Immunogenetics (2019) 71:307–320
https://doi.org/10.1007/s00251-018-01102-9

ORIGINAL ARTICLE

Llama peripheral B-cell populations producing conventional and heavy chain-only IgG subtypes are phenotypically indistinguishable but immunogenetically distinct

Kevin A. Henry[1] · Henk van Faassen[1] · Doreen Harcus[2] · Anne Marcil[2] · Jennifer J. Hill[1] · Serge Muyldermans[3] · C. Roger MacKenzie[1,4]

Received: 29 September 2018 / Accepted: 11 December 2018 / Published online: 18 January 2019

Abstract

Camelid ungulates produce homodimeric heavy chain-only antibodies (**HCAbs**) in addition to conventional antibodies consisting of paired heavy and light chains. In the llama, HCAbs are made up by at least two subclasses (long-hinge IgG2b and short-hinge IgG2c HCAbs vs. conventional heterotetrameric IgG1s). Here, we generated murine monoclonal antibodies (mAbs) specific for the hinge-C_H2 boundary of llama IgG2b (mAb 1C10) and the Fc of llama IgG2c HCAbs (mAb 5E4). Flow cytometric analysis of llama peripheral blood lymphocytes revealed that $IgG1^+$, $IgG2b^+$ and $IgG2c^+$ B cells could be distinguished using mAbs 1C10/5E4 but had equivalent expression of three other cell-surface markers. MiSeq sequencing of the peripheral B cell repertoires of three llamas showed that (i) IgG2b and IgG2c HCAbs were present in similar proportions in the repertoire, (ii) a subset of IgG2b and IgG2c HCAbs, but not IgG1s, entirely lacked a hinge exon and showed direct V_HH-C_H2 splicing; these "hingeless" HCAbs were clonally expanded, somatically mutated and derived from hinged HCAb precursors, (iii) substantial repertoire overlap existed between IgG subclasses, especially between IgG2b and IgG2c HCAbs, (iv) the complementarity-determining region (**CDR**)-H3 length distributions of IgG2b and IgG2c HCAbs were broader and biased towards longer lengths compared with IgG1s due to increased N-nucleotide addition, (v) IgG2b and IgG2c HCAbs used a more restricted set of IGHV genes compared with IgG1s, and (vi) IgG2b and IgG2c HCAbs had elevated somatic mutations rates of both CDRs and framework regions (FRs) compared with IgG1s, especially of CDR-H1 and FR3. The distinct molecular features of llama IgG1, IgG2b and IgG2c antibodies imply that these subclasses may have divergent immunological functions and suggest that specific mechanisms operate to diversify HCAb repertoires in the absence of a light chain.

Keywords Heavy chain-only antibody · Single-domain antibody · V_HH · Llama · Immunogenetics · Next-generation DNA sequencing

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00251-018-01102-9) contains supplementary material, which is available to authorized users.

✉ Kevin A. Henry
kevin.henry@nrc-cnrc.gc.ca

[1] Human Health Therapeutics Research Centre, National Research Council Canada, 100 Sussex Drive, Ottawa, ON K1A 0R6, Canada

[2] Human Health Therapeutics Research Centre, National Research Council Canada, 6100 Royalmount Avenue, Montréal, QC H4P 2R2, Canada

[3] Laboratory of Cellular and Molecular Immunology, Vrije Universiteit Brussel, Brussels, Belgium

[4] School of Environmental Sciences, University of Guelph, 50 Stone Road East, Guelph, ON N1G 2W1, Canada

Introduction

Homodimeric antibodies (Abs) or Ab-like proteins lacking light chains have evolved through convergent evolution at least twice (Greenberg et al. 1995; Hamers-Casterman et al. 1993) or possibly three or more times (Rast et al. 1998) subsequent to the appearance of the paired heavy/light chain Ab system of vertebrates. In camelid ungulates, heavy chain-only Abs (**HCAbs**) arose through acquisition of a point mutation that disrupts splicing of the C_H1 exon (Nguyen et al. 1999; Woolven et al. 1999) along with solubilizing substitutions in IGHV genes at the former interface with the V_L domain (Muyldermans et al. 1994; Nguyen et al. 1998). Although the specific immunological functions of HCAbs in host defense remain unclear, their variable binding domains (**V_HHs**)

are highly useful in therapeutic, diagnostic, and research applications (Muyldermans 2013).

At the serological level, camelid HCAbs make up 25–75% of total IgG (Hamers-Casterman et al. 1993; Muyldermans 2013); the extent of inter-animal and inter-species variation in this figure has not been well studied. Camelid serum IgG can be separated using protein G (homodimeric and heterotetrameric fractions) and protein A (two distinct homodimeric fractions) chromatography (Lauwereys et al. 1998). At least two subtypes of heterotetrameric conventional IgG (IgG1a and IgG1b) and two subtypes of homodimeric HCAb IgG (short-hinge and long-hinge) exist in all camelid species (Muyldermans 2013). Throughout this manuscript, we consistently use the nomenclature system for IgG subtypes originally proposed by the discoverers of camelid HCAbs (Conrath et al. 2003) on the basis of hinge sequence (Table 1). Confusingly, dromedary long-hinge HCAbs (IgG2) supposedly correspond to the non-protein G-reactive serum IgG fractions and short-hinge HCAbs (IgG3) correspond to the protein G-reactive fraction, while in alpacas and llamas, the opposite may be true (non-protein G-reactive short-hinge IgG2cs; protein G-reactive long-hinge IgG2bs) (van der Linden et al. 2000). Further subdivisions within these subtypes have been suggested (Daley et al. 2005), but challenges in purification and subtype-allotype discrimination have complicated interpretation of these studies. Secondary reagents have been developed against each major subtype of camelid IgG (Daley et al. 2005; Holzlohner et al. 2018) but are only offered commercially.

At the immunogenetic level, the camelid *igh* locus located on chromosome 4 is interspersed with dedicated IGHV genes used in conventional Abs and HCAbs (Achour et al. 2008), although a subset of IGHV genes may be used in both types of Abs (Deschacht et al. 2010). All IGHV genes are rearranged with a common set of IGHD and IGHJ genes. Most camelid IGHV genes are most closely homologous to human IGHV3-family genes (Klarenbeek et al. 2015), although a subset of IGHV genes used in the HCAbs of some camelid species are closer in sequence to human IGHV1 and IGHV4 genes (Achour et al. 2008; Deschacht et al. 2010). The full complement of germline IGHV, IGHD, and IGHJ genes has yet to be completely described in most camelid species. Based on Sanger sequencing of relatively small numbers of rearranged HCAb IGHV-D-J regions, HCAbs have been suggested to have longer complementarity-determining region (CDR)-H3s (Harmsen et al. 2000; Muyldermans et al. 1994; Vu et al. 1997) and potentially elevated rates of somatic hypermutation (**SHM**), insertion/deletion and gene replacement (Nguyen et al. 2000; Nguyen et al. 2002; Vu et al. 1997) compared with camelid and human conventional Abs (Griffin et al. 2014). One recent study used Illumina MiSeq sequencing to sample the peripheral repertoires of three Bactrian camels and found drastically longer CDR-H3s and modestly elevated SHM rates in the HCAb vs. the conventional Ab compartment (Li et al. 2016).

The primary aim of this study was to characterize the molecular immunogenetic features of peripheral HCAb and conventional Ab repertoires in the llama (*Lama glama*) using next-generation DNA sequencing (**NGS**), and to assess compartmentalization of these features by isotype. A secondary goal was to develop murine monoclonal anti-isotypic Abs specific for camelid HCAbs as secondary reagents in immunoassays and for immunophenotypic characterization of llama peripheral B cells.

Materials and methods

Antibodies and reagents

Horseradish peroxidase (**HRP**)-conjugated polyclonal goat anti-llama IgG was from Cedarlane Labs (Burlington, Canada; Cat. No. A160-100P), and HRP-conjugated polyclonal donkey anti-mouse IgG was from Jackson ImmunoResearch (West Grove, PA; Cat. No. 715-035-150). Fluorescein-conjugate polyclonal anti-camelid IgG was from Triple J Farms (Bellingham, WA), and murine monoclonal antibodies LH41A, TH14B, and LT79A were from the Washington State University Monoclonal Antibody Center (Pullman, WA). Streptavidins conjugated to allophycocyanin (**APC**), APC-cyanine7 (**APC-Cy7**), phycoerythrin (PE), PE-cyanine7 (**PE-Cy7**), and peridinin-chlorophyll protein-cyanine5.5 (**PerCP5.5**) were from BD Biosciences (San Jose, CA). EZ-Link™ *N*-hydroxysulfosuccinimide (NHS)-LC-LC-biotin was from Thermo Fisher Scientific

Table 1 Nomenclature of camelid IgG subtypes used throughout this study

Subtype	Light chain	Hinge length	Archetypal hinge sequence
IgG1a	Yes	19 aa	ELKTPQPQSQPECRCPKCP
IgG1b	Yes	12 aa	EPHGGCTCPQCP
IgG2a	No	35 aa	EPKIPQPQPKPQPQPQPQPKPQPKPEPECTCPKCP
IgG2b	No	27–29 aa	EPKTPKPQPQPQPQ(PQ)PNPTTESKCPKCP
IgG2c	No	15 aa	AHHSEDPSSKCPKCP
IgG3	No	12 aa	GTNEVCKCPKCP

(Waltham, MA). Bovine serum albumin (**BSA**), human serum albumin (**HSA**), skim milk, and Tween-20 were from Sigma-Aldrich (St. Louis, MO), and tetramethylbenzidine (**TMB**) substrate was from Mandel Scientific (Guelph, Canada). *Clostridium difficile* lipoteichoic acid (**LTA**) and LTA-HSA conjugate were produced as previously described (Cox et al. 2013).

