

RESEARCH ARTICLE

# A comprehensive analysis of immunoglobulin heavy chain genes in the Bactrian camel (*Camelus bactrianus*)

Zuoxiang LIANG<sup>1,2\*</sup>, Tao WANG<sup>2\*</sup>, Yi SUN<sup>3</sup>, Wenlong YANG<sup>1</sup>, Zhihong LIU<sup>4</sup>, Jing FEI<sup>2</sup>, Ying GUO<sup>2</sup>,  
Qingwei MA<sup>5</sup>, Qingjie PAN (✉)<sup>1</sup>, Liming REN (✉)<sup>2</sup>

<sup>1</sup> Key Laboratory of Animal Reproduction and Germplasm Enhancement in Universities of Shandong, College of Animal Science and Technology, Qingdao Agricultural University, Qingdao 266109, China

<sup>2</sup> State Key Laboratory of Agrobiotechnology, China Agricultural University, Beijing 100193, China

<sup>3</sup> Laboratory of Animal Molecular Genetics, College of Animal Science and Veterinary Medicine, Shandong Agricultural University, Taian 271018, China

<sup>4</sup> Key Laboratory of Animal Genetics, Breeding and Reproduction, Inner Mongolia Agricultural University, Hohhot 010018, China

<sup>5</sup> Bioyong Technology Inc., Beijing 100080, China

**Abstract** Heavy chain only antibodies (HCAbs) represent a rare type of antibody that is devoid of light chains and the CH1 domain that have been reported in cartilaginous fish and camelids. By analyzing transcript data and genome sequences, we conducted a comprehensive analysis of Bactrian camel immunoglobulin heavy chain genes. Based on the transcript data, one  $\mu$  gene, five  $\gamma$  genes, one  $\alpha$  gene and one  $\varepsilon$  gene were found. Additionally, the variable region of HCAbs (VHH) and the conventional antibodies (VH) sequences associated with the  $\gamma 3$ ,  $\gamma 1a/b$  and  $\mu$  genes were amplified. Based on these genome sequences, seven *DH*, six *JH*,  $\mu$ ,  $\gamma 2a$ ,  $\gamma 2c$ ,  $\alpha$ , and  $\varepsilon$  genes and a portion of a  $\gamma 3$  gene were observed. Different Kozak sequences within different VH families were found in our analysis, and the variability index differed between the VHH3 and VH3 families. Phylogenetic analysis of the constant regions of the camelid immunoglobulin genes indicates that these genes appeared before the evolutionary divergence of Bactrian camels and dromedaries.

**Keywords** Bactrian camel, heavy-chain antibodies, VHH,  $\gamma 3$

## 1 Introduction

Innate immunity is the first line of defense against pathogenic microorganisms. However, when pathogens break this early defense, the adaptive humoral immune system produces specific antibodies (Abs) from terminally differentiated B lymphocytes, which bind to specific targets or antigens derived from the pathogen to defend against further infection. At the same time, long-term memory is developed for this pathogen, such that it will be recognized upon reinfection. The structure of conventional antibodies, which are assembled from two identical large molecular weight heavy (H)-chain and two identical small molecular weight light (L)-chain polypeptides, is highly conserved. The two heavy chains are linked to each other by disulfide bonds, and each heavy chain is linked to a single light chain by another disulfide bond<sup>[1]</sup>. In camelids, rather unusually, HCAbs are found, which are completely devoid of light chains. The heavy chain is also missing its first constant domain (CH1)<sup>[2]</sup>. These HCAbs have also been found to occur in humans in the early years, where large and various parts of the VH and CH1 can be absent, leading to antibodies devoid of light chains<sup>[3]</sup>. New antigen receptors (NARs) are similar to camelid HCAbs and have been reported in nurse and wobbegong sharks and in spotted ratfish<sup>[4]</sup>. These IgNARs exhibit structural and functional convergent evolution with camelid HCAbs<sup>[5]</sup>.

HCAbs are highly stable and are more stable than conventional antibodies in some cases. These antibodies are able to respond to antigens at high temperatures and at high concentrations of denaturants<sup>[6]</sup>. Sequence and structure analyses have revealed several HCAb features

Received April 16, 2015; accepted May 19, 2015

Correspondences: 13969675699@126.com, renlm@cau.edu.cn

\*These authors contributed equally to the work

that are distinct from those of their conventional counterparts. First, the highly conserved hydrophobic amino acids, Val42, Gly49, Leu50 and Trp52, which occur at the second framework (FR2) in typical VH domains, are frequently substituted by the residues Phe42 (or Tyr42), Glu49, Arg50 (or Cys50), and Gly52, in the VHH domain<sup>[7]</sup>. The side chains of these four residues in the FR2 region of the VHH domain deviate from the corresponding positions in the VH domain, and this more hydrophilic surface abrogates the conserved, hydrophobic binding with a VL<sup>[8]</sup>. Furthermore, these amino acid substitutions might enhance the solubility of the isolated VHH domain<sup>[6]</sup>. The absence of a light chain in VHH is compensated for by an extended complementarity-determining region-1 (CDR1) and a longer CDR3. Due to the increased length of the H3 loop, the antigen binding surface is enlarged in the VHH domain. This added flexibility enables greater diversity despite the lack of VL<sup>[9]</sup>. In addition, unlike in VHs, an interloop disulfide bond, composed of two cysteines, is frequently observed between the CDR1 and CDR3 in dromedary VHHs. This interloop disulfide bond plays an important role in reducing the conformational flexibility of the long CDR3 loop<sup>[10]</sup>. Finally, CH1, which typically forms covalent bonds with CL domains in conventional Abs (cAbs), is present in the genome of the heavy chain isotype for the HCAs but is absent after mRNA splicing. This finding might be due to the presence of a point mutation at the conserved GT in the splice signal at the 3' end of the CH1 exon to AT<sup>[11]</sup>. Due to a relatively long and protruding H3 loop and the absence of light chains, HCAs can interact with concave antigen binding surfaces and might be better adapted to target epitopes that are normally inaccessible for conventional antibodies<sup>[12]</sup>. Due to the long CDR3, HCAs can also insert into the active site of certain enzymes to inhibit their activity<sup>[13]</sup>.

Bactrian camels are phylogenetically very close to dromedaries and are important livestock in north-west China. Whole-genome sequencing of the Bactrian camel has been reported<sup>[14,15]</sup>. The existence of partial immunoglobulin genes and a comparative analysis of conventional Abs in the Bactrian camel and the dromedary have also been reported<sup>[16]</sup>, but the analysis of HCAs in Bactrian camels was not comprehensive. In this study, we provide a more thorough analysis of Bactrian camel IgH genes based on our own sequence data and on previous studies.

