catalyzes deamination of deoxycytidine (dC) to deoxyuracil (dU) in either single-stranded DNA *in vitro* (32) or double-stranded DNA during transcription *in vivo* (33). APOBEC3, another DNA deaminase, fights intruding retroviruses by deaminating dC to dU on the reverse transcribed minus strand of viral DNA (34–37).

AID is encoded by the *AICDA* gene in humans, mice, and cows (38). The existence of AID is likely ancient because *AICDA* homologs have been identified in jawless vertebrates, in which immunoglobulin-type antigen receptors have not evolved (39–42). Indeed, lampreys use variable lymphocyte receptors, which contain leucine-rich repeat (VLR) modules that are structurally distinct from immunoglobulins, as the antigen receptors to carry out adaptive immune responses (39, 40). The diversity of the VLR repertoire comes from somatic insertions and rearrangements of different LRR modules from a large LRR module pool into the germline VLR gene, and AID was proposed to contribute to this GC-like process (39, 40, 42–44). The ability of AID to induce SHM evolved early in lower vertebrates (e.g., cartilaginous and bony fish) (45). Furthermore, although AID-induced CSR evolved later in tetrapods, fish AID was pre-adapted to drive CSR as ectopic expression of zebrafish AID in AID knockout mice is able to mediate equivalent CSR and SHM in murine immunoglobulin genes compared to murine AID (46, 47). In contrast, APOBEC1 and 3 evolved recently only in mammals (29, 48). APOBEC2 and 4 are likely ancient members in the family as well because they also exist in jawless vertebrates (but not in lamprey), similar to AID (48). In addition, phylogenetic relationships, gene structures, and

A

```
              V         N?                              D                              J           L (#Cys)
IGHV1-7    CTTVHQ
IGHD8-2               SCPDGYSYGYGCGY-GYGCSGYDCYGYGGY-GGYGGYGYSSYSYSYTYEY
IGHJ10                                                                    YVDAWGQGLLVTVSS

Blv1h12    CTSVHQ    ETKKYQ    SCPDGYRERSDCSN-RPACGTSDCCRVSVF-GNCLTTLPVSYSYTYNYEW    HVDVWGQGLLVTVSS    61 (6)
Blv5b8     CTTVHQ    ETRK      TCSDGYIAVDSCGRGQSDGCVNDCNS-----CYYGWRNCRRQPAIHSYEF    HVDAWGRGLLVTVSS    56 (6)
A01        CTIVHQ    ETSR      RGPDGYSWICECSSGTYTCDADNCGNLCPSDWQLTLHCHRLDSSTYTYDW    HVETWGQGLRVTVSS    61 (6)
B11        CVTVYQ    KTTQKK    NCPDDYTECYGGACDGTGCCSGSCGGASACRDWWPYRSICSSDNTYTYEF    HVDAWGQGLLVTVSS    63 (8)
E03        CTTVHQ    KTLEVR    SCPDGSRLIG----------NDCRNED-------GDDVNYITTFDYEW       YVDAWGQGLLVTVSS    44 (2)
AAB68841   CSSVTQ    RTHVSR    SCPDGCSDGDGCVD-GCCCSAYRCYTPGVR-----DLSCTSYSITYTYEW    NVDAWGQGLLVTVSS    57 (8)
AGN29255   CTAVHQ    ETKK      SCPEGYRDGGGCCC-KDDCSAYNCIYYRNN-GCIPRGTCMRVDYTYTYEW    HVDGWGQGSLITVSS    59 (8)
AGN29273   CTTVHQ    RSAET     SCPDGYIWLSGC----------CRRSDYR-----IRPCSPYGAAYLYEF     DVDAWGQGLLVTVSS    46 (4)
AGN29237   CATVYQ    KTNQKK    SCPDGYSDRDGCRY-ICSCSVFDCGGGGGC-CWAGYTDCTSYTLSNTYEW    FVDVWGQGLLVTVSS    61 (8)
AGN29238   CTTVLQ    KTNTQMQK  TCPDGFSDRAPCGVNIYWCKTAGCCRVRGIGVAGGVGNCADYVNAETYEL   HVEAWGQGLLVTVSS    65 (6)
AGN29256   CTTVHQ    KTQQKR    SCPAGQDCGYDCCY-RSDCSTYDSSRRY-DCDVVG-YSRYTYSYTYEF      HVDAWGQGLLVTVSS    60 (6)
AGN29257   CSTVYQ    TTHTLR    SCPDGYGEDGDYCV-RSGCTARDCRRSASG--CRARNHCSDFITSAAYDL    HIDTWGQGVLVTVSS    60 (6)
AGN29254   CSTVYQ    RTEKRQ    TCPDGWPNRSNCWA---DGSDRCCLTSGSD--CC--GPYQSPLDTYTYEF    YVDAWGQGVLVTVSS    56 (6)
ABF60644   CTTVYQ    STSR      GCPDGYSEGHRCRV-ILGSNYYNCGGYKPG------GSCGGTDYIYTYEW    HVDAWGQGLLVTVSS    54 (4)
AGN29295   CATVYQ    KTEP      ICPDGFTWGVDCGA-DCFCSRNQCGEEGCY-GCGRYGVCERRTYRHTYQF    YVHAWGQGLLVTVSS    59 (8)
AGN29276   CSTLHQ    ETKREQA   RCPTGYRSNSDCGE-RDGCSGDGCV-YDCCYGRCP--SYTYNYNW        FVDAWGQGLLVTVSS    60 (8)
AGN29298   CSSVIQ    KTRTQTR   ECPDGCSYDDRCTWGGCGCSGDGCDGHGGYYGDAGWITCSRCIDTYSYEF   YVDAWGQGLLVTVSS    64 (8)
AGN29296   CTTVHQ    DTLKRI    SCPDGSTLRGSCST-CPEDSRSDCFRNGGA-CVSSSGSCSPYRHIYTYQF   HVSAWGQGIPVIVSS    61 (6)
AGN29277   CVIVHQ    QTRK      NCPGPFTYTYGCGVGASGCSSNECCCPGFHGRCSDRAGGRGINVIYFSEF   YVDAWGQGFLVTVSS    61 (7)
AGN29274   CTTVSE    RSRR      ACPDGYTWRDSCWDGGSYCGLSGCFASGGKDRCDCVN---RGTEIDRFEF   YVEDWGQGLLVTVSS    58 (6)
```

```
             stalk                          knob                        stalk
```

