

METHODS AND RESOURCES

Isolation of antigen-specific, disulphide-rich knob domain peptides from bovine antibodies

Alex Macpherson^{1,2*}, Anthony Scott-Tucker¹, Anastasios Spiliotopoulos¹, Catherine Simpson¹, Justin Staniforth¹, Adam Hold¹, James Snowden¹, Leah Manning¹, Jean van den Elsen², Alastair D. G. Lawson¹

1 UCB, Slough, United Kingdom, 2 Department of Biology and Biochemistry, University of Bath, Bath, United Kingdom

* alex.macpherson@ucb.com



OPEN ACCESS

Citation: Macpherson A, Scott-Tucker A, Spiliotopoulos A, Simpson C, Staniforth J, Hold A, et al. (2020) Isolation of antigen-specific, disulphide-rich knob domain peptides from bovine antibodies. *PLoS Biol* 18(9): e3000821. <https://doi.org/10.1371/journal.pbio.3000821>

Academic Editor: David Nemazee, Scripps Research Institute, UNITED STATES

Received: March 27, 2020

Accepted: August 10, 2020

Published: September 4, 2020

Copyright: © 2020 Macpherson et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was funded by UCB-Celltech, a branch of UCB Pharma S.A., as part of an internal PhD program. The work was performed by A.M. as partial fulfillment of the requirements for a PhD from the University of Bath. Employees of UCB-Celltech were therefore involved in the study design, data collection and analysis, the decision to publish, and the preparation of the manuscript.

Abstract

As a novel alternative to established surface display or combinatorial chemistry approaches for the discovery of therapeutic peptides, we present a method for the isolation of small, cysteine-rich domains from bovine antibody ultralong complementarity-determining regions (CDRs). We show for the first time that isolated bovine antibody knob domains can function as autonomous entities by binding antigen outside the confines of the antibody scaffold. This yields antibody fragments so small as to be considered peptides, each stabilised by an intricate, bespoke arrangement of disulphide bonds. For drug discovery, cow immunisations harness the immune system to generate knob domains with affinities in the picomolar to low nanomolar range, orders of magnitude higher than unoptimized peptides from naïve library screening. Using this approach, knob domain peptides that tightly bound Complement component C5 were obtained, at scale, using conventional antibody discovery and peptide purification techniques.

Introduction

To date, the smallest autonomous, naturally occurring, functional antibody domains reported have been the variable regions of camelid heavy-chain antibodies (VHH) [1] and the variable regions of the immunoglobulin new antigen receptor (VNAR), derived from sharks [2], resulting in heavy-chain variable region fragments, of some 12–15 kDa [3, 4]. In 1997, the first report of bovine antibodies featuring an ultralong heavy-chain complementarity-determining region 3 (CDRH3) was published [5], with subsequent crystal structures revealing a conserved anti-parallel β -ribbon ‘stalk’ presenting a disulphide-stabilised ‘knob’ domain of 3–6 kDa [6]. The knob domain, shown in Fig 1, comprises several small loops stapled by 2 to 5 disulphide bonds and is held some 25–45 Å clear of the surface of the other complementarity-determining region (CDR) loops [6–8]. Mutation studies with whole IgG have shown that binding can occur predominantly through the knob domain with limited input from neighbouring CDRs [6], but the surrounding architecture of the antibody scaffold is still thought to play a critical

Competing interests: I have read the journal's policy and the authors of this manuscript have the following competing interests: All authors apart from J.v.d.E are current or previous employees of UCB-Celltech and may hold shares and/or stock options.

Abbreviations: AB, assay buffer; aHUS, atypical haemolytic uremic syndrome; BP, band pass; BSA, bovine serum albumin; CDR, CDR, complementarity-determining region; CDRH3, heavy-chain complementarity-determining region 3; CMV, cytomegalovirus; EDTA, ethylenediaminetetraacetic acid; FA, formic acid; FAC, final assay concentration; FACS, fluorescence-activated cell sorting; FCS, foetal calf serum; FDA, Food and Drug Administration; FRET, Förster resonance energy transfer; His, histidine; IC_{50} , half maximal inhibitory concentration; K_D app, apparent K_D ; IGHD8-2, immunoglobulin heavy chain diversity region 8–2; IGHV1-7, immunoglobulin heavy chain variable region 1–7; LC/MS, liquid chromatography mass spectrometry; PBS, phosphate buffered saline; PDB, Protein Data Bank; PE, Phycoerythrin; PNH, paroxysmal nocturnal haemoglobinuria; RP-HPLC, reversed-phase high-performance liquid chromatography; RPMI, Roswell Park Memorial Institute; RT-PCR, reverse transcription polymerase chain reaction; ScFc, single-chain Fc; SEC, size exclusion chromatography; SHM, somatic hypermutation; SPR, surface plasmon resonance; tc, transfer constant; TEV, Tobacco etch virus; TFA, trifluoroacetic acid; VHH, variable regions of camelid heavy-chain antibodies; VNAR, variable regions of the immunoglobulin new antigen receptor.

supporting role [6–10]. Recently, the first co-crystal structures of an ultralong CDRH3 Fab in complex with antigen—in this case, an anti-HIV broadly neutralising antibody [11] bound to a soluble portion of the envelope glycoprotein GP120—found that antigen recognition was contained entirely within the ultralong CDRH3 [12]. The study notes a nuanced role for the stalk, with alanine mutations in this region maintaining binding affinity but attenuating neutralisation efficiency [12]. In this paper, we show that, when isolated from the parent antibody infrastructure and the stalk, the bovine knob domain itself may be able to bind to antigen with high affinity and so can function as an autonomous entity.

Bovine ultralong CDRH3 arise due to the limited repertoire of antibody gene segments in domesticated cattle. Within their antibody heavy chain repertoire, cows have approximately $10\times$ highly homologous V_H , $10\times$ D_H , and just $4\times$ J_H segments, giving a theoretical heavy chain diversity of only approximately 4,000 [14]; by contrast, humans have $55\times$ V_H , $23\times$ D_H , and $6\times$ J_H segments, giving a theoretical diversity of up to 7,590 [15]. Cows exploit other mechanisms for generating antibody diversity. In addition to relying heavily on somatic hypermutation (SHM), bovines have evolved a longer CDRH3, even amongst the non-ultralong repertoire, maximising the value from rearrangement of their limited immune gene subset. Bovines have also evolved a mechanism for overcoming the conformational entropy penalty associated with long loops by using disulphide bonds to staple the CDRH3 in the biologically active conformer.

To facilitate disulphide bond formation, the introduction of cysteines is achieved through priming of a germline D-gene segment, immunoglobulin heavy chain diversity region 8–2 (IGHD8-2), which exclusively encodes the central portion of ultralong CDRH3. The germline peptide contains 4 cysteines, but glycine, tyrosine, and serine residues are also prominent, each of which can be mutated to cysteine by a single nucleotide change during SHM [16]. Interestingly, codon use within IGHD8-2 seems to favour glycine and tyrosine codons that are otherwise infrequently used in the bovine genome but are closer to the TGT/TGC encoding cysteine, hardwiring for this specific mutation [10]. Extended CDRH3 length and complexity has also been described in *Ornithorhynchus anatinus*, the Australian duck-billed platypus, but stabilisation occurs predominantly through inter-CDR disulphide bonds, again, in response to a paucity of V-gene segments [17]. Currently, complex stabilisation of CDRH3 knob domains with intra-CDR disulphide bonds appears unique to cattle.

In order to enable the generation of isolated knob domains, we performed bovine antibody discovery against Complement component C5, a 188-kDa humoral protein of the complement cascade, as the test antigen. C5 is the target of the Food and Drug Administration (FDA)-approved monoclonal antibody eculizumab in diseases such as paroxysmal nocturnal haemoglobinuria (PNH) [18, 19] and atypical haemolytic uremic syndrome (aHUS) [20] and remains an enduring focus of drug discovery efforts to prevent complement-induced autoinflammation, with a number of second-generation therapeutics in clinical and preclinical development [21]. Fig 1 shows how isolated knob domains would effectively bridge a molecular weight gap between camelid-derived VHH antibody domains and chemical macrocycles, with potential for therapeutic utility.

In this paper, we present a method for the preparation of an isolated knob domain peptide from bovine antibodies. This comprises immunisation of cattle, fluorescence-activated cell sorting (FACS) of antigen-specific B cells, targeted sequencing of CDRH3 cDNA, reformatting knob and stalk domains for screening, and finally, excision and purification of the isolated knob domains (Fig 2). Most peptides are currently derived by phage display [22] or combinatorial chemistry [23] methods, but the approach described herein is unique in exploiting the bovine immune system to enable affinity maturation and derive peptides with complex networks of disulphide bonds. Even if mammalian display may be used, the structural complexity

