A Rational Engineering Strategy for Designing Protein A-Binding Camelid Single-Domain Antibodies

Kevin A. Henry[1], Traian Sulea[2], Henk van Faassen[1], Greg Hussack[1], Enrico O. Purisima[2], C. Roger MacKenzie[1,3], Mehdi Arbabi-Ghahroudi[1,3,4]*

1 Human Health Therapeutics Portfolio, National Research Council Canada, 100 Sussex Drive, Ottawa, Ontario, Canada, K1A 0R6, 2 Human Health Therapeutics Portfolio, National Research Council Canada, 6100 Royalmount Avenue, Montreal, Quebec, Canada, H4P 2R2, 3 School of Environmental Sciences, University of Guelph, 50 Stone Road East, Guelph, Ontario, Canada, N1G 2W1, 4 Department of Biology, Carleton University, 1125 Colonel By Drive, Ottawa, Ontario, Canada, K1S 5B6

* mehdi.arbabi@nrc-cnrc.gc.ca

Abstract

Staphylococcal protein A (**SpA**) and streptococcal protein G (**SpG**) affinity chromatography are the gold standards for purifying monoclonal antibodies (**mAbs**) in therapeutic applications. However, camelid V_HH single-domain Abs (**sdAbs** or **V_HHs**) are not bound by SpG and only sporadically bound by SpA. Currently, V_HHs require affinity tag-based purification, which limits their therapeutic potential and adds considerable complexity and cost to their production. Here we describe a simple and rapid mutagenesis-based approach designed to confer SpA binding upon *a priori* non-SpA-binding V_HHs. We show that SpA binding of V_HHs is determined primarily by the same set of residues as in human mAbs, albeit with an unexpected degree of tolerance to substitutions at certain core and non-core positions and some limited dependence on at least one residue outside the SpA interface, and that SpA binding could be successfully introduced into five V_HHs against three different targets with no adverse effects on expression yield or antigen binding. Next-generation sequencing of llama, alpaca and dromedary V_HH repertoires suggested that species differences in SpA binding may result from frequency variation in specific deleterious polymorphisms, especially Ile57. Thus, the SpA binding phenotype of camelid V_HHs can be easily modulated to take advantage of tag-less purification techniques, although the frequency with which this is required may depend on the source species.

Introduction

Therapeutic antibodies (**Abs**) represent the fastest-growing class of biologic drugs, with expanding applications in cancer, chronic diseases and autoimmunity (reviewed in [1–3]). Currently licensed biologics are most commonly fully human or humanized monoclonal Abs (**mAbs**), with antigen-binding fragments such as Fab, F(ab')$_2$ and scFv making up a

smaller proportion of the market [4]. Next-generation Ab therapeutics will likely exploit the improved functional properties of molecularly engineered Abs, including bispecific Abs, Ab-drug conjugates and Fc-variant Abs [1,4,5]. V$_H$H single-domain antibodies (**V$_H$Hs**), the variable domains of heavy-chain-only Abs produced naturally by camelid ungulates, may be useful in the design of next-generation biologics as a result of their small size, stability and modularity [6].

Full-length mAbs of all IgG subtypes can be easily purified by affinity chromatography, taking advantage of the high-affinity interactions between the IgG Fc and either staphyloccal protein A (**SpA**) or streptococcal protein G (**SpG**) [7]. Purification of Ab fragments bearing light chains can be accomplished using protein L resins but purification of V$_H$Hs can be more complex, often requiring recombinant fusion to an affinity tag sequence [8]. The presence of affinity tags may impact protein folding, stability, solubility and aggregation state either directly or indirectly *via* the elution process [9–14]; although such tags can be removed proteolytically after purification, this process adds time and cost to the production pipeline. The gold standard for tag-less purification of Ab fragments and V$_H$Hs is affinity chromatography using staphyloccal protein A (**SpA**; reviewed in [15]), which contains binding sites for both the IgG Fc region and variable regions of Ig heavy chains, with the specificity of the latter being restricted to human IGHV3 gene family products [16–20] and their homologues in experimental animal species [21,22]. The structural basis for binding of IGHV3 Abs by SpA was initially ascribed broadly to residues in FR3, FR1 and possibly CDR2 [23,24] and later defined more precisely using a co-crystal structure of a human IGHV3-encoded IgM in complex with SpA [25]. Binding appears to depend on thirteen Ig variable region residues at the direct interface with SpA, which are almost universally conserved in germline IGHV3 genes, but can be lost as a result of somatic hypermutation [17–19]. The importance of these residues has been inferred from structural studies and comparison with sequences of non-SpA-binding Abs, but never formally tested by mutagenesis.

Both Old World and New World camelids produce both conventional and heavy-chain-only Abs, which are rearranged during B-cell development from separate sets of germline variable genes (V$_H$ and V$_H$H genes, respectively [26]). Most camelid V$_H$H genes are homologous to human IGHV3-family genes and fall into 4–7 subfamilies [26,27], although recent work suggests that human IGHV4 homologues may also exist that can be rearranged as both conventional and heavy-chain only Abs [28]. Despite their homology to human IGHV3-family genes, it has been reported that the majority of camelid V$_H$Hs are not bound by SpA [29,30], although this may depend to some degree on species origin. At the amino acid level, the major sequence differences between camelid V$_H$Hs and human IGHV3-family variable domains are (*i*) the presence of hallmark V$_H$H solubilizing residues located in FR2 at the interface with the absent V$_L$ domain [27], (*ii*) an elevated frequency of intradomain disulfide bridges between CDR3 and other regions, which may constrain V$_H$H structure [31], (*iii*) the presence of various other germline polymorphisms, depending on the individual V$_H$H gene and (*iv*) high somatic mutation rates [26]. However, it is unclear whether and how any of these sequence characteristics should impact SpA binding. Thus, the purpose of this work was: (*i*) to identify V$_H$H residues required for SpA binding and evaluate their overlap with those defined for human IGHV3 Abs [25], and (*ii*) to develop a strategy to confer SpA binding upon *a priori* non-binding V$_H$Hs without compromising antigen binding. Our data confirm that SpA binding by V$_H$Hs is determined primarily by the same set of contact positions as in human IGHV3 mAbs, but also provide a comprehensive description of IGHV domain polymorphisms that are permissive and non-permissive for SpA binding. These data can be used to engineer SpA-binding V$_H$Hs using limited site-specific mutagenesis, with no adverse effects on expression yield or antigen binding.

Materials and Methods

Phage-displayed libraries and isolation of V$_H$Hs

The V$_H$Hs described in this report are directed against a variety of antigens and were initially isolated from phage-displayed V$_H$H libraries for reasons other than the study of SpA binding, as described previously [32–38].

V$_H$H mutagenesis

The amino acid sequences of five non-SpA-binding camelid V$_H$Hs (ICAM11-4, ICAM34-1 [39], IGF1R-4, IGF1R-5 [40] and AFAI [35]) were altered at specific FR positions, either individually or in combination, either for humanization purposes or to confer SpA binding. Constructs encoding the engineered V$_H$Hs were synthesized commercially in the pSJF2H expression vector ([31]; GenScript, Piscataway, NJ or Life Technologies, Carlsbad, CA).