B

FIGURE 2 | Sequence diversity and genetic mutation in ultralong complementarity determining region (CDR) H3 regions. **(A)** Alignment showing diversity of bovine ultralong CDR H3. Residues encoded by germline IGHV1-7, IGHD8-2, and IGHJ10 are shown on the top of the alignment for reference. The start and the end of the CDR H3 are defined by the conserved cysteine and tryptophan which are colored turquoise. N? indicates residues possibly encoded by terminal deoxynucleotidyl transferase catalyzed N additions or conserved short nucleotide sequence insertions at the V–D junction. Cysteine residues in the D region are colored yellow and the unmutated germline-encoded cysteines are also underlined. The conserved CPDG and YxYxY motifs near the beginning and the end of the D region are enclosed in blue boxes, respectively. The length of each CDR H3 and the number of cysteine residues (in parenthesis) it contains is shown on the right of the alignment. The stalk and knob regions are indicated below the alignment. The alignment was performed using MAFFT, then manually adjusted to align the conserved residues. The first five sequences with crystal structures available are named according to published nomenclature, and the remaining sequences are named with their GenBank accession numbers. **(B)** Mutation induction by AID. AID deaminates cytosine to uracil at an AID hotspot (boxed in red). This can be followed by three different processes that cause point mutations. DNA polymerase replicates in both strands of parental DNA, resulting in two daughter strands, one with a C to T mutation and one without (Left). Uracil-DNA glycosylase removes uracil, creating an abasic site (Middle). Error-prone base excision repair can then replace this site with any dNTP (N = A, C, G, or T). Mismatch repair removes several nucleotides around and including the uracil (right). This is then followed by error-prone DNA polymerases, which can cause a mutation at any of the sites, where nucleotides have been removed (Ovals designate single enzymes and squares designate multi-enzyme processes). Mutated nucleotides are shown in red.

chromosome loci of AID/APOBEC family members suggest that APOBEC1 and 3 likely arose from gene duplications of AID in mammalian species, and RNA editing deaminase activity likely evolved from ancestral DNA deaminase activity (48). Therefore, the activities of AID on immune system genes may have been the most ancient activity of this family of enzymes, and is important in apparently being the most ancient diversification enzyme for immune receptors.

TISSUE AND CELLULAR EXPRESSION OF AID

AID was first discovered in CH12F3-2 murine B lymphoma cells (49). Much early work on AID also used the DT40 chicken bursa lymphoma cell line (50). Some of the first experiments on AID found that it is expressed in Peyer's patches and lymph nodes, and to a lesser extent in the spleen and bone marrow (49). Microscopically, AID is expressed in germinal centers, which are regions where B cells proliferate and differentiate in response to antigen (49). This expression pattern is consistent with AID's role in affinity maturation, GC, and CSR, processes that take place in B cells in the lymphatic system and spleen.

FUNCTION AND ACTIVITY OF AID

AID is a remarkable enzyme because of the many activities that it is responsible for, which include SHM, CSR, and GC. In SHM, AID introduces point mutations into the variable region of an antibody gene by deaminating cytosine to uracil. Several processes can take place following this C to U mutation (**Figure 2B**). In one case, DNA replication can replace the uracil with a thymidine, resulting in two different daughter strands: one with a T-A base pair, and one with the original G-C base pair. In another case, uracil-DNA glycosylase (UNG) can remove the U, creating an abasic site, which can then be replaced by any nucleotide (A, G, C, or T) through the process of short patch base excision repair (BER). These two processes combined account for about 40% of AID-mediated mutations (51). A third possibility is that mismatch repair (MMR) can excise several nucleotides (nt) around and including the U. When this is followed by an error-prone polymerase, such as Polζ, Rev1, Polη, or Polι (52, 53), mutations can occur at any of the excised positions on the DNA (54). This third process accounts for about 60% of AID-mediated mutations (51). Thus, AID catalyzes a C to U mutation that can ultimately lead to additional mutations near the original deaminated cytosine. These point mutations can in turn alter the amino acid sequence of the antibody, changing its affinity for its antigen. The CDR H3 region of mature cow ultralong CDR H3 antibodies are enormously mutated, thus any or all of these activities of AID may play a role in shaping the repertoire.

The mutation rate of AID is in the range of 0.78–10.4 mutations per 10^4 base pairs (55, 56). Importantly, SHM is up to six times as likely to occur in SHM hotspot regions compared to other regions of DNA (57). Canonical SHM hotspots have the sequence motif RGYW (R = purine, Y = pyrimidine, W = A or T) or its reverse complement WRCY (56), although DGYW (D = A, G, or T)/WRCH (H = A, C, or T) may be a better predictor of SHM

(58). The significant concentration of SHM hotspots in the bovine IGHD8-2 gene segment suggest that AID may preferentially act on this gene segment, thereby creating diversity in the bovine ultralong antibody repertoire. Indeed, IGHD8-2 contains 19 AID hotspots, including 38 out of 48 (79%) amino acid codons that can mutate to cysteine with a single nucleotide mutation. Thirty of those codons are at least partially within AID hotspots (9) (**Figure 1B**), suggesting that AID plays an important role in creating cysteine diversity in the ultralong D region.