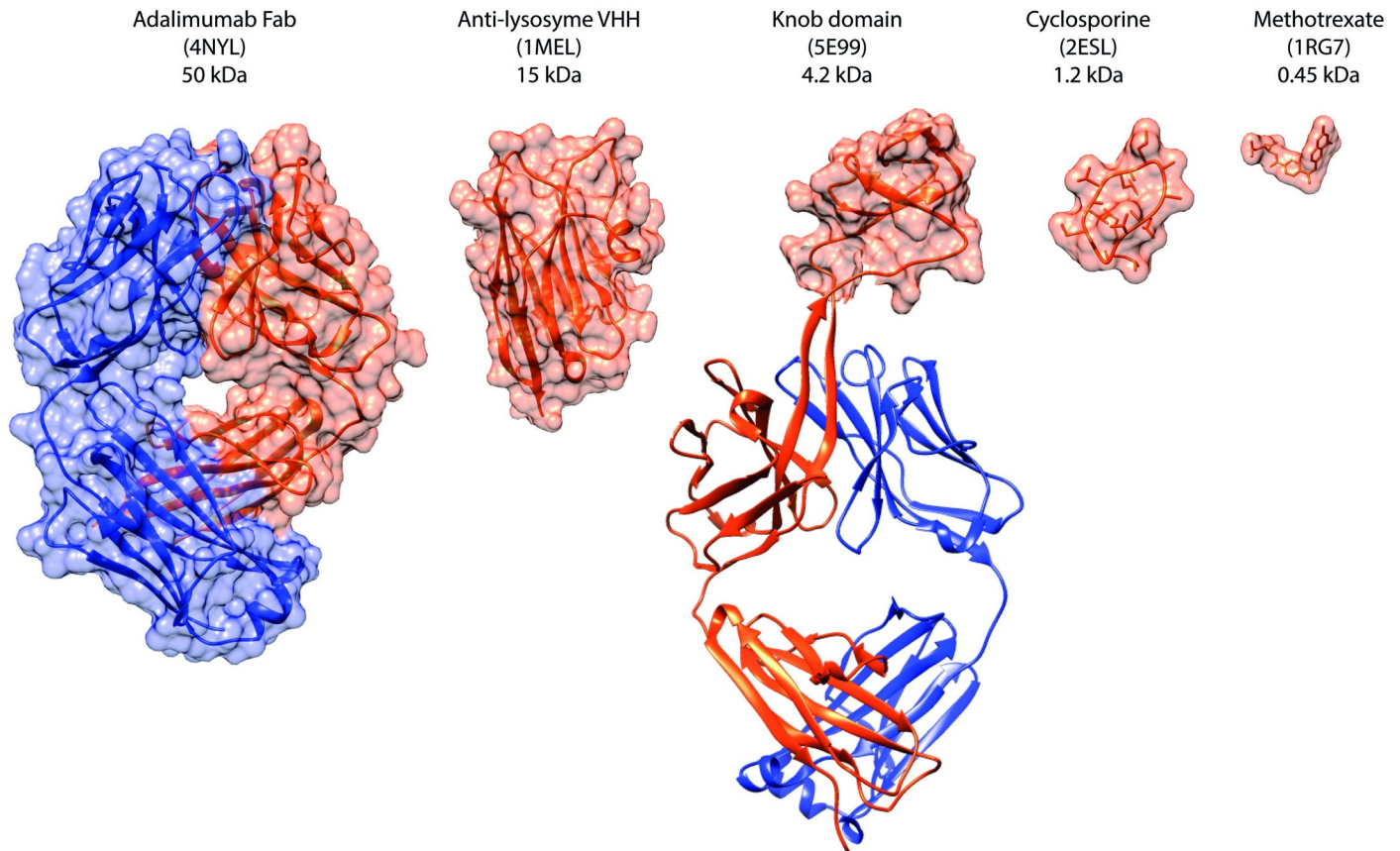


Fig 1. Therapeutic molecules by size. Molecular size comparison of therapeutic molecules, ranging from the 50 kDa Adalimumab Fab fragment to the 450 Da small-molecule methotrexate. A VHH antibody fragment (MW 15 kDa), derived from the heavy-chain-only subset of the camelid repertoire, is shown alongside the knob domain peptide of a bovine ultralong CDRH3 (MW 3–5 kDa), that bridges the gap between the antibody fragments and cyclic peptides, such as cyclosporin (MW 1.2 kDa). Ribbon diagrams and molecular surface representations were produced using UCSF Chimera [13]. PDB accession codes for the molecular structures used in this figure are indicated. CDRH3, heavy-chain complementarity-determining region 3; MW, molecular weight; PDB, Protein Data Bank; VHH, variable regions of camelid heavy-chain antibodies.

<https://doi.org/10.1371/journal.pbio.3000821.g001>

of disulphide-rich peptides makes their display nontrivial, with the first reported mammalian display of cysteine-rich peptide fragments estimating that just 17% of the library was correctly folded on the surface [24].

For knob domain peptides, the abundance of disulphide bonds may confer improved drug-like properties and render them especially suited for therapeutic applications. The method here is therefore highly applicable to the discovery of peptide therapeutics.

Results

Immunisation with C5 yielded an immune serum titre and antigen-specific B cells

We immunised 2 adult Holstein Friesian cattle with human C5 and obtained an immune serum titre of $> 1/10,000$ (S1 Fig and S1 Data). We isolated immune material from a draining lymph node, taken proximal to the site of immunisation, and performed cell sorting of antigen-specific B cells. Flow cytometry was used to identify cells that were double-positive for 2 fluorescently labelled populations of C5 (Fig 3), and a polyclonal mixture of antigen-enriched immune cells was collected. Similar to camelid VHH discovery from heavy-chain-only

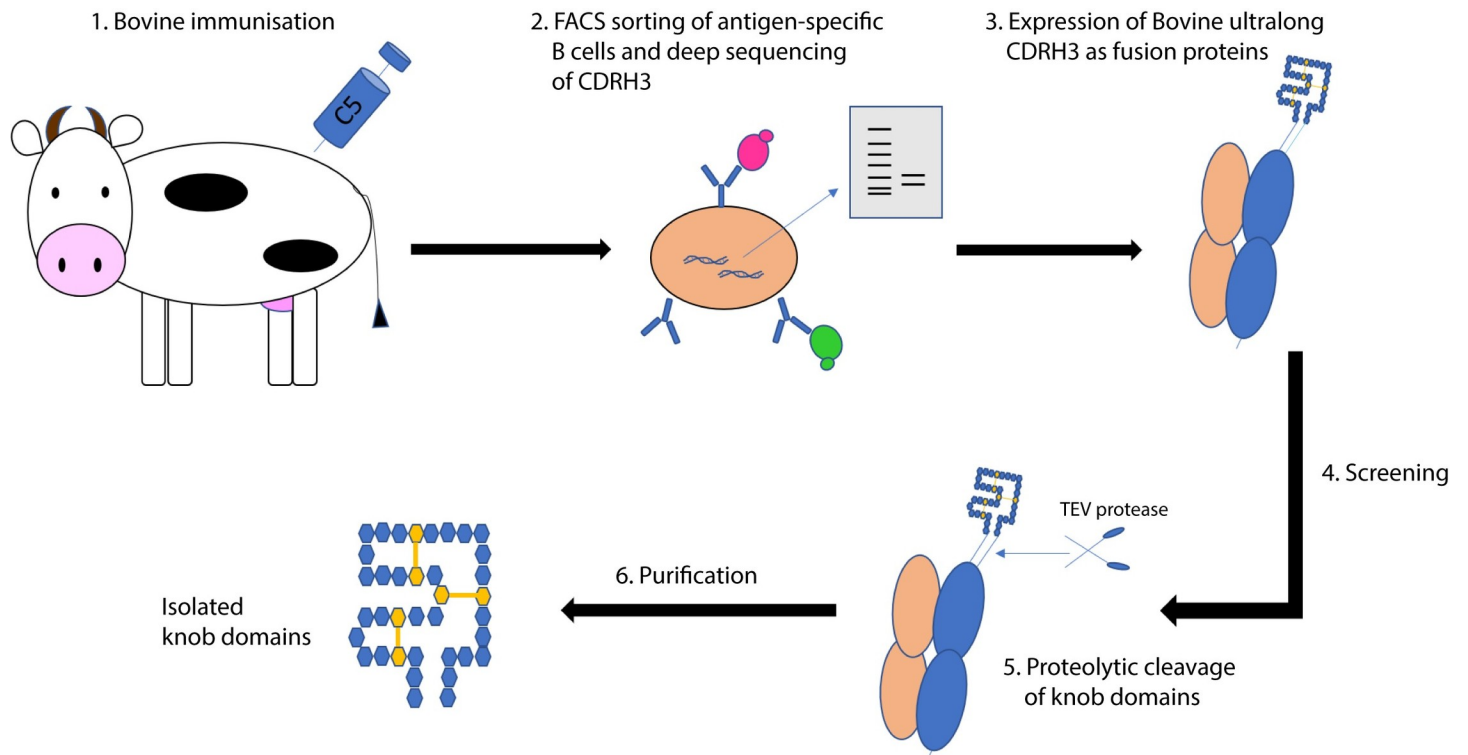


Fig 2. Schematic representation of the method for the isolation of knob domain peptides. Immunisation of cattle with human C5 is used to elicit an immune response (step 1). An antigen-specific response is isolated through FACS of immune cells, using 2 populations of fluorescently labelled C5 (step 2). Cells that are double positive for C5 are sorted into a polyclonal mixture. After RT-PCR, PCR primers specific to CDRH3 are used to create a CDRH3 library. Ultralong CDRH3 are identified using deep sequencing and expressed as cleavable fusion proteins (CDRH3-ScFc or Fab-CDRH3) (step 3). After screening (step 4), TEV protease can be used to excise the knob domain peptides (step 5), which are purified using RP-HPLC (step 6). CDRH3, heavy-chain complementarity-determining region 3; FACS, fluorescence-activated cell sorting; RP-HPLC, reversed-phase high-performance liquid chromatography; RT-PCR, reverse transcription polymerase chain reaction; ScFc, single-chain Fc; TEV, Tobacco etch virus.

<https://doi.org/10.1371/journal.pbio.3000821.g002>

antibodies, there was no requirement to pair heavy and light chains, since only the extended CDRH3 knob domains were of interest in this study. Consequently, sequencing of polyclonal libraries of antigen-enriched CDRH3 can be used as a rapid approach to discover knob domains.

Deep sequencing of target enriched CDRH3 library revealed ultralong CDRH3 clonotypes

The antigen-enriched pool of immune cells was lysed, and reverse transcription polymerase chain reaction (RT-PCR) was performed directly on the lysate. A primary PCR with primers flanking CDRH3, annealing to the conserved framework 3 and framework 4 regions of the V_H , was used to amplify all CDRH3 sequences, irrespective of their length or amino acid sequence. At this stage, Sanger sequencing of 96 colonies did not reveal any ultralong CDRH3 sequences, potentially as a result of shorter sequences being more readily incorporated into the plasmid vectors. A second round of PCR was used to barcode the CDRH3 sequences for ion torrent sequencing.

Deep sequencing of the CDRH3 library revealed ultralong CDRH3 sequences at 4.3% of the total sequences. Ultralong CDRH3 were easily identified by their length (>90 bp) and by a characteristic duplication of a portion of the immunoglobulin heavy chain variable region 1–7 (IGHV1-7) gene segment that has been reported as a universal feature of ultralong CDRH3 [14]. After filtering, 3,559 unique CDRH3 sequences were obtained from 2 samples taken

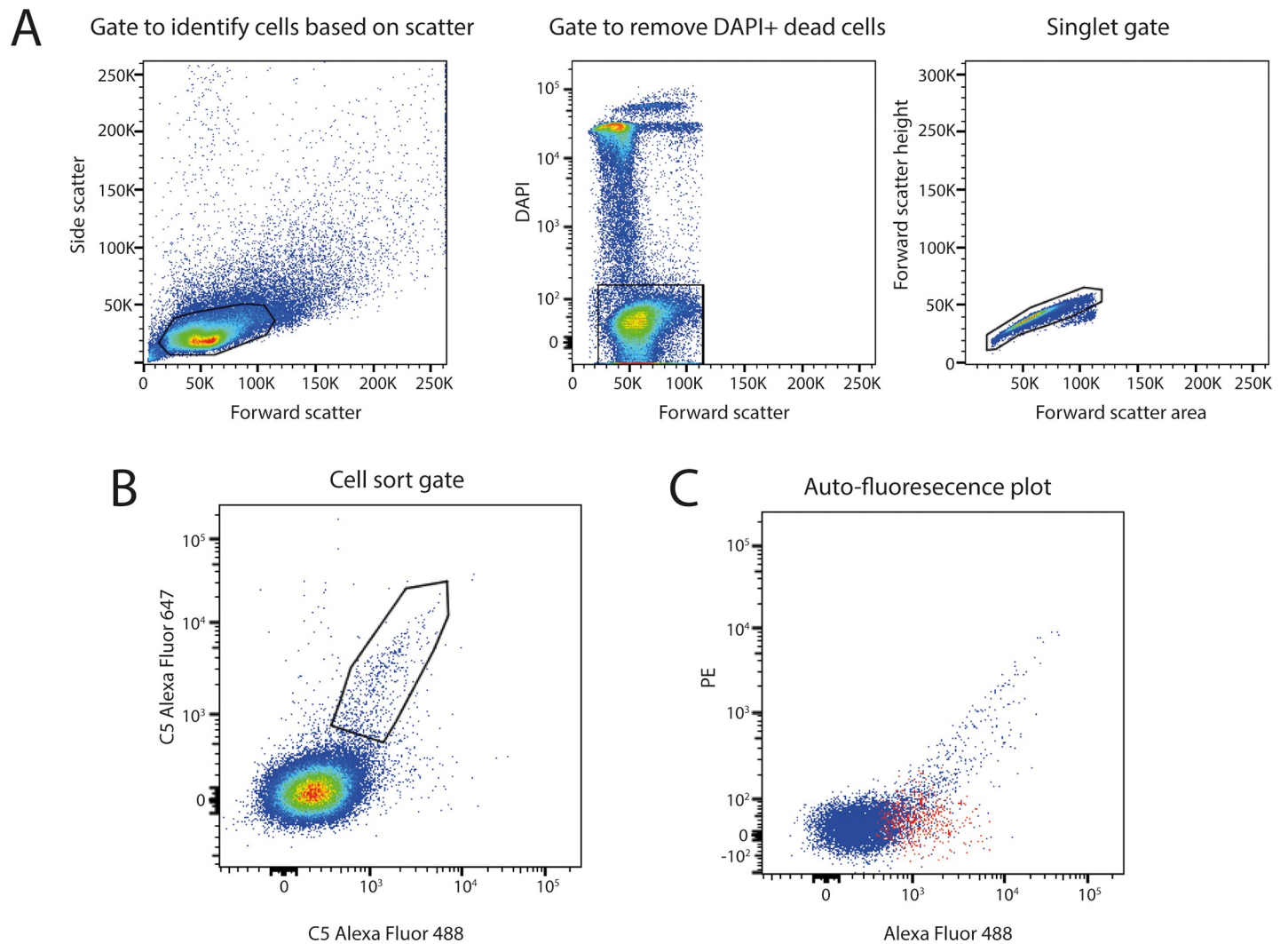


Fig 3. Generation of C5 enriched B cell pool. To identify antigen-specific B cells from immunised animals, FACS of live (DAPI negative) singlet cells was performed (panel A). Two populations of fluorescently labelled C5, C5-AlexaFluor 647 and C5-AlexaFluor 488, were used to identify immune cells that were double positive for C5 (panel B), indicating the presence of surface immunoglobulins against C5. C5 double-positive cells, shown in red, were checked for autofluorescence which may arise in dying or senescent cells (panel C). FACS, fluorescence-activated cell sorting; PE, Phycoerythrin.