## 2 Materials and methods

### 2.1 Animals, RNA isolation, and reverse transcription

The spleen of a single camel (*Camelus bactrianus*) was obtained from Xilinhaote abattoir, Inner Mongolia, China. Total RNA was isolated from spleen tissue using a TRIzol

kit (CWBio, Beijing, China) following the manufacturer's instructions and was treated with RNase-free DNase I (Qiagen, Beijing, China). Reverse transcription was conducted using Moloney Murine Leukemia Virus Reverse Transcriptase according to the manufacturer's instructions (Promega, Madison, WI, USA). The oligo (dT) adapter primer *Not* I-d (T)18 was used (Appendix A, Table S1).

### 2.2 Amplification of the IgM portion of the CH1 domain using gene-specific primers

Using the previously published IgM CH1 genomic sequences of dromedary and alpaca<sup>[14,17]</sup>, we designed two gene-specific primers from conserved sequences in both species. RT-PCR using these primers, plus Bactrian camel spleen cDNA and *Taq* DNA polymerase (CWBio, Beijing, China) was conducted under the following conditions: 94°C for 5 min; 35 cycles of 94°C for 30 s, 60°C for 30 s, and 72°C for 30 s; and a final extension at 72°C for 7 min. The resultant PCR product (approximately 200 bp) was cloned into the pMD19-T vector (Takara) for sequencing.

### 2.3 5' and 3' RACE PCR reaction and construction of Bactrian camel mini-IgH cDNA libraries

Primers derived from IgHM (IgM heavy chain) constant regions were used for 5' RACE amplification, with spleen RNA as the template. The 5' RACE System kit was used (Version 2.0, Invitrogen, NY, USA). After screening with the most frequently used IgH (immunoglobulin gene heavy chain) joining-segment coding sequence, 3' RACE (Invitrogen, NY, USA) was used. Every expressed IgG sub-isotype, IgM and IgE was screened with specific primers that were designed based on their unique regions. A second 5' RACE System (Invitrogen, NY, USA) was used to acquire diversity specific non-redundant VH and VHH sequences, which belong to IgG1a/b and IgG3. The resultant PCR products were cloned into a pMD19-T vector and sequenced. First-strand cDNA was synthesized using *Not* I-d (T) 18 primers (Appendix A, Table S1). One pair of primers, IgHMC1SP and IgHMC1AP, was designed using alpaca conserved IgM CH regions. Primers IgHMGSP1 and IgHMGSP2 were used to perform the 5' RACE, whereas primers J<sub>H</sub>SP and RT-P1 were used to perform the 3' RACE. Primers IgHMCSP and IgHMCAP were used to screen for IgM clones, and primers IgHGCAP, IgHG1aHSP, IgHG1bHSP, IgHG2aHSP, IgHG2cHSP and IgHG3HSP were used to screen for different isotypes of IgG.

According to the previously published constant domain sequences of alpaca and dromedary IgE, J<sub>H</sub>SP, IgHECAP, IgHECSP and RT-P1 were used to amplify the IgE constant region.

To investigate the different variable region libraries derived from HCabs and cAbs, two additional 5' RACE experiments were performed. Primer IgHGVGSP1 was designed to amplify the CH2 region. In one experiment, the  $\gamma 1a$  and  $\gamma 1b$  gene constant sequences were aligned, and a primer was designed based on the CH1 identity region (IgHG1a/bVGSP2). In the other 5' RACE experiment, the specific primer IgHG3VGS-P2 was designed based on the hinge region encoded by the  $\gamma 3$  gene.

## 2.4 Genome sequence confirmation

High molecular weight genomic DNA was isolated from spleen using a standard phenol-chloroform extraction method. A sense primer (IgHGC1SP) was designed based on the Bactrian camel conserved IgG CH1 region (Appendix A, Table S1); IgHG1aHAP, IgHG1bHAP, IgHG2aHAP, IgHG2cHAP and IgHG3VGSP2 (anti-sense primers) were used to amplify part of the CH1 exon and part of the hinge exon. The resultant PCR products were cloned into pMD19-T and sequenced.

## 2.5 Sequence computations

Homologous sequences were identified using the Basic Local Alignment Search Tool (BLAST) and the National Center for Biotechnology Information database. DNA and protein sequence editing, alignments, and comparisons were performed using Bioedit software. Multiple sequence alignments were performed using ClustalW, and MEGA 5.0 was used to assist the software. FUZZNUC software was used to search for the RSS signal on the genome sequences. Statistical analysis was performed using the R language and SPSS 21. Images were prepared using Origin 8.6, Photoshop 12.0.1 and GeneDoc 2.6.0.2. The online software ExPASy and WEBLOGO (<http://weblogo.berkeley.edu/logo.cgi>) were used to analyze the acquired database.

## 2.6 Phylogenetic analysis

Phylogenetic trees were generated using MrBayes3.1.2 and were viewed in FigTree1.4.1. A phylogenetic tree of mammalian *VH* genes was constructed using the nucleotide sequences of FR1-FR3 as defined by the IMGT numbering system. One representative *VH* sequence was selected for each of the *VH* families based on our study of the IgH-specific mini-cDNA libraries in Bactrian camel. Functional or potentially functional germline sequences and cDNA sequences in other species were included. Nurse shark *VH1* was used as an outgroup (Appendix A, Table S2). Another phylogenetic tree of Bactrian camel IgH constant genes was constructed using amino acid sequences. Nurse shark IgM was used as an outgroup in this instance (Appendix A, Table S2).

# 3 Results and discussion

## 3.1 Construction of IgH-specific mini-cDNA libraries

### 3.1.1 Construction of the variable region mini-cDNA library of IgM

The variable region library of IgM was investigated using 5' RACE of spleen RNA that was isolated from one Bactrian camel. The conserved CH1 fragment was sequenced, and 98% identity was confirmed between alpaca and Bactrian camel by BLAST analysis. Two hundred and 12 unique clones were further analyzed after removing duplicate, incomplete and nonfunctional sequences.

### 3.1.2 Construction of immunoglobulin heavy chain gene constant region libraries

A 3' RACE experiment was performed to obtain the maximum number of different immunoglobulin heavy chain gene isotypes. The most common JH sequence, 5'-GGCCAGGGGACCCAGGTCACCGTCTCCTCAGAG-3', was used to design the sense primer for subsequent experiments. This experiment resulted in two PCR fragments (approximately 1000 and 1600 bp), which were subjected to further PCR analysis and sequencing. One  $\mu$ -encoding gene, five  $\gamma$ -encoding genes and one  $\alpha$ -encoding gene were screened from two IgH-specific mini-libraries, as described above. We also identified one  $\epsilon$ -encoding gene. According to the published literature and the BLAST alignment, five  $\gamma$ -encoding genes exhibit obvious differences in the hinge region and are considered to generate two conventional Abs ( $\gamma 1a$  and  $\gamma 1b$ ) and three HCabs ( $\gamma 2a$ ,  $\gamma 2c$  and  $\gamma 3$ ) because their cDNA was devoid of the entire CH1-exon<sup>[18]</sup>.

### 3.1.3 Construction of variable region libraries derived from HCab and cAb

One hundred and ninety-two unique clones representing IgG3 and 188 unique clones representing IgG1a/b were used for further analysis.