Serum fractionation

Total serum IgG from healthy male adult llamas and dromedary camels was fractionated as previously described (Baral et al. 2013; Hamers-Casterman et al. 1993). Briefly, 2 mL of serum was dialyzed overnight against 20 mM sodium phosphate buffer, applied to a HiTrap® protein G HP column (GE Healthcare, Piscataway, NJ) connected to an ÄKTA FPLC protein purification system (GE Healthcare), and then G1-fraction polyclonal HCAbs (IgG2b) were eluted with 100 mM citrate buffer, pH 3.5, and G2-fraction polyclonal conventional Abs (IgG1) were eluted with 100 mM glycine buffer, pH 2.7. The flow-through was applied to a HiTrap® protein A HP column (GE Healthcare), and A1-fraction polyclonal HCAbs (IgG2c) were eluted with 100 mM sodium acetate buffer, pH 4.5, and A2-fraction polyclonal HCAbs (also IgG2c) were eluted with 100 mM glycine buffer, pH 2.7. All fractions were immediately neutralized with 1 M Tris-HCl, pH 8.8, and dialyzed overnight against phosphate-buffered saline (PBS), pH 7.4. For mouse immunization and screening, fractionated serum IgGs were subjected to size-exclusion chromatography using a Superdex™ 200 Increase 10/300 GL column (GE Healthcare) as a final polishing step.

Hybridoma generation

Mouse immunization and hybridoma generation were performed as previously described (Manceur et al. 2017; Sun et al. 2016). Briefly, 6-week-old female A/J mice were immunized subcutaneously and intraperiteonally with 75–100 μg of an equimolar mixture of pooled polyclonal HCAb fractions (G1, A1, and A2) emulsified in TiterMax adjuvant (Cedarlane). The mice were immunized three times (days 1, 22, and 82), and fusions were performed 4 days later as described (Sun et al. 2016). Hybridoma supernatants were screened by ELISA for binding to purified polyclonal serum IgG fractions (G1, G2, A1, and A2). The rearranged IGHV and IGKV sequences of the 1C10 and 5E4 hybridomas were determined, cloned into pTT5 vectors upstream of mouse IGHG2A, IGHG2B, or IGKC genes, and produced in HEK293-6E cells as previously described (Durocher et al. 2002).

ELISA

The specificity of monoclonal (m)Abs 1C10 and 5E4 was assessed using three types of ELISA: (i) indirect two-step ELISA against purified polyclonal serum IgG fractions (G1, G2, A1 and A2), (ii) indirect two-step ELISA against recombinant proteins consisting of the EG2 V_HH linked N-terminally to various hinge sequences with or without human IgG1 C_H2/C_H3 domains; the former molecules were expressed from the pTT5 vector in HEK293 cells and purified using protein A chromatography (Durocher et al. 2002), while the latter 6×His-tagged molecules were expressed from the pSJF2H vector in *Escherichia coli* TG1 cells and purified using immobilized metal affinity chromatography (Baral et al. 2013; Henry et al. 2016a; Henry et al. 2015), and (iii) indirect three-step ELISA to detect binding of immune llama serum to HSA and *C. difficile* LTA. Llama immunization was performed as previously described (Henry et al. 2016a, Henry et al. 2015). For all ELISAs, wells of microtiter plates were coated overnight at 4 °C with 100 ng of each antigen in 35 μL of PBS, pH 7.4. The next day, wells were rinsed with PBS, blocked with 200 μL of PBS containing 2% (*w*/*v*) skim milk at 37 °C for 1 h, and then washed 3× with PBS containing 0.01% (*v*/*v*) Tween-20 (**PBS-T**). MAbs, purified polyclonal Abs, or sera were diluted in 35 μL of PBS-T containing 1% skim milk (**PBS-MT**) and added to wells for 2 h at room temperature. After washing 5× with PBS-T, 35 μL of secondary Ab in PBS-MT was added for 45 min at room temperature: for ELISA designs (i) and (ii), the secondary Ab was HRP-conjugated donkey anti-mouse IgG (subtracted and with minimal species cross-reactivity) diluted 1:3000, while for ELISA design (iii), the secondary Ab was either mAb 1C10 or 5E4 (5 μg/mL). For ELISA design (iii) only, wells were washed 5× with PBS-T, and 35 μL of tertiary Ab (HRP-conjugated donkey anti-mouse IgG) in PBS-MT was added for 45 min at room temperature. After a final wash (5× with PBS-T), wells were developed with 35 μL of TMB substrate, stopped with 35 μL of 1 M H_2SO_4, and the absorbance at 450 nm was measured using a Multiskan™ FC photometer (Thermo Fisher Scientific).

Flow cytometry

MAbs 5E4 (murine IgG2a), 1C10 (murine IgG2b), LH41A (murine IgM), TH14B (murine IgG2a), and LT79A (murine IgG3) were biotinylated with NHS-LC-LC-biotin according to the manufacturer's instructions and dialyzed overnight against PBS, pH 7.4. Each mAb (10 μg) was reacted with fluorescenated streptavidin (0.2 μg) for 1 h at room temperature, with no purification. Llama peripheral blood mononuclear cells were purified from whole blood using density gradient centrifugation as previously described (Baral et al. 2013). Approximately 2×10^5 cells in 0.2 mL PBS containing 1% BSA were stained with 2 μg of each fluorescenated Ab on

ice for 1 h and washed twice with PBS, and then data were collected on a BD FACSCanto™ instrument.

Illumina MiSeq sequencing

Illumina MiSeq amplicon sequencing (2 × 250 bp paired-end reads) was conducted essentially as previously described (Henry 2018; Henry et al. 2016a; Henry et al. 2015). Briefly, total RNA was extracted from ~ 1×10^7 llama peripheral blood mononuclear cells using a PureLink™ RNA Mini Kit (Life Technologies, Carlsbad, CA) and reverse transcribed using qScript® cDNA supermix (Quanta Biosciences, Beverly, MA). Amplicons were generated from cDNA using an equimolar mixture of three forward primers located in framework region (FR)1 (MJ1, MJ2, and MJ3) and reverse primers located either in C_H1 or C_H2 (Supplementary Table SII) and purified using gel extraction and solid-phase reverse immobilization with AMPure® XP beads (Beckman Coulter, Brea, CA). Depending on the analysis, forward and reverse reads were either merged using FLASH (Magoc and Salzberg 2011) with default parameters and quality-filtered using the FASTX toolkit (Schmieder and Edwards 2011) with a stringency of Q30 over ≥ 95% of each read, or reverse reads only were trimmed and filtered using Trimmomatic (Bolger et al. 2014); leading and trailing bases with quality scores below 3, as well as 3′ regions where the average quality per base dropped below 15 using a 4-base sliding window, were trimmed, and reads shorter than 230 bases as a result of trimming were dropped. The filtered data were analyzed using custom R scripts, available from the corresponding author by request. Analyses of IGHV gene usage and SHM were performed using IMGT/High-VQUEST (Alamyar et al. 2012) with alignment to alpaca (*Vicugna pacos*) germline genes. Circos analyses were performed using Circos Table Viewer (Krzywinski et al. 2009), and analyses of sequence diversity were performed using the R/Bioconductor packages "msa" and "BALCONY" (Bodenhofer et al. 2015; Pluciennik et al. 2018).

Mass spectrometry

Polyclonal llama and dromedary G1-fraction HCAbs were diluted to 0.5 mg/mL in 50 mM ammonium bicarbonate, reduced with 10 mM dithiothreitol (DTT), and alkylated with 37.5 mM iodoacetic acid. The samples were then quenched with DTT and further diluted to 0.1 mg/mL with 50 mM ammonium bicarbonate. Each sample was split and digested overnight with trypsin (Promega, 1:20, 37 °C) or chymotrypsin (Sigma C-3142, 1:40, 25 °C). The resulting peptides (0.3 μg) were separated by nano-LC on a Waters nanoAcquity system using a BEH C18 column (1.7 μm, 100 μm × 100 mm) and a linear gradient of 4 to 40% acetonitrile in 0.1% acetic acid over 45 min, with a flow rate of 0.5 μL/min. Eluting peptides were analyzed in-line by MS (MS) on a Synapt G2S mass spectrometer in positive electrospray ionization mode using a data-dependent acquisition method with dynamic exclusion and lock mass enabled. All MS/MS fragment ion spectra were searched by Mascot against a database containing a collection of sequences for llama and dromedary IgG molecules, along with several common contaminants (keratins, trypsin, etc.). The following search parameters were used: no enzyme, peptide mass tolerance = 1.2 Da, fragment mass tolerance = 0.5 Da, fixed modification = carbamidomethyl (on C), and variable modification = oxidation (on M). Peptides were subsequently filtered to retain only those with a parent mass within 15 ppm of the predicted mass and with a mascot score > 30.

Results

Development of murine anti-isotypic mAbs against llama IgG2b and IgG2c

After immunizing mice with purified llama polyclonal serum HCAbs, we isolated two murine mAbs, 1C10 and 5E4, specific for llama HCAb isotypes (IgG2b and IgG2c). These MAbs, and/or their V_H and V_L sequences, may be available from the corresponding author by request[1]. MAb 1C10 was selective for llama G1-fraction IgG2b (Fig. 1a and Supplementary Fig. S1) and cross-reacted with dromedary camel A1- and A2-fraction IgG2a, while mAb 5E4 was specific for llama A1- and A2-fraction IgG2c and bound no dromedary serum fraction. Binding by mAb 1C0 depended on the presence of a camelid long-hinge sequence (either llama IgG2b or dromedary IgG2a) and a C_H2 domain (Fig. 1b), while mAb 5E4 recognized an epitope in the Fc region that was at least partially conserved in human IgG1 (but clearly not in llama G1-fraction IgG2b or G2-fraction IgG1). The mAbs could be used to detect binding by each isotype of llama serum IgG after immunization with a LTA-HSA conjugate (eliciting an IgG2b/IgG2c HCAb and conventional IgG1 response against HSA but only an IgG1 against LTA), eliminating the need for serum fractionation (Fig. 1c). MAbs purified from hybridoma supernatants and produced recombinantly in HEK293 cells showed identical reactivities (data not shown).