Soluble V$_H$H monomer and pentamer expression and purification

Wild-type or engineered V$_H$H genes bearing *BbsI/BamHI* or *BbsI/ApaI* restriction sites were cloned into pSJF2H or pVT2 expression vectors (for monomeric and pentameric expression, respectively) as described [32,33,35]. 6×His- and c-Myc-tagged V$_H$H monomers and pentamers were expressed in *E. coli* TG1, extracted from the periplasm by osmotic shock and purified by immobilized metal affinity chromatography (**IMAC**) or protein A chromatography using using HisTrap HP or HiTrap Protein A HP columns, respectively (GE Healthcare, Piscataway, NJ; [32,33]). The integrity and aggregation status of soluble V$_H$H monomers and pentamers were assessed by SDS-PAGE, Western blotting and size exclusion chromatography [32,33].

Surface plasmon resonance (SPR)

For screening of V$_H$H monomers and pentamers for binding to immobilized SpA at a single concentration, analyte proteins were used directly after purification by immobilized metal affinity chromatography. For determination of binding affinities to immobilized antigen or SpA, V$_H$H monomers were purified by size exclusion chromatography using Superdex™ 75 or 200 10/300 GL columns (GE Healthcare) on an ÄKTA FPLC protein purification system (GE Healthcare), and the monomer peaks collected in HBS-EP buffer (10 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (**HEPES**), pH 7.4, 150 mM NaCl, 3mM ethylenediaminetetra-acetic acid (**EDTA**), 0.005% (v/v) surfactant P20; [32,33]). Briefly, either SpA (Thermo Fisher, Waltham, MA), human ICAM-1 ectodomain (R&D Systems, Minneapolis, MN) or human IGF1R ectodomain (R&D Systems) were immobilized on CM5 sensor chips using an amine coupling kit (GE Healthcare) in 10 mM acetate buffer, pH 4.5, with surface densities of 609, 1348 and 476 resonance units, respectively. V$_H$H monomers and pentamers were injected at 25°C in HBS-EP buffer at a flow rate of 20 μL/min on a Biacore 3000 instrument (GE Healthcare) at different concentration ranges, depending on the interaction (binding to SpA: 50 nM —25 μM; binding to ICAM-1: 0.5 nM– 200 nM; binding to IGF1R: 0.1 nM– 10 nM). All surfaces were regenerated using 10 mM glycine, pH 2.0. Data were analyzed using BIAevaluation 4.1 software (GE Healthcare) and for affinity determinations, fitted to a 1:1 binding model.

In silico scanning mutagenesis of the SpA:V$_H$H interaction

The co-crystal structure of a human IGHV3-encoded Fab in complex with domain D of SpA (PDB ID: 1DEE) was used as a starting point for virtual scanning mutagenesis. Only one copy of the IGHV3 domain (chain D) and the bound SpA fragment (chain G, corresponding to SpA

domain D positions Asp8-Lys58 as numbered by Graille *et al.* [25]) were retained. The IGHV3 domain was first "camelized" by introducing four mutations in FR2: Val37Phe, Gly44Glu, Leu45Arg and Trp47Ala. Hydrogen atoms were added to the resulting V$_H$H:SpA complex and adjusted to maximize H-bonding interactions. Structural refinement of the complex was then carried out by energy-minimization using the AMBER force-field [41,42] with a distance-dependent dielectric and infinite cutoff for non-bonded interactions. Non-hydrogen atoms were restrained at their crystallographic positions with harmonic force constants of 20 and 5 kcal/(mol·A^2) for the backbone and side-chain atoms, respectively. The resulting structure was then used for single-point scanning mutagenesis simulations at the following positions of the IGHV3 domain: 15, 17, 19, 57, 59, 64, 65, 66, 68, 70, 75, 81, 82a, and 82b (all positions using Kabat numbering). We used three protocols (SIE-SCWRL [43–45], FoldX [46,47] and Rosetta [48,49]) for modeling the structures and evaluating the energies of single-point substitutions of the other 17 naturally-occurring amino acids (Cys and Pro excluded) at each of these 14 positions relative to the wild-type sequence. A consensus approach over specific versions of these three protocols was applied for building and scoring IGHV3 mutants. Further technical and implementation details of this approach and its component methods can be found in Sulea *et al.* [50].

Next-generation DNA sequencing

Phagemid replicative form DNA from naïve and immune phage-displayed V$_H$H libraries was isolated from *E. coli* TG1 cells using QIAprep spin miniprep kits (QIAGEN, Valencia, CA). Next-generation sequencing libraries were generated by two-step PCR amplification of V$_H$H genes and purified as previously described [51,52]. The final amplicons were pooled and purified from 1% (w/v) agarose gels using a QIAquick® gel extraction kit (QIAGEN), desalted using Agencourt AMPure XP beads (Beckman-Coulter, Pasadena, CA), then sequenced on a MiSeq Sequencing System (Illumina, San Diego, CA) using a 500-cycle MiSeq Reagent Kit V2 and a 5% PhiX genomic DNA spike. From each sample, 0.1–2.1 million reads were generated, of which 0.4×10^5–6.0×10^5 were used for analysis after assembly using FLASH (default parameters; [53]) and quality filtering using the FAST-X toolkit with a stringency of Q30 over ≥95% of each read [54].

Results

Definition of V$_H$H residues involved in SpA binding

Previous structural work [25] suggested the existence of a set of seven core SpA binding residues in human IGHV3 heavy-chain variable regions (positions 19, 65, 66, 68, 70, 81 and 82a; all positions using Kabat numbering), with a lesser contribution of six additional contact residues (positions 15, 17, 57, 59, 64 and 82b). Parenthetically, these tend to be mostly conserved in SpA-binding human and murine Abs and altered *via* germline variation or somatic hypermutation in non-SpA-binding Abs [17–19], although neither the importance of each residue nor the spectrum of tolerated substitutions has been rigorously tested by mutagenesis studies. Overlay of the three-dimensional structures of a human IGHV3-encoded IgM heavy-chain variable domain [25] and a llama V$_H$H directed against *Clostridium difficile* toxin A [55] showed strong overall conservation of these immunoglobulin folds (Fig 1).