## 3.2 Analysis of heavy chain variable region sequences

### 3.2.1 Family classification and phylogenetic analysis of VH/VHH segments

Phylogenetic analysis revealed two *VH* families associated with IgM, according to the 75% identity criterion<sup>[19]</sup>. Among 212 unique sequences, 169 exhibited homology to the human *VH3* family, and the remaining 43 exhibited homology to the human *VH4* family. To analyze the variable region cDNA mini-library of IgG1a/b, 188 unique

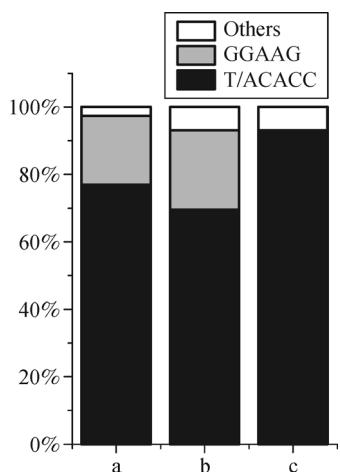


**Fig. 1** Five VH amino acid sequences were found among 192 HCABs using a cDNA database. The numbering used refers to the position in the locus in the 5' to 3' direction. V region sequences were analyzed using the multiple sequence alignment MEGA 5.0 software. CDR (reseda shading) and FR (black and white box) regions of the V genes are determined according to IMGT nomenclature and are indicated at the top. Purple and wathet shaded amino acid sequences represent hallmarks for the VHH and VH genes, respectively.

sequences were divided into two families; 139 sequences were allocated to the VH3 family, and the remaining sequences were allocated to the VH4 family.

In the variable region cDNA mini-library of IgHG3, all sequences exhibited homology to the human VH3 family. However, 187 sequences were considered VHH3 sequences because four hallmark amino acids were found in FR2<sup>[7]</sup>. The remaining 5 clones exhibited conventional V gene features (Fig. 1). A previous study reported that HCABs that contain normal VH sequences are functionally active against pathogens<sup>[20]</sup>.

When analyzing the VH3, VHH3 and VH4 genes, we found that the Kozak sequences were highly conserved in these sequences. Statistical analysis of the results revealed that 97.6%, 92.6% and 92.7% clones of these three gene families, respectively, contain T/ACACC or GGAAG in the corresponding positions (Fig. 2).



**Fig. 2** Five conserved bases located at the 5' UTR end of the heavy chain variable domains of IgM, IgG1a/b and IgG3. (a) Variable region sequences of IgHM; (b) variable region sequences of IgHG1a/b; (c) variable region sequences of IgHG3.

As reported in previous genomic analyses, the V-regions containing the MELG leader sequence belong to the VH3 family, whereas the MRLG leader sequence is associated with V-regions of VH4 family members<sup>[21]</sup>. This was also observed in our study (Appendix A, Table S3).

Furthermore, we observed that the Kozak sequence is unique to individual VH families and is significantly related to individual leader sequences. Clones containing the Kozak sequence T/ACACC were all VH3 family members and contained the MELG leader signal. Clones containing the GGAAG sequence were generally VH4 family members (only one member of the VH3 family contained GGAAG) and contained the MRLG leader signal (Appendix A, Table S4).

A phylogenetic tree of VH genes showed that the Bactrian VH3 family was closely related to the dromedary VH3 family and that both belong to clan III (Fig. 3). Bactrian VHH3 sequences also clustered most closely with clan III but were distinct from the Bactrian VH3 family. VH4 clones clustered most closely with the dromedary VH4 family and fell into clan II.

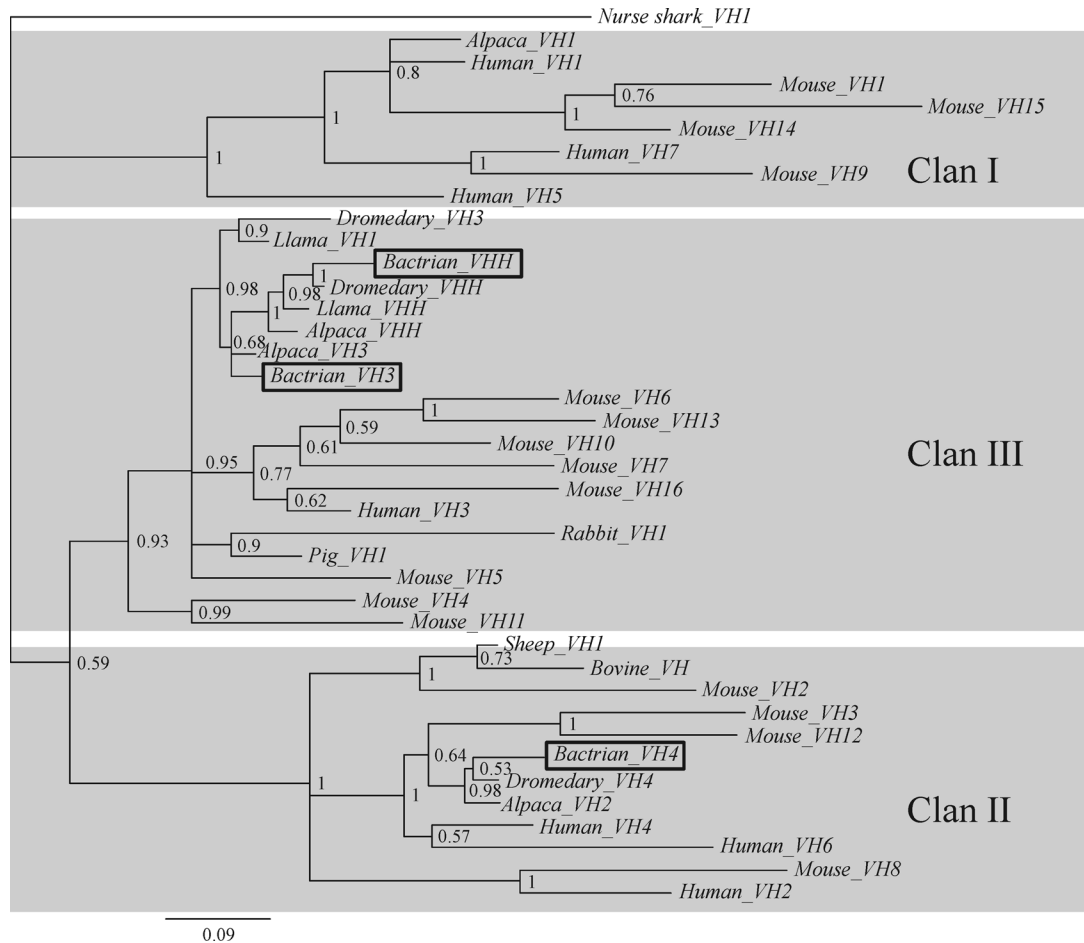
### 3.2.2 Hypervariability of Bactrian camel VH/VHH cDNA

Based on the special structure and function of VHH3 genes, the sequences of VH3 clones from IgM or IgG1a/b libraries were compared separately to VHH3 sequences from IgG3. The most frequently occurring sequences were chosen, and the sizes of both CDR1 and CDR2 were not more than eight residues. An additional highest variability index value was observed in the CDR1 of the VHH3 sequences was higher than that for the VH3 sequences derived from IgM and IgG1a/b at residues 27–30 (Fig. 4). This result was an agreement with previously published data that were based on the Kabat system of amino acid numbering, showing that an additional hypervariable region was present in the CDR1 of the VHH sequences<sup>[22]</sup>. However, no major differences were found in the VH3 sequences associated with IgM and IgG1a/b (Fig. 4). Two-way ANOVA analysis of the variability index revealed that the VHH3 sequences were significantly different from the VH3 sequences of IgM or IgG1a/b (Appendix A, Table S5).