Immunophenotyping of llama peripheral B cells

MAbs 1C10 (α-IgG2b), 5E4 (α-IgG2c), LH41A (unknown B cell marker), TH14B (MHC II), and LT79A (unknown B cell marker) were biotinylated and complexed with fluorescenated streptavidins (Supplementary Table SI). Immunophenotyping

[1] Requests for the anti-isotypic mAbs specific for camelid IgG subtypes described here should be directed to the corresponding author. The mAbs and/or their sequences should be available for non-commercial use under MTA.

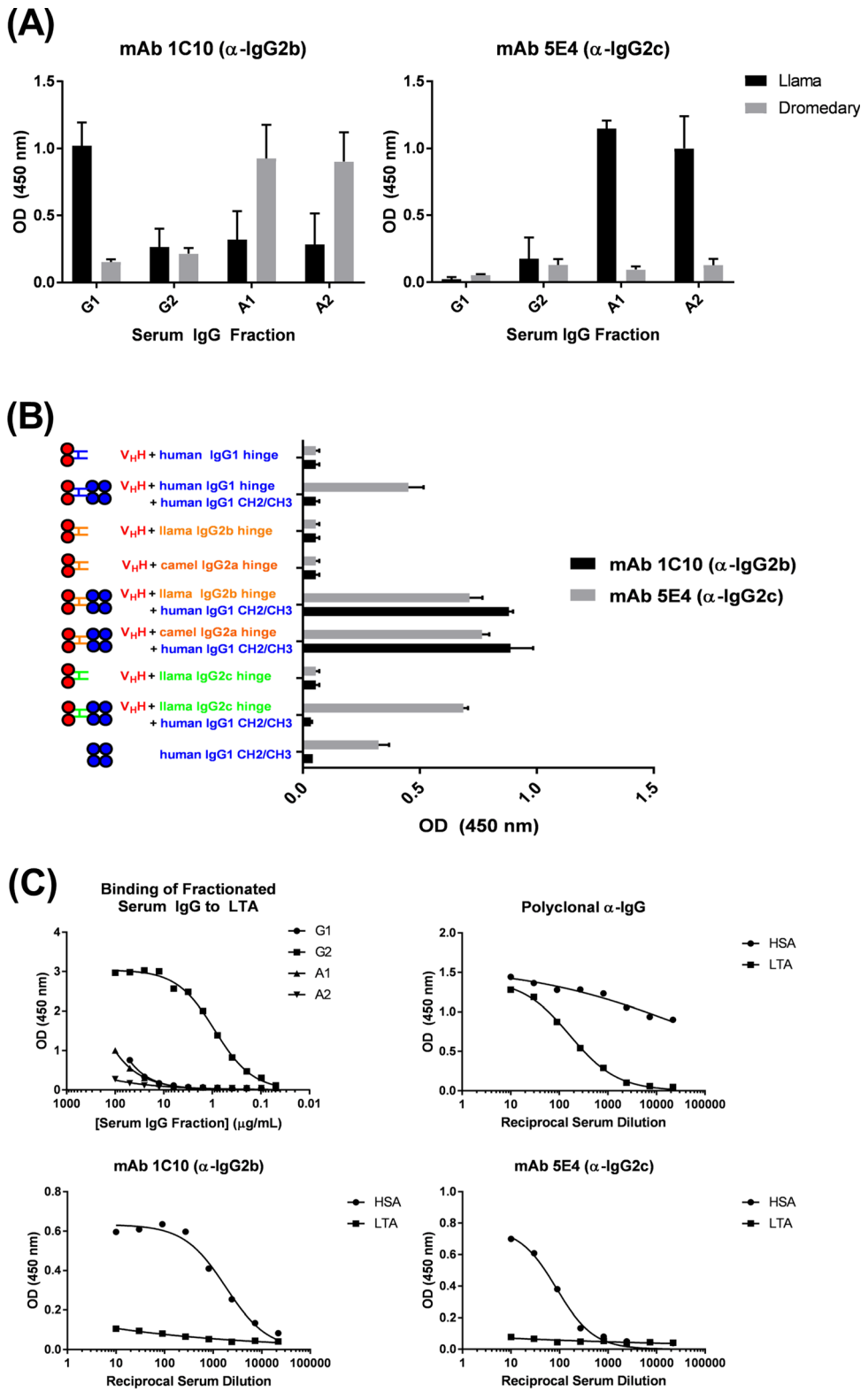

Fig. 1 Specificity of mAbs 1C10 (α-IgG2b) and 5E4 (α-IgG2c). **(a)** Binding of mAbs 1C10 and 5E4 (1 μg/mL) to polyclonal serum IgG fractions from five llamas and one dromedary. **(b)** Specificity of mAbs 1C10 and 5E4, as shown by their binding patterns at 1 μg/mL to recombinant protein antigens. Hinge sequences used were human IgG1 (EPKSCDKTHTCPPCP), llama IgG2b (EPKTPKPQPQPQPQPNPTTESKCPKCP), camel IgG2a (EPKIPQPQPKPQPQPQPQPKPQPKPEPECTCPKCP), and llama IgG2c (AHHSEDPSSKCPKCP). **(c)** Use of mAbs 1C10 and 5E4 as secondary antibodies to detect HSA- and LTA-binding by immune llama serum. Strong binding to HSA was observed for all serum fractions (not shown), but LTA binding was observed only for conventional Abs (G2 fraction). Results are representative of at least three independent experiments

of llama peripheral blood mononuclear cells was carried out using this panel, along with a fluorescenated polyclonal anti-camelid IgG that recognizes all surface-Ig$^+$ cells (Fig. 2). Staining by the three anti-Ig markers (mAb 1C10, mAb 5E4, and polyclonal anti-IgG) revealed three discrete B cell subpopulations (surface IgG1$^+$, IgG2b$^+$, and IgG2c$^+$ B cells) including one subpopulation (IgG2b$^+$ B cells) stained by both mAb 1C10 and 5E4. Double staining likely resulted from cross-reactivity of mAb 5E4 for mAb 1C10's murine IgG2b Fc region. Staining by mAbs LH41A, TH14B, and LT79A was equivalent for all three B cell subpopulations (data not shown).

Interrogation of llama HCAb and conventional Ab repertoires using next-generation DNA sequencing revealed substantial repertoire overlap between IgG subtypes and a unique population of hingeless HCAbs

We interrogated the repertoires of three llamas using primer sets located in C_H1 or C_H2 (Supplementary Table SII) to a depth of $1.1–9.8 \times 10^5$ reads per sample after quality filtering (Supplementary Table SIII). From the amplicons generated using C_H2-specific primers, we were able to assess relative isotype frequency among peripheral B cells (Fig. 3a, b). IgG2b- and IgG2c-producing B cells were present at roughly equal frequencies in the repertoire, with some degree of variability between animals, and IgG1a and IgG1b sequences were detected at low frequency (Fig. 3b; note that the size of the FR1-C_H2 amplicon is larger for conventional IgG1s due to the presence of the C_H1 exon, resulting in lower amplification efficiency). Two IgG2b variants (probably allotypes) were identified, distinguished by a single Thr/Ala hinge polymorphism (Fig. 3a). No non-IgG2b/IgG2c HCAb hinge sequences were detected in the dataset at frequency ≥ 0.1%.

Intriguingly, we detected a small proportion of HCAb sequences (~ 0.5–4% of the repertoire) from which the entire hinge exon was missing, with direct splicing of the V_HH and C_H2 exons (Fig. 3a, b). Hingeless IgG2b and IgG2c HCAbs were distinguishable based on their N-terminal C_H2 sequences. No hingeless conventional IgG1s with directly spliced C_H1 and C_H2 domains were detected (data not shown). The bulk of hingeless HCAbs were comprised of a relatively small number of clonally-expanded lineages with unusual properties, including very long CDR-H3s with unusual amino acid content and V_H-like FR2 sequences (Table 2). Hingeless HCAb sequences were derived from hinged precursors and showed evidence of SHM (Supplementary Fig. S2), suggesting their potential involvement in antigen-specific immune responses.

We also investigated the overlap between llama IgG1, IgG2b, and IgG2c repertoires more broadly. We found significant overlap between the CDR3 sequences of IgG2b and IgG2c HCAbs (Fig. 4a and Supplementary Fig. S3) and more minor overlap between conventional IgG1 and HCAb repertoires. This analysis also supported the ontogeny of hingeless HCAbs from hinged precursors. CDR3 clonotypes showing cross-subtype overlap (e.g., clonotypes shared between IgG2b and IgG2c HCAb repertoires; clonotypes shared between conventional IgG1 and HCAb repertoires; clonotypes shared between hinged and hingeless HCAb repertoires) were more clonally expanded than subtype-restricted clonotypes (Fig. 4b, c), suggesting their potential involvement in antigen-specific immune responses.

