> CHAPTER FOUR

The Unusual Genetics and Biochemistry of Bovine Immunoglobulins

Robyn L. Stanfield*, Jeremy Haakenson*, Thaddeus C. Deiss[†],
Michael F. Criscitiello[†], Ian A. Wilson*, Vaughn V. Smider*,[1]
*The Scripps Research Institute, La Jolla, CA, United States
[†]College of Veterinary Medicine and Biomedical Sciences, Texas A&M University, College Station, TX, United States
[1]Corresponding author: e-mail address: vvsmider@scripps.edu

Contents

Abstract

Antibodies are the key circulating molecules that have evolved to fight infection by the adaptive immune system of vertebrates. Typical antibodies of most species contain six complementarity-determining regions (CDRs), where the third CDR of the heavy chain (CDR H3) has the greatest diversity and often makes the most significant contact with antigen. Generally, the process of V(D)J recombination produces a vast repertoire of antibodies; multiple V, D, and J gene segments recombine with additional junctional diversity at the V-D and D-J joints, and additional combinatorial possibilities occur through heavy- and light-chain pairing. Despite these processes, the overall structure of the resulting antibody is largely conserved, and binding to antigen occurs predominantly through the CDR loops of the immunoglobulin V domains. Bovines have deviated from this general paradigm by having few VH regions and thus little germline combinatorial diversity, but their antibodies contain long CDR H3 regions, with

substantial diversity generated through somatic hypermutation. A subset of the repertoire comprises antibodies with ultralong CDR H3s, which can reach over 70 amino acids in length. Structurally, these unusual antibodies form a β-ribbon "stalk" and disulfide-bonded "knob" that protrude far from the antibody surface. These long CDR H3s allow cows to mount a particularly robust immune response when immunized with viral antigens, particularly to broadly neutralizing epitopes on a stabilized HIV gp140 trimer, which has been a challenge for other species. The unusual genetics and structural biology of cows provide for a unique paradigm for creation of immune diversity and could enable generation of antibodies against especially challenging targets and epitopes.

1. OVERVIEW OF COW ANTIBODIES

The immunology of domesticated species has been inextricably linked to humans throughout history. Cows have a unique place in the history of modern immunology research by being central to the discovery of the first vaccine. Indeed, the word "vaccine" itself is derived from the Latin "vacca" (cow). In the 1760s, Edward Jenner realized that dairy workers seemed to be resistant to the deadly smallpox virus because they had been infected with cowpox, which causes only mild disease in humans. Jenner then began inoculating humans with sera from cowpox-infected cows and found that he could prevent infection with smallpox (reviewed in Baxby, 1999). Despite the genetic divergence of cowpox and smallpox, conserved neutralizing epitopes could mediate an effective antibody response that could cross-neutralize these different viruses.

Interestingly, the discovery of immunologic tolerance also occurred in cattle; Ray D. Owen showed that dizygotic twin calves possess red blood cells of their twin, which were not self-immunoreactive (Owen, 1945). Similarly, Sir Peter Medawar's initial work on skin transplantation showed that dizygotic bovine twins would not reject grafts from their sibling (Anderson, Billingham, Lampkin, & Medawar, 1951). This latter work ultimately led to the series of classic experiments in mice and other species establishing key tolerance principles for which the Nobel Prize was awarded in 1960 (Billingham & Brent, 1956; Billingham, Brent, & Medawar, 1953). Despite these important historical experiments, cows have not been deeply studied at the molecular immunological level compared to many other model organisms, such as mice, rabbits, and macaques. Recent work, however, has shown that cow antibodies have unusually long CDR H3s with novel protruding "stalk" and "knob" structures (Fig. 1) and underlying genetics (Wang et al., 2013). Importantly, cows can mount particularly
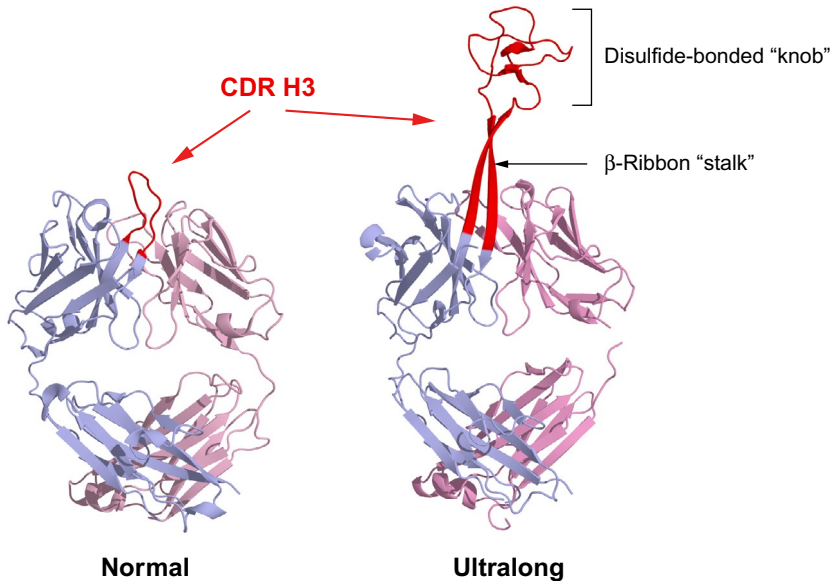
Fig. 1 Comparison of "normal" and "ultralong" CDR H3s. Structure of an antibody with a "normal" CDR H3 length, Ofatumumab (*left*, PDB: 3giz) and an ultralong CDR H3 cow antibody, BLV1H12 (*right*, PDB: 4k3d). The CDR H3s are colored *red* with the heavy chain in *light blue* and light chain in *gray*. The β-ribbon "stalk" and disulfide-bonded "knob" motifs are labeled.

robust responses against challenging viruses like HIV-1 (Sok et al., 2017), suggesting that these unusual antibodies may have particular utility in targeting certain antigenic epitopes and thereby warrant further investigation.

An initial clue to the unusual antibodies of cows came from the immunogenetic study of Berens et al., where several immunoglobulin heavy chain sequences were analyzed from fetal and adult lymphoid tissue, and one sequence termed F18M was found to have a CDR H3 of 51 amino acids (Berens, Wylie, & Lopez, 1997). Interestingly, this transcript was derived from fetal tissue and, in hindsight, appears to have an unmutated (germline) V-D-J rearranged sequence. Shortly thereafter, several more exceptionally long CDR H3 sequences were studied in more detail by Kaushik and colleagues who were investigating bovine leukemia virus (BLV)-infected B-cells from cows. A total of 44 heterohybridomas from BLV-infected B-cells were obtained and sequencing of the heavy chains revealed four unusually long CDR H3s (9% incidence), ranging from 56 to 61 amino acids. Further studies by several different groups found the ultralong sequences at a frequency of 1%–22% of heavy chains, with some of the

frequency variation possibly arising from different procedural methods, tissues analyzed, breeds, and immune status of the cattle (Deiss et al., 2017; Liljavirta, Niku, Pessa-Morikawa, Ekman, & Iivanainen, 2014; Ma et al., 2016; Walther, Czerny, & Diesterbeck, 2013; Wang et al., 2013). Cattle of different breeds all appear to have these unusually long CDR H3s, but at slightly different frequencies when examined side-by-side (Walther, Tietze, Czerny, Konig, & Diesterbeck, 2016). Kaushik's group identified multiple cysteine residues in CDR H3, found that they usually occurred in even numbers, and predicted that these may form unusual structures (Saini, Allore, Jacobs, & Kaushik, 1999). Interestingly, these four sequences (termed BLV1H12, BLV5B8, BLV5D3, and BLV8C11), which were highly homologous outside of CDR H3, were predicted to utilize the same germline variable region gene (originally termed g1.110.20). Importantly, radiolabeled clones secreted properly formed IgM antibody, suggesting that the antibodies with long CDR H3s were functional (Saini et al., 1999). The antigens for these antibodies were unknown; however, some appeared polyspecific when tested on a panel of antigens (Saini et al., 1999). While originally identified in several IgM antibodies, ultralong CDR H3 sequences were ultimately identified in all other isotypes (Walther et al., 2013; Wang et al., 2013).

2. GENETICS OF ULTRALONG CDR H3 COW ANTIBODIES

2.1 Genomic Organization of the IGH Locus

The immunoglobulin heavy chain of cattle was originally thought to be encoded in two distinct loci (Hayes & Petit, 1993; Hosseini, Campbell, Prorocic, & Aitken, 2004; Tobin-Janzen & Womack, 1992), on autosomes BTA11 and BTA21. However, more recently, it has been shown that chromosome 11 only harbors a truncated μCH2, while the complete functional locus has been localized by fluorescent in situ hybridization to BTA21q24 (Ma et al., 2016). All five immunoglobulin heavy chain isotypes known to eutherian (i.e., placental) mammals are also expressed in cattle and encoded on chromosome 21: IgM (Mousavi, Rabbani, Pilstrom, & Hammarstrom, 1998), IgD (Zhao et al., 2002), IgG (Kacskovics & Butler, 1996; Rabbani, Brown, Butler, & Hammarstrom, 1997; Symons, Clarkson, & Beale, 1989), IgE (Mousavi, Rabbani, & Hammarstrom, 1997), and IgA (Brown, Rabbani, Butler, & Hammarstrom, 1997). There are three subisotypes of bovine IgG, and more interestingly two IgM (Fig. 2). Both of the μ genes appear functional as they both must be inactivated to ablate B lymphopoiesis
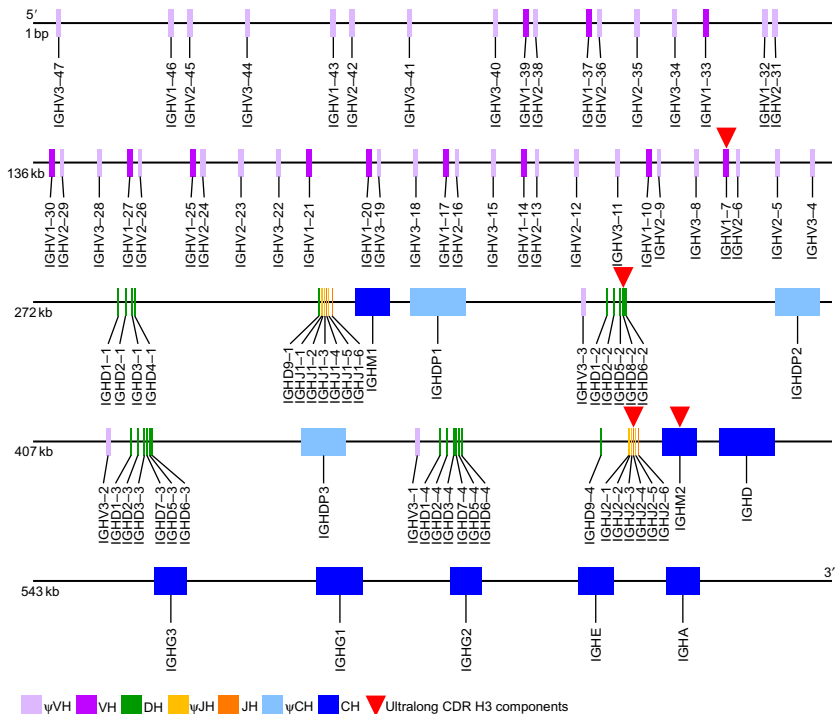
Fig. 2 Bovine IGH locus genomic organization. Schematic of the orientation of coding elements of the cattle immunoglobulin heavy chain locus. Key at *bottom left* shows the *purple* variable (V) segments, *green* diversity (D), *orange* joining (J), and *blue* constant (C) genes. Pseudo V, J, and C genes are *paler tones*. Elements used preferentially in the ultralong CDR H3 antibodies are marked by *red triangles*.