<https://doi.org/10.1371/journal.pbio.3000821.g003>

from a single draining lymph node. Of these, 154 were ultralong. Sequence analysis permitted clustering of the knob domains into clonotypes, based on the homology of their amino acid sequence (S1 Table). A total of 20 clonotypes, displaying sequence identity greater than 75%, were observed. Approximately half of the knob sequences could be classified into clonotypes with the remainder being singletons. A range of cysteine arrangements were present with an even number of cysteines preferred, consistent with the formation of disulphide bonds, and observations in earlier studies [6, 7, 14, 25, 26].

Ultralong CDRH3 and knob domains were expressed as functional fusion proteins

We attempted to directly express knob domain peptides in isolation using Expi293F cells. From our library of CDRH3 sequences, a small panel of knob domains were synthesised into a

mammalian expression vector containing an in-frame 10× histidine (His) tag. Based on the available structures, we designated the first residue of the D_H2 gene segment, immediately preceding the first germline Cys, as the N-terminal of the knob domain. The C-terminal was the final aliphatic residue preceding the alternating aromatic-aliphatic motif of the descending stalk. In Expi293F cells, knob domains expressed poorly with a C-terminal 10× His tag, so fusion constructs were designed to facilitate expression and purification.

We took the entire ultralong CDRH3 (defined as residues H93–H102 Kabat), comprising the knob and stalk, and expressed with a Tobacco etch virus (TEV)-cleavable, poly-His-tagged, single-chain Fc (ScFc) fused at the C-terminus. A ScFc, where the heavy chains of the Fc homodimer are fused by the incorporation of a linker sequence and expressed as a single polypeptide chain (S1 Text and S2 Fig), was chosen in preference to a conventional Fc to create a monovalent construct [27]. A representative set of 52 ultralong CDRH3, comprising a range of sequences with diverse numbers and arrangements of cysteines, were expressed as CDRH3-ScFc fusions as 2-mL transient transfections of Expi293F cells (S2 Table and S1 Text). Cell supernatants were screened in a C5 binding ELISA, and 14 binders were identified (S3 Table), translating to a 27% hit rate. These binders showed that it was possible to generate ultralong CDRH3 that were viable even in the absence of the supporting infrastructure of the parent antibody.

To assess the importance of the cognate β-ribbon stalk, knob domains were separated from the stalk and inserted into the CDRH3 of a Fab, flanked by TEV protease sites to allow the knob domain peptide to be excised. The human Fab PGT121 was selected as the acceptor framework [28]. The CDRH3 of PGT121, which binds complex-type N-glycans within the gp120 envelope of HIV and has no intrinsic affinity for C5, contains a slightly extended anti-parallel β-ribbon stalk that was an ideal platform on which to present the knob domain peptide.

From our ELISA hits, 6 diverse isolated knob domains, with different numbers and arrangements of cysteines, were reformatted as cleavable PGT121-knob fusion proteins (Fig 4 and S2 Text). Binding to C5 was measured using surface plasmon resonance (SPR) single-cycle kinetics. C5 has been shown to be activated by extremes of pH [29], and so single-cycle kinetics—in which sequential injections of increasing concentrations of analyte are performed in the absence of harsh regeneration steps—was employed to maintain the integrity of the C5 protein on the sensor chip surface. By SPR, five PGT121-knobs bound C5 with affinities in the picomolar to nanomolar range (Fig 5, S4 Table and S1 Data), with only the K60 knob domain found to be nonfunctional as a consequence of reformatting. Of note, the PGT121-K92 and PGT121-K136 fusion proteins were exceptionally tight binders, displaying off rates that exceeded the $1.0 \times 10^{-5} \text{ s}^{-1}$ limit for k_{off} determination by Biacore [30].

The PGT121-knob fusion proteins were counter-screened for binding to human complement component C3 (S3 Fig and S1 Data). C3 is the closest human homologue to C5, sharing a conserved fold [31] and a sequence homology of 26.5%. In these experiments, no cross reactivity was detected.

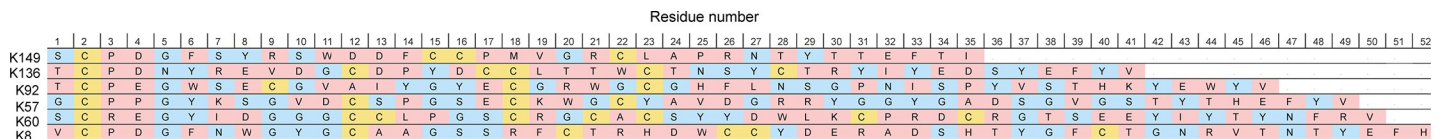


Fig 4. The D_H gene segment residues for the knob domain peptides exemplified in this study show the diversity of peptides resulting from bovine immunisation with C5. Residue numbering has been arbitrarily assigned from the beginning of the D_H gene segment. To highlight the mechanism for cysteine enrichment within the IGHD8-2 gene segment, cysteine residues are coloured yellow, while residues which may be single DNA base pair change from becoming a cysteine are coloured blue. Of note, the knob domain peptides are different lengths with different numbers and arrangements of cysteines.

<https://doi.org/10.1371/journal.pbio.3000821.g004>

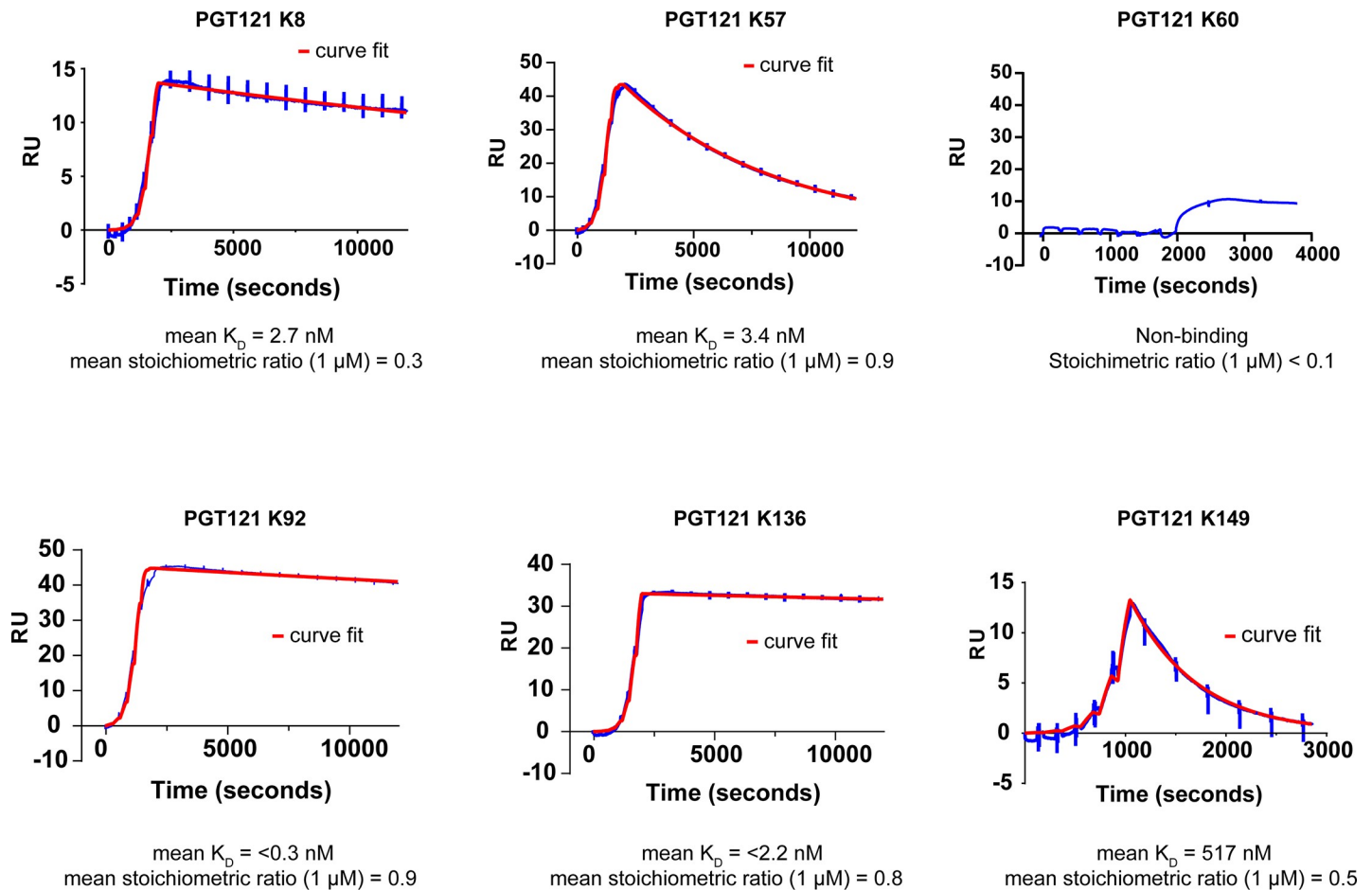


Fig 5. SPR single-cycle kinetics of PGT121 Fab-knob domain fusion proteins binding to C5. As fusion proteins, the knob domains confer high-affinity C5 binding to the PGT121 Fab. A representative sensorgram from Biacore single-cycle kinetics is displayed for each PGT121 Fab-knob domain fusion protein (curve in red), with accompanying curve fits for a 1:1 binding model (in blue). The mean equilibrium dissociation constants (K_D), determined from 3 experiments, show affinities in the <0.3 nM to 517 nM range for the fusion proteins, except for PGT121-K60, which did not bind C5. RU, response units; SPR, surface plasmon resonance.