To investigate whether the conserved SpA-binding residues in human IGHV3 Abs were also important for V$_H$H binding by SpA, we first determined the sequences of 55 V$_H$Hs obtained in our lab along with an IGHV3-encoded human autonomous V$_H$ domain, HVH430 [56], and assessed their binding to immobilized SpA at a single concentration (250 nM) by SPR (Fig 2; S1 Table). Many, but not all, camelid V$_H$Hs shared the human IGHV3 consensus
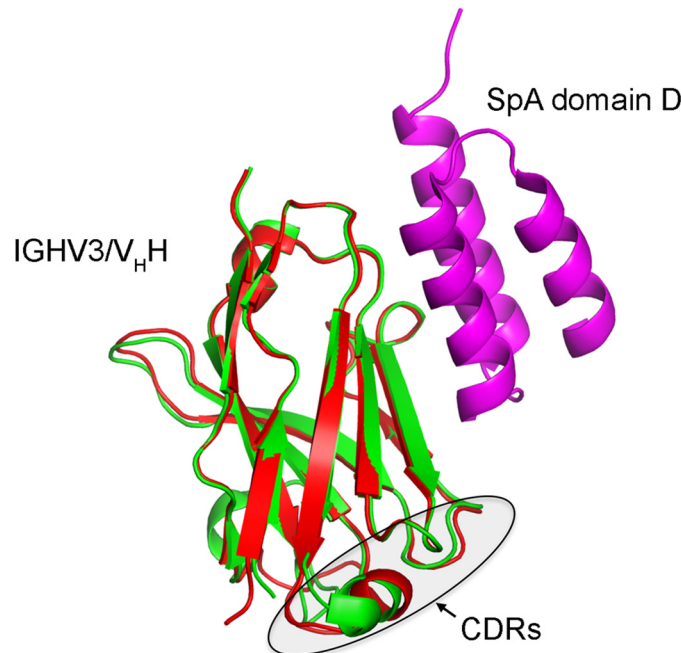
Fig 1. Overlay of three-dimensional structures of a llama V$_H$H (green, PDB ID: 4NC0) and a human IGHV3 heavy chain variable domain (red, PDB ID: 1DEE) in complex with SpA domain D. Antigen-binding CDRs are indicated.

doi:10.1371/journal.pone.0163113.g001

sequence at all thirteen SpA contact residues, suggesting that they may interact in a similar way with SpA. Complete or near-complete sequence conservation at some V$_H$H positions (Gly15, Ser17, Arg66, Ser70 and Gln81) among both SpA-binding V$_H$Hs and non-SpA-binding V$_H$Hs prevented the assessment of their importance. In agreement with the conclusions of structural studies [25] and a previous mutagenesis study [57], the salt bridge formed between the Asp residue at position 36 of SpA domain D and the conserved Arg residue at V$_H$H FR1 position 19 was indispensable, as its replacement with Lys, Ser or Thr abrogated SpA binding. Similarly, replacement of core Gly65 with negatively charged Asp, or of core Asn82a with Asp or Ser, had a destructive effect on SpA binding; the latter result is consistent with a previous mutagenesis study [57] which found that substitution of Asn82a with Ala abrogated SpA binding. Surprisingly, one V$_H$H bearing a substitution of core Thr68 with Ala (V$_H$H36) showed residual SpA binding. However, since V$_H$Hs encoding Ala68 in combination with other substitutions did not bind SpA, and the more conservative replacement of Thr68 with Ser reduced or ablated SpA binding, we infer that that the limited polymorphisms tolerated at this position confer a partial destabilizing effect on SpA binding.

In agreement with previous work [18,23], we found that a variety of residues (Thr, Arg and Lys) were tolerated at the non-core V$_H$H CDR2 position 57, although the full spectrum of tolerated residues could not be conclusively identified due to the co-occurrence of substitutions at other positions (Fig 2; S1 Table). Conversely, the presence of Ala57 clearly ablated SpA binding, and surprisingly, two V$_H$Hs bearing Ile57, ICAM11-4 and V$_H$H55, did not bind SpA; Ile57 is present in some germline human IGHV3 genes and was previously judged as permissive for SpA binding by some groups [23,25] but not others [18]. There was some indication that V$_H$H Tyr59, which forms a H-bond with SpA Asp37 that was not considered a core interaction by Graille *et al.* [25], might be essential, as its substitution with His or Val (in combination with other substitutions in V$_H$H26 and IGF1R-4) ablated SpA binding. Both reversal of charge at
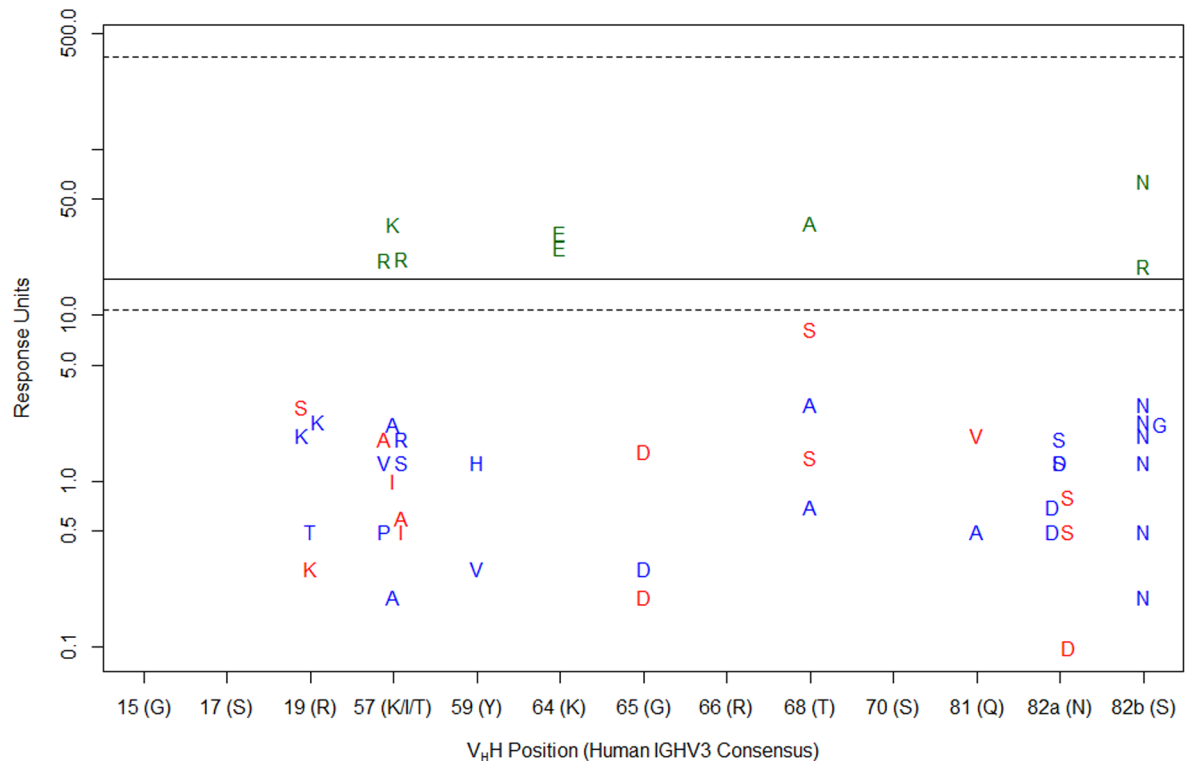
Fig 2. Identification of permissive and non-permissive residues for SpA binding at V$_H$H SpA contact positions. 55 V$_H$Hs of known sequence (FR sequences listed in S1 Table) were assayed for binding to immobilized SpA at a single concentration (250 nM) by SPR and the number of response units bound at the end of the injection was recorded. The solid line represents binding of HVH430 (17 RUs), an IGHV3-encoded human autonomous domain. For V$_H$Hs bearing the human IGHV3 consensus residue (shown on the X-axis) at all 13 SpA contact positions, no data are plotted on the graph; instead, dotted lines are shown representing the 95% confidence interval (CI) for mean SpA binding of "wild-type" V$_H$Hs bearing this consensus sequence. For V$_H$Hs bearing single amino acid substitutions at any one of the 13 SpA contact positions, the relevant substitution is plotted on the graph in green (substitution tolerated) if SpA binding fell within the 95% CI for wild-type V$_H$Hs (10–360 RUs), and red (substitution not tolerated) if it fell below. For V$_H$Hs bearing multiple amino acid substitutions at SpA contact sites, substitutions are plotted on the graph in blue.

doi:10.1371/journal.pone.0163113.g002

the non-core V$_H$H position 64 (Lys→Glu) and substitution of non-core Ser82b with Asn or Arg appeared to be well tolerated, with no consistent detrimental effect on SpA binding.