(Kuroiwa et al., 2009). One μCH (IGHM1) has a pseudogene δCH (psi-IGHD1) 3′ from it and is one of three δCH pseudogenes. The IgD (psi-IgD1) just 3′ of IGM1 has a frameshift in C_H2 and thus has been considered nonfunctional. However, the frameshift is a single base insertion in a homopolymer run, which is notoriously difficult to resolve on any sequencing platform. The other two δ pseudogenes are missing a 5′ splice site, >200 bases of C_H1, and a hinge exon in addition to large intronic deletions (ranging from 500 to 1 kb). The second μCH (IGHM2) precedes a functional δCH (Ma et al., 2016). The functional consequence of two IgM is unclear; however, it is interesting that ultralong CDR H3 antibodies appear to exclusively utilize IgM2 (Ma et al., 2016). The constant genes are ordered 5′-IGHM-IGHD-IGHG3-IGHG1-IGHG2-IGHE-IGHA-3′, with the upstream functional IGHM2 and three pseudo IGHD genes interspersed between clusters of V–D-J elements (Fig. 2).

The bovine immunoglobulin heavy chain locus contains a paucity of functional variable gene segments compared to most vertebrates (Berens et al., 1997; Liljavirta et al., 2014) [although pig (Eguchi-Ogawa et al., 2010) and rabbit (DiPietro et al., 1992) also have a low variable gene count (Jung, Giallourakis, Mostoslavsky, & Alt, 2006)]. Only 12 potentially functional V gene segments exist in cows, all belonging to a single subgroup (IGHV1) (Lopez, Perez, & Wylie, 1998; Niku, Liljavirta, Durkin, Schroderus, & Iivanainen, 2012; Saini, Hein, & Kaushik, 1997), which is most closely homologous to the human IGHV4 subgroup. More pseudogenes exist in this and two other variable subgroups (IGHV2 and IGHV3) (Niku et al., 2012). As discussed later, this initial dearth of rearrangement possibilities is compensated for by substantial activation–induced cytidine deaminase (AID)-mediated somatic hypermutation in the periphery in a second diversification (preantigen exposure) in the gut-associated lymphoid tissue of the Peyer's patches. Four of 12 joining genes are functional at the bovine IGHJ locus. A total of 23 diversity gene segments, 16 of which appear to be functional, are found within 4 clusters of DH segments (Ma et al., 2016). Thus, although restricted in number and diversity, there are adequate canonical V(D)J segments for generation of a heavy chain repertoire at this locus, yet some segments (IGHV1–7, IGHD8–2, IGHJ2–4; marked by red in Fig. 2) have evolved specifically for use in the ultralong CDR H3 antibodies. The variable gene IGHV1–7 (by IMGT nomenclature, previously termed g1.110.20, IGHV10/34, IGHV1S3, IGHV1–1, and VHBUL), whose $3'$ end contains a protracted block of nucleotides that encodes an extended F β-strand, is the most $3'$ of the functional variable gene segments. This variable segment recombines with the exceptionally long IGHD8–2 (previously termed DH2 Shojaei, Saini, & Kaushik, 2003) to encode the "stalk" and "knob" of the ultralong CDR H3 antibodies, as well as the descending strand that connects to the J segment encoded G β-strand, as described later.

2.2 Germline Basis for Ultralong CDR H3 Antibody Formation

The subset of antibodies with ultralong CDR H3 appears to be generated through a V-D-J recombination event that only utilizes one V_H (IGHV1–7), one D_H (IGHD8–2), and one J_H (IGHJ2–4) (Deiss et al., 2017; Kaushik, Shojaei, & Saini, 2002; Koti, Kataeva, & Kaushik, 2008; Wang et al., 2013). While junctional diversity at the V-D and D-J joints in the shorter cow antibody gene repertoire is significant (Liljavirta et al., 2014), for ultralong CDR H3 antibody genes, there appears to be little

diversity at the D-J joint, and only limited diversity characterized by long insertions at the V–D joint. Koti et al. suggested that atypical nucleotide insertions termed conserved short nucleotide sequences (primarily of adenine stretches) play a role in these V–D junctions (Koti, Kataeva, & Kaushik, 2010). Of note, the structural features encoded at the V–D and D-J joints are the ascending and descending β–strands of the protruding stalk, respectively. Therefore, it is likely that only a subset of junctional residues can encode amino acids that maintain the β–strand structures. In this regard, a significant stretch of adenines in this V–D junctional diversity encodes a relatively conserved stretch of amino acids in the ascending strand. It is unclear whether these junctional residues are added by terminal deoxynucleotidyl transferase (TdT), and then the conserved residues are selected during B-cell development (Liljavirta et al., 2014), or whether atypical nucleotide insertional processes (Koti et al., 2010) occur to specifically add these residues during V–D recombination. In any event, there appears to be little recombinational diversity in antibodies with ultralong CDR H3 sequences, with only minor junctional diversity at the V–D junction. Compared to antibody genes from other species, this mechanism to produce functional variable regions is unusual in not enabling significant diversity through V(D)J recombination.

The IGHV1–7 gene is utilized in nearly every ultralong CDR H3 antibody analyzed; however, it is also found rearranged to shorter D_H segments in the shorter repertoire (Deiss et al., 2017). Interestingly, CDR H1 and CDR H2 of IGHV1–7 seem to retain a greater level of germline conservation than other V_Hs, which suggests they act as a scaffold for the ultralong CDR H3 microdomain, in lieu of accruing somatic hypermutations during affinity maturation. IGHV1–7 is unusual among all vertebrate V_H regions in that it encodes three additional amino acids at the C-terminal region (Fig. 3A). All immunoglobulin domains encode two conserved cysteines, which form a disulfide bond at the core of the immunoglobulin β–sandwich structure. The second cysteine is typically three amino acids from the C-terminus of the variable region encoded by the V_H gene. In most vertebrates, an amino acid motif of CAR or CAK is readily recognized in this region. Notably, the positively charged arginine or lysine in CAR or CAK nearly always forms a salt bridge with a conserved aspartate encoded by the J_H region, thus stabilizing the base of the CDR H3 loop. Ultralong CDR H3 antibodies do not have CAR/K but instead have a CTTVHQ motif encoded by IGHV1–7 that does not participate in a salt bridge, but instead initiates the ascending β–strand of the stalk, which protrudes from the antibody surface. Genetically, an 8 bp duplication at the 3′ end of the

A

```
           T  A  T  Y  Y  C              A  R
Typical IGHV  ACGGCCACAT[ACTACTGT]--------GCAAGAGAcacagtgaggggaagtcagtgtgagcccagacaaaaacc

              S  A  T  Y  Y  C  T  T  V  H  Q
IGHV1-7     TCGGCCACAT[ACTACTGT][ACTACTGT]GCACCAGAcacagtgaggggaaatcagtgtgagcccagacaaaaacc
                                                    └─────────────────────────────────┘
                                                              RSS
```

B

```
    S   C   P   D   G   Y   S   Y   G   Y   G   C   G   Y   G   Y   G   C   S   G
GT [AGT T]GT CCT GAT [GGT T]AT [AGT T]AT [GGT T]AT [GGT T]GT [GGT T]AT [GGT T]AT [GGT T]GT AGT [GGT]

    Y   D   C   Y   G   Y   G   G   Y   G   G   Y   G   G   Y   G   Y   S   S   Y
[T]AT GAT TGT TAT [GGT T]AT GGT [GGT T]AT GGT [GGT T]AT GGT [GGT T]AT [GGT T]AT [AGT A]GT TAT

    S   Y   S   Y   T   Y   E   Y
[AGT T]AT [AGT T]A[T ACT] TAC GAA TAT AC
```

C

					L	#Cys
IGHV1-7	CTTVHQ					
IGHD8-2			SCPDGYSYGYGCGYGYGCSGYDCYGYGGYGGYGGYGYSSYSYSYTYEY			
IGHJ2-4				YVDAWGQGLLVTVSS		
BLV1H12	CTSVHQ	ETKKYQ	SCPDGYRERSDCSNRPACGTSDCCRVSVFGNCLTTLPVSYSYTYNYEW	HVDVWGQGLLVTVSS	61	6
BLV5B8	CTTVHQ	ETRKT	CSDGYIAVDSCGRGQSDGCVNDCNSCYYGWRNCRRQPAIHSYEF	HVDAWGRGLLVTVSS	56	6
E03	CTTVHQ	KTLEVR	SCPDGSRLIGNDCRNEDGDDVNYITTFDYEW	YVDAWGQGLLVTVSS	44	2
B11	CVTVYQ	KTTQKK	NCPDDYTECYGGACDGTGCCSGSCGGASACRDWWPYRSICSSDNTYTYEF	HVDAWGQGLLVTVSS	63	8
A01	CTIVHQ	ETSRR 100	GPDGYSWICECSSGTYTCDADNCGNLCPSDWQLTLHCHRLDSSTYTYDW D2	HVETWGQGLRVTVSS 101	61	6

Fig. 3   See figure legend on opposite page.