Llama IgG2b and IgG2c HCAbs have longer CDR3s, elevated SHM rates, and restricted IGHV gene usage compared with IgG1s

We compared the CDR-H3 length distributions of IgG1 (conventional Ab), IgG2b (long-hinge HCAb), and IgG2c (short-hinge HCAb) sequences (Fig. 5). The amplicons used in this analysis

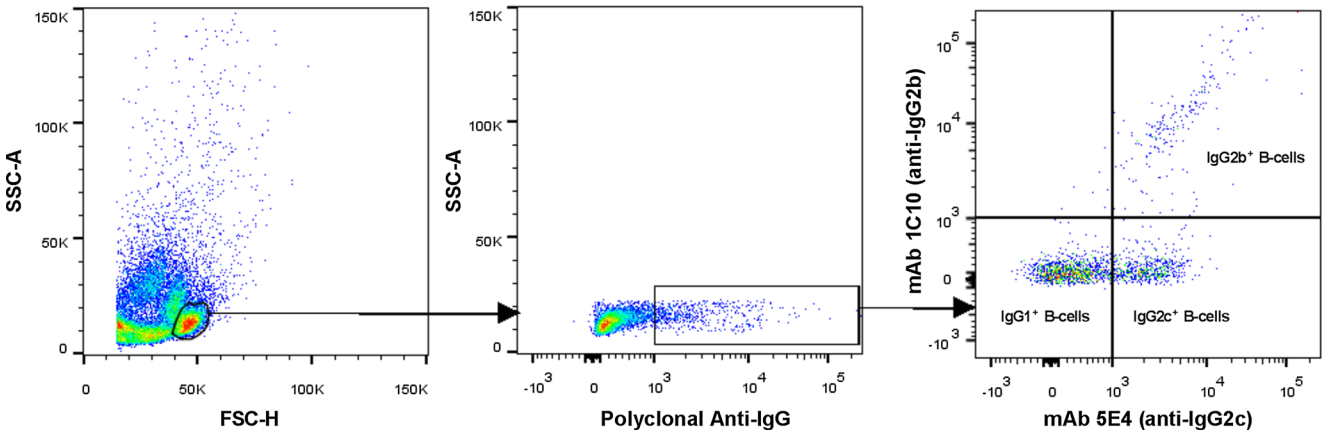

Fig. 2 Flow cytometric analysis of llama peripheral B cells. The gating strategy for separation of IgG1$^+$, IgG2b$^+$ and IgG2c$^+$ B cells is shown. Lymphocytes were gated based on FSC and SSC parameters, and then all surface IgG$^+$ B cells recognized by a FITC-conjugated polyclonal anti-IgG were gated. In the rightmost panel, conventional IgG1$^+$ B cells are 5E4$^-$1C10$^-$ double-negative, IgG2c$^+$ B cells are 5E4$^+$1C10$^-$ single-positive, and IgG2b$^+$ B cells are 5E4$^+$1C10$^+$ double-positive due to mAb 5E4's cross-reactivity with the murine IgG2b Fc of mAb 1C10

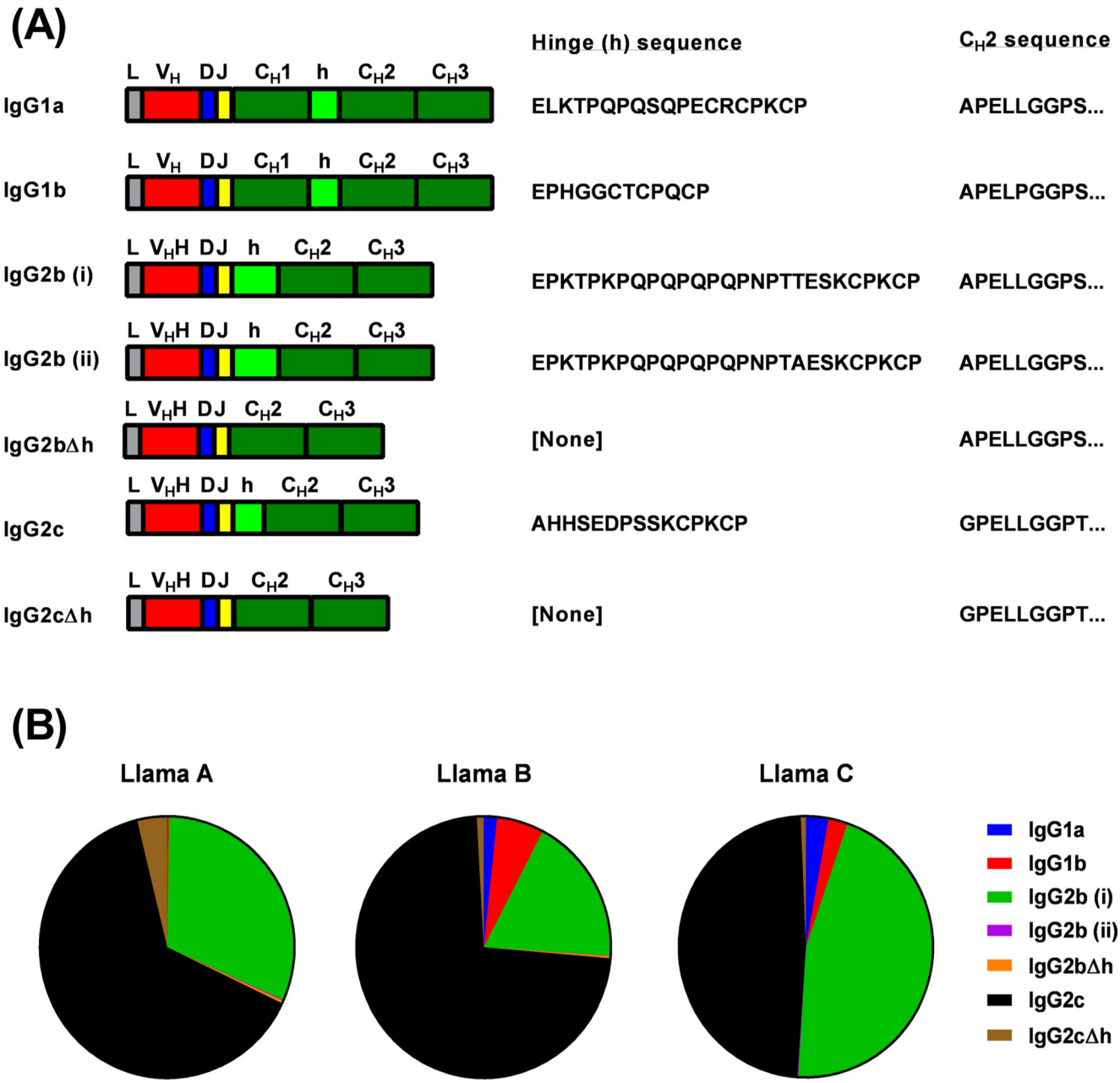

Fig. 3 IgG subtype frequency among llama peripheral B cells. **(a)** Pictorial depiction of rearranged llama IgG transcripts and amino acid sequences of their hinge regions. **(b)** IgG subtype frequency determined via Illumina MiSeq sequencing of amplicons produced using seqR-CH2-b primer (Supplementary Table SII). The low frequency of IgG1 transcripts detected is almost certainly due to PCR and sequencing bias against the longer V_H-C_H1-hinge-C_H2 amplicon

were produced using FR1/C_H1-specific primers for IgG1 Abs and FR1/C_H2-specific primers for IgG2b and IgG2c HCAbs. CDR3 length was roughly normally distributed for IgG1 Abs, with an average length of ~ 13 residues. By contrast, the CDR-H3s of IgG2b and IgG2c HCAbs were distributed more broadly and had slightly longer average lengths (~ 15 residues). The long CDR-H3s of IgG2b and IgG2c HCAbs arose through an increased number of N-nucleotide additions in the junctions of these Abs (roughly twice the number present in IgG1s; Supplementary Fig. S4). For both IgG2b and IgG2c HCAbs, CDR-H3 length was less normally distributed, possibly reflecting oligoclonality of these repertoires, with a subpopulation of both IgG2b and IgG2c HCAbs having very long CDR-H3s (≥ 20 residues). To ensure that long-CDR-H3 Ab sequences were not lost during merging of paired-end reads, the same analysis was conducted using reverse reads only and showed an identical result (data not shown).

The IGHV gene usage of IgG1 (conventional Ab), IgG2b (long-hinge HCAb), and IgG2c (short-hinge HCAb) repertoires were compared by alignment to alpaca germline

Table 2 Properties of abundant hingeless HCAb lineages

	CDR3 sequence	CDR3 length (aa)[a]	IGHV gene[b]	Frequency (%)[c]
Llama A	SALPRYGSRCPTLTAGDY	18	IGHV3S66	21.4
	AKATSAAQAMGARPYDY	17	IGHV3-1	17.2
	NVVTS	5	IGHV3S53	16.9
	YAYVLEIGSFSEHPRMKDY	20	IGHV3S53	9.0
	AAGPPWAAGPSPNSHHMDS	20	IGHV3-3	7.1
	NTNLGNSDY	10	IGHV3S53	6.4
	AAVTGVNWCLHSQYEYDY	18	IGHV3S63	3.6
	NAQGYGAAATYKYDY	15	IGHV3S53	0.5
	AASVERLCRRVFDANVVEY	19	IGHV3S61	0.3
	AGDRF	5	IGHV3S53	0.3
			Total	82.7
Llama B	KPIALRAGPASES	13	IGHV3S53	4.3
	AAKSSLLDSSPYEY	14	IGHV3-3	3.2
	NAESSDTYYPWY	12	IGHV3-3	2.4
	AADRGLHSILVESTADY	17	IGHV3-3	2.2
	VSDENYDSGEVA	13	IGHV3S32	1.7
	AAEPGSRVSLIWMHAELYDY	20	IGHV3S53	1.7
	TSLFH	5	IGHV3S31	1.6
	YMSTINYSDYQLSDH	15	IGHV3S53	1.6
	AKYYSNNYSYDY	13	IGHV3S1	1.6
	AKDRQYSEYRYYDGMDY	17	IGHV3S1	1.4
			Total	21.8
Llama C	ALKTESYCGASRSGYDY	17	IGHV3-3	35.0
	NADIVSPQGDKGVEIRPH	18	IGHV3S53	18.0
	AADMWDRCFETIRVAEAQYDY	21	IGHV3-3	12.1
	TLAKFPGPVRVGRPALVKQQDF	22	IGHV3S61	7.2
	NAHISEAVGDLPYDY	16	IGHV3S53	6.6
	ARDPRDTNSWSFAS	15	IGHV3-2	3.2
	ATDYSQTVDYSEYDQDVKESCLLLPNRKYHY	31	IGHV3-3	1.4
	AADRWDDVRGDQYDY	15	IGHV3S53	0.3
	AARQGVAIRSGSYDY	15	IGHV3-3	0.2
	AASRYDTGYPVTRNAVAY	18	IGHV3-3	0.2
			Total	84.2