Thus, the overall picture emerging from these data was that V$_H$H interaction with SpA depended primarily on the same set of residues as human IGHV3 Abs, albeit with: (*i*) an unanticipated but minor degree of tolerance for variation at core position Thr68 and a broader tolerance at non-core positions Thr57, Lys64 and Ser82b; (*ii*) a potentially critical role for non-core V$_H$H position Tyr59; and (*iii*) a destructive effect of Ile at V$_H$H position 57, in contrast to the predictions of some previous studies. We attempted to corroborate these data using a set of V$_H$H pentamers [35] but found that the multivalent nature of the pentamer:SpA interaction made it difficult to discriminate very weakly-binding from non-binding V$_H$H pentamers (S1 Fig; S2 Table). We found no evidence to suggest that FR sequence polymorphisms outside the previously defined SpA interface [25] played any role in SpA binding, with one notable exception (V$_H$H39): this V$_H$H did not bind SpA despite bearing the human IGHV3 consensus sequence at all 13 SpA contact positions, but encodes a single-residue deletion at FR3 position 76. This provided the first preliminary evidence that FR3 positions 71–80 may influence structuring of nearby SpA contact residues and play an indirect role in modulating SpA binding.

Table 1. FR sequences of five non-SpA-binding V$_H$Hs at SpA contact positions.

V$_H$H	V$_H$H Position												
	15 (G)	17 (S)	19 (R)	57 (K/I/T)	59 (Y)	64 (K)	65 (G)	66 (R)	68 (T)	70 (S)	81 (Q)	82a (N)	82b (S)
ICAM11-4	G	S	R	I*	Y	K	G	R	T	S	Q	N	S
ICAM34-1	G	S	R	A*	Y	K	G	R	T	S	Q	N	G*
IGF1R-4	G	S	R	T	V*	K	D*	R	T	S	Q	N	S
IGF1R-5	G	S	R	A*	Y	K	G	R	T	S	Q	N	N*
AFAI	G	S	R	T	Y	E*	G	R	T	S	Q	N	D*

*Residues were reverted to the human IGHV3 germline consensus during humanization

SpA binding is reliably conferred to V$_H$Hs by humanization

The simplest possible strategy for rescuing SpA binding in non-SpA-binding V$_H$Hs would be to revert any discrepancies at the 13 contact positions defined by Graille *et al.* [25] to the human IGHV3 consensus sequence. This is always accomplished, by definition, through the process of humanization, in which a llama V$_H$H's FRs are replaced almost entirely with those of the most closely homologous human IGHV3 gene. To confirm that this strategy would be successful, we took four *a priori* non-SpA-binding V$_H$Hs (ICAM11-4 and ICAM34-1, directed against human ICAM-1; IGF1R-4 and IGF1R-5, directed against human IGF1R) and humanized them, thus reverting their FR sequences to human IGHV3 germline both at SpA contact positions ([Table 1](#)) and elsewhere ([S3 Table](#)). A fifth example, AFAI (a V$_H$H directed against CEACAM6), was not humanized, but instead had two reversions incorporated *via* site-directed mutagenesis (Glu64Lys, Asp82bSer).

SpA binding by these five camelid V$_H$Hs was initially extremely weak or undetectable, as measured using single-concentration injections of 250 nM V$_H$H over immobilized SpA in SPR. However, in each case, V$_H$H humanization (or in the case of AFAI, reversion of both Glu64 to Lys and Asp82b to Ser) conferred SpA binding ([S2 Fig](#)) that was sufficient to enable their purification by SpA affinity chromatography ([S3 Fig](#)). In three of four cases, expression yields of the humanized V$_H$Hs after IMAC purification were reduced by 50% or more compared to their wild-type llama counterparts ([Table 2](#), left *vs.* middle columns), and in one case (ICAM11-4), the humanized V$_H$H had a ~20-fold loss of affinity for its cognate antigen ([Table 3](#), left *vs.* middle columns). Thus, perhaps unsurprisingly, SpA binding could be successfully introduced into five *a priori* non-SpA-binding V$_H$Hs by humanization, although this came at the cost of reduced expression yield, and in one case, impaired antigen recognition.

Table 2. Expression yields (mg/L) in *E. coli* of wild-type llama V$_H$Hs *vs.* their humanized or SpA-engineered counterparts.

V$_H$H	Wild-type	Humanized	SpA-engineered
ICAM11-4	20.0	4.5	17.0
ICAM34-1	21.0	7.0	25.0
IGF1R-4	9.0	4.5	6.0
IGF1R-5	14.0	11.0	12.0
AFAI	6.0	n.d.	4.0

n.d., not determined

Table 3. Affinities for cognate antigen (pH 7.4, 25°C) of wild-type llama V$_H$Hs _vs._ their humanized or SpA-engineered counterparts.

V$_H$H	Wild-type	Humanized	SpA-engineered
ICAM11-4	0.9 nM	17 nM	1.3 nM
ICAM34-1	32 nM	26 nM	n.d.[a]
IGF1R-4	0.3 nM	0.4 nM	n.d.[a]
IGF1R-5	0.8 nM	0.9 nM	n.d.[a]
AFAI	+[b]	n.d.	+[b]

n.d., not determined

[a]Affinities were not determined on the assumption that since no loss of affinity was observed in the humanized V$_H$H, SpA-engineered V$_H$Hs (bearing fewer changes from wild-type) would also retain full affinity for antigen.

[b]Affinities were not determined using monomeric V$_H$H; rather, binding of wild-type and SpA-engineered V$_H$H pentamers are shown in S4 Fig.

doi:10.1371/journal.pone.0163113.t003

SpA binding can be conferred to V$_H$Hs by limited site-directed mutagenesis without negatively impacting expression yield or antigen binding

To determine whether SpA binding could be restored in non-SpA binding V$_H$Hs without incurring the negative consequences of humanization on expression yield and antigen binding, and to better understand the relative impacts of individual substitutions on SpA binding, we took the same five non-SpA binding V$_H$Hs as above and reverted any discrepancies at the 13 SpA contact positions (Table 1) to the human IGHV3 consensus by site-directed mutagenesis. In one case (ICAM34-1), we also reverted an unusual Pro at position 75 to the human IGHV3 consensus Lys, based on the hypothesis that this Pro residue might affect structuring of surrounding SpA contact residues. Substitutions were incorporated either singly or in all possible combinations, and the SpA-binding affinities of the resulting engineered V$_H$Hs were then determined by SPR.