IGHV1–7 region shifts the reading frame and encodes three additional amino acids (Fig. 3A). As the recombination signal sequence (RSS) immediately follows the inserted sequence, the frame shift has no further consequences on the protein sequence; the residues after CTTVHQ will be then encoded by junctional residues and IGHD8–2. In addition to the germline 8 bp duplication, IGHV1–7 also encodes amino acids in CDR H2, which interact with the ultralong CDR H3 stalk, and may play stabilizing roles. For example, AsnH50 in CDR H2 forms a hydrogen bond to GlnH97 in CDR H3. To summarize, bovine IGHV1–7 is unusual compared to other V_H regions in vertebrates in having a germline frameshift insertion which lengthens the V_H region and encodes a protruding β–ribbon, and additionally encodes potentially stabilizing residues that interact with the ultralong CDR H3.

The IGHD8–2 region is particularly unusual because of its long length; it encodes 48 amino acids (Fig. 3B) (Shojaei et al., 2003) that include the entire knob domain and a portion of the descending strand of the stalk. However, several other features are also apparent. A CPDG motif at the N-terminal end includes the first cysteine that is nearly always conserved in ultralong CDR H3 sequences (Fig. 3). This cysteine forms a disulfide bond that is spatially conserved, despite the fact that its bonding partner is not conserved in the germline (Stanfield et al., 2016). The CPDG motif also forms a β–turn, which leads into the first short β–strand that is conserved in the core of the

Fig. 3 Genetic basis for encoding ultralong CDR H3s. (A) A germline insertion in germline IGHV1–7. The nucleotide and encoded amino acid sequence of the 3′ end of a typical vertebrate IGVH region is shown above and IGHV1–7 shown below. The coding region nucleotides are in *uppercase letters*. The recombination signal sequence (RSS) is shown in *small letters*, with the heptamer and nonamer in *brown*. IGHV1–7, which is used in ultralong CDR H3 antibodies, has an 8 bp duplication in the germline (*boxed*). This duplication shifts the reading frame and extends the amino acid sequence, allowing the initial portion of the "stalk" β-strand to be encoded. (B) Nucleotide and amino acid sequence of the germline IGHD8–2 gene. Nucleotides which can be mutated to a residue that encodes a cysteine are shown in *red*, as are the corresponding amino acids. RGYW/WRCY mutational hotspots for AID are *boxed*. Note that a large fraction of codons of IGHD8–2 are mutational hotspots that can be mutated to cysteine. (C) Sequences of the five Fabs with ultralong CDR H3s whose structures have been determined. The germline regions of the IGHV1–7, IGHD8–2, and IGHJ2–4 are shown above, with the CDR H3 lengths and number of cysteines indicated to the *right*. The cysteines are highlighted in *yellow*, and the disulfide connectivities of the knob region are indicated. The cysteine and tryptophan, which define the boundaries of CDR H3, are highlighted in *yellow* and *cyan*, respectively. The amino acid numbers, using the system of Stanfield et al. are shown below (Stanfield, Wilson, & Smider, 2016).

knob domain (see later). The 3′ end of IGHD8–2 encodes several alternating tyrosines, which form stacking interactions at the C-terminus of CDR H3 and stabilize the descending strand of the stalk.

While we may infer that the knob encoded by the germline IGHD8–2 forms a distinct β-sheet core structure, as seen in five affinity-matured cow antibody structures (described in Section 3), this motif is not obviously encoded by IGHD8–2. Internally IGHD8–2 encodes a repeating pattern of GYG amino acids. This repeating motif is also seen in other bovine DH regions (though of shorter lengths) and is also found in other proteins in nature including keratins and prisilkin, which function in the extracellular matrix. In this regard, a BLAST search using multiple GYG repeats did not yield any obvious results suggesting a structure for the germline D_H; GYG-repeat containing proteins include a HET-domain containing protein (GenBank: EIW77936.1), an ankyrin-repeat domain containing protein (GenBank: EZA54628.1), keratins (e.g., GenBank: KFQ43969.1), the RNA-binding region of RNP-1 (GenBank: APR83217.1), the glycine/tyrosine rich eggshell protein (GenBank: ABF13207.1), and several pre-dicted and so-called hypothetical proteins. Thus, there currently is no significant insight into the role and importance of the GYG repeats in the germline structure of the knob. Although the small β-sheet has been seen in the crystallographic studies of five affinity-matured Fab fragments (Stanfield et al., 2016), the amino acid identities of the β-strands in these five structures are diverse, and the residues encoding the three β-strands cannot be readily mapped to the germline IGHD8–2 gene. Thus, the structural features encoded by IGHD8–2 in the knob region Gly-Tyr-Gly repeats are not entirely clear. IGHD8–2 does encode four cysteines, so presumably two disulfide bonds can form in the germline knob. The IGHJ2–4 region appears to be used in all bovine heavy chains (shorter and ultralong) and encodes the short β-strand at the end of CDR H3. This region is similar in sequence and structure to J_H regions of other vertebrates.

For ultralong CDR H3 antibodies, few germline light chains appear to be utilized in pairing with IGHV1–7 heavy chains. Adult bovines nearly exclusively use λ light chains (Arun, Breuer, & Hermanns, 1996; Pasman, Saini, Smith, & Kaushik, 2010), and the germline V30 (previously termed Vλx1) was used in the four original ultralong CDR H3 antibodies found by the Kaushik group (Saini, Farrugia, Ramsland, & Kaushik, 2003), as well as several of the recently described anti–gp140 SOSIP trimer antibodies (Sok et al., 2017). The germline V30 light chain has structural features (see later) that may optimize its utility in ultralong CDR H3 antibodies. This gene is

the most highly expressed V_L in the fetal repertoire (Liljavirta et al., 2014) and maintains expression throughout adult life (Pasman, Merico, & Kaushik, 2017). Interestingly, although λ light chains predominate in adults, kappa light chains were identified in up to 22% of the neonatal repertoire (Pasman et al., 2017), although it is unclear whether these kappa light chains can pair with heavy chains containing ultralong CDR H3s. Interestingly, IGHV1–7 and IGHD8–2 are also highly expressed in neonatal calves, suggesting that ultralong CDR H3 antibodies may be a predominant component of the neonatal repertoire (Pasman et al., 2017). The genetic locus of the λ light chain has been assembled and analyzed in some detail; approximately 17 functional Vλ genes were identified with preferential expression of the vλ1 family (Pasman et al., 2010), and particularly V30 (Liljavirta et al., 2014); however, detailed analysis of light chain usage and V_H–V_L pairing in the context of an immune response and ultralong CDR H3 heavy chains has not yet been examined.

2.3 Somatic Factors Impacting Diversity

In addition to encoding the CPDG β-turn motif and the descending β-strand of ultralong CDR H3s, the IGHD8–2 region has high mutagenic potential based on the number of RGYW (or in reverse, WRCY) hotspots (where R = A/G, W = A/T, Y = T/C). These sequences are substrates for the AID, which is the key enzyme involved in somatic hypermutation of antibody genes (Muramatsu et al., 2000). Indeed, there are 19 hotspots in the 148 bp IGHD8–2 region, covering nearly all of the protein coding sequence (Fig. 3). Furthermore, the codon usage of the Gly-Tyr-Gly repeating motif is (i) biased toward comprising a hotspot as well as (ii) being optimized for cysteine mutations. The codon encoding glycine is GGT, and for tyrosine is TAT. GGT is the most underrepresented (i.e., the rarest) of the four codons encoding glycine in the bovine genome, yet is used nearly exclusively to encode multiple glycines in the IGHD8–2 germline gene. Similarly, TAT is rarer than TAC, the other codon that encodes tyrosine in the genome. However, the Gly-Tyr-Gly repeats, encoded by GGTTATGGT invariably include the GGTT hotspot for AID. Remarkably, both GGT and TAT codons can be mutated to cysteine with just 1 bp change (both GGT and TAT to TGT). Thus, the IGHD8–2 gene appears to be "hard coded" for potential mutation, through multiple hotspots, and specifically toward mutation to cysteine. Of note, other shorter bovine D_H regions also have Gly-Tyr-Gly repeats and, although they do

not have the remarkable length as IGHD8–2, they may also have hotspots and proclivity to mutate to cysteine. Thus, these genetic characteristics may be general features of bovine germline CDR H3s.

2.4 Sequence Features of Matured Ultralong CDR H3s

Unlike humans or mice, several species appear to activate AID immediately following V(D)J recombination and utilize somatic hypermutation to create diversity for the primary immunoglobulin repertoire (Liljavirta et al., 2013; Reynaud, Garcia, Hein, & Weill, 1995; Yasuda, Jenne, Kennedy, & Reynolds, 2006; Zhao, Jackson, & Aitken, 2006). AID is highly expressed in bovine fetal spleen and Peyer's patch (prior to antigen exposure), and serves to diversify the primary repertoire in cows (Liljavirta et al., 2013). While this mechanism may appear fundamentally different compared to other vertebrates who primarily develop their B-cells in the bone marrow, species like bovines primarily develop B-cells in gut-associated lymphoid tissue, which appears to be an analog of the chicken's bursa of fabricius (Yasuda et al., 2006, 2002). Recently, however, evidence of AID activity was found in antigen naïve B-cells derived from the gut in well-controlled studies in transgenic mice (Yeap et al., 2015), suggesting that, in this compartment, AID-mediated diversification may indeed occur to diversify the primary repertoire in *Mus musculus*. Indeed, bovine ultralong CDR H3 sequences that are not fetally derived are very heavily mutated in their CDR H3 (Kaushik et al., 2009; Wang et al., 2013). Thus, a potential mechanism to reconcile these species differences is that the primary repertoire derived from gut B-cell development utilizes AID for primary diversification, whereas those B-cells derived from bone marrow development do not. The species differences, therefore, may be due to the relative proportion of the repertoire derived from these different compartments used for B-cell maturation. Of considerable note, the ability to mutate prior to antigen exposure dramatically increases the potential repertoire diversity and has implications in vaccinology where germline/naïve vs mutated sequences in the repertoire may be important (Yeap et al., 2015).