<https://doi.org/10.1371/journal.pbio.3000821.g005>

Isolated knob domains demonstrated high-affinity binding to C5

To obtain knob domain peptides, TEV protease was used to excise the knob domain from a Fab scaffold (Fig 4 and S5 Table). We postulated that the absence of secondary structure and the presence of a stabilising network of disulphide bonds may permit purification by solvent chromatography methods, such as reversed-phase high-performance liquid chromatography (RP-HPLC). RP-HPLC affords greater resolution of minor species compared to size exclusion chromatography (SEC) but is infrequently used for preparative purification of monoclonal antibodies, where milder aqueous-based chromatography techniques are typically favoured. After cleavage, the knob domain peptides were purified by RP-HPLC and freeze-dried. When analysed by liquid chromatography mass spectrometry (LC/MS), masses consistent with the predicted isotope patterns—based on peptide amino acid sequences and formation of disulphide bonds—were present (S6 Table, S4 Fig and S5 Fig).

We report individual rate constants (k_{on} and k_{off}) and equilibrium dissociation constants (K_D) for knob domain peptides binding to C5. The peptides display high-affinity binding, with K_D ranging from below 1 nM to 40 nM, when measured by SPR single-cycle kinetics (Fig 6, S7 Table and S1 Data). Similar to our experiments with the PGT121 Fab, the only knob domain

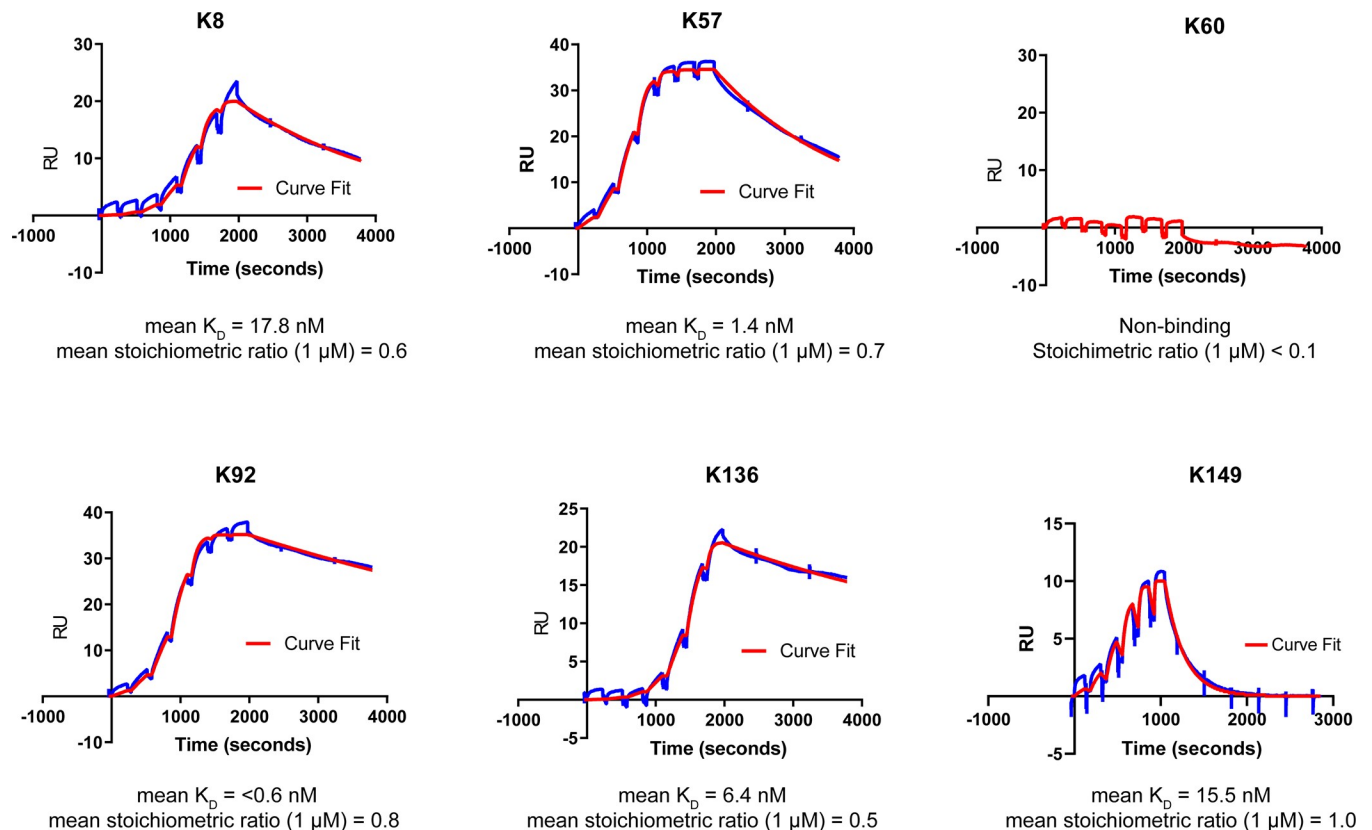


Fig 6. SPR single-cycle kinetics of isolated knob domain binding to C5. The isolated knob domain peptides retain high-affinity binding to C5 when proteolytically cleaved from a Fab scaffold. A representative sensorgram from Biacore single-cycle kinetics is displayed for each knob domain (curve in red), with accompanying curve fits for a 1:1 binding model (in blue). The mean binding affinities equilibrium dissociation constants (K_D), determined from 3 experiments, show affinities in the <604 pM to 36.5 nM range for all the knob domain peptides, except for K60, which remained inactive after cleavage from the PGT121 Fab. RU, response units; SPR, surface plasmon resonance.

<https://doi.org/10.1371/journal.pbio.3000821.g006>

peptide found not to bind was K60. The knob domain peptides were counter-screened for cross reactivity to human complement component C3 (S6 Fig and S1 Data), and to ovalbumin (S7 Fig and S1 Data), with no binding to either protein detected.

To provide further evidence for high-affinity binding, we developed steady state Förster resonance energy transfer (FRET) assays. This was achieved by labelling C5 with a terbium chelate donor fluorophore and each of the active PGT121 knob fusion proteins with an AlexaFluor 647 acceptor fluorophore. Titrations of the PGT121 knob fusion proteins—in the presence and absence of a saturating concentration of unlabelled C5 protein—were used to derive apparent K_D (K_D *app*) for the interaction with C5, after an incubation of 24 hours. Despite modification of both proteins by labelling, the PGT121 fusion proteins bound C5 with high affinity (Fig 7, S8 Table and S1 Data), in a manner consistent with our SPR experiments. Although the PGT121-K92 and PGT121-K136 fusion proteins—which displayed particularly tight binding in the Biacore—displayed lower affinities in the steady-state method.

Next, we tested titrations of knob domain peptides in a competition assay format, using displacement of the parent PGT121 knob domain fusion protein as a readout for binding. For each peptide, the concentrations of PGT121 fusion protein were fixed at the K_D *app* and incubated with a titration of peptide for 24 hours. We measured half maximal inhibitory concentration (IC_{50}) values and derived inhibition constants (K_i) for each of the knob domain

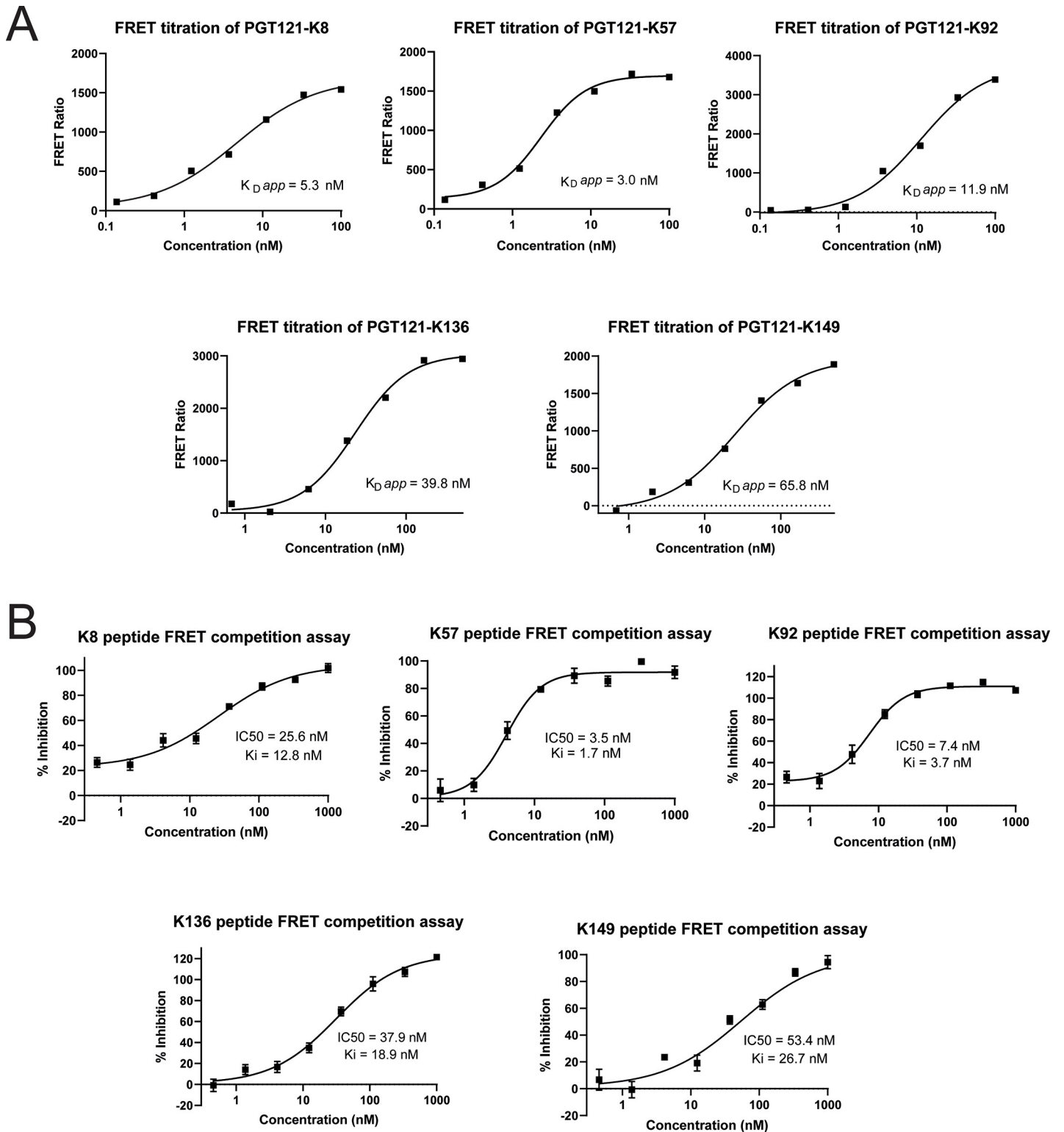


Fig 7. FRET assays. C5 protein was labelled with a Terbium chelate and PGT121 knob domain fusion protein with AlexFluor 647. Panel A shows titrations of labelled knob domain fusion proteins to derive an $K_D app$ (geomean from 3 experiments). These curves have had background subtracted using quenching concentrations of unlabelled C5. Panel B shows IC_{50} curves obtained by titrating unlabelled knob domain peptide against its respective PGT121 knob domain fusion protein, with the assay performed at the $K_D app$ concentration. Data are available in [S1 Data](#). FRET, Förster resonance energy transfer; IC_{50} , half maximal inhibitory concentration; $K_D app$, apparent K_D .

<https://doi.org/10.1371/journal.pbio.3000821.g007>

peptides (Fig 7, S9 Table and S1 Data). These data provide further evidence for high-affinity—sub-50 nM—interactions and consolidate our observations by SPR. We note that in all our FRET experiments, the Hill slopes were within the expected range (0.5–2); this is indicative of reversible binding as defined by the law of mass action, while nM displacement of the PGT121 knob domain fusion proteins suggests that the interactions are indeed site specific.