[a] IMGT definition

[b] Nearest alpaca germline IGHV gene

[c] Frequency of CDR3 among all hingeless HCAb sequences

sequences available in the IMGT database (Fig. 6). Alignment with alpaca germline genes provides only a rough overview of IGHV gene usage, as the categorization of rearranged and expressed llama V_HHs according to similarity with alpaca genes may group some sequences inappropriately. As expected, the nearest germline alpaca IGHV genes of llama IgG1 conventional Abs and IgG2b/IgG2c HCAbs showed little overlap, reflecting the dedicated sets of IGHV genes for each Ab type. Llama IgG1 repertoires showed broad IGHV gene usage but were most frequently homologous to four alpaca germline IGHV genes (IGHV3-1, IGHV3S1, IGHV3S32, and IGHV3S42). The bulk of IgG2b and IgG2c HCAb sequences were most closely homologous to two alpaca germline genes (IGHV3-3 and IGHV3S53), with IgG2c repertoires showing slightly more diverse IGHV gene usage and a higher proportion of V_H-like sequences than IgG2b repertoires. Despite apparently restricted IGHV gene usage, amino acid sequence diversity in both FRs and CDRs was higher in

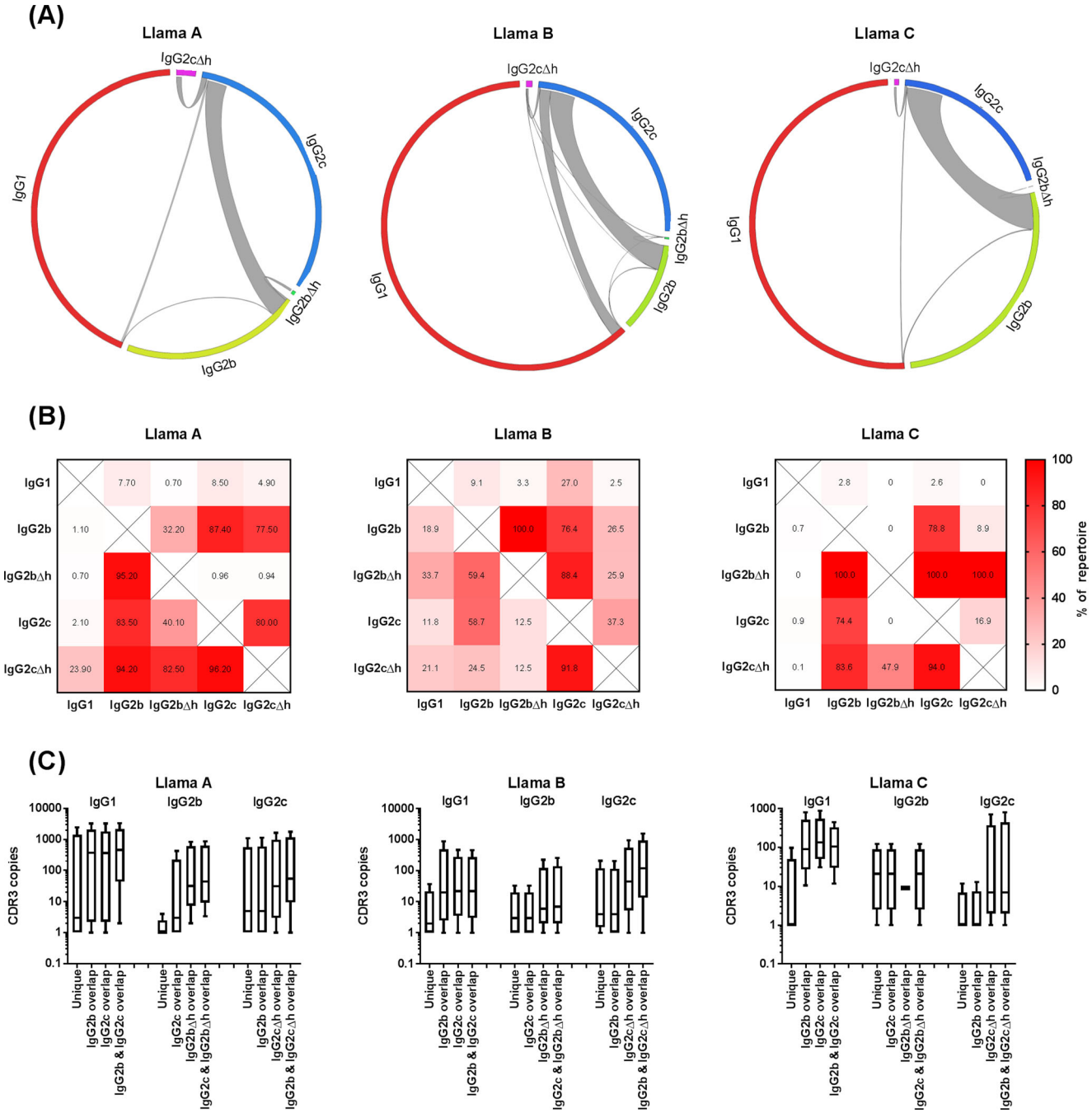

Fig. 4 Cross-talk between llama peripheral IgG1, IgG2b, and IgG2c repertoires. **(a)** Circos plot of CDR3 sharing between each llama IgG subtype (see Fig. 3). The size of each subtype along the circle's circumference is determined by the number of unique CDR3 amino acid sequences, and repertoire overlap (shown in gray) is determined by the number of shared unique CDR3 sequences, irrespective of the number of CDR3 copies for each unique CDR3. **(b)** Global overlap between IgG subtype repertoires, taking into account clonal expansion and using all CDR3 sequences, not unique CDR3 sequences, as the denominator. For each row, the heatmap shows the percentage of all CDR3 sequences in the row subtype's repertoire shared with each other subtype. For instance, for llama A, 7.7% of IgG1 CDR3s were also found among IgG2b CDR3s. The discrepancy between parts A and B of this figure is due to clonal expansion (i.e., in llama A, a small number of unique IgG1 CDR3 sequences are shared with the IgG2b repertoire, but these are disproportionately clonally expanded and make up 7.7% of all IgG1 CDR3 sequences). **(c)** Degree of clonal expansion of subtype-restricted and shared CDR3 clonotypes. For each subtype, the boxplot shows the median (line), interquartile range (box), and 95% confidence interval (whiskers) of the number of CDR3 copies present for each unique CDR3 sequence. CDR3 clonotypes shared between conventional IgG1 and HCAb repertoires, between IgG2b and IgG2c HCAb repertoires, and between hinged and hingeless HCAb repertoires showed elevated clonal expansion compared with subtype-restricted clonotypes