In addition to the introduction of point mutations, insertions and deletions also occur during SHM. These insertions and deletions typically occur in CDRs and are twice as likely to occur at SHM hotspots compared to other regions of DNA (59), implicating AID in this process. Insertions typically duplicate the preceding codons (e.g., AGC to AGC AGC) (60) and usually range from 2 to 23 nt in length (60, 61), while deletions are usually 1 to 51 nt long (61). Frameshift insertions and deletions (i.e., those not divisible by 3) can result in premature stop codons, so that often only in-frame insertions and deletions survive clonal selection. Ultralong CDR H3 bovine antibodies often contain deletions ranging from 3 to 54 nt in the D region (10). Deletions in bovine ultralong antibodies are easier to detect than in short antibodies because in short antibodies, it is difficult to ascertain whether an insertion/deletion has occurred or whether a different V, D, or J gene segment has been used, particularly when the CDR H3 has been highly mutated. Bovine ultralong antibodies, on the other hand, appear to always use the IGHD8-2 germline gene segment, which are 48 amino acids long. By comparison, the next longest bovine DH gene segment (IGHD7-4) is only 23 amino acids long. The rate of deletions in bovine ultralong antibodies has been reported to be 48% (10), compared to 2.6% in human IgG antibodies (62). However, this latter number might be an underestimate due to difficulty inherent in detecting insertions and deletions in short CDR H3 antibodies. Clearly, the ability of AID-mediated insertions and deletions to increase antibody variability further highlights the importance of AID and SHM in creating diversity in bovine ultralong antibodies.

In addition to causing SHM, AID is also responsible for CSR, in which the constant region of an antibody switches from one class to another (e.g., IgM to IgG) while the variable region remains the same. Indeed, mice and humans deficient in AID exhibit hyper-IgM syndrome, in which the immune system is unable to make IgG, IgA, and IgE antibodies (27, 63). CSR takes place during transcription of antibody genes. While antibodies are being transcribed, R-loops form, in which one strand of DNA is hybridized to the resulting mRNA and the other strand of DNA is in a single-stranded state (ssDNA). During transcription, AID targets R-loops and deaminates C to U on both strands of DNA. UNG then removes the U, creating an abasic site that is converted into a single-strand nick by apurinic/apyrimidinic endonuclease. In effect, nicks on both strands of DNA create a double-strand break (64). When double-strand breaks occur in the switch regions that flank each constant region, the existing constant region is removed and replaced by a new constant region through a DNA synaptic process involving non-homologous end joining (NHEJ) (64–66). Of note, hotspots occur in close proximity (e.g., separated by only two base pairs) or even overlap

in the cow IGHD8-2 (**Figure 1B**), suggesting the possibility that double-strand breaks, and associated deletions/insertions could occur through NHEJ within the ultralong D region.

Besides catalyzing SHM and CSR, AID also mediates GC, a process found in chickens and rabbits but not apparently in humans and mice (50). In GC, which takes place after V(D)J recombination but before SHM, pseudogenes on the same chromosome as the antibody genes are used as a template for part of the antibody variable region during DNA replication, thus adding diversity to the antibody (50). Bovine HCs and LCs have been reported to undergo GC (67, 68). However, some researchers suggest that AID-mediated SHM alone accounts for bovine Ig diversification after V(D)J recombination (69). As there are several V_H pseudogenes in the cow, it will be important to definitively determine whether GC plays a role in diversity generation in bovines.

REGULATION OF AID

AID is primarily located in the cytoplasm (70) but clearly functions in the nucleus of B cells (71). Similar to its close relative, APOBEC1, AID is also a nuclear–cytoplasm shuttle protein that contains a nuclear localization signal on its N-terminus and a nuclear export signal on its C-terminus. Both signals are indispensable for maintaining the dynamic equilibrium of AID distribution between the nucleus and the cytoplasm (71, 72). As a potential oncogenic mutator which can also target non-Ig genes in cells other than B cells, the cellular level of AID has to be tightly controlled to prevent diseases caused by its abnormal expression (72, 73). AID expression is induced through various B cell signaling pathways, including B cell receptor, IL-4R, toll-like receptor, and CD40 (38). Several levels of regulation then engage to control the AID expression process. First, AID has very long transcriptional regulatory elements including a 1.6 kb promoter with which transcription factors NF-κB, HOXC4, and Pax5, etc., interact, thus regulating its transcription (74–76). Second, after transcription, cellular AID mRNA levels are modulated by two microRNAs (miR-155 and miR181b) (77–79). Third, the amount of AID is also regulated at the post-translational level in both the cytoplasm and the nucleus. For example, Hsp90 is able to stabilize AID in the cytoplasm of B cells by binding to and preventing it from being degraded through polyubiquitination (80). Translation elongation factor 1α also forms a complex with AID in the cytoplasm to prevent excessive transport into the nucleus (81), while REG-γ mediates degradation of overexpressed AID in the nucleus through proteasomes (82). Moreover, activities of AID are further regulated through phosphorylation. For example, protein kinase A phosphorylates AID at Ser-38, which is crucial for the ability of AID to induce CSR (83, 84). Given that bovines induce SHM in the primary repertoire, these regulatory mechanisms are likely important in cow B-cell development, however, detailed study of AID regulation, and potential differences between human, mouse, and cow, remain to be elucidated.