In the case of ICAM11-4, a single substitution (Ile57Thr) was sufficient to restore 9.0 μM affinity for SpA (Table 4; in this table and hereafter, SpA-engineered V$_H$Hs (underlined) are defined as the variants bearing the minimal substitutions from wild-type sequences required to restore SpA binding), near to the wild-type affinity range of 1–5 μM described for human IGHV3-encoded V$_H$ domains [56]. In the case of ICAM34-1, two reversions (Ala57Thr and Pro75Lys) were necessary to recover any degree of SpA binding (9.6 μM), which was improved slightly by a third reversion (Gly82bSer, 4.5 μM). Likewise, two substitutions (Val59Tyr and Asp65Gly) were required to restore SpA binding (1.5 μM) to IGF1R-4, confirming the critical nature of the consensus residue at both of these positions for interaction with SpA. In the case of IGF1R-5, a single substitution (Ala57Thr) conferred 3.7 μM affinity for SpA, with no further affinity improvement obtained by reversion of position 82b from Asn to Ser. Finally, reversion of Asp82b to Ser in AFAI restored weak SpA binding (11.0 μM), which was further improved with the additional reversion of Glu64 to Lys (0.6 μM). This suggested a cumulative defect of both substitutions in the wild-type V$_H$H, with Asp82b conferring more significant binding impairment than Glu64.

In summary, at least in the five examples above, SpA binding could be conferred to non-SpA binding V$_H$Hs using limited numbers (1–3) of amino acid substitutions. The resulting engineered V$_H$Hs bound SpA with affinities similar to those of human IGHV3-encoded Fabs, could be purified by SpA affinity chromatography (S3 Fig), and had expression yields after IMAC purification that were indistinguishable from their non-SpA binding counterparts (Table 2, left _vs._ rightmost column). In the case of ICAM11-4, the engineered SpA-binding V$_H$H retained full affinity for ICAM-1, unlike the humanized V$_H$H (Table 3, middle _vs._ right

Table 4. Affinities for SpA (pH 7.4, 25°C) of wild-type llama V$_H$Hs *vs.* their SpA-engineered counterparts.

V$_H$H	Wild-type	SpA-engineered (K$_D$)
ICAM11-4	n.b.	I57T (9.0 μM)*
ICAM34-1	n.b.	A57T (n.b.)
		P75K (n.b.)
		G82bS (n.b.)
		A57T, P75K (9.6 μM)
		A57T, G82bS (n.b.)
		P75K, G82bS (n.b.)
		A57T, P75K, G82bS (4.5 μM)*
IGF1R-4	n.b.	V59Y (n.b.)
		D65G (n.b.)
		V59Y, D65G (1.5 μM)*
IGF1R-5	n.b.	A57T (3.7 μM)*
		N82bS (n.b.)
		A57T, N82bS (3.5 μM)
AFAI	n.b.	E64K (n.b.)
		D82bS (11.0 μM)
		E64K, D82bS (0.6 μM)*

n.b., no binding

*SpA-engineered V$_H$Hs (defined as the variants bearing the minimal substitutions from wild-type sequences required to restore SpA binding) are underlined

columns). The data provided from these mutagenesis studies also expanded our understanding of the spectrum of permissive and non-permissive V$_H$H residues for SpA binding, which are listed in Table 5, and clearly supported the hypothesis that residues outside the SpA contact interface such as Pro75 can indirectly modulate SpA binding.

3.4. *In silico* modeling of the SpA:V$_H$H interaction by virtual mutagenesis

To support and extend the experimental results above, we carried out a computational assessment of the effects of virtual mutagenesis of V$_H$H SpA contact residues on SpA binding (Fig 3 and Table 5). We used a consensus approach for mutant building and scoring that has been found to afford improved ranking of Ab:antigen binding affinities relative to various individual methods when applied to over 200 single-point antibody mutants of the SiPMAB database curated from the literature [50]. We used the published SpA:IGHV3 Fab structure [25] as the basis for *in silico* analyses, but prior to virtual mutagenesis excluded the light chain from consideration and "camelized" the human IGHV3 domain by substitution of FR2 Val37Phe, Gly44Glu, Trp47Ala and Leu45Arg. The results of *in silico* mutagenesis provided an explanation for the almost complete conservation of Gly15, core Gly65 and core Arg66 in V$_H$Hs, since substitution of these residues was predicted to drastically destabilize V$_H$H folding; the backbone phi-psi dihedral angles of Gly15 and Gly65 are (92, -18) and (92, -20), respectively, in the 4NC0 crystal structure and are high-energy regions for non-glycine amino acids. Conversely, Ser17 and core Ser70 were predicted to tolerate substitution with a variety of residues despite being almost totally conserved in V$_H$Hs. In the case of core V$_H$H Gly65, experimental evidence contradicted computational predictions of fold destabilization by substitution with Asp, as this residue was observed in three V$_H$H monomers and one V$_H$H pentamer (S1 and S2 Tables). In agreement with experimental data, core V$_H$H Arg19 was critical for interaction with SpA, and

Table 5. Spectrum of residues tolerated at V$_H$H SpA contact positions, as demonstrated by experimental data or as predicted by *in silico* mutagenesis.

V$_H$H Position	Experimentally Validated		Computationally Predicted	
	Tolerated	Not tolerated	Tolerated	Not tolerated
15 (G)	G, D[a]	-	G	All others
17 (S)	S, A[a]	-	Many	-
19 (R)	R	K, S, A[a], Q[a]	R	All others
57 (K/I/T)	K, R, T	A, I	K, R, T	All others
59 (Y)	Y	V	Y	All others
64 (K)	K (E)*	-	Many	-
65 (G)	G	D	G	All others
66 (R)	R	-	R	All others
68 (T)	(A)	S	I, M, V	Many
70 (S)	S	-	Many	I, T, V
75 (K)	A, E, K, Q, R	P	Many	I, T, V
81 (Q)	Q	-	M, R	Many
82a (N)	N	D, S, A[a]	I, L, M, V	Many
82b (S)	S, N (G)	D	Many	Many

Residues in brackets are tolerated but result in some loss of binding

[a]Data from Fridy *et al*. [57]

could not be substituted even conservatively by Lys. Also in line with experimental data, moderate but potentially non-destructive reductions of SpA binding were predicted when V$_H$H core Thr68 was substituted with Ser, Ala and several other residues; similar trends were observed for core Gln81 and Asn82a. Virtual mutagenesis predicted a critical role for non-core V$_H$H positions 57 (Thr, Lys, and Arg) and Tyr59, a negligible role for non-core V$_H$H position Lys64 and variable effects of Ser82b substitution, with minor loss of binding resulting from