### 3.2.3 Analysis of CDR3 length

According to the IMGT CDR definition, the CDR3 length derived from VHH3 ranged from 9 to 31 aa (one clone





**Fig. 3** Phylogenetic analysis of Bactrian camel VH families. Nurse shark VH1 was used as the outgroup. The scale bar shows the genetic distance, and the credibility value is noted beside each node. Shaded boxes represent the clustering of sequences into clans. With the exception of the Bactrian camel sequences, all VH sequences were obtained from IMGT.

CDR3 length was 5 aa) but was most commonly between 18 and 22 aa (Fig. 5). The average CDR3 derived from HCAs (20 aa) was longer than the corresponding region of VH3 sequences derived from IgM (12.7 aa) or IgG1a/b (13.8 aa). The average length of the Bactrian camel CDR3 domains in VHH3 was longer than that of the VHHs in dromedary (15 aa), llama (14.9 aa) or alpaca (17.8 aa)<sup>[23]</sup>. No major differences were detected in CDR3 length between the VH4 sequences from IgM (ranging from 10 to 23 aa) and IgG1a/b (ranging from 10 to 24 aa).

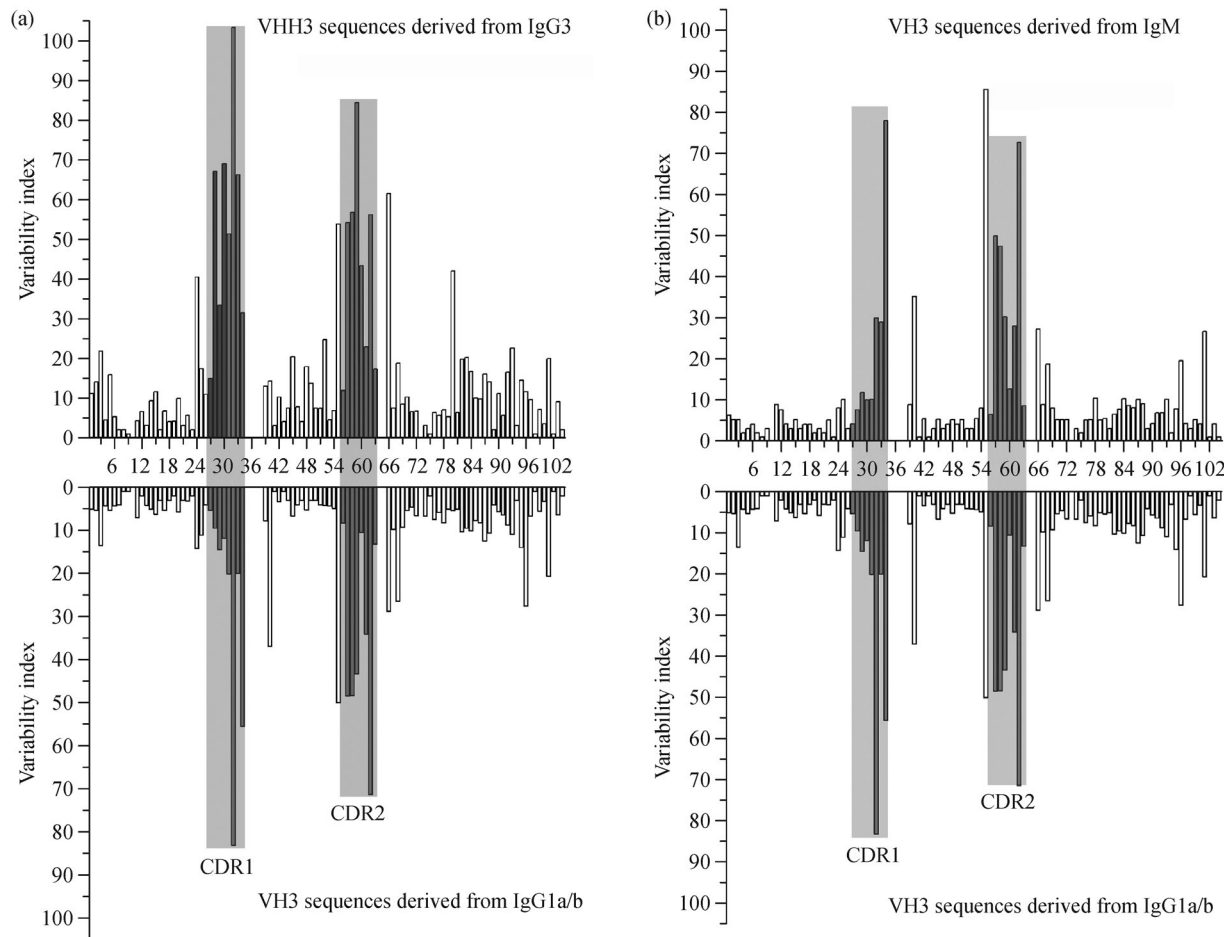
### 3.2.4 Analysis of the potential interloop disulfide bond

Two cysteine residues (one of which is usually located in the CDR1 region and the other of which is located in the CDR3 region) were included in a number of VHH3 sequences and are known to form an interloop disulfide bond that enhances structural stability. Platypus and shark antibodies also contain an interloop disulfide bond, which is located in the variable domains, to restrict the flexibility of the CDR loops<sup>[24,25]</sup>; however, this extra disulfide bond

is observed less often in llama VHH<sup>[26]</sup>. We found that 133 sequences included two conserved Cys at the corresponding positions and that 59 VHH3 sequences contained no more than one Cys. A chi-square test was used to investigate the variability of the CDR3 length in relation to the presence or absence of the interloop disulfide bond. No correlation was found between CDR3 length and the presence of the interloop disulfide bond (Appendix A, Table S6). Among 212 VH sequences obtained from IgM, only two sequences contained two conserved Cys at the corresponding positions. No potential interloop disulfide bond was found among 188 VH sequences from IgG1a/b.