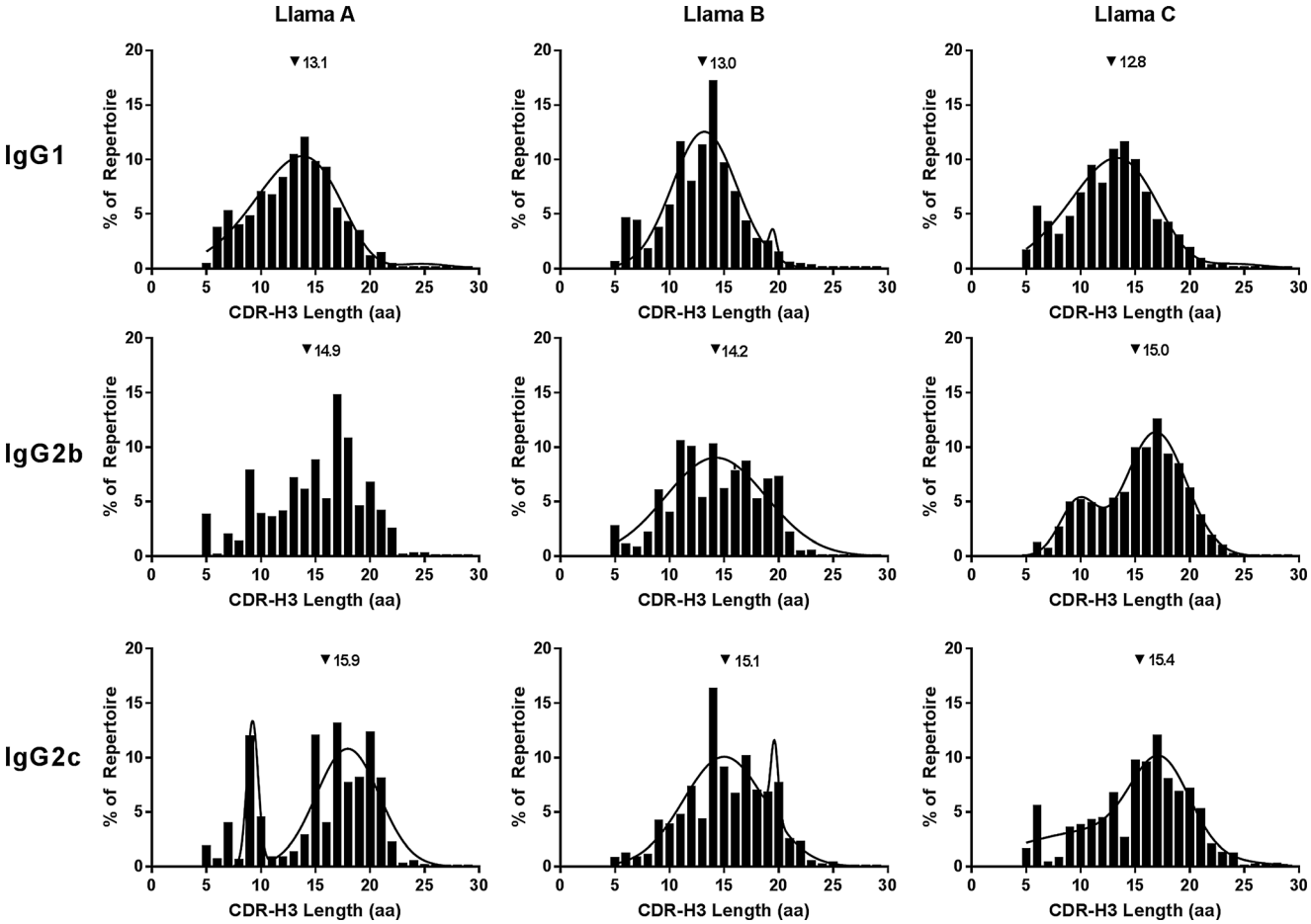

Fig. 5 CDR-H3 length distributions of llama IgG1, IgG2b, and IgG2c repertoires. Average CDR-H3 length is indicated by the triangle. Curves were fit to the data using equations describing the sum of two Gaussian distributions in GraphPad Prism 7

IgG2b and IgG2c HCAb repertoires compared with IgG1 repertoires, especially near the CDR1 N-terminus and in FR3, likely reflecting elevated SHM rates (Fig. 7). Using the number of amino acid substitutions with respect to the nearest alpaca germline gene as a surrogate for SHM rate, we found that IgG2b and IgG2c repertoires diverged from the IGHV germline, on average, at 5–10 additional positions compared with IgG1 repertoires (Supplementary Fig. S5). The additional substitutions occurred most frequently in CDR1 and in FR3.

Discussion

Our data argue that in contrast to the dromedary camel, probably only four major IgG subtypes exist in *L. glama* (IgG1a, IgG1b, IgG2b, and IgG2c; Fig. 3). We base this judgment on (i) our recovery of only two murine mAbs, 1C10 (long hinge/C_H2-specific; G1 serum IgG fraction) and 5E4 (Fc-specific; A1 and A2 serum IgG fractions), seemingly capable of recognizing the bulk of HCAbs in sera, and (ii) NGS data from amplicons produced from IgG transcripts, showing only four major IgG subtypes. This finding is in agreement with at least four previous studies in llamas and alpacas, including the only sequenced camelid *igh* locus (Achour et al. 2008; Maass et al. 2007; van der Linden et al. 2000; Woolven et al. 1999), but discordant with two other studies that identified IgG2as and IgG3s in the expressed HCAb repertoires of llamas (Saccodossi et al. 2012; Vu et al. 1997). Because of the potential conflict with previous literature, we also attempted to amplify each isotype by PCR using hinge-specific primers and could not detect the presence of IgG3 in cDNA even after 70 cycles of semi-nested PCR (Supplementary Fig. S6). Moreover, by mass spectrometry, we detected the presence of hinge peptides derived from IgG2b but not IgG3 in llama G1-fraction HCAbs, while the converse was true in dromedary camel G1-fraction HCAbs (Supplementary Fig. S7 and Supplementary Table SIV). However, it remains possible that additional subtypes bearing polymorphisms downstream from the hinge region may exist and/or that that the primers used here may not have annealed to some IgG transcripts. We did not assess binding of the anti-isotypic mAbs described here to every reported HCAb isotype/allotype sequence in llamas, alpacas, and camels, which may also vary across populations.

The apparent inverse relationship between HCAb isotype (IgG2b vs. IgG3) and protein G reactivity in llamas and camels is surprising, as is the apparent variability in HCAb

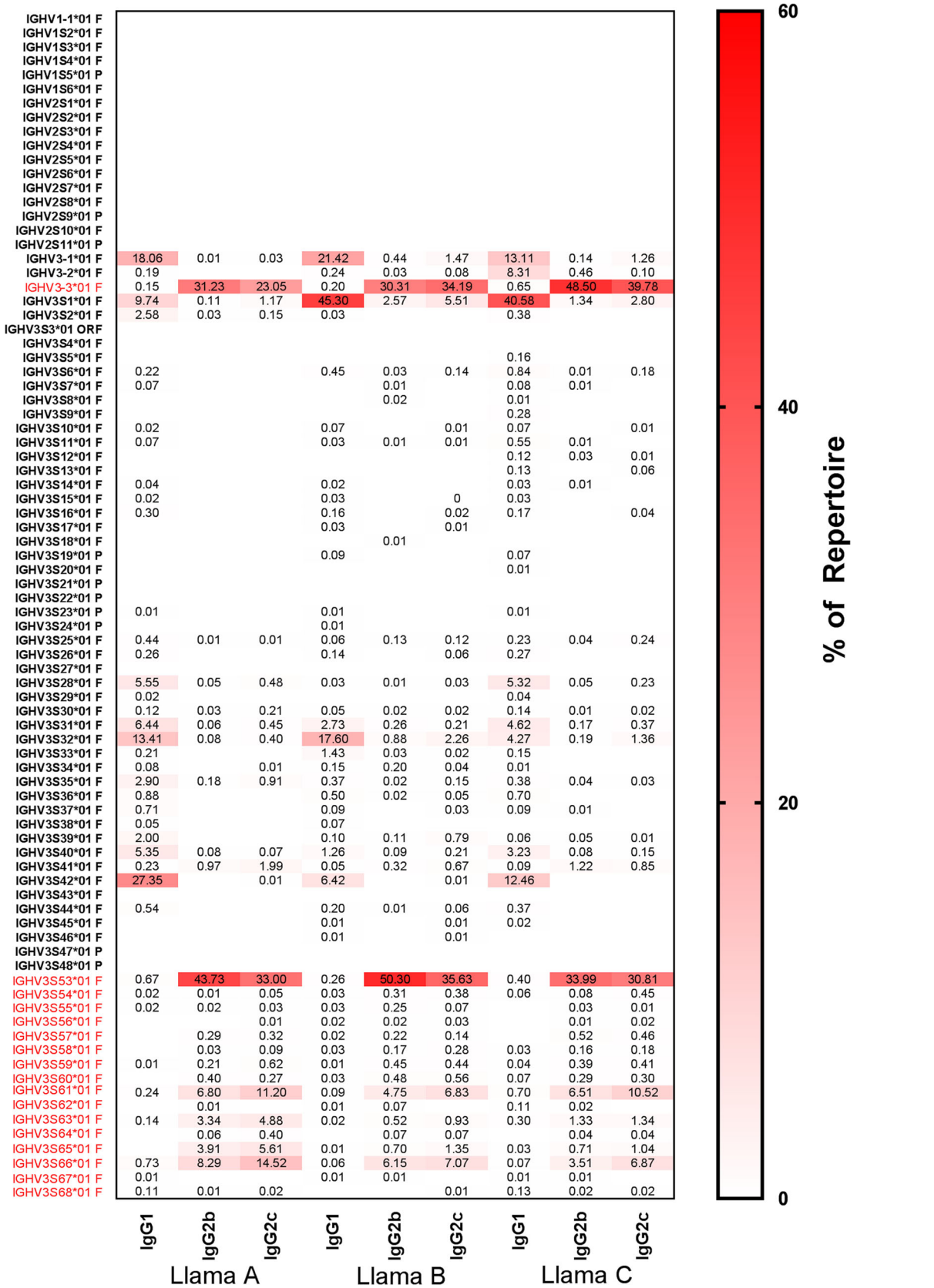

Fig. 6 IGHV gene usage of IgG1, IgG2b, and IgG2c repertoires in three llamas. Sequences were aligned to alpaca (*Vicugna pacos*) germline genes using IMGT/High-VQUEST. IGHV genes with frequency < 0.005 are not shown. HCAb V_HH genes are colored in red and conventional Ab V_H genes in black