The biophysical and biochemical experiments described here provide the first evidence that isolated knob domains of bovine ultralong CDRH3 can function autonomously, outside of the supporting infrastructure of the V_H domain.

Discussion

Here, we present a method for obtaining isolated knob domains from bovine antibodies, using standard FACS, PCR, and mammalian transient expression. In the present study, we found that knob domains alone were poorly expressed by an Expi293F cell line, but we have shown that they can be readily expressed, either with their accompanying stalk as ultralong CDRH3-ScFc or as just knob domains with cleavable linkers when inserted into a CDRH3 of a conventional Fab fragment. Recovery of isolated knob domain peptides from the expressed Fab fragments was efficient, and readily enabled purification and characterisation on milligram amounts from only a few hundred millilitres of transient cultures.

In the test system we used, 27% of the transposed knob domains from preliminary screening were able to bind to the C5 target protein with high affinity, showing for the first time that they are capable of interacting with antigen autonomously, with neighbouring CDRs not prerequisite for binding. Additionally, we have shown that the cognate stalk is not obligate to orientate or position the knob domain for every bovine antibody. As such, knob domain peptides in isolation constitute a practicable disulphide-rich peptide, being around a third of the molecular weight of a camelid VHH and approximately twice the size of a conventional therapeutic macrocycle.

After immunisation with C5, deep sequencing analysis allowed us to observe 154 distinct ultralong CDRH3 sequences out of a total of 3,559 total unique bovine CDRH3 sequences. In sequencing of naïve bovine antibody repertoires, ultralong CDRH3 are reported to occur at a frequency of approximately 10%, although ranges of 1%–20% have been reported [6, 10, 14, 32–35]. In our antigen-specific CDRH3 pool, the ultralong CDRH3 were observed at a frequency of 4.3%. For the discovery of knob domains, conventional PCR favours amplification of shorter amplicons due to their higher insertion rate [36], which could deselect for ultralong CDRH3. As an alternative, suppression PCR, which favours the amplification of longer amplicons [37], could be explored to try and bias the reaction toward amplification of longer CDRH3 sequences.

From our initial screening with CDRH3-ScFc constructs, we observed a 27% hit rate for C5 binding. While entirely practicable for knob domain discovery, the attrition at this stage may indicate that not all ultralong CDRH3 can function entirely independently of their parent antibody. Previous studies have suggested that the β -ribbon stalk of an ultralong CDRH3 is obligate to orientate the knob domain peptide [6–8, 38]. Inspection of bovine Fab structures in the Protein Data Bank (PDB) frequently reveals stabilising interactions between the knob domain peptide and the β -ribbon stalk, and numerous interactions between the β -ribbon stalk and neighbouring CDRs and V_H framework residues [6–8]. In one recent structure (BOV-7, PDB accession code: 6E9U [7]), a disulphide bond was observed between cysteine residues of the stalk and knob domain [7]. Therefore, for certain ultralong CDRH3, removal of the stalk may in turn remove critical interactions required to stabilise the knob domain. Despite this, we have shown that our method can be used to identify isolated knob domains that are capable of

binding antigen autonomously, in the absence of the ultralong CDR-H3 stalk region. For the K60 knob domain, which was the only knob domain peptide to become nonfunctional when removed from the β -ribbon stalk, the loss of stabilising interactions of the knob domain peptide and the β -ribbon stalk may have been decisive, markedly increasing the conformational entropy barrier to binding.

For the remaining knob domains, all 5 bound C5 when inserted in the CDRH3 of PGT121, and all were viable binders when proteolytically excised from a Fab scaffold. In the case of the tight binder, K136, a modest decrease in affinity was observed in the isolated knob domain peptide, mediated largely by an increase in the value of k_{off} . Conversely, binding of the K8- and K149-PGT121 constructs was slightly improved in the isolated peptides.

We note that although the K_{D} values between the PGT121 knob domain fusion proteins and the isolated knob domain are similar, there are differences in the rate constants. Typically, both the k_{on} and k_{off} are increased in the isolated knob domain compared to the PGT121 fusion protein. In the PGT121 fusion proteins, the N- and C- terminals of the knob domain are constrained by the polypeptide chain and held in proximity, in a manner similar to the native bovine antibody; this might reduce the conformational freedom of the domain and confer stability to the complex. For K_{on} , the smaller knob domains might have an advantage in avoiding the steric hinderance which is inherent with a larger scaffold, allowing faster diffusion to buried or transient epitopes.

Our FRET competition experiments provide further evidence that knob domain peptides bind C5 with affinities in the low nanomolar range. The K_{i} values from the FRET assay are in close agreement with the K_{D} values measured from our kinetic experiments, particularly for the K8, K57, and K149 knob domains, which are within 2-fold. For the K92 and K136 knob domains, which displayed the slowest K_{off} in SPR experiments, the FRET K_{i} values were approximately 3-fold less than the K_{D} values determined by SPR, but still indicate a high-affinity interaction with C5.

Our data suggest that, for the ultralong CDRH3 population that function autonomously, retaining activity when extracted from parent antibody, the degree of attrition when reformatting from ultralong CDRH3 to knob domain peptide is low. Through our biophysical and biochemical analysis, we have demonstrated that knob domain peptides bind their targets with high affinity akin to a conventional monoclonal antibody, but at a markedly reduced molecular weight, affording an improved ligand efficacy.

Unlike conventional antibody fragments, our study suggests that knob domain peptides can be readily purified by high-resolution RP-HPLC. This apparent robustness, most likely arising from an absence of denaturable secondary structure and the presence of stabilising disulphide bonds, could be advantageous for pharmaceutical development. For example, RP-HPLC can be performed over a wide pH range, allowing pH-driven amino acid modifications, such as deamidation of asparagine [39] and formation of pyroglutamic acid from N-terminal glutamine [40], to be minimised.

Isolated bovine-derived knob domain peptides are novel peptide molecules, with low molecular weight (4–6 kDa) and with intrinsic rigidity and thermal stability, due to extensive disulphide bonding [41]. Having undergone affinity maturation *in vivo*, knob domain peptides may be more specific for their target protein than peptides derived by naïve display methods. When considered as potential therapies, the reduced conformational freedom, imposed by the additional structural constraints of the disulphide bonds, may enhance metabolic stability [42] and bioavailability [43] and reduce off-target effects [44]. The method described here offers a tractable way to isolate these structures with attractive yields and purity from immunised bovines.

Methods

Ethics statement

Bovine immunisations were subject to University of Reading Animal Welfare Ethical Review Body approval (Personal Project Licence: 70/8108 “Antibody Production”) and were conducted in accordance with the UK Animals (Scientific Procedures) Act, 1986 and ARRIVE guidelines, whereby every effort was made to minimise pain and discomfort experienced by animals.

Immunisation of Holstein Friesians with Complement C5

Two adult Friesian cows were immunised with Complement C5 (CompTech). For early immunisations, 1.25 mg of C5 was mixed 1:1 (v/v) with Adjuvant Fama (GERBU Biotechnik). Three subcutaneous injections into the shoulder were performed at 1-month intervals. Ten days post immunisation, 10 mL of blood was taken to allow testing of the serum antibody titre. Due to a lower than anticipated serum titre, a further 2 immunisations were performed with revised protocols. For subsequent boosts, 1.25 mg C5 was emulsified 1:1 with Freund’s complete adjuvant (Sigma)—and, for the final shot, Montanide (Seppic)—immediately prior to subcutaneous injection into the shoulder. A serum bleed was taken 10 days later to check the serum antibody titre (S1 Fig).

Harvesting of immune material

Five-hundred-millilitre samples of whole blood were taken, post immunisation with C5. PBMCs were isolated using Leucosep tubes (Griener Bio-one), as per the manufacturer’s instructions. After immunisation, the animals were euthanized, and a draining lymph node from the neck—adjacent to the site of immunisation—and a portion of spleen were collected. The tissues were homogenised using a gentle MACS tissue dissociator (Miltenyi), passed through a 40 μ m cell strainer, and collected in Roswell Park Memorial Institute (RPMI) 10% foetal calf serum (FCS). Cells were frozen in FCS, 10% dimethyl sulfoxide.

Sorting of antigen-specific immune cells by flow cytometry

A sample of draining lymph node was thawed at 37°C and re-suspended in RPMI 1640 Medium, 10% FCS (v/v), 1 mM ethylenediaminetetraacetic acid (EDTA). The cells were centrifuged for 5 min at 400g and the supernatant removed. The cell pellet was disrupted and resuspended in Assay Buffer (AB): phosphate buffered saline (PBS), 1 mM EDTA, 1% bovine serum albumin (BSA) (w/v), 25 mM HEPES. The cells were centrifuged as before and resuspended in 2 mL ice-cold AB containing 2 μ g/mL each of C5-AlexaFluor 488 and C5-AlexaFluor 647 and incubated for 30 min on ice. The cells were then centrifuged, the supernatant removed, and the pellet washed in ice-cold AB. An aliquot was taken for counting. The cells were centrifuged again, the supernatant removed, and the cells resuspended to 5×10^6 /mL in ice-cold AB before filtering through a 40 μ m mesh; 4',6-diamidino-2-phenylindole (DAPI) was added at a final concentration of 1 μ g/mL just before acquisition on a BD Biosciences FACSAria III (San Jose) cell sorter. Cells were identified by forward and side scatters, and DAPI-positive dead cells were removed from the analysis. Single cells were identified by pulse processing of height and area scatter parameters. Cells positive for both C5-AlexaFluor 488 and C5-AlexaFluor 647 were then identified and sorted into 1.5 mL Eppendorf tubes containing 1 mL PBS, 20% FCS, 25 mM HEPES kept at 4°C.

The AlexaFluor 488 was excited by a 488 nm laser and collected through a 530/30 nm band pass (BP) filter, and the AlexaFluor 647 was excited by a 640 nm laser and collected through a

660/20 nm BP filter. DAPI was excited by a 407 nm laser and collected through a 450/40 nm BP filter.

RT-PCR on B cell lysate

The C5-specific B cells from the FACS were pelleted by centrifugation at 10,000g in a 4°C centrifuge. The cells were resuspended and lysed with 120 µL of an ice-cold solution of NP-40 detergent (0.5% v/v) and RNasin (Promega) at 1 U/µL. An RT-PCR mix was prepared using Super Script IV vilo Master Mix (Invitrogen), comprising 32 µL of cell lysate and 8 µL of Master Mix. The reaction mix was incubated at 25°C for 10 minutes, 50°C for a further 10 minutes, and finally, heated to 85°C for 5 minutes.

Primary PCR

A primary PCR was used to specifically amplify IgG CDR3 cDNA sequences. The forward primer anneals to the conserved framework 3 sequence of the V_H, and the reverse primer sequence anneals to the conserved framework 4 V_H sequence. The PCR product when read from 5' to 3' therefore encodes the CDRH3 sequence irrespective of length, amino acid sequence, or composition of V-, D-, J- gene segments. The PCR mix was prepared using a Hot-start KOD master mix kit (Merck Millipore), as per the manufacturer's instructions. The primers used were 5'-GGACTCGGCCACMTAYTACTG-3' and 5'-GCTCGAGACGGTGAYCAG-3', and 2 µL of cDNA template was used per 50 µL PCR. The reaction mix was heated for 2 minutes at 96°C and then subjected to 30 cycles of: 96°C for 30 seconds, 55°C for 30 seconds, and 68°C for 60 seconds. Finally, the mix was heated at 68°C for 5 minutes.