The molecular mechanism of how AID is specifically targeted to the Ig variable regions to induce SHM is still under exploration. The previously described SHM hotspot motif ("RGYW") alone is not enough to explain why Ig loci are preferentially targeted by AID, because this motif also exists in many other functional non-Ig genes (85). One early study investigating AID targeting on different segments of Ig genes concluded that the reason for the first 100 bp at the 5′ end and C regions of the Ig genes free of SHM is due to spatial constraints which prevent AID's access to these regions (85, 86). Another proposed positive regulation for AID targeting to the Ig variable regions is the E-box motif ("CANNTG"), a cis-regulatory element (CRE) that exists at multiple locations in the Ig loci (87). E-box motifs provide binding sites for the E2A trans-acting factors, which stimulate SHM during transcription (79, 85, 87, 88). However, it has been reported that E2A proteins may also interact with other unknown cis-acting elements, not the E-box motif alone (88, 89). To date, the only confirmed SHM-related CRE is the one that targets AID to the Ig light chain locus in chicken DT40 cells to induce SHM. It is located in the 3′ regulatory region and is different from previously known CREs (79, 90–92). However, how AID interacts with this CRE and whether similar CREs exist in chicken HC locus as well as in other mammalian B cells such as cows still await further characterization. One latest research effort made use of a DNA capture library to identify 275 AID targets at the genomic level in mouse germinal center B cells and successfully revealed more characteristics of AID targets (93). For example, AID targets have higher transcription levels and rates than non-targets, higher binding density of RNAPolII and Spt5, and more transcription and elongation marks. AID targets are also close to superenhancers as well as within convergent transcription regions (93). In addition, this research also confirms the roles of BER and MMR in repairing AID-mistargeted genes in germinal center B cells (93, 94). The findings and methods in this research contribute substantially to the understanding of AID targeting specificity, and is likely applicable to bovine germinal centers.

POTENTIAL TARGET OF AID IN BOVINE ULTRALONG CDR H3 IG GENES

Because bovine antibodies with ultralong CDR H3 almost solely use IGHV1-7, which encodes antibodies with low variation in CDR H1 and H2, AID has been hypothesized to preferentially target IGHD8-2, which encodes the entire knob region, to induce SHM (9, 10). This targeting occurs in the pre-immune repertoire (before antigen exposure). As discussed in previous sections, evidence supporting this hypothesis includes the prevalence of significant hotspot motifs ("RGYW") in the germline IGHD8-2 sequence (9), diverse amino acid content, length variation, and different cysteine numbers and positions in the knob regions encoded by IGHD8-2 (9, 10). Experiments on AID's ability to induce mutations, insertions, and/or deletions into the bovine germline IGHD8-2 gene will be important to ascertain its role in development of the cow ultralong CDR H3 repertoire.

SUMMARY

A subset of bovine antibodies contains ultralong CDR H3s ranging from 40–70 amino acids in length. These ultralong CDR H3s form a unique stalk and knob structure that likely allows them

to bind epitopes that would be inaccessible to shorter antibodies. Ultralong CDR H3s always contain the same V and D gene segments, implying that a significant component of their diversity must result from processes other than V(D)J recombination. Indeed, the germline D gene segment found in all ultralong antibodies contains 19 AID hotspots, suggesting that AID is largely responsible for the unusual cysteine-based diversity seen in bovine ultralong antibodies, possibly through a combination of point mutations, insertions, deletions, and GC.

NOMENCLATURE

ADCC	Antibody-dependent cell-mediated cytotoxicity
AID	Activation-induced cytidine deaminase
APE1	Apurinic, apyrimidinic endonuclease
BER	Base excision repair
C	Cytidine
CDC	Complement-dependent cytotoxicity
CDR	Complementarity determining region
CSR	Class switch recombination
CRE	cis-regulatory element
D	Diversity
FR	Framework
GC	Gene conversion
HC	Heavy chain
Ig	Immunoglobulin
IgG	Immunoglobulin gamma

(Continued)

IgNAR	Immunoglobulin new antigen receptor
J	Joining region
LC	Light chain
LRR	Leucine-rich repeat
mAb	Monoclonal antibody
MMR	Mismatch repair
NHEJ	Non-homologous end joining
nt	Nucleotide
Polη	DNA polymerase eta
Polι	DNA polymerse iota
Polζ	DNA polymerase zeta
SHM	Somatic hypermutation
TdT	Terminal deoxynucleotidyl transferase
U	Uracil
UNG	Uracil-DNA glycosylase
V	Variable
V_H	Heavy chain variable region
V_L	Light chain variable region
VLR	Variable lymphocyte receptor

AUTHOR CONTRIBUTIONS

JH, RH, and VS wrote the manuscript and generated the figures.

FUNDING

REFERENCES

1. Chehimi J, Valiante NM, D'andrea A, Rengaraju M, Rosado Z, Kobayashi M, et al. Enhancing effect of natural killer cell stimulatory factor (NKSF/interleukin-12) on cell-mediated cytotoxicity against tumor-derived and virus-infected cells. *Eur J Immunol* (1993) 23:1826–30. doi:10.1002/eji.1830230814

2. Lu LL, Suscovich TJ, Fortune SM, Alter G. Beyond binding: antibody effector functions in infectious diseases. *Nat Rev Immunol* (2018) 18:46–61. doi:10.1038/nri.2017.106

3. Berens SJ, Wylie DE, Lopez OJ. Use of a single VH family and long CDR3s in the variable region of cattle Ig heavy chains. *Int Immunol* (1997) 9:189–99. doi:10.1093/intimm/9.1.189