	V$_H$H Position	R	K	Q	N	S	T	H	W	Y	F	M	L	I	V	A	G	E	D
Core SpA-binding	19 (R)	0.0	4.8	5.3	5.3	5.6	5.4	5.6	4.4	5.7	5.6	5.7	5.7	6.0	6.1			6.4	6.4
	65 (G)																0.0		
	66 (R)	0.0																	
	68 (T)	4.4	1.5	2.7	1.8	2.6	0.0	1.6	18.4	19.8	18.6	0.9	3.3	0.3	1.0	2.8	3.7	5.9	6.2
	70 (S)	-0.5	0.1	0.8	-1.9	0.0		0.0	0.7	0.8	1.9	-0.1	0.8			0.8	1.8	1.6	1.9
	81 (Q)	1.0	2.5	0.0	1.8	2.7	1.9	6.7	12.2	27.5	24.4	1.3	2.3	3.7	3.5	3.2		2.1	5.1
	82a (N)	3.8	2.5	3.6	0.0	3.4	2.4	8.5	15.5	23.2	20.4	0.5	1.3	0.2	0.7	3.5	4.9	4.1	2.1
Non-core SpA-binding	15 (G)																0.0		
	17 (S)	-1.4	0.4	-0.9	-0.6	0.0	-1.3	-1.3	0.7	-2.3	-1.9	-1.7	-1.0	-2.0	-1.9	0.0	0.6	-0.6	1.6
	57 (K)	0.7	0.0	3.1	3.5	2.2	1.5					2.5	2.6	3.5	3.9	3.5		6.0	5.2
	59 (Y)									0.0	2.8	3.8	3.9						
	64 (K)	0.5	0.0	0.3	0.2	0.9	0.5	-0.2	-0.3	0.0	-0.3	0.3	0.4	0.5	0.7	1.1	0.9	0.4	0.3
	82b (S)	0.0	-0.2	3.9	1.3	0.0	2.0	-0.2	-1.6	0.1	-0.2	0.1	10.2	2.5	3.2	2.3	3.3	3.1	2.3
	75 (K)	0.1	0.0	0.1	0.0	0.0		0.0	0.0	0.0	0.0	0.1	0.1			0.0	0.0	0.1	0.1

Fig 3. Effect of *in silico* scanning mutagenesis of V$_H$H SpA contact residues on V$_H$H folding and SpA binding. Consensus Z-scores for SpA binding were calculated over the SIE-SCWRL, FoldX and Rosetta protocols for mutations to 18 amino acids (shown at the top) at 14 V$_H$H positions (shown on the left). Predicted change to ΔG (free energy of binding, kCal/mol) for SpA binding to V$_H$H are shown numerically and colored in shades of red or blue for single-point mutations predicted to strengthen or weaken SpA binding, respectively, relative to the wild-type amino acid corresponding to the human IGHV3 sequence. Mutations predicted by FoldX to destabilize V$_H$H folding by more than 10-fold relative to parental IGHV3 sequence are highlighted in black shading. See Materials and Methods and Sulea *et al*. [50] for further details.

Asn substitution and more pronounced defects resulting from substitution with Gly or Asp. Interestingly, V$_H$H position 75 was predicted to tolerate substitution with most residues except those with beta-branched side chains, which probably disrupt the H-bonded turn of the FR3 loop at this location. We speculate that the effect of Pro substitution at this position has a similar effect and the incurred local misfolding may be propagated to adjacent SpA contact residues.

Overall, the experimental and computational data were congruent and together supported the classification of SpA contact residues of V$_H$Hs into three categories: (*i*) critical for folding (Gly15, Gly65, Arg66) and/or binding (Arg19, Tyr59, Gly65); (*ii*) partially permissive to specific substitutions (Thr68, Gln81, Asn82a, Thr/Lys/Arg57, Ser82b) and generally tolerant to many substitutions (Ser17, Ser70, Lys64). For V$_H$H residues falling into the second and third categories, however, the spectrum of amino acids predicted computationally to be tolerated at each position could not be validated experimentally using the available data.

Next-generation DNA sequencing of camelid V$_H$H repertoires suggests that species differences in SpA binding may arise from differential frequency of Ile57 polymorphism

SpA binding is observed rarely in V$_H$Hs of dromedary origin and more commonly in those of llama origin, although a detailed comparison of the frequency of SpA binding by species has yet to be published [30]. To identify potential explanations at the level of protein sequences that might account for this observation, we used next-generation sequencing technology to interrogate four V$_H$H repertoires to variable depths (0.4–6.0×10^5 reads; S4 Table). The source of the V$_H$H repertoires were either lymphocytes derived from three individual llamas or a single pooled sample of alpaca, camel and llama lymphocytes [34].

As shown in Fig 4A, there was no general defect in SpA binding evident from the sequences of V$_H$H repertoires of camels and alpacas. The proportion of V$_H$H sequences bearing the human IGHV3 consensus sequence at SpA contact positions was broadly similar in all four repertoires, except at position 57, where Ile was much more frequently present in dromedary and/or alpaca V$_H$Hs (llama repertoire: ~5% vs. mixed llama, alpaca and camel repertoires: ~22%). A major difference was observed in the frequency of putative CDR1-CDR3 intradomain disulfide bonds, as indicated by the simultaneous presence of Cys residues at any position within both these regions, which were almost entirely absent in the repertoires of llamas but very frequent in the mixed-species repertoire (Fig 4B). To confirm that CDR1-CDR3 disulfide bridging had no effect on SpA binding, we ablated this disulfide from a non-SpA-binding dromedary V$_H$H as well as introduced it into the engineered SpA-binding variant of the same V$_H$H (Ile57Thr; S5 Table). Neither introduction nor ablation of the CDR1-CDR3 disulfide bridge had any impact on SpA binding, at least in the case of the single dromedary V$_H$H tested here.