In cows, the ability to generate further diversity in the primary repertoire through somatic hypermutation is very important—in the ultralong subset of antibodies, only single V, D, and J genes appear to produce the entire repertoire. Even in shorter CDR H3 antibodies, the homology between the remaining V_H regions is so high that only limited diversity can be obtained through the use of alternative D_H regions. Thus, it appears that, in cattle, a
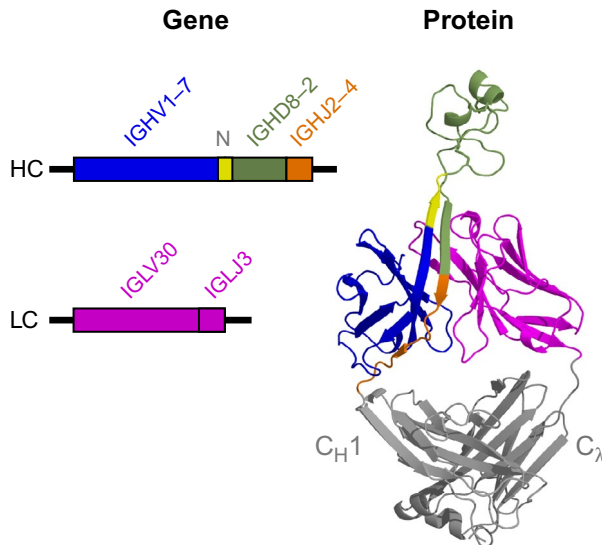
Fig. 4 Genetic components mapped onto the Fab structure of BLV5B8. The rearranged genes (*left*) and their corresponding structural components (*right*) are color coded (*blue*, IGHV1–7; *green*, IGHD8–2; *orange*, IGHJ2–4, *pink*, both IGLV30 and IGLJ3). In *yellow* is the area of V-D junction (labeled N) that encodes the ascending β-strand. This region is conserved and it is unclear whether it is derived from typical TdT catalyzed N-region insertions or other mechanisms like nucleotide capture. The C_H1 and C_λ regions are indicated on the *right*.

primary driver of antibody gene diversification is AID–mediated somatic mutation. When these ultralong CDR H3 antibody genes in the periphery are analyzed, they have unusual characteristics. First, several conserved nongermline-encoded residues resulting from junctional diversity at the V-D joint are observed (Figs. 3C and 4). As mentioned, there are typically 5–6 nontemplated amino acid residues in this region that encode the distal portion of the ascending β–strand of the stalk. Next, a nearly universally conserved cysteine in the knob is present. This cysteine, typically in the context of a CPDG motif, allowed for the derivation of a numbering system for bovine ultralong CDR H3 (Fig. 3C). Due to the exceptionally long length of these CDR H3 regions, standard numbering schemes like Kabat and Chothia (Al-Lazikani, Lesk, & Chothia, 1997; Kabat & Wu, 1991) do not easily apply (Stanfield et al., 2016). When mature sequences are aligned with this conserved cysteine (at position D2), other cysteines conserved with the germline IGHD8–2 region can be identified. Interestingly, it appears that, as one moves from the N- to C-terminal direction in CDR H3, the

conservation of the cysteines becomes increasingly less. Thus, the first germline-encoded cysteine is nearly completely conserved, whereas the fourth cysteine is the least conserved. Furthermore, presumably as a result of somatic hypermutation, nongermline-encoded cysteines are clearly evident and can be found throughout the CDR H3 knob region. As mentioned earlier, the Gly-Tyr-Gly repeating motif is comprised of an AID hotspot, and each of these residues can be mutated to cysteine with only one nucleotide change.

Although there is not extensive junctional diversity in ultralong CDR H3 transcripts, there is considerable length variation within CDR H3. Given the potential structural constraints imposed by the ascending and descending strands of the β-ribbon stalk, which are each encoded at a recombinational junction, other sources of length diversity have been analyzed (Deiss et al., 2017). Surprisingly, deletions internal to IGHD8–2 (e.g., not at the V-D or D-J junctions) were identified in nearly 50% of recombined mature ultralong CDR H3 transcripts. These deletions range from 3 to 54 bp (1–18 codons) (Deiss et al., 2017). While AID is known to induce deletions, the number of nucleotides deleted, as well as the unusually high frequency of deletions, seem unique to bovine ultralong CDR H3 sequences. These deletions presumably resulted from AID-mediated processes and could potentially be enhanced through the multiple repeats in IGHD8–2 that can allow DNA slippage events to occur. From the standpoint of a functional protein, and creation of diversity, deletions between cysteines potentially dramatically alter the surface loops on the knob, as well as the disulfide bonding patterns. Thus, in addition to SH providing an ability to alter amino acid content and mutation to and from cysteines, the feature of deleting residues can additionally alter protein secondary structure by changing loop length and disulfide bonds.

3. STRUCTURAL BIOLOGY OF ULTRALONG CDR H3 ANTIBODIES

Currently, five bovine Fab structures have been deposited to the Protein Data Bank (PDB), Fabs BLV1H12 (4K3D, 1.85 Å), BLV5B8 (4K3E, 2.2 Å) (Wang et al., 2013), B11 (5IHU, 2.06 Å), E03 (5IJV, 2.2 Å), and A01 (5ILT, 2.0 Å) (Stanfield et al., 2016). These Fabs contain CDR H3 regions that range in length from 44 to 63 residues, and that contain anywhere from 1 to 4 disulfides (Fig. 5 and Table 1). While the long CDR H3s all adopt the stalk and knob architecture, the other five CDR loops have
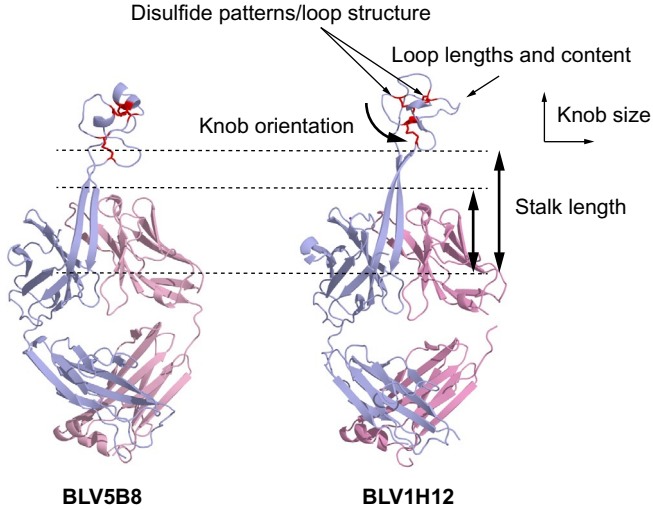
Fig. 5 Structural diversity in the bovine ultralong CDR H3 repertoire. These include diverse disulfide patterns, driven by cysteine mutations, different loop lengths, and amino acid content formed through AID point mutations as well as diversity in deletions, knob orientation, size, and stalk length.

Table 1 Diverse Structural Features of Ultralong CDR H3 Fabs

Fab (PDB ID)	CDR H3 AA Length	Protrusion Distance (Å)[a]	No. Cys (Disulfide Patterns)	Stalk Length (Residues Pairs)[b]
BLV1H12 (4K3D)	61	41	6 (1–4, 2–6, 3–5)	12
BLV5B8 (4K3E)	56	36	6 (1–3, 2–4, 5–6)	12
E03 (5IJV)	44	37	2 (1–2)	12
B11 (5IHU)	63	36	8 (1–4, 2–7, 3–8, 5–6)	8
A01 (5ILT)	61	32	6 (1–4, 2–5, 3–6)	8

[a]Approximate distance from the tips of CDRs L1–L3, H1, H2 to the distal tip of the H3 knob.
[b]Length of hydrogen-bonded stalk domain in residues pairs (the number of amino acids forming the two–stranded β–ribbon divided by two).

limited sequence diversity and have structures largely consistent with the canonical structures seen previously in human and murine Fabs (Al-Lazikani et al., 1997; Chothia & Lesk, 1987; Chothia et al., 1992), except for CDR L3 that has a bend in its tip to accommodate and likely support
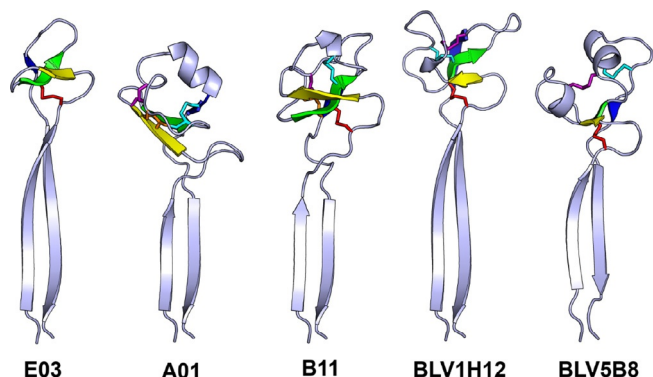
Fig. 6 The stalk and knob domains of the five currently known bovine Fab ultralong CDR H3 regions. The three strands in the core knob domain are colored *yellow*, *green*, and *blue*, and the disulfide bonds are shown as *sticks*. The *orange* disulfide is always formed from the first Cys in the CPDG motif to a Cys in the second core β-strand.

the C-terminal strand of the stalk. The five currently determined structures all contain the same λ light chain (germline V30).

The knob domains all contain three short antiparallel β-strands, connected by two loops of variable length and secondary structure (Fig. 6). Each knob domain is initiated by a type-I β turn at the "CPDG" sequence motif (C/G P/S D G/XD in the five current structures) encoded by the 5′ end of the D_H gene. The D and G are the $i+1$ and $i+2$ residues in the β-turn. While this sequence motif is usually recognizable, there is no conserved sequence within the rest of the knob domains, and even the three β-strands vary in length and sequence. There is one spatially conserved disulfide from the first Cys residue in the CPDG motif to a Cys located on the second β-strand; however, the number of residues between this disulfide-bonded pair can vary. Fab A02 lacks the highly conserved Cys in the CPDG motif, but nevertheless maintains the type I β-turn found in the other knob domains. The other disulfide bonds in the knob domain occur at different positions and with different disulfide connectivities (Table 1). The knob domains have limited structural homology with the small cyclic peptide Kalata B1 (PDB: 1N1U, Daly, Clark, & Craik, 2003), WW domain FE65 (PDB: 2HO2, Meiyappan, Birrane, & Ladias, 2007), and part of the tumor necrosis factor receptor superfamily member 13C (PDB: 1OQE, Liu et al., 2003).