### 3.2.5 Analysis of positions that are critical for the variable domain immunoglobulin fold

A statistical analysis of the key positions that shape the structure of antibodies<sup>[27]</sup> was performed using WebLogo<sup>[28]</sup>. An alignment of the amino acid sequences of VH3DJ and VHH3DJ revealed that nearly all residues



**Fig. 4** Comparison of the variability of the VH3 sequences derived from IgM or IgG1a/b with IgG3 VHH3 sequences. (a) Amino acid sequence variability plots for the cDNA sequences of VHH3 derived from IgG3 (upper left panel) and VH3 derived from IgG1a/b (lower panels). The variability index was calculated from FR1 to FR3 as described in the results section. Grey and open bars represent amino acids in the CDR and FR regions, respectively, and dark gray bars represent residues 27–30 in the VHH, which we postulate is the extra hypervariable region that has been reported in the literature using Kabat nomenclature. CDR1 and CDR2, as defined by IMGT nomenclature, are shaded and placed between the dotted lines; (b) amino acid sequence variability plots comparing the sequences predicted from cDNA sequences of VH3 derived from IgM (upper right panel) and VH3 derived from IgG1a/b (lower panels).

that are important for the variable domain structure are conserved. As indicated in Fig. 6, there is no major difference between the VH3DJ sequences from IgM and IgG1a/b. However, a clear difference was observed between the VH3DJ and VHH3DJ sequences (Fig. 6), in particular at positions 12, 15, 24, 30, 42, 49, 50, 52, 54, 96 and 106. Previous reports have shown that VH3 and VHH3 are distinguished by substitutions from highly conserved hydrophobic amino acids (Val42, Gly49, Leu50, Trp52) in classical Abs to amino acids (Phe or Tyr 42, Glu49, Arg or Cys 50, Gly or Phe52) in HCAs; this is consistent with our results (Fig. 6).

### 3.3 Analysis of *DH* and *JH* genes

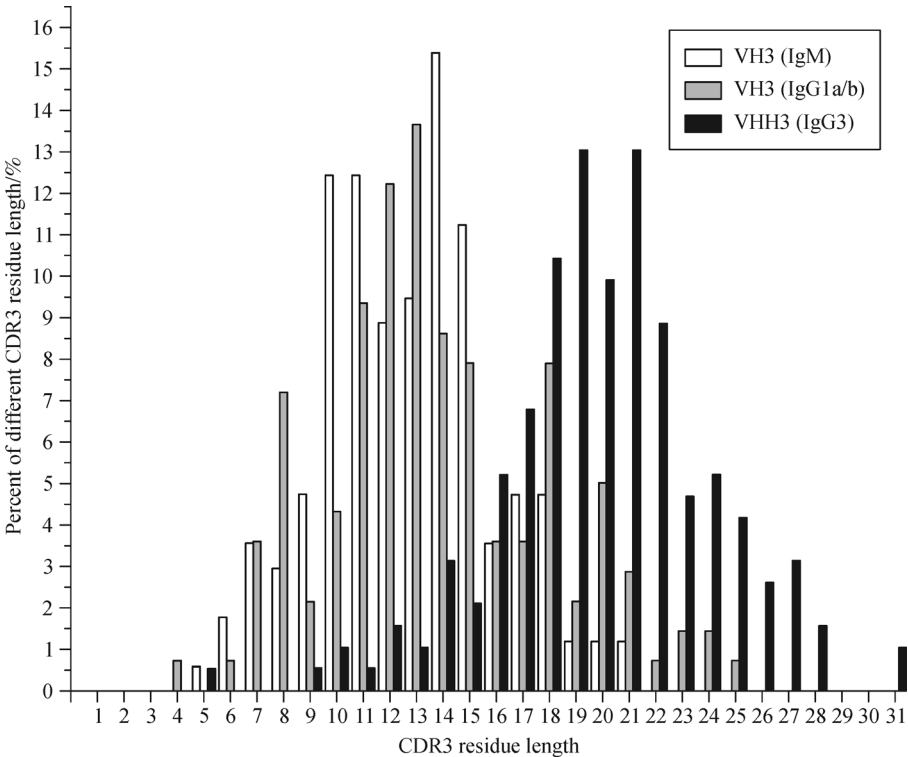
Recently, the whole genome of the Bactrian camel has been published<sup>[15]</sup>, and alpaca *DH* and *JH* genes and Bactrian *JH* sequences were used to search the Bactrian

camel genome<sup>[17]</sup>. Seven *DH* sequences and six *JH* genes were located in scaffold 131 (accession number: NW\_011509943) (Fig. 7). Analysis of these *DH* and *JH* sequences indicated that a high homology exists between Bactrian camel and alpaca, and the same *DH* and *JH* genes were observed (Fig. 7). We speculated that these genes might be derived from a common ancestor. Due to the occurrence of somatic hypermutation in these regions, the discrimination of specific *DH/JH* genes was difficult. However, *JH3* and *JH5* genes occurred most frequently.

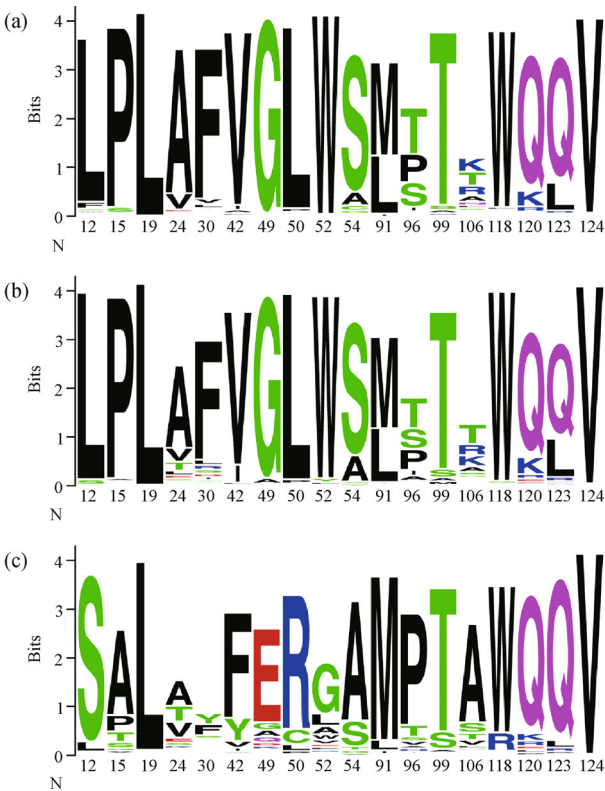
### 3.4 Structural and phylogenetic analysis of the heavy chain constant region