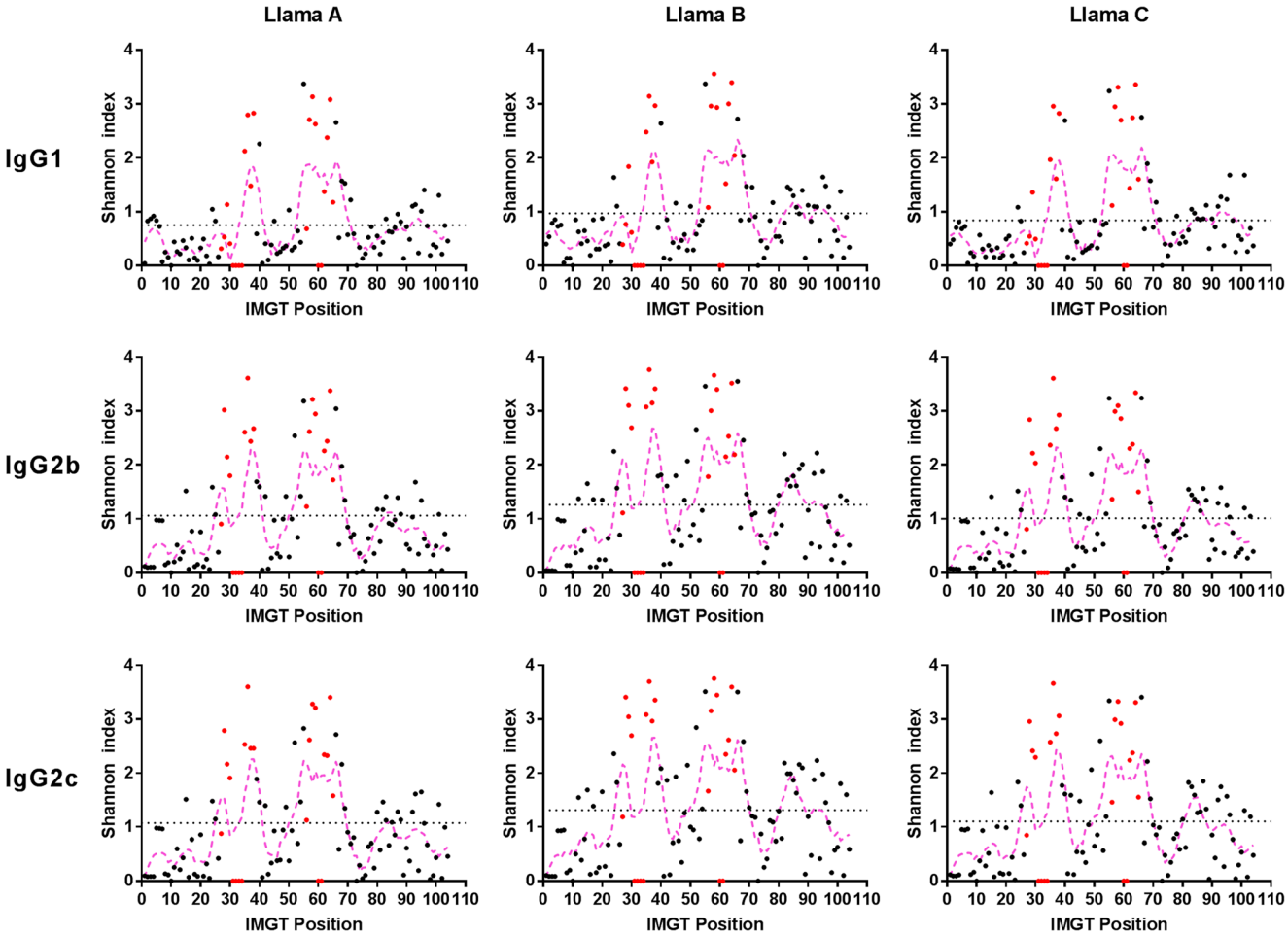

Fig. 7 Sequence diversity over the length of the IGHV genes of llama IgG1, IgG2b, and IgG2c transcripts as a surrogate for SHM rate. For each animal's repertoire, the set of unique IGHV amino acid sequences were gapped using IMGT/HighV-QUEST, and 5000 sequences were randomly sampled from the repertoire. These sequences were aligned using the R package "msa" (ClustalW, default parameters) and then the Shannon index for each position was calculated using the R package "BALCONY." The Shannon index for each position is shown with a dot, with CDR1 and CDR2 positions colored in red. The pink line shows the moving average, and the horizontal line indicates the average Shannon index over the entire IGHV sequence. Results are reflective of three independent analyses with sequence re-sampling for each analysis

isotypes among and between camelid populations. Although a complete reconciliation of all previous studies is not currently possible, one possibility suggested by the available data is that there are at least two configurations of camelid HCAb constant region genes (a haplotype encoding γ2a, γ2c, γ3, and possibly γ2b in camels and potentially some llamas, and a haplotype encoding only γ2b and γ2c in llamas). Thus, clearly, there has been rapid evolution of these constant region genes in the camelid lineage, and isotypes not detectable in the expressed repertoires of some animals may still be present in the genome as pseudogenes. However, the generally accepted explanation for pH-dependent elution of two fractions from protein A columns in dromedary camels (subisotypic variation) is unlikely to apply to *L. glama* and may instead be explained by V_HH domain binding (Henry et al. 2016b).

In camelids, surface IgG^+ B cells likely belong to the memory compartment, given that V_H-IgM and V_HH-IgM transcripts including a CH1 exon have been detected (Achour et al. 2008). It remains unclear whether V_HH-IgMs pair with light chains or whether light chains are rearranged in these cells. Given the apparently minor differences in cell-surface phenotypes of $IgG1^+$, $IgG2^+$, and $IgG3^+$ human memory B cells (de Jong et al. 2017), it would perhaps not be surprising that the division of labor between IgG isotypes was likewise not reflected in the surface markers of llama IgG^+ B cells. More work remains to be done in this area, especially in developing and validating Ab panels against camelid B cell markers.

The novel population of hingeless HCAbs reported here (Fig. 3, Table 2) are presumably derived either from SHM of hinge exon splice sites (at the DNA level) or from errors in transcript processing whereby splicing donor or acceptor sites become cryptic (at the RNA level). Other genetic mechanisms, such as exon deletion or class-switch recombination (CSR) defects, may also be responsible (Zou et al. 2007); these hypotheses would need to be verified at the genomic DNA level. Similar types of unusual HCAb splicing variants have been previously observed

at the cDNA level, including clones in which the V_HH was linked directly to C_H3, although these variant HCAbs could not be identified at the protein level despite significant effort (personal communication, N. Deschacht). However, the genetic features of hingeless HCAbs (clonal expansion; somatic mutation) strongly imply that B cells expressing these molecules may be involved in antigen-specific immune responses, rather than cells in which SHM has compromised the B cell receptor. The importance of HCAb hinges in antigen recognition is poorly understood (Henry and MacKenzie 2018), including the role of very long and flexible hinges of some HCAbs and the lack of a defined hinge in shark IgNARs (Feige et al. 2014). One possibility is that loss of the hinge region would serve to minimize the distance between V_HH paratopes for optimal recognition of closely spaced antigens (Brooks et al. 2018).

Compared with a previous study of Bactrian camel Ab repertoires (Li et al. 2016), we found that *L. glama* HCAb repertoires had only modestly elongated CDR-H3 length distributions compared with conventional Ab repertoires (~ two to three residues longer vs. five residues longer; Fig. 5). Thus, very long CDR-H3 loops appear not to be a general feature of HCAbs related to their structure and function, and differences in CDR-H3 length among camelid species may reflect normal species-to-species variation (e.g., between human and murine repertoires). However, it appears that the longer CDR3 of HCAbs compared with conventional IgG1s are generated at the "birth" of these Abs during V-D-J rearrangement by increased N-nucleotide addition, contributing to elevated non-templated junctional diversity in HCAbs. If long CDR3s were required for the stability or function of HCAbs but were not generated through specific mechanisms, one might expect to observe survivorship bias of HCAb rearrangements involving long D genes, which was not the case.

Elevated SHM rates of camelid HCAbs compared with conventional Abs have been suggested previously on the basis of relatively small numbers of Sanger reads, and typically without reference to camelid conventional Abs. Our results here support higher SHM rates of llama IgG2b and IgG2c HCAbs compared with IgG1s, although the lack of publicly available llama germline IGHV gene sequences complicates interpretation of these data. Ours is at least the third study to show that the bulk of the camelid peripheral IgG repertoire (conventional and HCAb) may be mutated from germline (Achour et al. 2008; Li et al. 2016), indicating that either memory cells are the major circulating IgG^+ B cell type or that the mechanisms of SHM regulation in camelids differ from those of humans and mice. The significant overlap of CDR3 sequences among IgG1, IgG2b, and IgG2c repertoires, especially between IgG2b and IgG2c repertoires, also suggests that sequential CSR may occur antigen independently in the llama, at least during some stage of B cell development. One possibility is that both SHM and CSR (both of which require expression of activation-induced cytidine deaminase) occur antigen independently in camelids as a mechanism of primary HCAb repertoire diversification. This could be investigated in future studies.

In conclusion, we used serological, immunophenotypic, and NGS analyses to assess the compartmentalization of llama Ab repertoires by isotype. We provide preliminary evidence describing the features of llama IgG1, IgG2b, and IgG2c repertoires as well as two mAbs that may be useful for future studies in this area. However, we caution that mAb 5E4's recognition of an Fc epitope that is at least partially conserved in some human and murine Abs complicates its use in multiplexed assays involving multiple Abs. The immunogenetic features of HCAbs (restricted IGHV gene usage; long CDR3s with high N-nucleotide content; potentially antigen-independent SHM and CSR; potentially functional "hingeless" HCAbs with closely-spaced antigen-combining sites) suggest that HCAbs use a unique set of genetic mechanisms and rely more heavily on non-genomically templated sequence diversity, to expand the antigen-binding repertoire in the absence of a paired light chain.

Acknowledgments We gratefully acknowledge the excellent technical assistance of Christine Gadoury and Mike Lowden. We thank Tammy-Lynn Tremblay for preparing LC-MS samples.