4. Lopez O, Perez C, Wylie D. A single VH family and long CDR3s are the targets for hypermutation in bovine immunoglobulin heavy chains. *Immunol Rev* (1998) 162:55–66. doi:10.1111/j.1600-065X.1998.tb01429.x

5. Saini SS, Allore B, Jacobs RM, Kaushik A. Exceptionally long CDR3H region with multiple cysteine residues in functional bovine IgM antibodies. *Eur J Immunol* (1999) 29:2420–6. doi:10.1002/(SICI)1521-4141(199908)29:08<2420::AID-IMMU2420>3.0.CO;2-A

6. Shojaei F, Saini SS, Kaushik AK. Unusually long germline DH genes contribute to large sized CDR3H in bovine antibodies. *Mol Immunol* (2003) 40:61–7. doi:10.1016/S0161-5890(03)00098-1

7. Kaushik AK, Kehrli ME Jr, Kurtz A, Ng S, Koti M, Shojaei F, et al. Somatic hypermutations and isotype restricted exceptionally long CDR3H contribute to antibody diversification in cattle. *Vet Immunol Immunopathol* (2009) 127:106–13. doi:10.1016/j.vetimm.2008.09.024

8. Koti M, Kataeva G, Kaushik AK. Novel atypical nucleotide insertions specifically at VH-DH junction generate exceptionally long CDR3H in cattle antibodies. *Mol Immunol* (2010) 47:2119–28. doi:10.1016/j.molimm.2010.02.014

9. Wang F, Ekiert DC, Ahmad I, Yu W, Zhang Y, Bazirgan O, et al. Reshaping antibody diversity. *Cell* (2013) 153:1379–93. doi:10.1016/j.cell.2013.04.049

10. Deiss TC, Vadnais M, Wang F, Chen PL, Torkamani A, Mwangi W, et al. Immunogenetic factors driving formation of ultralong VH CDR3 in *Bos taurus* antibodies. *Cell Mol Immunol* (2017) 14:1–12. doi:10.1038/cmi.2017.117

11. Sok D, Le KM, Vadnais M, Saye-Francisco KL, Jardine JG, Torres JL, et al. Rapid elicitation of broadly neutralizing antibodies to HIV by immunization in cows. *Nature* (2017) 548:108–11. doi:10.1038/nature23301

12. Vu KB, Ghahroudi MA, Wyns L, Muyldermans S. Comparison of llama VH sequences from conventional and heavy chain antibodies. *Mol Immunol* (1997) 34:1121–31. doi:10.1016/S0161-5890(97)00146-6

13. Tanha J, Dubuc G, Hirama T, Narang SA, Mackenzie CR. Selection by phage display of llama conventional V(H) fragments with heavy chain antibody V(H)H properties. *J Immunol Methods* (2002) 263:97–109. doi:10.1016/S0022-1759(02)00027-3

14. Diaz M, Stanfield RL, Greenberg AS, Flajnik MF. Structural analysis, selection, and ontogeny of the shark new antigen receptor (IgNAR): identification of a new locus preferentially expressed in early development. *Immunogenetics* (2002) 54:501–12. doi:10.1007/s00251-002-0479-z

15. Dooley H, Flajnik MF, Porter AJ. Selection and characterization of naturally occurring single-domain (IgNAR) antibody fragments from immunized sharks by phage display. *Mol Immunol* (2003) 40:25–33. doi:10.1016/S0161-5890(03)00084-1

16. Stanfield RL, Dooley H, Flajnik MF, Wilson IA. Crystal structure of a shark single-domain antibody V region in complex with lysozyme. *Science* (2004) 305:1770–3. doi:10.1126/science.1101148

17. Prabakaran P, Chen W, Singarayan MG, Stewart CC, Streaker E, Feng Y, et al. Expressed antibody repertoires in human cord blood cells: 454 sequencing and IMGT/HighV-QUEST analysis of germline gene usage, junctional diversity, and somatic mutations. *Immunogenetics* (2012) 64:337–50. doi:10.1007/s00251-011-0595-8

18. Weinstein PD, Anderson AO, Mage RG. Rabbit IgH sequences in appendix germinal centers: VH diversification by gene conversion-like and hypermutation mechanisms. *Immunity* (1994) 1:647–59. doi:10.1016/1074-7613(94)90036-1

19. Mehr R, Edelman H, Sehgal D, Mage R. Analysis of mutational lineage trees from sites of primary and secondary Ig gene diversification in rabbits and chickens. *J Immunol* (2004) 172:4790–6. doi:10.4049/jimmunol.172.8.4790

20. 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. doi:10.1038/mi.2012.132

21. Ekman A, Niku M, Liljavirta J, Iivanainen A. *Bos taurus* genome sequence reveals the assortment of immunoglobulin and surrogate light chain genes in domestic cattle. *BMC Immunol* (2009) 10:22. doi:10.1186/1471-2172-10-22

22. Butler JE. Immunoglobulin diversity, B-cell and antibody repertoire development in large farm animals. *Rev Sci Tech* (1998) 17:43–70. doi:10.20506/rst.17.1.1096

23. Walther S, Czerny CP, Diesterbeck US. Exceptionally long CDR3H are not isotype restricted in bovine immunoglobulins. *PLoS One* (2013) 8:e64234. doi:10.1371/journal.pone.0064234

24. Koti M, Kataeva G, Kaushik AK. Organization of D(H)-gene locus is distinct in cattle. *Dev Biol (Basel)* (2008) 132:307–13. doi:10.1159/000317176

25. Stanfield RL, Wilson IA, Smider VV. Conservation and diversity in the ultralong third heavy-chain complementarity-determining region of bovine antibodies. *Sci Immunol* (2016) 1:aaf7962. doi:10.1126/sciimmunol.aaf7962