The β-ribbon stalk regions can also vary in length. E03 and BLV1H12 have long, regular β-ribbon stalks with 12 residues in each strand (e.g., 12 residue pairs; Table 1), but A01, B11, and BLV5B8 have shorter
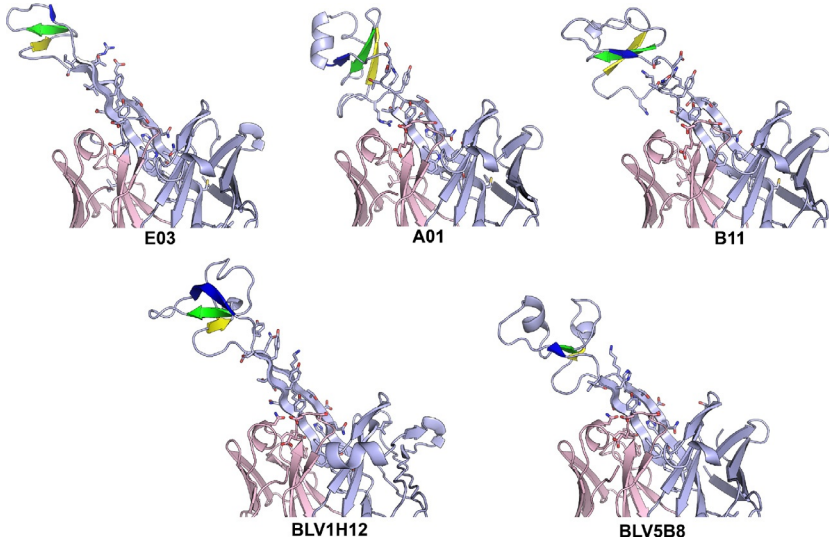
Fig. 7 The bovine stalk and surrounding CDR residues. The Fab heavy and light chains are shown in *light blue* and *pink,* respectively. The stalk regions all have very similar take-off angles from the body of the Fabs and appear to be structurally stable. Ladders of alternating aromatic residues may add to the stabilization of the stalks. The non-H3 CDR loops are well conserved and may also aid in support for the long stalk.

β-ribbons of 8 residues on each side, with the B11 ribbon topped by short non-β linkers to the knob (Fig. 6). The descending strands of the stalks generally contain stacking hydrophobic residues, likely adding stability to the stalk (Fig. 7). Superposition of the Fab variable heavy chain domains shows that the stalk/knob units all extend at approximately the same angle from the Fabs, suggesting a fairly high degree of rigidity for the stalk. But the knobs can twist by as much as almost 60 degrees with respect to their stalks, so that superposition of the stalks does not then superpose the knobs and vice versa. The combination of differing ribbon length and knob position results in different overall lengths for the stalk/knob units with respect to the antibody surface. The distance between the top of the knob domain to a plane roughly bisecting the tips of the other five CDR loops ranges between 32 and 41 Å (Table 1). Residue conservation is found at the N-terminal base of the stalk (C, T/V, T/S/I, V, H/Y, Q, E/K) and in the descending strand in the ladder of alternating aromatic residues. Along with several hydrogen bonds within the stalk regions, the surrounding CDR residues may play a role in the support and stabilization of the stalk domain, especially the descending strand of the stalk (Fig. 7). The tip of

CDR L3 (AEDSS, residues L92–95) stacks against a Tyr residue in the descending strand, 7 residues before the conserved TrpH103 (the CDR H3's are too large for conventional Kabat numbering, so we will describe residue positions with respect to conserved residue TrpH103, as described in Stanfield et al. (2016). The main chain of this Tyr also hydrogen bonds to AsnL30. TyrL32 stacks against a Glu/Asp found 6 residues before H103 and TyrL49 stacks against a His/Tyr found 4 residues upstream of H103. Thus, there are key residues both within CDR H3, and in other CDRs which appear to provide structural support to the unusual architecture of the heavy chain ultralong CDR H3.

4. BIOCHEMISTRY AND POTENTIAL FUNCTION OF ULTRALONG CDR H3 ANTIBODIES

4.1 Antigen Binding

Unlike any other vertebrates examined, cows are able to mount a rapid broadly neutralizing antibody response against the stable engineered HIV gp140 SOSIP Env trimer (Sok et al., 2017). Several cow monoclonal antibodies were isolated and their epitopes mapped to the CD4 binding site of the gp120 component of the Env glycoprotein. Of note, all of the antibodies reported in this study had ultralong CDR H3s of 60 or 61 amino acids. One antibody termed NC-Cow1 was able to neutralize 72% of the strains in a 117 strain panel at high potency. The likely mechanism for the neutralization of these antibodies is that their long CDR H3 is able to navigate through the glycans in the glycan shield surrounding the CD4 binding site to contact conserved the protein epitopes on the glycoprotein that interact with CD4 receptor. Indeed, electron microscopy was able to visualize the long CDR H3 binding to the CD4 binding site, without evidence of direct antigen contact by the other CDRs. In this regard, current evidence suggests that some of the other five CDRs, as opposed to contacting antigen, play a structural role through supporting the conformation of the long CDR H3 loop at its base. In addition to the direct EM evidence with NC-Cow1, an antibody targeting bovine viral diarrheal virus (BVDV) was identified and removal of its knob domain completely abrogated binding to the virus (Wang et al., 2013). Alanine-scanning mutagenesis of the entire CDR H3 was able to identify the key residues involved in binding. Although these studies do not rule out the possibility that some cow ultralong CDR H3 antibodies may be found that have binding contributions from the other CDRs, the fact that little diversity is seen in the other CDRs

(Deiss et al., 2017), with only a single V_H region and utilization of a limited number of light chains, which are often in nearly germline configuration, suggests that the main driver of antigen binding is the ultralong CDR H3. Further crystallographic information on a diversity of antigens with this unusual antibody class will shed more light on their binding and recognition mechanisms at the atomic level.

A key feature of any antibody is its affinity and specificity. While early experiments with the ultralong antibodies derived from BLV-infected B-cells suggested that some ultralong CDR H3 antibodies had polyspecific binding properties (Saini et al., 1999), ELISAs and cell-based binding assays of the anti-HIV and anti-BVDV monoclonal antibodies show that these particular monoclonal antibodies bind with high specificity to their cognate antigens (Sok et al., 2017; Wang et al., 2013). Neutralization potency of antibodies targeting HIV BG505 SOSIP suggests their binding affinities are in the low nanomolar range. Indeed, surface plasmon resonance analysis showed subnanomolar K_ds for NS-Cow1 and other anti-HIV cow antibodies (V. Smider and M. Vadnais, unpublished). Thus, the unusual antibodies with ultralong CDR H3s derived from cattle appear to have similar antigen binding characteristics in terms of specificity and affinity as other antibodies derived by affinity maturation from other vertebrate species.

4.2 Potential Function and Evolution

It is presently unclear why bovines have evolved the ultralong CDR H3 system, which differs dramatically from other vertebrate repertoires at both the genetic and structural levels. At least two germline genetic events occurred during evolution to enable creation of this system: (i) the in-frame 8 bp duplication at the 3′ end of IGHV1–7 and (ii) the generation of the ultralong IGHD8–2 diversity gene. On one hand, it is possible that this system developed as an alternative mechanism to create structural diversity in the repertoire. In this vein, whereas multiple V_H, D_H, and J_H segments can form enormous combinatorial diversity in some organisms, bovines, and perhaps other mammals, may have minimized combinatorial diversity in favor of a mechanism whereby structural diversity is created prior to antigen exposure through AID-driven processes that alter loops and structures by varying the disulfide connectivities and amino acid content of the knob domain. An alternative hypothesis, however, is that specific microorganisms drove evolution of this system. The digestive physiology of cows is distinct in utilizing

a four-chambered stomach, with the rumen compartment acting as a large fermenter (80 L for a full grown cow). In order to process cellulose and other feedstuff, the rumen contains a highly diverse and symbiotic mixture of both prokaryotic and eukaryotic microorganisms, particularly bacteria and protozoa (Bryant, 1959; Hungate, 1975; Hungate, Bryant, & Mah, 1964; Mao, Zhang, Liu, & Zhu, 2015; Sharpe, Latham, & Reiter, 1969), of which the latter may present an unusual antigenic burden to which other vertebrates are not typically exposed. Indeed, cows must contain these microorganisms within the rumen compartment, as dissemination of these microorganisms via the blood or tissue can cause life-threatening disease (Miesner & Reppert, 2017; Ward & Ducharme, 1994; Zinicola et al., 2015). Serum from cattle is known to possess antibodies reactive to rumen microorganisms, suggesting that an immune response against their rumen antigens is common (Sharpe et al., 1969).

An alternative hypothesis is that certain infections (as opposed to rumen symbionts) have driven development of the unusual architecture of bovine ultralong CDR H3s. As a species, cows are exposed to several retroviruses including BLV, which can cause substantial morbidity within a herd, bovine syncytial virus (BSV), and bovine immunodeficiency-like virus (BIV), which is analogous to HIV but only causes mild disease in bovines and does not lead to acquired immunodeficiency syndrome (AIDS) (Carpenter et al., 1992; Flaming et al., 1993; Gonda et al., 1990; Jacobs, Smith, Gregory, Valli, & Whetstone, 1992; Onuma et al., 1992). A recent observation in studying broadly neutralizing antibodies to certain human viruses was that these antibodies often had long CDR H3s; human antibodies typically average 12–15 residues, whereas in broadly neutralizing antibodies against HIV, the CDR H3 can be >30 amino acids (reviewed in Burton & Hangartner, 2016; Burton & Mascola, 2015).