Funding information This work was funded by the National Research Council Canada.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

Achour I, Cavelier P, Tichit M, Bouchier C, Lafaye P, Rougeon F (2008) Tetrameric and homodimeric camelid IgGs originate from the same *IgH* locus. J Immunol 181:2001–2009

Alamyar E, Duroux P, Lefranc MP, Giudicelli V (2012) IMGT(®) tools for the nucleotide analysis of immunoglobulin (IG) and T cell receptor (TR) V-(D)-J repertoires, polymorphisms, and IG mutations: IMGT/V-QUEST and IMGT/HighV-QUEST for NGS. Methods Mol Biol 882:569–604

Baral TN, MacKenzie R, Arbabi Ghahroudi M (2013) Single-domain antibodies and their utility. Curr Protoc Immunol 103:Unit 2:17

Bodenhofer U, Bonatesta E, Horejs-Kainrath C, Hochreiter S (2015) msa: an R package for multiple sequence alignment. Bioinformatics 31:3997–3999

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120

Brooks CL, Rossotti MA, Henry KA (2018) Immunological functions and evolutionary emergence of heavy-chain antibodies. Trends Immunol 39:956–960

Conrath KE, Wernery U, Muyldermans S, Nguyen VK (2003) Emergence and evolution of functional heavy-chain antibodies in *Camelidae*. Dev Comp Immunol 27:87–103

Cox AD, St Michael F, Aubry A, Cairns CM, Strong PC, Hayes AC, Logan SM (2013) Investigating the candidacy of a lipoteichoic acid-based glycoconjugate as a vaccine to combat *Clostridium difficile* infection. Glycoconj J 30:843–855

Daley LP, Gagliardo LF, Duffy MS, Smith MC, Appleton JA (2005) Application of monoclonal antibodies in functional and comparative investigations of heavy-chain immunoglobulins in new world camelids. Clin Diagn Lab Immunol 12:380–386

de Jong BG, H IJ, Marques L, van der Burg M, van Dongen JJ, Loos BG, van Zelm MC (2017) Human IgG2- and IgG4-expressing memory B cells display enhanced molecular and phenotypic signs of maturity and accumulate with age. Immunol Cell Biol 95:744–752

Deschacht N, De Groeve K, Vincke C, Raes G, De Baetselier P, Muyldermans S (2010) A novel promiscuous class of camelid single-domain antibody contributes to the antigen-binding repertoire. J Immunol 184:5696–5704

Durocher Y, Perret S, Kamen A (2002) High-level and high-throughput recombinant protein production by transient transfection of suspension-growing human 293-EBNA1 cells. Nucleic Acids Res 30:E9–E99

Feige MJ, Grawert MA, Marcinowski M, Hennig J, Behnke J, Auslander D, Herold EM, Peschek J, Castro CD, Flajnik M, Hendershot LM, Sattler M, Groll M, Buchner J (2014) The structural analysis of shark IgNAR antibodies reveals evolutionary principles of immunoglobulins. Proc Natl Acad Sci U S A 111:8155–8160

Greenberg AS, Avila D, Hughes M, Hughes A, McKinney EC, Flajnik MF (1995) A new antigen receptor gene family that undergoes rearrangement and extensive somatic diversification in sharks. Nature 374:168–173

Griffin LM, Snowden JR, Lawson AD, Wernery U, Kinne J, Baker TS (2014) Analysis of heavy and light chain sequences of conventional camelid antibodies from *Camelus dromedarius* and *Camelus bactrianus* species. J Immunol Methods 405:35–46

Hamers-Casterman C, Atarhouch T, Muyldermans S, Robinson G, Hamers C, Songa EB, Bendahman N, Hamers R (1993) Naturally occurring antibodies devoid of light chains. Nature 363:446–448

Harmsen MM, Ruuls RC, Nijman IJ, Niewold TA, Frenken LG, de Geus B (2000) Llama heavy-chain V regions consist of at least four distinct subfamilies revealing novel sequence features. Mol Immunol 37:579–590

Henry KA (2018) Next-generation DNA sequencing of V_H/V_L repertoires: a primer and guide to applications in single-domain antibody discovery. Methods Mol Biol 1701:425–446

Henry KA, Hussack G, Collins C, Zwaagstra JC, Tanha J, MacKenzie CR (2016a) Isolation of TGF-β-neutralizing single-domain antibodies of predetermined epitope specificity using next-generation DNA sequencing. Protein Eng Des Sel 29:439–443

Henry KA, MacKenzie CR (2018) Antigen recognition by single-domain antibodies: structural latitudes and constraints. MAbs 10:815–826

Henry KA, Sulea T, van Faassen H, Hussack G, Purisima EO, MacKenzie CR, Arbabi-Ghahroudi M (2016b) A rational engineering strategy for designing protein A-binding camelid single-domain antibodies. PLoS One 11:e0163113

Henry KA, Tanha J, Hussack G (2015) Identification of cross-reactive single-domain antibodies against serum albumin using next-generation DNA sequencing. Protein Eng Des Sel 28:379–383

Holzlohner P, Butze M, Maier N, Hebel N, Schliebs E, Micheel B, Funer J, Heidicke G, Hanack K (2018) Generation of murine monoclonal antibodies with specificity against conventional camelid IgG1 and heavy-chain only IgG2/3. Vet Immunol Immunopathol 197:1–6

Klarenbeek A, El Mazouari K, Desmyter A, Blanchetot C, Hultberg A, de Jonge N, Roovers RC, Cambillau C, Spinelli S, Del-Favero J, Verrips T, de Haard HJ, Achour I (2015) Camelid Ig V genes reveal significant human homology not seen in therapeutic target genes, providing for a powerful therapeutic antibody platform. MAbs 7:693–706

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA (2009) Circos: an information aesthetic for comparative genomics. Genome Res 19:1639–1645

Lauwereys M, Arbabi Ghahroudi M, Desmyter A, Kinne J, Holzer W, De Genst E, Wyns L, Muyldermans S (1998) Potent enzyme inhibitors derived from dromedary heavy-chain antibodies. EMBO J 17:3512–3520

Li X, Duan X, Yang K, Zhang W, Zhang C, Fu L, Ren Z, Wang C, Wu J, Lu R, Ye Y, He M, Nie C, Yang N, Wang J, Yang H, Liu X, Tan W (2016) Comparative analysis of immune repertoires between Bactrian camel's conventional and heavy-chain antibodies. PLoS One 11:e0161801

Maass DR, Sepulveda J, Pernthaner A, Shoemaker CB (2007) Alpaca (*Lama pacos*) as a convenient source of recombinant camelid heavy chain antibodies (V_HHs). J Immunol Methods 324:13–25

Magoc T, Salzberg SL (2011) FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics 27:2957–2963

Manceur AP, Zou W, Marcil A, Paquet E, Gadoury C, Jaentschke B, Li X, Petiot E, Durocher Y, Baardsnes J, Rosa-Calatrava M, Ansorge S, Kamen AA (2017) Generation of monoclonal pan-hemagglutinin antibodies for the quantification of multiple strains of influenza. PLoS One 12:e0180314

Muyldermans S (2013) Nanobodies: natural single-domain antibodies. Annu Rev Biochem 82:775–797

Muyldermans S, Atarhouch T, Saldanha J, Barbosa JA, Hamers R (1994) Sequence and structure of V_H domain from naturally occurring camel heavy chain immunoglobulins lacking light chains. Protein Eng 7:1129–1135

Nguyen VK, Hamers R, Wyns L, Muyldermans S (1999) Loss of splice consensus signal is responsible for the removal of the entire C_H1 domain of the functional camel IgG2a heavy-chain antibodies. Mol Immunol 36:515–524

Nguyen VK, Hamers R, Wyns L, Muyldermans S (2000) Camel heavy-chain antibodies: diverse germline V_HH and specific mechanisms enlarge the antigen-binding repertoire. EMBO J 19:921–930

Nguyen VK, Muyldermans S, Hamers R (1998) The specific variable domain of camel heavy-chain antibodies is encoded in the germline. J Mol Biol 275:413–418

Nguyen VK, Su C, Muyldermans S, van der Loo W (2002) Heavy-chain antibodies in *Camelidae*; a case of evolutionary innovation. Immunogenetics 54:39–47

Pluciennik A, Stolarczyk M, Bzowka M, Raczynska A, Magdziarz T, Gora A (2018) BALCONY: an R package for MSA and functional compartments of protein variability analysis. BMC Bioinformatics 19:300

Rast JP, Amemiya CT, Litman RT, Strong SJ, Litman GW (1998) Distinct patterns of *IgH* structure and organization in a divergent lineage of chrondrichthyan fishes. Immunogenetics 47:234–245

Saccodossi N, De Simone EA, Leoni J (2012) Structural analysis of effector functions related motifs, complement activation and hemagglutinating activities in *Lama glama* heavy chain antibodies. Vet Immunol Immunopathol 145:323–331

Schmieder R, Edwards R (2011) Quality control and preprocessing of metagenomic datasets. Bioinformatics 27:863–864

Sun Y, Gadoury C, Hirakawa MP, Bennett RJ, Harcus D, Marcil A, Whiteway M (2016) Deletion of a Yci1 domain protein of *Candida albicans* allows homothallic mating in MTL heterozygous cells. MBio 7:e00465–e00416

van der Linden R, de Geus B, Stok W, Bos W, van Wassenaar D, Verrips T, Frenken L (2000) Induction of immune responses and molecular cloning of the heavy chain antibody repertoire of *Lama glama*. J Immunol Methods 240:185–195

Vu KB, Ghahroudi MA, Wyns L, Muyldermans S (1997) Comparison of llama V_H sequences from conventional and heavy chain antibodies. Mol Immunol 34:1121–1131

Woolven BP, Frenken LG, van der Logt P, Nicholls PJ (1999) The structure of the llama heavy chain constant genes reveals a mechanism for heavy-chain antibody formation. Immunogenetics 50:98–101

Zou X, Osborn MJ, Bolland DJ, Smith JA, Corcos D, Hamon M, Oxley D, Hutchings A, Morgan G, Santos F, Kilshaw PJ, Taussig MJ, Corcoran AE, Bruggemann M (2007) Heavy chain-only antibodies are spontaneously produced in light chain-deficient mice. J Exp Med 204:3271–3283