26. Liljavirta J, Niku M, Pessa-Morikawa T, Ekman A, Iivanainen A. Expansion of the preimmune antibody repertoire by junctional diversity in *Bos taurus*. *PLoS One* (2014) 9:e99808. doi:10.1371/journal.pone.0099808

27. 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. doi:10.1016/S0092-8674(00)00078-7

28. Arakawa H, Saribasak H, Buerstedde JM. Activation-induced cytidine deaminase initiates immunoglobulin gene conversion and hypermutation by a common intermediate. *PLoS Biol* (2004) 2:E179. doi:10.1371/journal.pbio.0020179

29. 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. doi:10.1093/molbev/msi026

30. Navaratnam N, Bhattacharya S, Fujino T, Patel D, Jarmuz AL, Scott J. Evolutionary origins of apoB mRNA editing: catalysis by a cytidine deaminase that has acquired a novel RNA-binding motif at its active site. *Cell* (1995) 81:187–95. doi:10.1016/0092-8674(95)90328-3

31. Moris A, Murray S, Cardinaud S. AID and APOBECs span the gap between innate and adaptive immunity. *Front Microbiol* (2014) 5:534. doi:10.3389/fmicb.2014.00534

32. Dickerson SK, Market E, Besmer E, Papavasiliou FN. AID mediates hypermutation by deaminating single stranded DNA. *J Exp Med* (2003) 197:1291–6. doi:10.1084/jem.20030481

33. Chaudhuri J, Tian M, Khuong C, Chua K, Pinaud E, Alt FW. Transcription-targeted DNA deamination by the AID antibody diversification enzyme. *Nature* (2003) 422:726–30. doi:10.1038/nature01574

34. Sheehy AM, Gaddis NC, Choi JD, Malim MH. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature* (2002) 418:646–50. doi:10.1038/nature00939

35. Suspene R, Sommer P, Henry M, Ferris S, Guetard D, Pochet S, et al. APOBEC3G is a single-stranded DNA cytidine deaminase and functions independently of HIV reverse transcriptase. *Nucleic Acids Res* (2004) 32:2421–9. doi:10.1093/nar/gkh554

36. Yu Q, Konig R, Pillai S, Chiles K, Kearney M, Palmer S, et al. Single-strand specificity of APOBEC3G accounts for minus-strand deamination of the HIV genome. *Nat Struct Mol Biol* (2004) 11:435–42. doi:10.1038/nsmb758

37. Chiu YL, Greene WC. The APOBEC3 cytidine deaminases: an innate defensive network opposing exogenous retroviruses and endogenous retroelements. *Annu Rev Immunol* (2008) 26:317–53. doi:10.1146/annurev.immunol.26.021607.090350

38. Xu Z, Pone EJ, Al-Qahtani A, Park SR, Zan H, Casali P. Regulation of aicda expression and AID activity: relevance to somatic hypermutation and class switch DNA recombination. *Crit Rev Immunol* (2007) 27:367–97. doi:10.1615/CritRevImmunol.v27.i4.60

39. Pancer Z, Amemiya CT, Ehrhardt GR, Ceitlin J, Gartland GL, Cooper MD. Somatic diversification of variable lymphocyte receptors in the Agnathan sea lamprey. *Nature* (2004) 430:174–80. doi:10.1038/nature02740

40. Alder MN, Rogozin IB, Iyer LM, Glazko GV, Cooper MD, Pancer Z. Diversity and function of adaptive immune receptors in a jawless vertebrate. *Science* (2005) 310:1970–3. doi:10.1126/science.1119420

41. Nagawa F, Kishishita N, Shimizu K, Hirose S, Miyoshi M, Nezu J, et al. Antigen-receptor genes of the Agnathan lamprey are assembled by a process involving copy choice. *Nat Immunol* (2007) 8:206–13. doi:10.1038/ni1419

42. Rogozin IB, Iyer LM, Liang L, Glazko GV, Liston VG, Pavlov YI, et al. Evolution and diversification of lamprey antigen receptors: evidence for involvement of an AID-APOBEC family cytosine deaminase. *Nat Immunol* (2007) 8:647–56. doi:10.1038/ni1463

43. Pancer Z, Saha NR, Kasamatsu J, Suzuki T, Amemiya CT, Kasahara M, et al. Variable lymphocyte receptors in hagfish. *Proc Natl Acad Sci U S A* (2005) 102:9224–9. doi:10.1073/pnas.0503792102

44. De Los Rios M, Criscitiello MF, Smider VV. Structural and genetic diversity in antibody repertoires from diverse species. *Curr Opin Struct Biol* (2015) 33:27–41. doi:10.1016/j.sbi.2015.06.002

45. Magor BG. Antibody affinity maturation in fishes—our current understanding. *Biology (Basel)* (2015) 4:512–24. doi:10.3390/biology4030512

46. Barreto VM, Pan-Hammarstrom Q, Zhao Y, Hammarstrom L, Misulovin Z, Nussenzweig MC. AID from bony fish catalyzes class switch recombination. *J Exp Med* (2005) 202:733–8. doi:10.1084/jem.20051378

47. Wakae K, Magor BG, Saunders H, Nagaoka H, Kawamura A, Kinoshita K, et al. Evolution of class switch recombination function in fish activation-induced cytidine deaminase, AID. *Int Immunol* (2006) 18:41–7. doi:10.1093/intimm/dxh347

48. Conticello SG. The AID/APOBEC family of nucleic acid mutators. *Genome Biol* (2008) 9:229. doi:10.1186/gb-2008-9-6-229