#### 3.4.1 $\mu$ gene

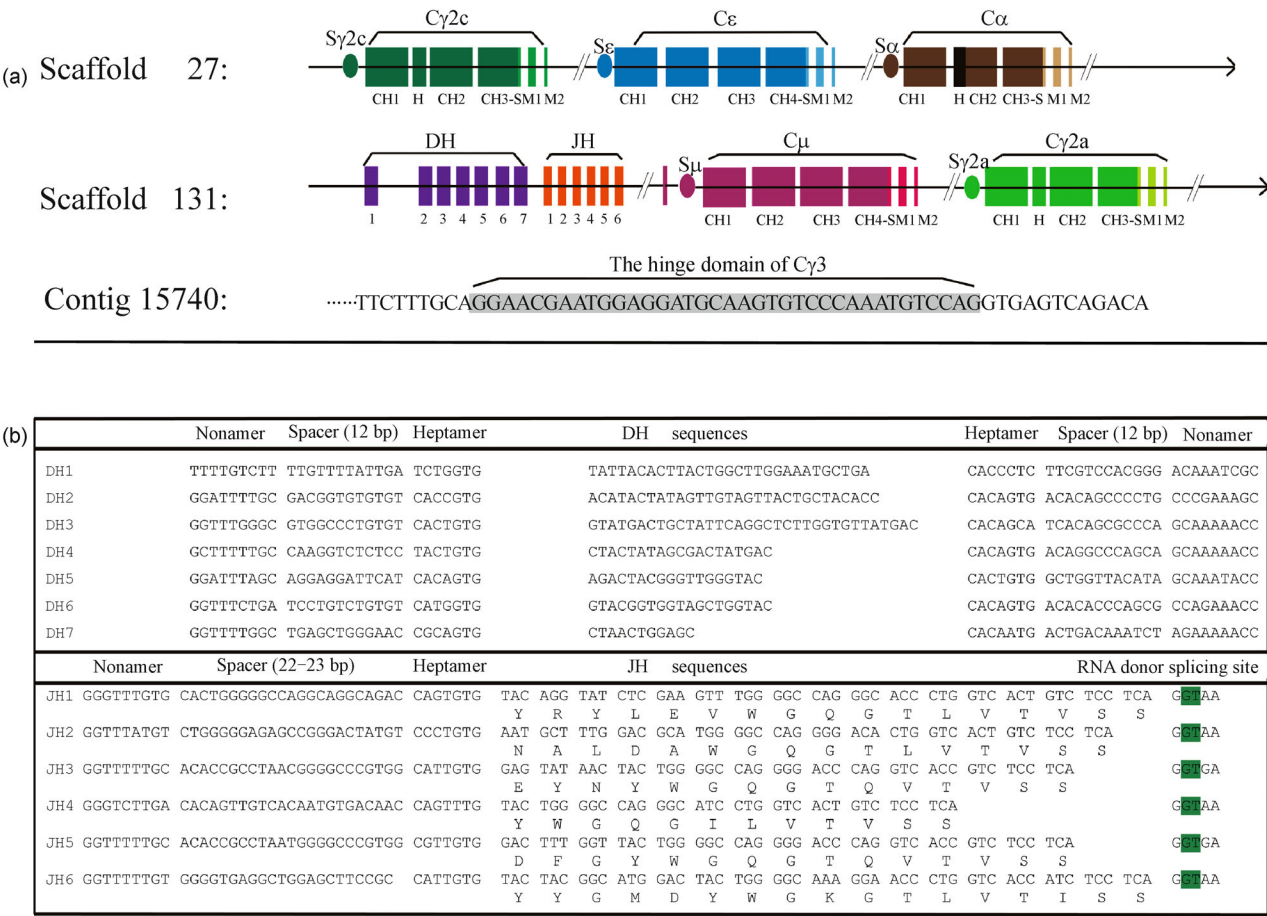
The Bactrian camel  $\mu$  gene was found in scaffold 131 (Fig. 7) of the whole-genome sequence, and the secreted



**Fig. 5** Frequency distribution histograms of heavy chain CDR3 lengths of the VH3 and VHH3 families. CDRs are defined according to the IMGT nomenclature.



**Fig. 6** WebLogo plots presenting sequence differences between the VHH3 domains derived from IgG3 and the VH3 domains derived from IgM and IgG1a/b. (a) VH3 sequences derived from IgM; (b) VH3 sequences derived from IgG1a/b; (c) VHH3 sequences derived from IgG3.



**Fig. 7** Bactrian camel immunoglobulins: structure and gene organization. (a) Details of the Bactrian camel *IgH* gene organization and the domain structure of each immunoglobulin heavy chain gene constant region. (b) Nucleotide/amino acid sequences of Bactrian camel *DH* and *JH* genes.

form was obtained using 3' RACE. Sequence alignment showed that this gene had overall amino acid sequence identities of 64% and 60% to the human and mouse  $\mu$  genes and of 97% and 95% to the dromedary and alpaca  $\mu$  genes, respectively. The identity of the Bactrian camel  $\mu$  gene was also confirmed by phylogenetic analysis (Fig. 8), the results of which suggested that the Bactrian camel  $\mu$  gene was most closely related to its counterparts in dromedary and alpaca.

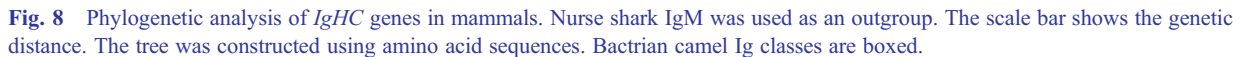
The alignment of the  $\mu$  gene between Bactrian camel and other mammals showed that the distribution of cysteine and tryptophan residues is conserved. The Cys residues that form intra- and inter-H chain disulfide bonds and the covalent bonding of the L chains were also conserved (Supplemental Fig. S1). The Bactrian camel  $\mu$  gene also presented conserved Cys residues that were found in CH3 and a secretory tail that covalently polymerizes pentameric IgM in mammals<sup>[29]</sup>. Utilizing ExPASy software, five putative N-linked glycosylation sites (N-X-S/T) were discovered at the same position in Bactrian camel and dromedary, and one of these sites was highly conserved in camelids.

3.4.2  $\gamma$  genes

Five isotypes of  $\gamma$  genes, which can be distinguished based on their hinge regions, were screened from 531 clones within the CH mini-libraries. Comparison of the Bactrian camel  $\gamma 1a$  gene with the corresponding alpaca sequence showed 94.4% nucleotide sequence identity. Comparison of the Bactrian camel  $\gamma 1b$  gene with the corresponding alpaca sequences showed 95.8% sequence identity. By analyzing the genome sequences of the Bactrian camel,  $\gamma 1a$  and  $\gamma 1b$  were not located in any scaffolds<sup>[15]</sup>; however,  $\gamma 2a$  and  $\gamma 2c$  were discovered in scaffold 27 (NW\_011517098) and in scaffold 131. The hinge domains of Bactrian camel IgG2a and IgG2c were identical to those of dromedary IgG2a and IgG2c, respectively (Fig. 9).

The hinge exon of IgG3 was found in contig 15740 (JARL01015740) (Fig. 7). One hundred percent nucleotide sequence identity was found with the  $\gamma 3$  sequence that was screened from the cDNA mini-library. Comparison of Bactrian camel  $\gamma 3$  gene with the corresponding dromedary and alpaca genes indicated that two amino acids differed (GG to EV) in the hinge region.