In light of the importance of these rare human antibodies, the HIV immunization experiments described earlier were performed. Four cows were immunized with a soluble, native-like gp140 Env trimer (BG505 SOSIP), and the neutralizing antibody response was impressive (Sok et al., 2017). Immunization of one cow led to a remarkable 96% serum neutralization breadth (of 117 cross-clade isolates), with isolation of the monoclonal antibody NS-Cow1 with an ultralong CDR H3 of 60 amino acids that could potently neutralize 72% of cross-clade isolates with an IC_{50} of 0.028 μg/mL. Broad neutralization properties developed rapidly; breadth (20%) could even be observed after a single prime and homologous boost. Such responses are unprecedented in animal immunization against

HIV, and could make cows a novel model for evaluating vaccine candidates because of their unusual, and potentially optimized, repertoire for such types of antigen. Of note, cross-reactive immune responses between HIV and BIV were noted decades ago (Jacobs et al., 1992), as was crossreactivity of serum antibodies against different bovine retroviruses. Thus, a potential hypothesis for evolution of the unusual ultralong CDR H3 cow antibody repertoire is that it was selected to combat certain classes of viruses, and potentially retroviruses, as a biologically economic means to cross-react against several related and perhaps even poorly related strains.

5. PERSPECTIVES

One of the most fundamental questions in the history of immunology was how a limited amount of genetic material in the genome could encode an apparently limitless array of antibodies in its repertoire to protect the host from the vast numbers of possible infections as well as cancer. The discovery of V(D)J recombination several decades ago (Hozumi & Tonegawa, 1976) and the resulting understanding of the combinatorial nature of antibody diversity development provided a fundamental and compelling answer to this question. Humans, for example, have a variety of variable light- and heavy-chain germline genes that can recombine with different (D) and J regions. The ultralong CDR H3 antibody system of cows, however, appears to utilize a different mechanism to achieve similar ends (Fig. 8). Only a single V-D-J recombination event leads to the precursor "naïve" antibody molecule. This naïve molecule has the motifs associated with "stalk" and "knob" formation with four germline-encoded cysteines that presumably form two disulfide bonds. The V-D and D-J junctions can provide limited junctional diversity (as occurs in other mammalian systems), which may alter the length of the "stalk" domain. This single genetic unit then undergoes somatic hypermutation that can potentially "reshape" the knob domain through (i) amino acid point mutations; (ii) mutations to and from cysteine, altering loop structures and disulfide connectivities; and (iii) nucleotide deletions which can further alter loop lengths, as well as disulfide bond patterns (Fig. 8). While detailed molecular and cellular studies, including the tissue compartment(s) involved (such as bone marrow vs Peyer's patch) or the temporal nature of these events, have not been performed, all of these events can be mediated by AID, which serves to shape the primary repertoire (before antigen exposure) in bovines. Thus, although cows perform V(D)
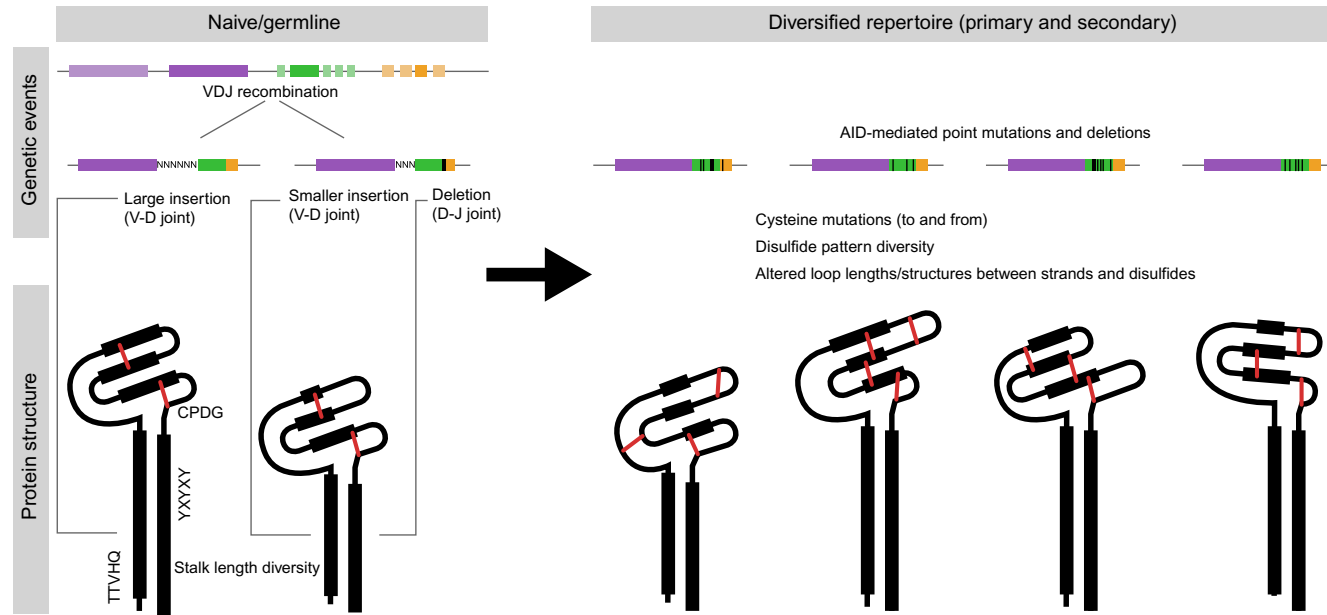
Fig. 8 Genetic and structural model for ultralong CDR H3 antibody repertoire formation. One or more germline V-D-J recombination events using IGHV1–7 (*dark purple*), IGHD8–2 (*dark green*), and IGHJ2–4 (*dark orange*) is shown in the *upper left*. Junctional diversity, including insertions and deletions, can potentially alter the length of the stalk, as the V-D and D-J joints occur in regions encoding the ascending and descending strands, respectively. Presumably, two disulfide bonds can form in the knob of the naïve/germline antibody (shown in *red*). The primary repertoire is further diversified in the gut-associated lymphoid tissue by AID-mediated processes (*upper right*). Nucleotide changes that alter amino acid content, particularly to cysteines, can change disulfide patterns and loops (*red lines*). Larger deletions (*black rectangles*) in the D$_H$ region can also potentially change loop lengths and disulfide patterns. The β-strands of the knob core and stalk domain are represented by *black arrows*.

J recombination, the primary driver of genetic and ultimately structural diversity may be somatic mutation by the AID enzyme.

From a structural view, it is intriguing that a single germline protein can apparently evolve, through mutation, into a myriad of different disulfide-bonded "knob" minidomains. In this regard, it has not yet been determined whether the germline precursor antibody can adopt one or many conformations in its knob domain, or whether it possesses a defined (single) structure. Crystal structures of five affinity-matured Fab fragments show a conserved short β-sheet core; however, neither the amino acid sequences of the β-strands nor their linear positions are conserved across different knob sequences. Thus, although it seems likely that the β-sheet core should be encoded in the germline, the details regarding its formation are unclear. Thus, a paradox appears to exist, while this presumably important β-sheet feature is conserved between several affinity-matured knobs, the genetic basis for this conserved structure is not readily apparent in the germline IGHD8–2 DNA. In contrast, the underlying genetic basis for the other conserved structural motifs of the ultralong CDR H3, including the ascending and descending strands of the stalk, and β-turn of the knob, is readily identifiable in IGHV1–7 or IGHD8–2. The germline D_H region encodes the β-turn, the four cysteines with intervening Gly-Tyr-Gly repeats, and the descending strand of the stalk. With regard to the cysteines, while we presume that two disulfide bonds form in the germline structure, there are actual several theoretical possibilities. These two disulfide bonds may adopt one of three different configurations (e.g., 1–2, 3–4; 1–3, 2–4; or 1–4, 2–3 patterns). Alternatively, it is possible that some cysteines are unpaired; there are six configurations of four cysteines where one disulfide bond forms but the remaining two cysteines are unpaired. A final possibility is that all four cysteines are unpaired, and that disulfides only form during the SH process, when the knob is fully folded or adopts an alternate conformation. In the context of these possibilities, the Gly-Tyr-Gly repeats are likely important, as either (i) the Gly-Tyr-Gly repeats can form a β-sheet core allowing at least the first highly conserved cysteine to form a disulfide bond, or (ii) perhaps less likely, the knob itself is partially stabilized through formation of the disulfide bonds and does not form a β-sheet core until the requisite mutations from AID occur within the Gly-Tyr-Gly regions. An intriguing question related to the first possibility (discussed later) is whether more than one β-sheet structure/disulfide combination could exist in the germline, which would enable further diversity in the structural repertoire.

The structure of the germline knob should give further insights into the generation of diversity in the ultralong CDR H3 repertoire. The germline knob encoded by the IGHD8–2 region is not obviously intrinsically disordered; in fact, predictive algorithms like IUPred or DisEMBL (http://iupred.enzim.hu, and http://dis.embl.de) score the germline D_H as more likely to be highly ordered. Coupled with the likelihood of constraints on the N- and C-termini on the knob from the close proximity to the β-strands forming the stalk, as well as two probable disulfide bonds, this suggests that one or more stable structures of the germline knob may exist. Indeed, some germline CDRs have been reported to be able to form more than one conformation for the same sequence (Wedemayer, Patten, Wang, Schultz, & Stevens, 1997; Yin, Beuscher, Andryski, Stevens, & Schultz, 2003). This "structural plasticity" has been suggested to provide added diversity to the naïve repertoire and explain the tendency for germline antibodies to have oligospecific or polyspecific antigen binding properties. Furthermore, germline CDR H3s have been shown to be flexible compared to their affinity-matured counterparts in certain cases that have been examined (Adhikary et al., 2015; Jimenez, Salazar, Yin, Joo, & Romesberg, 2004; Thielges, Zimmermann, Yu, Oda, & Romesberg, 2008; Zimmermann et al., 2006). It has been postulated that germline CDR amino acid content has been evolved for structural flexibility (Babor & Kortemme, 2009), and "induced fit" of the antibody with antigen has been proposed (Rini, Schulze-Gahmen, & Wilson, 1992; Stanfield, Dooley, Verdino, Flajnik, & Wilson, 2007). CDR H3s in all species, and their corresponding germline D_H regions, are overrepresented by tyrosine and glycine (Babor & Kortemme, 2009; Birtalan et al., 2008; Fellouse, Wiesmann, & Sidhu, 2004). Thus, if the knob region encoded by bovine IGHD8–2 behaves like other germline CDR H3s, it may have substantial flexibility. The presence of four cysteines, and the potential to form different disulfide patterns, could "lock in" different knob loop patterns as part of the repertoire during somatic hypermutation.