49. 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. doi:10.1074/jbc.274.26.18470

50. Arakawa H, Buerstedde JM. Immunoglobulin gene conversion: insights from bursal B cells and the DT40 cell line. *Dev Dyn* (2004) 229:458–64. doi:10.1002/dvdy.10495

51. Peled JU, Kuang FL, Iglesias-Ussel MD, Roa S, Kalis SL, Goodman MF, et al. The biochemistry of somatic hypermutation. *Annu Rev Immunol* (2008) 26:481–511. doi:10.1146/annurev.immunol.26.021607.090236

52. Diaz M, Lawrence C. An update on the role of translesion synthesis DNA polymerases in Ig hypermutation. *Trends Immunol* (2005) 26:215–20. doi:10.1016/j.it.2005.02.008

53. Daly J, Bebenek K, Watt DL, Richter K, Jiang C, Zhao ML, et al. Altered Ig hypermutation pattern and frequency in complementary mouse models of DNA polymerase zeta activity. *J Immunol* (2012) 188:5528–37. doi:10.4049/jimmunol.1102629

54. King DJ, Bowers PM, Kehry MR, Horlick RA. Mammalian cell display and somatic hypermutation in vitro for human antibody discovery. *Curr Drug Discov Technol* (2014) 11:56–64. doi:10.2174/15701638113109990037

55. Martin A, Bardwell PD, Woo CJ, Fan M, Shulman MJ, Scharff MD. Activation-induced cytidine deaminase turns on somatic hypermutation in hybridomas. *Nature* (2002) 415:802–6. doi:10.1038/nature714

56. Martin A, Scharff MD. Somatic hypermutation of the AID transgene in B and non-B cells. *Proc Natl Acad Sci U S A* (2002) 99:12304–8. doi:10.1073/pnas.192442899

57. Michael N, Martin TE, Nicolae D, Kim N, Padjen K, Zhan P, et al. Effects of sequence and structure on the hypermutability of immunoglobulin genes. *Immunity* (2002) 16:123–34. doi:10.1016/S1074-7613(02)00261-3

58. Rogozin IB, Diaz M. Cutting edge: DGYW/WRCH is a better predictor of mutability at G:C bases in Ig hypermutation than the widely accepted RGYW/WRCY motif and probably reflects a two-step activation-induced cytidine deaminase-triggered process. *J Immunol* (2004) 172:3382–4. doi:10.4049/jimmunol.172.6.3382

59. Reason DC, Zhou J. Codon insertion and deletion functions as a somatic diversification mechanism in human antibody repertoires. *Biol Direct* (2006) 1:24. doi:10.1186/1745-6150-1-24

60. Wilson PC, De Bouteiller O, Liu YJ, Potter K, Banchereau J, Capra JD, et al. Somatic hypermutation introduces insertions and deletions into immunoglobulin V genes. *J Exp Med* (1998) 187:59–70. doi:10.1084/jem.187.1.59

61. Sale JE, Neuberger MS. TdT-accessible breaks are scattered over the immunoglobulin V domain in a constitutively hypermutating B cell line. *Immunity* (1998) 9:859–69. doi:10.1016/S1074-7613(00)80651-2

62. Briney BS, Willis JR, Crowe JE Jr. Location and length distribution of somatic hypermutation-associated DNA insertions and deletions reveals regions of

antibody structural plasticity. *Genes Immun* (2012) 13:523–9. doi:10.1038/gene.2012.28

63. 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. doi:10.1016/S0092-8674(00)00079-9

64. Chaudhuri J, Alt FW. Class-switch recombination: interplay of transcription, DNA deamination and DNA repair. *Nat Rev Immunol* (2004) 4:541–52. doi:10.1038/nri1395

65. Gu H, Zou YR, Rajewsky K. Independent control of immunoglobulin switch recombination at individual switch regions evidenced through Cre-loxP-mediated gene targeting. *Cell* (1993) 73:1155–64. doi:10.1016/0092-8674(93)90644-6

66. Lieber MR, Yu K, Raghavan SC. Roles of nonhomologous DNA end joining, V(D)J recombination, and class switch recombination in chromosomal translocations. *DNA Repair (Amst)* (2006) 5:1234–45. doi:10.1016/j.dnarep.2006.05.013

67. Parng CL, Hansal S, Goldsby RA, Osborne BA. Gene conversion contributes to Ig light chain diversity in cattle. *J Immunol* (1996) 157:5478–86.

68. Walther S, Tietze M, Czerny CP, Konig S, Diesterbeck US. Development of a bioinformatics framework for the detection of gene conversion and the analysis of combinatorial diversity in immunoglobulin heavy chains in four cattle breeds. *PLoS One* (2016) 11:e0164567. doi:10.1371/journal.pone.0164567

69. Verma S, Aitken R. Somatic hypermutation leads to diversification of the heavy chain immunoglobulin repertoire in cattle. *Vet Immunol Immunopathol* (2012) 145:14–22. doi:10.1016/j.vetimm.2011.10.001

70. Rada C, Jarvis JM, Milstein C. AID-GFP chimeric protein increases hypermutation of Ig genes with no evidence of nuclear localization. *Proc Natl Acad Sci U S A* (2002) 99:7003–8. doi:10.1073/pnas.092160999

71. Ito S, Nagaoka H, Shinkura R, Begum N, Muramatsu M, Nakata M, et al. Activation-induced cytidine deaminase shuttles between nucleus and cytoplasm like apolipoprotein B mRNA editing catalytic polypeptide 1. *Proc Natl Acad Sci U S A* (2004) 101:1975–80. doi:10.1073/pnas.0307335101