Gel purification

The PCR product contains a polyclonal mixture of CDR3 sequences, comprising regular and ultralong CDR3, which were visualised on an analytical gel. An excision was taken with spanning between approximately 250 and 500 bp, based on the marker. A QiaQUICK gel extraction kit (Qiagen) was used to extract the DNA from the excised portion of the gel, as per the manufacturer's instructions.

Barcoding PCR and gel purification

A secondary PCR was performed to barcode sequences for ion torrent sequencing. The primers were as used before but with the addition of adaptor (*italics*) and barcoding sequence (denoted with N): 5'-*CCATCTCATCCCTGCGTGTCTCCGACTCAGNNNNNNNNN*CGGACTCGGCCACMTAYTACTG-3' and 5'-*CCTCTCTATGGGCAGTCGGTGATGCTCGAGACGGTGAYCAG*-3'. The secondary PCR was performed using a KOD Master Mix kit, as described for the primary PCR. The secondary PCR product was gel purified and concentrated. Finally, the samples were purified using Beckman coulter AMP magnetic beads, as per the manufacturer's instructions.

Ion Torrent deep sequencing of CDR3 libraries

Two samples were prepared for deep sequencing from a single draining lymph node sample from a single animal. The purified DNA samples were diluted to 20 ng/µL (1 µg total) and stored at -20°C. Deep sequencing of all samples was performed using an Ion Torrent PGM technology commercial service by MACROGEN on a 318 chip as previously described [45, 46].

Briefly, approximately 1.1 Gb of data were obtained per chip, which translates to 6,613,415 raw reads of a 170-bp mean length. The first processing step consisted of conversion of the FASTQ to FASTA followed by demultiplexing of the latter to sort the sequences according to their barcodes and to generate individual FASTA files. Each FASTA file was then translated in all 3 reading frames and concatenated into one file containing all frames. Using Perl scripts, the amino acid motifs of the translated primers (DSATYY and LL[V,I]TVSS), which correspond to the conserved framework 3 and framework 4 regions either side of the CDRH3, were used to identify sequences of interest. The number of meaningful reads, after processing, for each sample was found to be approximately 680,000 and 530,000.

Screening of knob-TEV-HKH-ScFc supernatants

To screen for C5 binding knob domains, ultralong CDR3 sequences were selected based on clonotype, copy number, and cysteine pattern to make a representative set of 52 sequences. Cloning into a pMH expression vector, containing a human cytomegalovirus (CMV) promoter and an in-frame ScFc expression tag (S1 Text and S2 Fig), was performed by Twist Biosciences. For this, and all subsequent protein expressions, a *Mus musculus* immunoglobulin heavy chain leader sequence was used: MEWSWVFLFFLSVTTGVHS (accession number: A0N1R4_MOUSE).

Plasmid DNA for each construct was amplified using miniprep kits (Qiagen). Individual 2.0 mL Expi293F cell cultures (Thermo Scientific), at 3×10^6 cells/mL, per construct, were set up in 48-well culture blocks using Expifectamine 293 Transfection kits (Invitrogen), as per the manufacturer's instructions. The cells were cultured for 4 days and centrifuged at 2,500 rpm for 30 minutes.

C5 binding ELISA

Ninety-six-well ELISA plates (Nunc Maxisorp) were coated with a 2 µg/mL solution of C5, purified from serum as previously described [47], in carbonate-bicarbonate buffer (Sigma). All washing steps comprised 4 wash cycles with PBS, 0.05% Tween 20. Blocking buffer was PBS, 1% BSA (w/v). Cell supernatants were plated as 1:10 and 1:100 dilutions in AB (PBS, 0.05% Tween 20, 0.1% BSA [w/v]). To reveal, a 1/5,000 dilution of a goat anti-human Fc, Horse Radish Peroxidase (Thermo Scientific) secondary antibody was used with 'One-Step' 3,3',5,5' Tetramethylbenzidine (Thermo Scientific). The reaction was stopped with the addition of a 2% (w/v) NaF solution and the OD measured at 630 and 390 nm wavelengths using a BMG labtech plate reader.

Protein expression and purification

To enable expression of the PGT121-knob fusion proteins, plasmid DNA was amplified using QIAGEN Plasmid Plus Giga Kits and quantified by A260. Individual Expi293F cell cultures, at 3×10^6 cells/mL, per construct, were set up using Expifectamine 293 Transfection kits (Invitrogen), as per the manufacturer's instructions. Both heavy-chain and light-chain plasmid DNA were co-transfected at 1 mg/L. The cells were cultured for 4 days, centrifuged at 4,000 rpm for 1 hour, and filtered through a 0.22 µm filter. Using an Akta pure (GE Healthcare), Hi-Trap Nickel excel columns (GE Healthcare) were equilibrated with 10 column volumes (CV) of PBS. Cell supernatants were loaded at 1.0 mL/min, and the column was washed with 7× CV of PBS, 0.5 M NaCl. The column was then washed with 7× CV of Buffer A (0.5 M NaCl, 0.02 M imidazole, PBS [pH 7.3]). Protein samples were eluted by isocratic elution with 10× CV of Buffer B (0.5 M NaCl, 0.25 M imidazole, PBS [pH 7.3]), as 1.0 mL fractions. The column was washed with 0.1 M NaOH and re-equilibrated into PBS prior to subsequent loading. Post

elution, the protein containing fractions were pooled and buffer exchanged into PBS, using PD-10 columns (GE Healthcare).

Cleavage and purification of knob domain peptides

Fab-knob peptide fusion proteins (10 mg/mL) were incubated with TEV protease, at a ratio of 100:1 (*w/w*), for a minimum of 2 hours at room temperature. Peptides were purified using a Waters UV-directed FractionLynx system with a Waters XBridge Protein BEH C4 OBD Prep Column (300 Å, 5 µm, 19 mm × 100 mm). An aqueous solvent of water, 0.1% trifluoroacetic acid (TFA) and an organic solvent of 100% MeCN was used. The column was run at 20 mL/min at 40°C with a gradient of 5%–50% organic solvent, over 11.0 minutes. The column was cleaned with 3 sharp ramps of 5%–95% organic solvent. Fractions containing knob peptide were pooled and lyophilised using a Labconco Freezone freeze drier. For –80°C storage and subsequent analysis, peptides were resuspended with 20 mM Tris (pH 7.4).

Biacore single-cycle kinetics

Using a Biacore 8K (GE Healthcare), C5 purified from serum [47] was immobilized on a CM5 chip by amine coupling. For PGT121 fusion proteins, flow cells were activated using a minimal immobilisation protocol: EDC/NHS was mixed at 1:2 ratio (flow rate, 10 µL/min; contact time, 30 s). C5, at 1 µg/mL in pH 4.5 sodium-acetate buffer, was immobilized on flow cell 2 only (flow rate, 10 µL/min; contact time, 420 s). Finally, ethanolamine was applied to both flow cells (flow rate, 10 µL/min; contact time, 420 s). A final immobilization level of approximately 100–160 response units was obtained for PGT121 knob domain fusion protein samples.

For knob domain peptides, flow cells were activated using a standard immobilisation protocol: EDC/NHS was mixed at 1:1 ratio (flow rate, 10 µL/min; contact time, 30 s). C5, at 1 µg/mL in pH 4.5 sodium-acetate buffer, was immobilized on flow cell 2 only (flow rate, 10 µL/min; contact time, 420 s). Finally, ethanolamine was applied to both flow cells (flow rate, 10 µL/min; contact time, 420 s). A final immobilization level of approximately 1,500–2,000 response units was obtained for peptide samples. For the PGT121-K149 fusion protein and K149 knob domain, the same immobilisation protocols were used but with a 750 ng/mL solution of C5, instead of 1 µg/mL. This resulted in immobilisation level of 50–85 RU for the minimal immobilisation protocol and 500–570 RU for the standard protocol.

For counter screen experiments, the same minimal and standard immobilisation protocols were used with 1 µg/mL of either C3 (CompTech) or ovalbumin (Sigma), in place of C5. For C3, the minimal immobilisation protocol achieved 40–55 RU of immobilisation. The standard immobilisation protocol used for the knob domain peptides achieved 2,300–2,650 RU of immobilisation for C3 and 330–350 RU for ovalbumin.

One-micromolar stock solutions of the analytes (PGT121 knob domain fusion proteins and free knob domain peptides) were prepared in HBS-EP buffer. A series of six 3-fold dilutions of each stock solution was prepared in the same buffer. The result was 7 solutions for each analyte, at concentrations ranging from 1 µM down to 1.37 nM. A high flow rate of 40 µL/min was used, with a contact time of 230 s and a dissociation time of 1,800 s for knob domain peptides and 10,000 s for PGT121 knob domain fusion proteins. For the PGT121-K149 fusion protein and K149 knob domain, the flow rate was further increased to 100 µL/min. Binding to the reference surface was subtracted, and the data were fitted to a single-site binding model using Biacore evaluation software. Data were inspected for evidence of mass transport limitation using the flow-rate-independent component of the mass transfer constant (*tc*).

LC/MS

Knob domain peptides were diluted to 0.1 mg/mL in PBS. Protein LC/MS was performed on a Waters Xevo G2 QToF with a Waters Acquity UPLC.

For chromatography, a Waters BioResolve RP mAB Polyphenyl column (450 Å, 2.7 μm, 2.1 × 150 mm + VanGuard) was used, with an aqueous mobile phase of water, 0.02% TFA, and 0.08% formic acid (FA) and an organic mobile phase of 95% acetonitrile, 5% water, 0.02% TFA, and 0.08% FA. The column was run at 0.6 mL/min at 80°C with a gradient of 5%–50% organic solvent, over 8.8 minutes. At the end of the injection, a sharp ramp to 95% organic solvent and a single cycle of 5%–95% organic solvent used to clean the column.

For mass spectrometry, the system was configured as follows: ion mode: ESI positive ion mode; acquisition mode: resolution; mass range: 400–5,000 m/z; cone voltage: 30 V; capillary voltage: 3.2 kV; desolvation temperature: 350°C; desolvation gas: 1,000 L/h and source temperature: 150°C. Data were analysed with MassLynx software.

FRET

C5 purified from serum [47] was labelled with an amine reactive terbium chelate (molecular probes, Life Technologies), as per the manufacturer's instructions. Briefly, C5 at 1.15 mg/mL was buffer exchanged into a 50 mM Bicine, 100 mM NaCl (pH 8.2) buffer using a Zeba column (Thermo Scientific). Terbium was reconstituted in DMSO at 5 mM and added to 1% final (v/v) to create an approximate 10-fold molar excess to C5. After a 1-hour incubation at room temperature, unbound dye was removed by 2 sequential buffer exchanges into 20 mM Tris, 100 mM NaCl (pH 7.4), again using Zeba columns (Thermo Scientific). The labelling ratio was quantified by UV spectroscopy; the final molar ratio of dye to protein was 4:1. Next, PGT121 knob domain fusion proteins were labelled with an amine reactive AlexaFluor 647 (AF647) dye (Molecular Probes, Life Technologies), using the same protocol as above but with a 30-minute incubation with dye. After removal of unbound dye, UV spectroscopy quantified the final molar ratio of dye to protein as 2:1.

For determination of PGT121 knob domain fusion K_D *app*, C5 Tb was plated into a black, low-volume 384-well assay plate (Corning) to give a final assay concentration (FAC) of 1 nM; either HBS-EP buffer (GE Healthcare) or unlabelled C5 (1 μM FAC) was added. Serial dilutions of PGT121 knob domain fusion proteins were prepared in HBS-EP buffer to give a range of 100 nM–0.046 nM or 500 nM–0.22 nM (FAC). The plates were wrapped in foil and incubated for 24 hours, with shaking. The plates were read on an Envision plate reader (Perkin Elmer) at 2- and 24-hour intervals (HTRF laser, Excitation 330 nm and Emission 665/615 nm). For fitting, background was subtracted, and curves were fitted using prism software to a 4-parameter logistic model.