Discussion

SpA affinity chromatography has become the purification method of choice for most antibody manufacturers over the last 15 years, especially in the therapeutic Ab pipeline [8]. For many applications, V$_H$Hs are fused to mouse or human F$_c$ regions for mammalian cell production and easily purified using conventional methods [58]. However, in some circumstances (*e.g.*, *in vivo* imaging [59]; co-crystallization [60]; tumor targeting requiring tissue penetration [61]), the small size of the V$_H$H molecule is essential, and the addition of affinity tags for purification may detract from the stability and homogeneity of the final product. The same is true of V$_H$H pentamers, and the starting point for this work was the observation that His- and/or Myc-
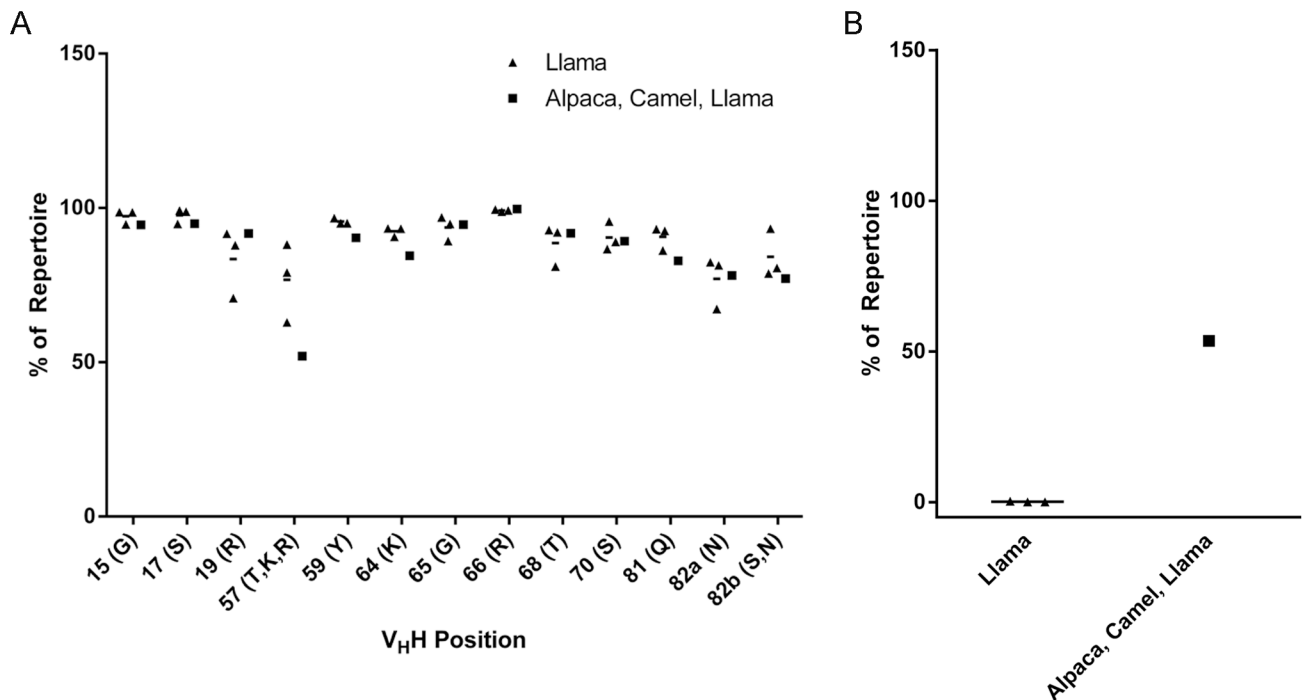
Fig 4. Next-generation DNA sequencing of llama and mixed-species (alpaca, camel, llama) V$_H$H repertoires and analysis of SpA contact residues. (A) Frequency of SpA-permissive residues at contact positions in llama *vs.* mixed-species V$_H$H repertoires. Triangles represent individual llama repertoires, squares represent a single pooled mixed-species repertoire, and horizontal lines represent the mean. Residues permissive for SpA binding are indicated on the X-axis. **(B)** Frequency of putative CDR1-CDR3 intradomain disulfide bond in llama *vs.* mixed-species V$_H$H repertoires. Horizontal lines represent the averages.

doi:10.1371/journal.pone.0163113.g004

tagged pentamers had increased propensity for aggregation and non-specific binding than their untagged counterparts. Thus, tagless strategies for V$_H$H purification are highly desirable.

Here, we present a detailed description of the sequence features of SpA-binding and SpA-non-binding V$_H$Hs, and by extension, the features governing the interaction between SpA and other immunoglobulin V$_H$ domains. Our results have important distinctions from those inferred from the SpA:Fab crystal structure [25]. First, we found that several core positions in the V$_H$H:SpA interface can tolerate a limited degree of polymorphism without ablating SpA binding (shown experimentally: Thr68; predicted computationally: Thr68, Ser70, Gln18, Asn82a). Second, on the basis of both experimental data and computational structural modeling, we found that non-core V$_H$H Thr/Lys/Arg57 and Tyr59 are indispensable for SpA binding, and that Ile57 had a destructive effect on SpA binding. Third, we found that V$_H$H Pro75 exerts a destabilizing effect on SpA binding, potentially by altering the conformation of FR3 sufficiently to displace SpA contact residues some distance away. Given this surprising finding, a potential role for additional V$_H$H polymorphisms outside the SpA interface in determining SpA binding cannot be ruled out. However, the fact that a three amino acid insertion in the FR3 of V$_H$H52 (S1 Table) did not impair SpA binding suggests that epistatic effects involving "action at a distance" may be rare.

In the analyses presented here, we did not attempt to rank the relative effects of individual V$_H$H substitutions on SpA binding, instead opting to categorize them dichotomously as permissive or non-permissive for SpA binding (Table 5). The rationale for this decision was two-fold. First, several substitutions were observed experimentally in only a single V$_H$H, and their effects may depend on the sequence background and presence of other polymorphisms.

Second, the weak overall affinity of the SpA:V$_H$H interaction makes determinations of monovalent binding strength challenging below a certain threshold, and weak residual binding is difficult to rule out experimentally; this was clearly evident in comparisons of SpA binding of V$_H$H monomers and pentamers bearing similar sequences. Nevertheless, the overall consistency between experimental data and computational predictions, as well as with the limited mutagenesis data of Fridy *et al.* [57], provides a strong degree of confidence in the general effects of many of the substitutions described here. We caution, however, that computational predictions of minor or moderate reductions in SpA binding were not always accurate in the degree of their effects, and that the safest course is to revert all V$_H$H contact positions back to the human IGHV3 consensus, even if some of the original polymorphisms might have been tolerated.

Using next-generation DNA sequencing of llama, alpaca and dromedary V$_H$H repertoires, we found that the most likely explanation for non-SpA binding of dromedary V$_H$Hs was the presence of non-permissive residues at SpA contact positions, especially Ile57. Since the comparison in our analysis was between llama and mixed-species repertoires, the frequency of deleterious polymorphisms detrimental to SpA binding, including Ile57, is likely even higher than shown here in the repertoires of alpacas and/or dromedaries. It remains unclear whether such differences arise through germline polymorphism or somatic mutation, although comparison of germline [26] and rearranged [27] camelid V$_H$H sequences favours the latter hypothesis. Also unclear are the reasons why Ile57 should be a hotspot for mutation in dromedaries and/or alpacas, but not llamas.

On the basis of these data, we propose the following general strategy for conferring SpA binding upon camelid V$_H$Hs: (*i*) ensure that FR1 residues Gly15, Ser17 and Arg19 are present and revert any discrepancies to this consensus; (*ii*) ensure that CDR2 residue Thr/Lys/Arg57 is present and revert any discrepancies to this consensus; (*iii*) ensure that FR3 residues Tyr59, Lys64, Gly65, Arg66, Thr68, Ser70, Gln81, Asn82a, Ser/Asn82b are present and revert any discrepancies to this consensus; and (*iv*) closely examine FR3 positions 71–80 for Pro residues (especially at position 75) and unusual deletions and revert these to the nearest human IGHV3 germline residue. We have found no evidence to suggest that any of the substitutions introduced following these rules affect the expression yield, solubility, stability or aggregation status of V$_H$Hs, and while some may not be essential for SpA binding, they are also not harmful. While there were no affinity penalties resulting from these substitutions for any of the five V$_H$Hs shown here, the necessity of restricting CDR2 position 57 diversity to Thr, Lys or Arg may compromise the affinity of other V$_H$Hs.

In conclusion, we have identified the sequence hallmarks responsible for determining camelid V$_H$H binding by SpA, which provide an explanation for species differences in V$_H$H SpA reactivity. We used this information to develop a strategy for engineering V$_H$Hs to introduce SpA binding and enable their tagless purification by SpA chromatography. This strategy may also apply to Ab fragments of other species, or at least those that share homology with human IGHV3 Abs.