The recent discovery of two IgM genes in cattle is interesting in that ultralong CDR H3 antibodies appear to utilize IgM2 (Ma et al., 2016). Each IGHM gene has upstream V, D, and J sequences, which appear to rearrange and splice to IGHM1 or IGHM2. Thus, while ultralong CDR H3 antibodies use IGHM2, other V-D-J rearrangements with shorter CDR H3s can also use this isotype. No mammalian species other than bovines have been reported to have more than one IGHM gene; however, crocodilians interestingly have three (Cheng et al., 2013). This duplicative arrangement of gene segments suggests that independent V-D-J events could occur in a single B-cell, resulting in expression of two heavy chains; however, this has

not been investigated. Furthermore, both IGHM genes have μ switch regions, suggesting that a V–D–J recombination event upstream of IGHM1 could class switch to IGHM2. Indeed, Ma et al. determined that class switch recombination to IGHM2 from IGHM1 does occur in splenocytes. Of course, during an immune response, class switch recombination from either IGHM1 or IGHM2 to downstream γ, ε, or α constant region genes would also occur. Quantitative PCR of IGHM1 vs IGHM2 gene expression from a variety of lymphoid tissues at different developmental stages did not show any tissue-specific patterns for the two genes, although IGHM2 had slightly higher expression levels in all tissues examined (Ma et al., 2016). Thus, the evolutionary significance of two IGHM genes and their relationship to the unusual ultralong CDR H3 antibodies of cattle remains a mystery.

In conclusion, the ultralong CDR H3 subset of bovine antibodies illuminates unexpected mechanisms to produce a structurally diverse immune repertoire. These antibodies contain CDR H3s that project out far from the typical antibody paratope using a β-ribbon stalk and a disulfide-bonded knob containing a small conserved β-sheet core. The knob loops and disulfide patterns are quite diverse in sequence and structure, and we are just beginning to understand the genesis of this diversity that appears to occur through heavy SH that includes significant deletions. The intrigue of these antibodies is heightened by cow immunization experiments with the Envelope glycoprotein of HIV, where potent broadly neutralizing antibodies with ultralong CDR H3s were rapidly induced—a finding so far not observed in any other species by immunization with the HIV Env antigen. The potential of cow antibodies to penetrate through glycan-coated antigens in viruses (as in HIV), or bind to cryptic or recessed epitopes, such as those in enzymes, pores, or channels, is likely to be the subject of further investigation, as antibodies to these types of epitopes could have important applications in research and medicine.

ACKNOWLEDGMENTS

REFERENCES

Adhikary, R., Yu, W., Oda, M., Walker, R. C., Chen, T., Stanfield, R. L., et al. (2015). Adaptive mutations alter antibody structure and dynamics during affinity maturation. *Biochemistry*, *54*, 2085.

Al-Lazikani, B., Lesk, A. M., & Chothia, C. (1997). Standard conformations for the canonical structures of immunoglobulins. *Journal of Molecular Biology*, *273*, 927.

Anderson, D., Billingham, R. E., Lampkin, G. H., & Medawar, P. B. (1951). The use of skin grafting to distinguish between monozygotic and dizygotic twins in cattle. *Heredity*, *5*, 379.

Arun, S. S., Breuer, W., & Hermanns, W. (1996). Immunohistochemical examination of light–chain expression (lambda/kappa ratio) in canine, feline, equine, bovine and porcine plasma cells. *Zentralblatt für Veterinärmedizin. Reihe A*, *43*, 573.

Babor, M., & Kortemme, T. (2009). Multi-constraint computational design suggests that native sequences of germline antibody H3 loops are nearly optimal for conformational flexibility. *Proteins*, *75*, 846.

Baxby, D. (1999). Edward Jenner's inquiry; a bicentenary analysis. *Vaccine*, *17*, 301.

Berens, S. J., Wylie, D. E., & Lopez, O. J. (1997). Use of a single V_H family and long CDR3s in the variable region of cattle Ig heavy chains. *International Immunology*, *9*, 189.

Billingham, R. E., & Brent, L. (1956). Acquired tolerance of foreign cells in newborn animals. *Proceedings of the Royal Society of London—Series B: Biological Sciences*, *146*, 78.

Billingham, R. E., Brent, L., & Medawar, P. B. (1953). Actively acquired tolerance of foreign cells. *Nature*, *172*, 603.

Birtalan, S., Zhang, Y., Fellouse, F. A., Shao, L., Schaefer, G., & Sidhu, S. S. (2008). The intrinsic contributions of tyrosine, serine, glycine and arginine to the affinity and specificity of antibodies. *Journal of Molecular Biology*, *377*, 1518.

Brown, W. R., Rabbani, H., Butler, J. E., & Hammarstrom, L. (1997). Characterization of the bovine Cα gene. *Immunology*, *91*(1), 1–6.

Bryant, M. P. (1959). Bacterial species of the rumen. *Bacteriological Reviews*, *23*, 125.

Burton, D. R., & Hangartner, L. (2016). Broadly neutralizing antibodies to HIV and their role in vaccine design. *Annual Review of Immunology*, *34*, 635.

Burton, D. R., & Mascola, J. R. (2015). Antibody responses to envelope glycoproteins in HIV-1 infection. *Nature Immunology*, *16*, 571.

Carpenter, S., Miller, L. D., Alexandersen, S., Whetstone, C. A., VanDerMaaten, M. J., Viuff, B., et al. (1992). Characterization of early pathogenic effects after experimental infection of calves with bovine immunodeficiency-like virus. *Journal of Virology*, *66*, 1074.

Cheng, G., Gao, Y., Wang, T., Sun, Y., Wei, Z., Li, L., et al. (2013). Extensive diversification of IgH subclass-encoding genes and IgM subclass switching in crocodilians. *Nature Communications*, *4*, 1337.

Chothia, C., & Lesk, A. M. (1987). Canonical structures for the hypervariable regions of immunoglobulins. *Journal of Molecular Biology*, *196*, 901.

Chothia, C., Lesk, A. M., Gherardi, E., Tomlinson, I. M., Walter, G., Marks, J. D., et al. (1992). Structural repertoire of the human V_H segments. *Journal of Molecular Biology*, *227*, 799.

Daly, N. L., Clark, R. J., & Craik, D. J. (2003). Disulfide folding pathways of cystine knot proteins. Tying the knot within the circular backbone of the cyclotides. *The Journal of Biological Chemistry*, *278*, 6314.

Deiss, T. C., Vadnais, M. L., Wang, F., Chen, P. L., Torkamani, A., Mwangi, W., et al. (2017). Immunogenetic factors driving formation of ultralong V_H CDR3 in *Bos taurus* antibodies. *Cellular & Molecular Immunology*, *14*, 1–12.

DiPietro, L. A., Sethupathi, P., Kingzette, M., Zhai, S. K., Suter, M., & Knight, K. L. (1992). Limited repertoire of utilized V_H gene segments in a VHa3-allotype-suppressed rabbit. *International Immunology*, *4*, 555.

Eguchi-Ogawa, T., Wertz, N., Sun, X. Z., Piumi, F., Uenishi, H., Wells, K., et al. (2010). Antibody repertoire development in fetal and neonatal piglets. XI. The relationship of variable heavy chain gene usage and the genomic organization of the variable heavy chain locus. *Journal of Immunology*, *184*, 3734.

Fellouse, F. A., Wiesmann, C., & Sidhu, S. S. (2004). Synthetic antibodies from a four-amino-acid code: A dominant role for tyrosine in antigen recognition. *Proceedings of the National Academy of Sciences of the United States of America*, *101*, 12467.

Flaming, K., van der Maaten, M., Whetstone, C., Carpenter, S., Frank, D., & Roth, J. (1993). Effect of bovine immunodeficiency-like virus infection on immune function in experimentally infected cattle. *Veterinary Immunology and Immunopathology, 36*, 91.

Gonda, M. A., Oberste, M. S., Garvey, K. J., Pallansch, L. A., Battles, J. K., Pifat, D. Y., et al. (1990). Development of the bovine immunodeficiency-like virus as a model of lentivirus disease. *Developments in Biological Standardization, 72*, 97.

Hayes, H. C., & Petit, E. J. (1993). Mapping of the β–lactoglobulin gene and of an immunoglobulin m heavy chain-like sequence to homoeologous cattle, sheep, and goat chromosomes. *Mammalian Genome, 4*, 207.

Hosseini, A., Campbell, G., Prorocic, M., & Aitken, R. (2004). Duplicated copies of the bovine J_H locus contribute to the Ig repertoire. *International Immunology, 16*, 843.

Hozumi, N., & Tonegawa, S. (1976). Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions. *Proceedings of the National Academy of Sciences of the United States of America, 73*, 3628.

Hungate, R. E. (1975). The rumen microbial ecosystem. *Annual Review of Ecology and Systematics, 6*, 39.

Hungate, R. E., Bryant, M. P., & Mah, R. A. (1964). The rumen bacteria and protozoa. *Annual Review of Microbiology, 18*, 131.

Jacobs, R. M., Smith, H. E., Gregory, B., Valli, V. E., & Whetstone, C. A. (1992). Detection of multiple retroviral infections in cattle and cross-reactivity of bovine immunodeficiency-like virus and human immunodeficiency virus type 1 proteins using bovine and human sera in a western blot assay. *Canadian Journal of Veterinary Research, 56*, 353.

Jimenez, R., Salazar, G., Yin, J., Joo, T., & Romesberg, F. E. (2004). Protein dynamics and the immunological evolution of molecular recognition. *Proceedings of the National Academy of Sciences of the United States of America, 101*, 3803.

Jung, D., Giallourakis, C., Mostoslavsky, R., & Alt, F. W. (2006). Mechanism and control of V(D)J recombination at the immunoglobulin heavy chain locus. *Annual Review of Immunology, 24*, 541.

Kabat, E. A., & Wu, T. T. (1991). Identical V region amino acid sequences and segments of sequences in antibodies of different specificities. Relative contributions of V_H and V_L genes, minigenes, and complementarity-determining regions to binding of antibody-combining sites. *Journal of Immunology, 147*, 1709.

Kacskovics, I., & Butler, J. E. (1996). The heterogeneity of bovine IgG2—VIII. The complete cDNA sequence of bovine IgG2a (a2) and an IgG1. *Molecular Immunology, 33*, 189.

Kaushik, A. K., Kehrli, M. E., Jr., Kurtz, A., Ng, S., Koti, M., Shojaei, F., et al. (2009). Somatic hypermutations and isotype restricted exceptionally long CDR3H contribute to antibody diversification in cattle. *Veterinary Immunology and Immunopathology, 127*, 106.