To confirm whether the mutations that have occurred on the bactrian camel genome corresponded to regions that have been reported in dromedary<sup>[11]</sup>, the CH1-Hinge genomic sequences of all five  $\gamma$  genes were amplified from a bactrian camel. As indicated in Fig. 8, a comparison of two conventional Abs ( $\gamma 1a$  and  $\gamma 1b$ ) with three other Bactrian camel genes ( $\gamma 2a$ ,  $\gamma 2c$  and  $\gamma 3$ ) showed that a mutation has occurred at the canonical splice site that is located at the 3' CH1/intron border, changing GTAAG to ATAAG (Fig. 9).

Of the sequenced 3' RACE products, some were highly homologous to human and alpaca IgA, with 62.8% and 91.3% nucleotide sequence identities, respectively. The phylogenetic tree also indicated that this sequence represented the Bactrian camel  $\alpha$  gene (Fig. 8).

Based on a previously published alpaca IgE heavy chain constant sequence<sup>[17]</sup>, we confirmed that the Bactrian camel  $\varepsilon$  gene shared 98.1% and 55.2% nucleotide identity with the corresponding regions in alpaca and humans, respectively. According to the phylogenetic tree, the Bactrian camel  $\varepsilon$  gene is most closely related to the alpaca  $\varepsilon$  gene.

Species	Gene	CH1 3' end	Intron	Hinge region
Alpaca	<i>γ1a</i>			: ELKTP-----QPQSQP-----ECRCPKCP : (19)
Llama	<i>γ1a</i>			: ELKTP-----QPQSQP-----ECRCPKCP : (19)
<b>Bactrian</b>	<b><i>γ1a</i></b>	<b>AAGCGTGTGG</b>	<b>CTAAGTAG</b>	: <b>ELEIS</b> ----- <b>EPQQP</b> ----- <b>GCTCPKCP</b> : (19)
Dromedary	<i>γ1a</i>			: ELEIS-----EPQQP-----GCTCPKCP : (19)
Alpaca	<i>γ1b</i>			: -----EPHG-----GCTCPQCP : (12)
Llama	<i>γ1b</i>			: -----EPHG-----GCTCPQCP : (12)
<b>Bactrian</b>	<b><i>γ1b</i></b>	<b>AAGCGTGTGG</b>	<b>CTAAGTAG</b>	: ----- <b>EPHG</b> ----- <b>GCPCKPCK</b> : (12)
Dromedary	<i>γ1b</i>			: -----EPHG-----GCTCPKCP : (12)
<b>Bactrian</b>	<b><i>γ2a</i></b>	<b>AAGCGTGTGG</b>	<b>ATAAGTAG</b>	: <b>EPKIPQPKPKPQPKPQPKPKPKEPECTCPKCP</b> : (35)
Dromedary	<i>γ2a</i>			: EPKIPQPKPKPQPKPQPKPKPKEPECTCPKCP : (35)
Llama	<i>γ2a</i>			: EPKIPQPKPKPQPKPQPKPKPKEPECTCPKCP : (35)
Alpaca	<i>γ2b</i>			: EPKTP---KPQPKPQPKPQPKPNTTES--KCPKCP : (29)
Llama	<i>γ2b</i>			: EPKTP---KPQPKPQPKP--NPTAES--KCPKCP : (27)
Alpaca	<i>γ2c</i>			: -AHHS-----EDPSS-----KCPKCP : (15)
Llama	<i>γ2c</i>			: -AHHS-----EDPSS-----KCPKCP : (15)
<b>Bactrian</b>	<b><i>γ2c</i></b>	<b>AAGCGTGTGG</b>	<b>ATAAGTAG</b>	: <b>AHHP</b> ----- <b>EDPSS</b> ----- <b>QCPKCP</b> : (15)
Dromedary	<i>γ2c</i>			: -AHHP-----EDPSS-----QCPKCP : (15)
<b>Bactrian</b>	<b><i>γ3</i></b>	<b>AAGCGTGTGG</b>	<b>ATAAGTAG</b>	: -----GTNG-----GCKCPKCP : (12)
Dromedary	<i>γ3</i>			: -----GTNE-----VCKCPKCP : (12)
Llama	<i>γ3</i>			: -----GTNE-----VCKCPKCP : (12)

**Fig. 9** The 3' CH1 boundary and hinge region analysis in  $\gamma$  genes. The point mutation destroying the conventional splicing signal is shaded in gray. The amino acids encoded by the hinge region are listed at the right, and the size of the hinge (number of amino acids) is given within brackets. Bactrian camel sequences are in bold font. The  $\gamma1a$  and  $\gamma1b$  genes encode conventional Abs, whereas the  $\gamma2a$ ,  $\gamma2b$ ,  $\gamma2c$  and  $\gamma3$  genes encode HCAs.

## 4 Conclusions

In summary, we analyzed the Bactrian camel immunoglobulin heavy chain genes and obtained new knowledge regarding their structure. Phylogenetic analysis indicated that Bactrian camel immunoglobulin genes are similar to dromedary genes.

**Note** The sequences presented in this article have been submitted to the National Center for Biotechnology Information's GenBank database (<http://www.ncbi.nlm.nih.gov/nuccore>) under accession numbers KP999944 – KP999951.

**Acknowledgements** The authors would like to thank Laihui Lv for help with the statistical analysis.

**Supplementary materials** The online version of this article at <http://dx.doi.org/10.15302/J-FASE-2015056> contains supplementary materials (Appendix A and B).

**Compliance with ethics guidelines** Zuoxiang Liang, Tao Wang, Yi Sun, Wenlong Yang, Zhihong Liu, Jing Fei, Ying Guo, Qingwei Ma, Qingjie Pan and Liming Ren declare that they have no conflict of interest or financial conflicts to disclose.

All applicable institutional and national guidelines for the care and use of animals were followed.

## References

- Edelman G M. Antibody structure and molecular immunology. *Scandinavian Journal of Immunology*, 1991, **34**(1): 4–22
- Hamers-Casterman C, Atarhouch T, Muyldermans S, Robinson G, Hammers C, Songa E B, Bendahman N, Hammers R. Naturally occurring antibodies devoid of light chains. *Nature*, 1993, **363** (6428): 446–448
- Seligmann M, Mihaesco E, Preud'homme J L, Danon F, Brouet J C. Heavy chain diseases: current findings and concepts. *Immunological Reviews*, 1979, **48**(1): 145–167
- Dooley H, Flajnik M F. Antibody repertoire development in cartilaginous fish. *Developmental and Comparative Immunology*, 2006, **30**(1–2): 43–56
- Flajnik M F, Deschacht N, Muyldermans S. A case of convergence: why did a simple alternative to canonical antibodies arise in sharks and camels? *PLoS Biology*, 2011, **9**(8): e1001120
- De Genst E, Saerens D, Muyldermans S, Conrath K. Antibody repertoire development in camelids. *Developmental and Comparative Immunology*, 2006, **30**(1–2): 187–198
- Muyldermans S, Atarhouch T, Saldanha J, Barbosa J A, Hamers R. Sequence and structure of VH domain from naturally occurring camel heavy chain immunoglobulins lacking light chains. *Protein Engineering*, 1994, **7**(9): 1129–1135
- Muyldermans S, Cambillau C, Wyns L. Recognition of antigens by single-domain antibody fragments: the superfluous luxury of paired domains. *Trends in Biochemical Sciences*, 2001, **26**(4): 230–235
- Desmyter A, Transue T R, Ghahroudi M A, Dao Thi M H, Poortmans F, Hamers R, Muyldermans S, Wyns L. Crystal structure of a camel single-domain VH antibody fragment in complex with lysozyme. *Nature Structural Biology*, 1996, **3**(9): 803–811
- Govaert J, Pellis M, Deschacht N, Vincke C, Conrath K, Muyldermans S, Saerens D. Dual beneficial effect of interloop disulfide bond for single domain antibody fragments. *The Journal of Biological Chemistry*, 2012, **287**(3): 1970–1979
- Khong Nguyen V, Hamers R, Wyns L, Muyldermans S. Loss of