72. Okazaki IM, Hiai H, Kakazu N, Yamada S, Muramatsu M, Kinoshita K, et al. Constitutive expression of AID leads to tumorigenesis. *J Exp Med* (2003) 197:1173–81. doi:10.1084/jem.20030275

73. Zan H, Casali P. Regulation of Aicda expression and AID activity. *Autoimmunity* (2013) 46:83–101. doi:10.3109/08916934.2012.749244

74. Park SR, Zan H, Pal Z, Zhang J, Al-Qahtani A, Pone EJ, et al. HoxC4 binds to the promoter of the cytidine deaminase AID gene to induce AID expression, class-switch DNA recombination and somatic hypermutation. *Nat Immunol* (2009) 10:540–50. doi:10.1038/ni.1725

75. Stavnezer J. Complex regulation and function of activation-induced cytidine deaminase. *Trends Immunol* (2011) 32:194–201. doi:10.1016/j.it.2011.03.003

76. Keim C, Kazadi D, Rothschild G, Basu U. Regulation of AID, the B-cell genome mutator. *Genes Dev* (2013) 27:1–17. doi:10.1101/gad.200014.112

77. De Yebenes VG, Belver L, Pisano DG, Gonzalez S, Villasante A, Croce C, et al. miR-181b negatively regulates activation-induced cytidine deaminase in B cells. *J Exp Med* (2008) 205:2199–206. doi:10.1084/jem.20080579

78. Teng G, Hakimpour P, Landgraf P, Rice A, Tuschl T, Casellas R, et al. MicroRNA-155 is a negative regulator of activation-induced cytidine deaminase. *Immunity* (2008) 28:621–9. doi:10.1016/j.immuni.2008.03.015

79. Delker RK, Fugmann SD, Papavasiliou FN. A coming-of-age story: activation-induced cytidine deaminase turns 10. *Nat Immunol* (2009) 10:1147–53. doi:10.1038/ni.1799

80. Orthwein A, Patenaude AM, Affar El B, Lamarre A, Young JC, Di Noia JM. Regulation of activation-induced deaminase stability and antibody gene diversification by Hsp90. *J Exp Med* (2010) 207:2751–65. doi:10.1084/jem.20101321

81. Hasler J, Rada C, Neuberger MS. Cytoplasmic activation-induced cytidine deaminase (AID) exists in stoichiometric complex with translation elongation factor 1alpha (eEF1A). *Proc Natl Acad Sci U S A* (2011) 108:18366–71. doi:10.1073/pnas.1106729108

82. Uchimura Y, Barton LF, Rada C, Neuberger MS. REG-gamma associates with and modulates the abundance of nuclear activation-induced deaminase. *J Exp Med* (2011) 208:2385–91. doi:10.1084/jem.20110856

83. Basu U, Chaudhuri J, Alpert C, Dutt S, Ranganath S, Li G, et al. The AID antibody diversification enzyme is regulated by protein kinase A phosphorylation. *Nature* (2005) 438:508–11. doi:10.1038/nature04255

84. Basu U, Franklin A, Alt FW. Post-translational regulation of activation-induced cytidine deaminase. *Philos Trans R Soc Lond B Biol Sci* (2009) 364:667–73. doi:10.1098/rstb.2008.0194

85. Odegard VH, Schatz DG. Targeting of somatic hypermutation. *Nat Rev Immunol* (2006) 6:573–83. doi:10.1038/nri1896

86. Longerich S, Tanaka A, Bozek G, Nicolae D, Storb U. The very 5' end and the constant region of Ig genes are spared from somatic mutation because AID does not access these regions. *J Exp Med* (2005) 202:1443–54. doi:10.1084/jem.20051604

87. Michael N, Shen HM, Longerich S, Kim N, Longacre A, Storb U. The E box motif CAGGTG enhances somatic hypermutation without enhancing transcription. *Immunity* (2003) 19:235–42. doi:10.1016/S1074-7613(03)00204-8

88. Yang SY, Schatz DG. Targeting of AID-mediated sequence diversification by cis-acting determinants. *Adv Immunol* (2007) 94:109–25. doi:10.1016/S0065-2776(06)94004-8

89. Schoetz U, Cervelli M, Wang YD, Fiedler P, Buerstedde JM. E2A expression stimulates Ig hypermutation. *J Immunol* (2006) 177:395–400. doi:10.4049/jimmunol.177.1.395

90. Kothapalli N, Norton DD, Fugmann SD. Cutting edge: a cis-acting DNA element targets AID-mediated sequence diversification to the chicken Ig light chain gene locus. *J Immunol* (2008) 180:2019–23. doi:10.4049/jimmunol.180.4.2019

91. Blagodatski A, Batrak V, Schmidl S, Schoetz U, Caldwell RB, Arakawa H, et al. A cis-acting diversification activator both necessary and sufficient for AID-mediated hypermutation. *PLoS Genet* (2009) 5:e1000332. doi:10.1371/journal.pgen.1000332

92. Kothapalli NR, Fugmann SD. Targeting of AID-mediated sequence diversification to immunoglobulin genes. *Curr Opin Immunol* (2011) 23:184–9. doi:10.1016/j.coi.2010.12.009

93. Alvarez-Prado AF, Perez-Duran P, Perez-Garcia A, Benguria A, Torroja C, De Yebenes VG, et al. A broad atlas of somatic hypermutation allows prediction of activation-induced deaminase targets. *J Exp Med* (2018) 215:761–71. doi:10.1084/jem.20171738

94. Reynaud CA, Weill JC. Predicting AID off-targets: a step forward. *J Exp Med* (2018) 215:721–2. doi:10.1084/jem.20180231

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.