For competition assays, C5-Tb was plated to 1 nM (FAC) and serial dilutions of knob domain peptide prepared in HBS-EP buffer, to give a range of 1000 nM–0.46 nM (FAC). PGT121 knob domain fusion-AF647 proteins were prepared to give the following concentrations, which equate to the K_D *app* measured in the previous experiment: PGT121-K8 AF647 5 nM (FAC), PGT121-K57 AF647 3 nM (FAC), PGT121-K92 AF647 12 nM (FAC), PGT121-136 AF647 40 nM (FAC), and PGT121-149 AF647 66 nM (FAC). The plates were wrapped in foil, incubated for 24 hours, with shaking, and read on an Envision plate reader (HTRF laser, Excitation 330 nm and Emission 665/615 nm). Curves were fitted using prism software to a 4-parameter logistic model. IC_{50} values were converted to inhibitory constants (K_i) using the Cheng-Prusoff equation [48].

$$K_i = IC_{50} \div \left(1 + \frac{[R]}{K_D} \right)$$

in which [R] is the concentration of PGT121 knob domain AF647 used in the assay, and K_D is the apparent equilibrium dissociation constant of the PGT121 knob domain AF647 in the assay.

Supporting information

S1 Fig. Serum antibody titres in a C5 serum ELISA. The terminal bleed achieved a titre in excess of 1/10,000.

(PDF)

S2 Fig. Diagram of the ScFc construct. A cartoon schematic showing the domain and linker arrangement of the ScFc tag.

(PDF)

S3 Fig. Biacore single-cycle kinetics complement component C3 counter screen of PGT121 knob domain fusion proteins. No non-specific binding to C3 was observed with the PGT121 knob domain fusion proteins.

(PDF)

S4 Fig. LC/MS on purified peptides, comparison of predicted charge envelopes versus experimental data (1 of 2). 4⁺ charge envelopes are shown.

(PDF)

S5 Fig. LC/MS on purified peptides, comparison of predicted charge envelopes versus experimental data (2 of 2). 4⁺ charge envelopes are shown.

(PDF)

S6 Fig. Biacore single-cycle kinetics complement component C3 counter screen of knob domain peptides. No non-specific binding to C3 was observed with the knob domain peptides.

(PDF)

S7 Fig. Biacore single-cycle kinetics ovalbumin counter screen of knob domain peptides. No non-specific binding to ovalbumin was observed with the knob domain peptides.

(PDF)

S1 Text. Amino acid sequence of the cleavable C-terminal ScFc tag. The ScFc is composed of: CH2-CH3-linker-CH2-CH3. TEV site is shown in bold and poly-His tag in italics.

(DOCX)

S2 Text. Amino acid sequences of the PGT-121 Fab knob domain fusion. The heavy chain sequences of the PGT-121-knob domain fusions were as follows: the knob domain sequences are shown in italics, with the TEV protease cleavage sites shown in bold.

(DOCX)

S1 Table. Clonotyping of 154 ultralong CDRH3 sequences derived from deep sequencing of an antigen specific (C5⁺) pool of PBMCs. The sequences in this table have been derived from 2 samples prepared from the same draining lymph node from a single cow. The first and last residue of each sequence correspond to H93 to H102 of the Kabat numbering scheme, respectively, irrespective of CDRH3 length. Clonotypes are assigned on the basis of $\geq 75\%$ sequence homology. CDRH3 selected for transient expression and screening as ScFc fusion proteins are highlighted in bold. Sequences marked with a star were characterised as isolated knob domains.

(DOCX)

S2 Table. Panel of 52 ultralong CDRH3 sequences selected for reformatting as ScFc fusion proteins.

(DOCX)

S3 Table. List of 14 ultralong CDRH3 which bound C5 in a single-point ELISA screen.

Sequences chosen for reformatting as PGT121-knob domain fusion proteins are shown in bold.

(DOCX)

S4 Table. Biacore single-cycle kinetics data on PGT121 Fab–knob domain fusion proteins.

Summary of kinetics from $n = 3$. (for individual occasions see S4B Table, S4C Table and S4D Table).

(DOCX)

S5 Table. Knob domain sequences derived from Fab cleavage.

(DOCX)

S6 Table. LC/MS on purified peptides identifies masses consistent with the predicted isotope patterns, based on peptide amino acid sequences and formation of disulphide bonds.

Data are shown for the 4⁺ charge state. DSB, disulphide bonds

(DOCX)

S7 Table. Biacore single-cycle kinetics data on isolated knob domain peptides. Summary of kinetics from $n = 3$ occasions (for individual occasions see S7B Table, S7C Table and S7D Table).

(DOCX)

S8 Table. FRET assay K_D *app* for PGT121 fusion proteins binding to C5-Tb. Summary data table for $n = 3$ experiments. For 2-hour data, refer to S8A Table and for 24-hour data, refer to S8B Table.

(DOCX)

S9 Table. Competition FRET assays to derive IC_{50} and K_i values for knob domain peptides.

Summary data table for $n = 3$ experiments.

(DOCX)

S1 Data.

(XLSX)

Acknowledgments

We would like to thank Barney Jones from the Centre for Dairy Research and Professor Gary Stephens and Hong Lin from the University of Reading for their assistance with the bovine immunisations. We would also like to thank Chris Trower for his veterinary advice.

Author Contributions

Conceptualization: Alex Macpherson, Jean van den Elsen, Alastair D. G. Lawson.

Data curation: Alex Macpherson.

Formal analysis: Alex Macpherson, Jean van den Elsen, Alastair D. G. Lawson.

Methodology: Alex Macpherson, Anthony Scott-Tucker, Anastasios Spiliotopoulos, Catherine Simpson, Justin Staniforth, Adam Hold, James Snowden, Leah Manning, Alastair D. G. Lawson.

Supervision: Jean van den Elsen, Alastair D. G. Lawson.

Visualization: Alex Macpherson.

Writing – original draft: Alex Macpherson, Jean van den Elsen, Alastair D. G. Lawson.

Writing – review & editing: Alex Macpherson, Anthony Scott-Tucker, Anastasios Spiliotopoulos, Justin Staniforth, Adam Hold, James Snowden, Leah Manning, Jean van den Elsen, Alastair D. G. Lawson.

References

1. Hamers-Casterman C, Atarhouch T, Muyldermans S, Robinson G, Hamers C, Songa EB, et al. Naturally occurring antibodies devoid of light chains. *Nature*. 1993; 363(6428):446–8. Epub 1993/06/03. <https://doi.org/10.1038/363446a0> PMID: 8502296.
2. Greenberg AS, Avila D, Hughes M, Hughes A, McKinney EC, Flajnik MF. A new antigen receptor gene family that undergoes rearrangement and extensive somatic diversification in sharks. *Nature*. 1995; 374(6518):168–73. Epub 1995/03/09. <https://doi.org/10.1038/374168a0> PMID: 7877689.
3. Muyldermans S, Atarhouch T, Saldanha J, Barbosa JA, Hamers R. Sequence and structure of VH domain from naturally occurring camel heavy chain immunoglobulins lacking light chains. *Protein Eng*. 1994; 7(9):1129–35. Epub 1994/09/01. PMID: 7831284.
4. Nuttall SD, Krishnan UV, Hattarki M, De Gori R, Irving RA, Hudson PJ. Isolation of the new antigen receptor from wobbegong sharks, and use as a scaffold for the display of protein loop libraries. *Mol Immunol*. 2001; 38(4):313–26. Epub 2001/09/22. PMID: 11566324.
5. Berens SJ, Wylie DE, Lopez OJ. Use of a single VH family and long CDR3s in the variable region of cattle Ig heavy chains. *Int Immunol*. 1997; 9(1):189–99. Epub 1997/01/01. <https://doi.org/10.1093/intimm/9.1.189> PMID: 9043960.
6. Wang F, Ekiert DC, Ahmad I, Yu W, Zhang Y, Bazirgan O, et al. Reshaping antibody diversity. *Cell*. 2013; 153(6):1379–93. Epub 2013/06/12. <https://doi.org/10.1016/j.cell.2013.04.049> PMID: 23746848; PubMed Central PMCID: PMC4007204.
7. Dong J, Finn JA, Larsen PA, Smith TPL, Crowe JE Jr. Structural Diversity of Ultralong CDRH3s in Seven Bovine Antibody Heavy Chains. *Front Immunol*. 2019; 10:558. Epub 2019/04/11. <https://doi.org/10.3389/fimmu.2019.00558> PMID: 30967877; PubMed Central PMCID: PMC6440498.
8. Stanfield RL, Wilson IA, Smider VV. Conservation and diversity in the ultralong third heavy-chain complementarity-determining region of bovine antibodies. *Sci Immunol*. 2016; 1(1). Epub 2016/08/31. <https://doi.org/10.1126/sciimmunol.aaf7962> PMID: 27574710; PubMed Central PMCID: PMC5000368.
9. Zhang Y, Goswami D, Wang D, Wang TS, Sen S, Magliery TJ, et al. An antibody with a variable-region coiled-coil "knob" domain. *Angew Chem Int Ed Engl*. 2014; 53(1):132–5. Epub 2013/11/21. <https://doi.org/10.1002/anie.201307939> PMID: 24254636; PubMed Central PMCID: PMC3926434.
10. Stanfield RL, Haakenson J, Deiss TC, Criscitiello MF, Wilson IA, Smider VV. The Unusual Genetics and Biochemistry of Bovine Immunoglobulins. *Adv Immunol*. 2018; 137:135–64. Epub 2018/02/20. <https://doi.org/10.1016/bs.ai.2017.12.004> PMID: 29455846; PubMed Central PMCID: PMC5935254.
11. Sok D, Le KM, Vadnais M, Saye-Francisco KL, Jardine JG, Torres JL, et al. Rapid elicitation of broadly neutralizing antibodies to HIV by immunization in cows. *Nature*. 2017; 548(7665):108–11. Epub 2017/07/21. <https://doi.org/10.1038/nature23301> PMID: 28726771; PubMed Central PMCID: PMC5812458.
12. Stanfield RL, Berndsen ZT, Huang R, Sok D, Warner G, Torres JL, et al. Structural basis of broad HIV neutralization by a vaccine-induced cow antibody. *Sci Adv*. 2020; 6(22):eaba0468. Epub 2020/06/11. <https://doi.org/10.1126/sciadv.aba0468> PMID: 32518821; PubMed Central PMCID: PMC7253169.
13. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem*. 2004; 25(13):1605–12. Epub 2004/07/21. <https://doi.org/10.1002/jcc.20084> PMID: 15264254.
14. Deiss TC, Vadnais M, Wang F, Chen PL, Torkamani A, Mwangi W, et al. Immunogenetic factors driving formation of ultralong VH CDR3 in *Bos taurus* antibodies. *Cell Mol Immunol*. 2019; 16(1):53–64. Epub 2017/12/05. <https://doi.org/10.1038/cmi.2017.117> PMID: 29200193; PubMed Central PMCID: PMC6318308.
15. Giudicelli V, Duroux P, Ginestoux C, Folch G, Jabado-Michaloud J, Chaume D, et al. IMGT/LIGM-DB, the IMGT comprehensive database of immunoglobulin and T cell receptor nucleotide sequences. *Nucleic Acids Res*. 2006; 34(Database issue):D781–4. Epub 2005/12/31. <https://doi.org/10.1093/nar/gkj088> PMID: 16381979; PubMed Central PMCID: PMC1347451.