- splice consensus signal is responsible for the removal of the entire C (H)1 domain of the functional camel IGG2A heavy-chain antibodies. *Molecular Immunology*, 1999, **36**(8): 515–524
12. Wu Y, Eigenbrot C, Liang W C, Stawicki S, Shia S, Fan B, Ganesan R, Lipari M T, Kirchhofer D. Structural insight into distinct mechanisms of protease inhibition by antibodies. *Proceedings of the National Academy of Sciences of the United States of America*, 2007, **104**(50): 19784–19789
13. Lauwereys M, Arbabi Ghahroudi M, Desmyter A, Kinne J, Hölzer W, De Genst E, Wyns L, Muyldermans S. Potent enzyme inhibitors derived from dromedary heavy-chain antibodies. *The EMBO Journal*, 1998, **17**(13): 3512–3520
14. Jirimutu W Z, Wang Z, Ding G, Chen G, Sun Y, Sun Z, Zhang H, Wang L, Hasi S, Zhang Y, Li J, Shi Y, Xu Z, He C, Yu S, Li S, Zhang W, Batmunkh M, Ts B, Narenbatu, Unierhu, Bat-Ireedui S, Gao H, Baysgalan B, Li Q, Jia Z, Turigenbayila, Subudenggerile, Narenmanduhu, Wang Z, Wang J, Pan L, Chen Y, Ganerdene Y, Dabxilt, Erdemt, Altansha, Altansukh, Liu T, Cao M, Aruuntsever, Bayart, Hosblig, He F, Zha-ti A, Zheng G, Qiu F, Sun Z, Zhao L, Zhao W, Liu B, Li C, Chen Y, Tang X, Guo C, Liu W, Ming L, Temuulen, Cui A, Li Y, Gao J, Li J, Wurentaodi, Niu S, Sun T, Zhai Z, Zhang M, Chen C, Baldan T, Bayaer T, Li Y, Meng H. Genome sequences of wild and domestic bactrian camels. *Nature Communications*, 2012, **3**: 1202
15. Wu H, Guang X, Al-Fageeh M B, Cao J, Pan S, Zhou H, Zhang L, Abutarboush M H, Xing Y, Xie Z, Alshanqeeti A S, Zhang Y, Yao Q, Al-Shomrani B M, Zhang D, Li J, Manee M M, Yang Z, Yang L, Liu Y, Zhang J, Altammami M A, Wang S, Yu L, Zhang W, Liu S, Ba L, Liu C, Yang X, Meng F, Wang S, Li L, Li E, Li X, Wu K, Zhang S, Wang J, Yin Y, Yang H, Al-Swailem A M, Wang J. Camelid genomes reveal evolution and adaptation to desert environments. *Nature Communications*, 2014, **5**: 5188
16. Griffin L M, Snowden J R, Lawson A D G, Wernery U, Kinne J, Baker T S. Analysis of heavy and light chain sequences of conventional camelid antibodies from *Camelus dromedarius* and *Camelus bactrianus* species. *Journal of Immunological Methods*, 2014, **405**: 35–46
17. Achour I, Cavelier P, Tichit M, Bouchier C, Lafaye P, Rougeon F. Tetrameric and homodimeric camelid IgGs originate from the same IgH locus. *Journal of Immunology*, 2008, **181**(3): 2001–2009
18. Conrath K E, Wernery U, Muyldermans S, Nguyen V K. Emergence and evolution of functional heavy-chain antibodies in Camelidae. *Developmental and Comparative Immunology*, 2003, **27**(2): 87–103
19. Giudicelli V, Lefranc M P. Ontology for immunogenetics: the IMGT-ONTOLOGY. *Bioinformatics*, 1999, **15**(12): 1047–1054
20. Fu X, Gao X, He S, Huang D, Zhang P, Wang X, Zhang S, Dang R, Yin S, Du E, Yang Z. Design and selection of a camelid single-chain antibody yeast two-hybrid library produced *de novo* for the cap protein of porcine circovirus type 2 (PCV2). *PLoS ONE*, 2013, **8**(3): e56222
21. Deschacht N, De Groeve K, Vincke C, Raes G, De Baetselier P, Muyldermans S. A novel promiscuous class of camelid single-domain antibody contributes to the antigen-binding repertoire. *Journal of Immunology*, 2010, **184**(10): 5696–5704
22. Nguyen V K, Hamers R, Wyns L, Muyldermans S. Camel heavy-chain antibodies: diverse germline V(H)H and specific mechanisms enlarge the antigen-binding repertoire. *The EMBO Journal*, 2000, **19**(5): 921–930
23. Maass D R, Sepulveda J, Pernthaner A, Shoemaker C B. Alpaca (*Lama pacos*) as a convenient source of recombinant camelid heavy chain antibodies (VHHs). *Journal of Immunological Methods*, 2007, **324**(1–2): 13–25
24. Johansson J, Aveskog M, Munday B, Hellman L. Heavy chain V region diversity in the duck-billed platypus (*Ornithorhynchus anatinus*): long and highly variable complementarity-determining region 3 compensates for limited germline diversity. *Journal of Immunology*, 2002, **168**(10): 5155–5162
25. Stanfield R L, Dooley H, Flajnik M F, Wilson I A. Crystal structure of a shark single-domain antibody V region in complex with lysozyme. *Science*, 2004, **305**(5691): 1770–1773
26. Vu K B, Ghahroudi M A, Wyns L, Muyldermans S. Comparison of llama V<sub>H</sub> sequences from conventional and heavy chain antibodies. *Molecular Immunology*, 1997, **34**(16–17): 1121–1131
27. Sircar A, Sanni K A, Shi J, Gray J J. Analysis and modeling of the variable region of camelid single-domain antibodies. *Journal of Immunology*, 2011, **186**(11): 6357–6367
28. Crooks G E, Hon G, Chandonia J M, Brenner S E. WebLogo: a sequence logo generator. *Genome Research*, 2004, **14**(6): 1188–1190
29. Sørensen V, Sundvold V, Michaelsen T E, Sandlie I. Polymerization of IgA and IgM: roles of Cys309/Cys414 and the secretory tailpiece. *Journal of Immunology*, 1999, **162**(6): 3448–3455