Kaushik, A., Shojaei, F., & Saini, S. S. (2002). Novel insight into antibody diversification from cattle. *Veterinary Immunology and Immunopathology, 87*, 347.

Koti, M., Kataeva, G., & Kaushik, A. (2008). Organization of D_H–gene locus is distinct in cattle. *Developmental Biology, 132*, 307.

Koti, M., Kataeva, G., & Kaushik, A. K. (2010). Novel atypical nucleotide insertions specifically at V_H-D_H junction generate exceptionally long CDR3H in cattle antibodies. *Molecular Immunology, 47*, 2119.

Kuroiwa, Y., Kasinathan, P., Sathiyaseelan, T., Jiao, J. A., Matsushita, H., Sathiyaseelan, J., et al. (2009). Antigen-specific human polyclonal antibodies from hyperimmunized cattle. *Nature Biotechnology, 27*, 173.

Liljavirta, J., Ekman, A., Knight, J. S., Pernthaner, A., Iivanainen, A., & Niku, M. (2013). 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 Immunology, 6*, 942.

Liljavirta, J., Niku, M., Pessa-Morikawa, T., Ekman, A., & Iivanainen, A. (2014). Expansion of the preimmune antibody repertoire by junctional diversity in *Bos taurus*. *PLoS One*, *9*, e99808.

Liu, Y., Hong, X., Kappler, J., Jiang, L., Zhang, R., Xu, L., et al. (2003). Ligand-receptor binding revealed by the TNF family member Tall-1. *Nature*, *423*, 49.

Lopez, O., Perez, C., & Wylie, D. (1998). A single V$_H$ family and long CDR3s are the targets for hypermutation in bovine immunoglobulin heavy chains. *Immunological Reviews*, *162*, 55.

Ma, L., Qin, T., Chu, D., Cheng, X., Wang, J., Wang, X., et al. (2016). Internal duplications of D$_H$, J$_H$, and C region genes create an unusual IgH gene locus in cattle. *Journal of Immunology*, *196*, 4358.

Mao, S., Zhang, M., Liu, J., & Zhu, W. (2015). Characterising the bacterial microbiota across the gastrointestinal tracts of dairy cattle: Membership and potential function. *Scientific Reports*, *5*, 16116.

Meiyappan, M., Birrane, G., & Ladias, J. A. (2007). Structural basis for polyproline recognition by the FE65 WW domain. *Journal of Molecular Biology*, *372*, 970.

Miesner, M. D., & Reppert, E. J. (2017). Diagnosis and treatment of hardware disease. *The Veterinary Clinics of North America. Food Animal Practice*, *33*, 513–523. pii: S0749.

Mousavi, M., Rabbani, H., & Hammarstrom, L. (1997). Characterization of the bovine epsilon gene. *Immunology*, *92*, 369.

Mousavi, M., Rabbani, H., Pilstrom, L., & Hammarstrom, L. (1998). Characterization of the gene for the membrane and secretory form of the IgM heavy-chain constant region gene (Cμ) of the cow (*Bos taurus*). *Immunology*, *93*, 581.

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

Niku, M., Liljavirta, J., Durkin, K., Schroderus, E., & Iivanainen, A. (2012). The bovine genomic DNA sequence data reveal three IGHV subgroups, only one of which is functionally expressed. *Developmental and Comparative Immunology*, *37*, 457.

Onuma, M., Koomoto, E., Furuyama, H., Yasutomi, Y., Taniyama, H., Iwai, H., et al. (1992). Infection and dysfunction of monocytes induced by experimental inoculation of calves with bovine immunodeficiency-like virus. *Journal of Acquired Immune Deficiency Syndromes*, *5*, 1009.

Owen, R. D. (1945). Immunogenetic consequences of vascular anastomoses between bovine twins. *Science*, *102*, 400.

Pasman, Y., Merico, D., & Kaushik, A. K. (2017). Preferential expression of IGHV and IGHD encoding antibodies with exceptionally long CDR3H and a rapid global shift in transcriptome characterizes development of bovine neonatal immunity. *Developmental and Comparative Immunology*, *67*, 495.

Pasman, Y., Saini, S. S., Smith, E., & Kaushik, A. K. (2010). Organization and genomic complexity of bovine λ-light chain gene locus. *Veterinary Immunology and Immunopathology*, *135*, 306.

Rabbani, H., Brown, W. R., Butler, J. E., & Hammarstrom, L. (1997). Polymorphism of the IGHG3 gene in cattle. *Immunogenetics*, *46*, 326.

Reynaud, C. A., Garcia, C., Hein, W. R., & Weill, J. C. (1995). Hypermutation generating the sheep immunoglobulin repertoire is an antigen-independent process. *Cell*, *80*, 115.

Rini, J. M., Schulze-Gahmen, U., & Wilson, I. A. (1992). Structural evidence for induced fit as a mechanism for antibody-antigen recognition. *Science*, *255*, 959.

Saini, S. S., Allore, B., Jacobs, R. M., & Kaushik, A. (1999). Exceptionally long CDR3H region with multiple cysteine residues in functional bovine IgM antibodies. *European Journal of Immunology*, *29*, 2420.

Saini, S. S., Farrugia, W., Ramsland, P. A., & Kaushik, A. K. (2003). Bovine IgM antibodies with exceptionally long complementarity-determining region 3 of the heavy chain share unique structural properties conferring restricted $V_H + V_\lambda$ pairings. *International Immunology*, *15*, 845.

Saini, S. S., Hein, W. R., & Kaushik, A. (1997). A single predominantly expressed polymorphic immunoglobulin V_H gene family, related to mammalian group, I, clan, II, is identified in cattle. *Molecular Immunology*, *34*, 641.

Sharpe, M. E., Latham, M. J., & Reiter, B. (1969). The occurrence of natural antibodies to rumen bacteria. *Journal of General Microbiology*, *56*, 353.

Shojaei, F., Saini, S. S., & Kaushik, A. K. (2003). Unusually long germline D_H genes contribute to large sized CDR3H in bovine antibodies. *Molecular Immunology*, *40*, 61.

Sok, D., Le, K. M., Vadnais, M., Saye-Francisco, K. L., Jardine, J. G., Torres, J. L., et al. (2017). Rapid elicitation of broadly neutralizing antibodies to HIV by immunization in cows. *Nature*, *548*, 108.

Stanfield, R. L., Dooley, H., Verdino, P., Flajnik, M. F., & Wilson, I. A. (2007). Maturation of shark single-domain (IgNAR) antibodies: Evidence for induced-fit binding. *Journal of Molecular Biology*, *367*, 358.

Stanfield, R. L., Wilson, I. A., & Smider, V. V. (2016). Conservation and diversity in the ultralong third heavy-chain complementarity-determining region of bovine antibodies. *Science Immunology*, *1*, aaf7962.

Symons, D. B., Clarkson, C. A., & Beale, D. (1989). Structure of bovine immunoglobulin constant region heavy chain γ1 and γ2 genes. *Molecular Immunology*, *26*, 841.

Thielges, M. C., Zimmermann, J., Yu, W., Oda, M., & Romesberg, F. E. (2008). Exploring the energy landscape of antibody-antigen complexes: Protein dynamics, flexibility, and molecular recognition. *Biochemistry*, *47*, 7237.

Tobin-Janzen, T. C., & Womack, J. E. (1992). Comparative mapping of IGHG1, IGHM, FES, and FOS in domestic cattle. *Immunogenetics*, *36*, 157.

Walther, S., Czerny, C.-P., & Diesterbeck, U. S. (2013). Exceptionally long CDR3H are not isotype restricted in bovine immunoglobulins. *PLoS One*, *8*, e64234.

Walther, S., Tietze, M., Czerny, C. P., Konig, S., & Diesterbeck, U. S. (2016). 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*, *11*, e0164567.

Wang, F., Ekiert, D. C., Ahmad, I., Yu, W., Zhang, Y., Bazirgan, O., et al. (2013). Reshaping antibody diversity. *Cell*, *153*, 1379.

Ward, J. L., & Ducharme, N. G. (1994). Traumatic reticuloperitonitis in dairy cows. *Journal of the American Veterinary Medical Association*, *204*, 874.

Wedemayer, G. J., Patten, P. A., Wang, L. H., Schultz, P. G., & Stevens, R. C. (1997). Structural insights into the evolution of an antibody combining site. *Science*, *276*, 1665.

Yasuda, M., Jenne, C. N., Kennedy, L. J., & Reynolds, J. D. (2006). The sheep and cattle Peyer's patch as a site of B-cell development. *Veterinary Research*, *37*, 401.

Yasuda, M., Tanaka, S., Arakawa, H., Taura, Y., Yokomizo, Y., & Ekino, S. (2002). A comparative study of gut-associated lymphoid tissue in calf and chicken. *The Anatomical Record*, *266*, 207.

Yeap, L. S., Hwang, J. K., Du, Z., Meyers, R. M., Meng, F. L., Jakubauskaite, A., et al. (2015). Sequence-intrinsic mechanisms that target AID mutational outcomes on antibody genes. *Cell*, *163*, 1124.

Yin, J., Beuscher, A. E., 4th, Andryski, S. E., Stevens, R. C., & Schultz, P. G. (2003). Structural plasticity and the evolution of antibody affinity and specificity. *Journal of Molecular Biology*, *330*, 651.

Zhao, Y., Jackson, S. M., & Aitken, R. (2006). The bovine antibody repertoire. *Developmental and Comparative Immunology*, *30*, 175.

Zhao, Y., Kacskovics, I., Pan, Q., Liberles, D. A., Geli, J., Davis, S. K., et al. (2002). Artiodactyl IgD: The missing link. *Journal of Immunology*, *169*, 4408.

Zimmermann, J., Oakman, E. L., Thorpe, I. F., Shi, X., Abbyad, P., Brooks, C. L., 3rd, et al. (2006). Antibody evolution constrains conformational heterogeneity by tailoring protein dynamics. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 13722.

Zinicola, M., Lima, F., Lima, S., Machado, V., Gomez, M., Dopfer, D., et al. (2015). Altered microbiomes in bovine digital dermatitis lesions, and the gut as a pathogen reservoir. *PLoS One*, *10*, e0120504.