16. Stanfield RL, Wilson IA, Smider VV. Conservation and diversity in the ultralong third heavy-chain complementarity-determining region of bovine antibodies. *Sci Immunol*. 2016; 1(1):aaf7962. Epub 2017/08/08. <https://doi.org/10.1126/sciimmunol.aaf7962> PMID: 28783675.
17. 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. *J Immunol*. 2002; 168(10):5155–62. Epub 2002/05/08. <https://doi.org/10.4049/jimmunol.168.10.5155> PMID: 11994470.
18. Hill A. Eculizumab for the treatment of paroxysmal nocturnal hemoglobinuria. *Clin Adv Hematol Oncol*. 2005; 3(11):849–50. Epub 2006/02/24. PMID: 16491626.
19. Hill A, Hillmen P, Richards SJ, Elebute D, Marsh JC, Chan J, et al. Sustained response and long-term safety of eculizumab in paroxysmal nocturnal hemoglobinuria. *Blood*. 2005; 106(7):2559–65. Epub 2005/06/30. <https://doi.org/10.1182/blood-2005-02-0564> PMID: 15985537.
20. Nurnberger J, Philipp T, Witzke O, Opazo Saez A, Vester U, Baba HA, et al. Eculizumab for atypical hemolytic-uremic syndrome. *N Engl J Med*. 2009; 360(5):542–4. Epub 2009/01/31. <https://doi.org/10.1056/NEJMc0808527> PMID: 19179328.
21. Morgan BP, Harris CL. Complement, a target for therapy in inflammatory and degenerative diseases. *Nat Rev Drug Discov*. 2015; 14(12):857–77. Epub 2015/10/24. <https://doi.org/10.1038/nrd4657> PMID: 26493766.
22. Scott JK, Smith GP. Searching for peptide ligands with an epitope library. *Science*. 1990; 249(4967):386–90. Epub 1990/07/27. <https://doi.org/10.1126/science.1696028> PMID: 1696028.
23. Lam KS, Salmon SE, Hersh EM, Hruby VJ, Kazmierski WM, Knapp RJ. A new type of synthetic peptide library for identifying ligand-binding activity. *Nature*. 1991; 354(6348):82–4. Epub 1991/11/07. <https://doi.org/10.1038/354082a0> PMID: 1944576.
24. Crook ZR, Sevilla GP, Friend D, Brusniak MY, Bandaranayake AD, Clarke M, et al. Publisher Correction: Mammalian display screening of diverse cysteine-dense peptides for difficult to drug targets. *Nat Commun*. 2018; 9(1):1072. Epub 2018/03/11. <https://doi.org/10.1038/s41467-018-03350-5> PMID: 29523778; PubMed Central PMCID: PMC5844863.
25. Haakenson JK, Deiss TC, Warner GF, Mwangi W, Criscitiello MF, Smider VV. A Broad Role for Cysteines in Bovine Antibody Diversity. *Immunohorizons*. 2019; 3(10):478–87. Epub 2019/10/18. <https://doi.org/10.4049/immunohorizons.1900058> PMID: 31619454.
26. Saini SS, Allore B, Jacobs RM, Kaushik A. Exceptionally long CDR3H region with multiple cysteine residues in functional bovine IgM antibodies. *Eur J Immunol*. 1999; 29(8):2420–6. Epub 1999/08/24. [https://doi.org/10.1002/\(SICI\)1521-4141\(199908\)29:08<2420::AID-IMMU2420>3.0.CO;2-A](https://doi.org/10.1002/(SICI)1521-4141(199908)29:08<2420::AID-IMMU2420>3.0.CO;2-A) PMID: 10458755.
27. Newnham LE, Wright MJ, Holdsworth G, Kostarelos K, Robinson MK, Rabbitts TH, et al. Functional inhibition of beta-catenin-mediated Wnt signaling by intracellular VHH antibodies. *MAbs*. 2015; 7(1):180–91. Epub 2014/12/20. <https://doi.org/10.4161/19420862.2015.989023> PMID: 25524068; PubMed Central PMCID: PMC4622660.
28. Mouquet H, Scharf L, Euler Z, Liu Y, Eden C, Scheid JF, et al. Complex-type N-glycan recognition by potent broadly neutralizing HIV antibodies. *Proc Natl Acad Sci U S A*. 2012; 109(47):E3268–77. Epub 2012/11/02. <https://doi.org/10.1073/pnas.1217207109> PMID: 23115339; PubMed Central PMCID: PMC3511153.
29. Dessauer A, Rother U. The fifth component of complement (C5): purification without activation. *Immunobiology*. 1983; 164(5):370–9. Epub 1983/05/01. [https://doi.org/10.1016/S0171-2985\(83\)80033-3](https://doi.org/10.1016/S0171-2985(83)80033-3) PMID: 6347874.
30. Onell A, Andersson K. Kinetic determinations of molecular interactions using Biacore—minimum data requirements for efficient experimental design. *J Mol Recognit*. 2005; 18(4):307–17. Epub 2005/06/14. <https://doi.org/10.1002/jmr.745> PMID: 15952219.
31. Janssen BJ, Huizinga EG, Raaijmakers HC, Roos A, Daha MR, Nilsson-Ekdahl K, et al. Structures of complement component C3 provide insights into the function and evolution of immunity. *Nature*. 2005; 437(7058):505–11. Epub 2005/09/24. <https://doi.org/10.1038/nature04005> PMID: 16177781.
32. Liljavirta J, Niku M, Pessa-Morikawa T, Ekman A, Iivanainen A. Expansion of the preimmune antibody repertoire by junctional diversity in *Bos taurus*. *PLoS ONE*. 2014; 9(6):e99808. Epub 2014/06/14. <https://doi.org/10.1371/journal.pone.0099808> PMID: 24926997; PubMed Central PMCID: PMC4057420.
33. Ma L, Qin T, Chu D, Cheng X, Wang J, Wang X, et al. Internal Duplications of DH, JH, and C Region Genes Create an Unusual IgH Gene Locus in Cattle. *J Immunol*. 2016; 196(10):4358–66. Epub 2016/04/08. <https://doi.org/10.4049/jimmunol.1600158> PMID: 27053761.

34. Walther S, Czerny CP, Diesterbeck US. Exceptionally long CDR3H are not isotype restricted in bovine immunoglobulins. *PLoS ONE*. 2013; 8(5):e64234. Epub 2013/05/30. <https://doi.org/10.1371/journal.pone.0064234> PMID: 23717573; PubMed Central PMCID: PMC3661452.
35. Walther S, Tietze M, Czerny CP, König S, Diesterbeck US. Development of a Bioinformatics Framework for the Detection of Gene Conversion and the Analysis of Combinatorial Diversity in Immunoglobulin Heavy Chains in Four Cattle Breeds. *PLoS ONE*. 2016; 11(11):e0164567. Epub 2016/11/10. <https://doi.org/10.1371/journal.pone.0164567> PMID: 27828971; PubMed Central PMCID: PMC5102495.
36. Contreras PJ, Urrutia H, Sossa K, Nocker A. Effect of PCR amplicon length on suppressing signals from membrane-compromised cells by propidium monoazide treatment. *J Microbiol Methods*. 2011; 87(1):89–95. Epub 2011/08/09. <https://doi.org/10.1016/j.mimet.2011.07.016> PMID: 21821068.
37. Shagin DA, Lukyanov KA, Vagner LL, Matz MV. Regulation of average length of complex PCR product. *Nucleic Acids Res*. 1999; 27(18):e23. Epub 1999/09/03. <https://doi.org/10.1093/nar/27.18.e23> PMID: 10471753; PubMed Central PMCID: PMC148615.
38. Haakenson JK, Huang R, Smider VV. Diversity in the Cow Ultralong CDR H3 Antibody Repertoire. *Front Immunol*. 2018; 9:1262. Epub 2018/06/20. <https://doi.org/10.3389/fimmu.2018.01262> PMID: 29915599; PubMed Central PMCID: PMC5994613.
39. Peters B, Trout BL. Asparagine deamidation: pH-dependent mechanism from density functional theory. *Biochemistry*. 2006; 45(16):5384–92. Epub 2006/04/19. <https://doi.org/10.1021/bi052438n> PMID: 16618128.
40. Chelius D, Jing K, Lueras A, Rehder DS, Dillon TM, Vizel A, et al. Formation of pyroglutamic acid from N-terminal glutamic acid in immunoglobulin gamma antibodies. *Anal Chem*. 2006; 78(7):2370–6. Epub 2006/04/04. <https://doi.org/10.1021/ac051827k> PMID: 16579622.
41. Colgrave ML, Craik DJ. Thermal, chemical, and enzymatic stability of the cyclotide kalata B1: the importance of the cyclic cystine knot. *Biochemistry*. 2004; 43(20):5965–75. Epub 2004/05/19. <https://doi.org/10.1021/bi049711q> PMID: 15147180.
42. Tugyi R, Mezo G, Fellinger E, Andreu D, Hudecz F. The effect of cyclization on the enzymatic degradation of herpes simplex virus glycoprotein D derived epitope peptide. *J Pept Sci*. 2005; 11(10):642–9. Epub 2005/05/03. <https://doi.org/10.1002/psc.669> PMID: 15864815.
43. Cascales L, Henriques ST, Kerr MC, Huang YH, Sweet MJ, Daly NL, et al. Identification and characterization of a new family of cell-penetrating peptides: cyclic cell-penetrating peptides. *J Biol Chem*. 2011; 286(42):36932–43. Epub 2011/08/30. <https://doi.org/10.1074/jbc.M111.264424> PMID: 21873420; PubMed Central PMCID: PMC3196123.
44. Craik DJ. Chemistry. Seamless proteins tie up their loose ends. *Science*. 2006; 311(5767):1563–4. Epub 2006/03/18. <https://doi.org/10.1126/science.1125248> PMID: 16543448.
45. Naqid IA, Owen JP, Maddison BC, Spiliotopoulos A, Emes RD, Warry A, et al. Mapping polyclonal antibody responses to bacterial infection using next generation phage display. *Sci Rep*. 2016; 6:24232. Epub 2016/04/14. <https://doi.org/10.1038/srep24232> PMID: 27072017; PubMed Central PMCID: PMC4829855.
46. Spiliotopoulos A, Blokpoel Ferreras L, Densham RM, Caulton SG, Maddison BC, Morris JR, et al. Discovery of peptide ligands targeting a specific ubiquitin-like domain-binding site in the deubiquitinase USP11. *J Biol Chem*. 2019; 294(2):424–36. Epub 2018/10/31. <https://doi.org/10.1074/jbc.RA118.004469> PMID: 30373771; PubMed Central PMCID: PMC6333900.
47. Macpherson A, Liu X, Dedi N, Kennedy J, Carrington B, Durrant O, et al. The rational design of affinity-attenuated OmCI for the purification of complement C5. *J Biol Chem*. 2018; 293(36):14112–21. Epub 2018/07/22. <https://doi.org/10.1074/jbc.RA118.004043> PMID: 30030376; PubMed Central PMCID: PMC6130949.
48. Cheng Y, Prusoff WH. Relationship between the inhibition constant (K1) and the concentration of inhibitor which causes 50 per cent inhibition (I50) of an enzymatic reaction. *Biochem Pharmacol*. 1973; 22(23):3099–108. Epub 1973/12/01. [https://doi.org/10.1016/0006-2952\(73\)90196-2](https://doi.org/10.1016/0006-2952(73)90196-2) PMID: 4202581.