Supporting Information

S1 Fig. Binding of V$_H$H pentamers at 100 nM to immobilized SpA by SPR. Each pentamer (FR sequences listed in S2 Table) was injected for 2 min and the number of response units bound at the end of the injection was measured. For pentamers bearing the human IGHV3 consensus residue at all 13 SpA contact positions, no residues are plotted on the graph; instead, dotted lines are shown representing the 95% confidence interval (CI) for mean SpA binding of wild-type pentamers bearing this consensus sequence. For pentamers bearing single amino

acid substitutions at any one of the 13 SpA contact positions, the relevant substitution is plotted on the graph in green (substitution tolerated) if SpA binding fell within the 95% CI for wild-type pentamers, and red (substitution not tolerated) if not. For pentamers bearing multiple amino acid substitutions at SpA contact sites, substitutions are plotted on the graph in blue. We used a verotoxin B-irrelevant peptide fusion as a negative control to rule out potential interactions between SpA and the pentamerization domain (data not shown).
(TIF)

S2 Fig. SpA binding by SPR of four llama V$_H$H monomers (ICAM11-4, ICAM34-1, IGF1R-4 and IGF1R-5; dotted lines) and one llama V$_H$H pentamer (AFAI; dotted line) along with their humanized counterparts (solid line). V$_H$H monomers and pentamers (250 nM) were injected over immobilized SpA for 2 min and allowed to dissociate as described in methods.
(TIF)

S3 Fig. Representative chromatogram overlay of purification of llama, humanized and SpA-engineered ICAM11-4 V$_H$H. All three V$_H$Hs were produced in *E. coli* TG1 cells in 1L 2×YT overnight cultures grown at 37°C. V$_H$Hs were extracted from periplasmic space by osmotic shock and purified using a HiTrap Protein A HP column on an ÄKTA FPLC protein purification system (GE Healthcare).
(TIF)

S4 Fig. Binding of wild-type or SpA-engineered AFAI V$_H$H pentamers at either 1 nM (A) or 50 nM (B) to immobilized CEACAM6 N-terminal domain by SPR. V$_H$H pentamers were injected over immobilized CEACAM6 N-terminal domain for 3 min and allowed to dissociate as described in methods.
(TIF)

S1 Table. FR sequences of SpA-binding and non-SpA-binding V$_H$H monomers.
(DOCX)

S2 Table. FR sequences of SpA-binding and non-SpA-binding V$_H$H pentamers.
(DOCX)

S3 Table. FR sequences of five non-SpA-binding V$_H$Hs and their humanized SpA-binding counterparts.
(DOCX)

S4 Table. Metrics for Illumina MiSeq NGS data used in this study.
(DOCX)

S5 Table. FR and CDR sequences of a SpA-binding (Thr57) and non-SpA-binding (Ile57) dromedary V$_H$H and effect of CDR1-CDR3 disulfide bridge on SpA binding.
(DOCX)

Acknowledgments

Author Contributions

Conceptualization: MAG KAH CRM.

Formal analysis: KAH GH TS EOP MAG HVF.

Investigation: KAH HVF TS EOP.

Methodology: MAG KAH CRM TS EOP GH.

Project administration: KAH MAG.

Software: TS EOP.

Supervision: KAH MAG CRM GH.

Validation: KAH GH CRM MAG.

Visualization: KAH TS.

Writing – original draft: KAH.

Writing – review & editing: KAH CRM MAG GH TS EOP.

References

1. Beck A, Wurch T, Bailly C, Corvaia N. Strategies and challenges for the next generation of therapeutic antibodies. Nat Rev Immunol. 2010; 10: 345–352. doi: 10.1038/nri2747 PMID: 20414207

2. Scott AM, Wolchok JD, Old LJ. Antibody therapy of cancer. Nat Rev Cancer. 2012; 12: 278–287. doi: 10.1038/nrc3236 PMID: 22437872

3. Chan AC, Carter PJ. Therapeutic antibodies for autoimmunity and inflammation. Nat Rev Immunol. 2010; 10: 301–316. doi: 10.1038/nri2761 PMID: 20414204

4. Chames P, Van Regenmortel M, Weiss E, Baty D. Therapeutic antibodies: successes, limitations and hopes for the future. Br J Pharmacol. 2009; 157: 220–233. doi: 10.1111/j.1476-5381.2009.00190.x PMID: 19459844

5. Jain M, Kamal N, Batra SK. Engineering antibodies for clinical applications. Trends Biotechnol. 2007; 25: 307–316. PMID: 17512622

6. Muyldermans S. Nanobodies: natural single-domain antibodies. Annu Rev Biochem. 2013; 82: 775–797. doi: 10.1146/annurev-biochem-063011-092449 PMID: 23495938

7. Bjorck L, Kronvall G. Purification and some properties of streptococcal protein G, a novel IgG-binding reagent. J Immunol. 1984; 133: 969–974. PMID: 6234364

8. Low D, O'Leary R, Pujar NS. Future of antibody purification. J Chromatogr B Analyt Technol Biomed Life Sci. 2007; 848: 48–63. PMID: 17134947

9. Woestenenk EA, Hammarstrom M, van den Berg S, Hard T, Berglund H. His tag effect on solubility of human proteins produced in *Escherichia coli*: a comparison between four expression vectors. J Struct Funct Genomics. 2004; 5: 217–229. PMID: 15503425

10. Khan F, Legler PM, Mease RM, Duncan EH, Bergmann-Leitner ES, Angov E. Histidine affinity tags affect MSP1(42) structural stability and immunodominance in mice. Biotechnol J. 2012; 7: 133–147. doi: 10.1002/biot.201100331 PMID: 22076863

11. Hamilton S, Odili J, Pacifico MD, Wilson GD, Kupsch JM. Effect of imidazole on the solubility of a his-tagged antibody fragment. Hybrid Hybridomics. 2003; 22: 347–355. PMID: 14683594

12. Wu J, Filutowicz M. Hexahistidine (His6)-tag dependent protein dimerization: a cautionary tale. Acta Biochim Pol. 1999; 46: 591–599. PMID: 10698267

13. Chant A, Kraemer-Pecore CM, Watkin R, Kneale GG. Attachment of a histidine tag to the minimal zinc finger protein of the *Aspergillus nidulans* gene regulatory protein AreA causes a conformational change at the DNA-binding site. Protein Expr Purif. 2005; 39: 152–159. PMID: 15642465

14. Amor-Mahjoub M, Suppini JP, Gomez-Vrielyunck N, Ladjimi M. The effect of the hexahistidine-tag in the oligomerization of HSC70 constructs. J Chromatogr B Analyt Technol Biomed Life Sci. 2006; 844: 328–334. PMID: 16904956

15. Hober S, Nord K, Linhult M. Protein A chromatography for antibody purification. J Chromatogr B Analyt Technol Biomed Life Sci. 2007; 848: 40–47. PMID: 17030158

16. Romagnani S, Giudizi MG, del Prete G, Maggi E, Biagiotti R, Almerigogna F, et al. Demonstration on protein A of two distinct immunoglobulin-binding sites and their role in the mitogenic activity of *Staphylococcus aureus* Cowan I on human B cells. J Immunol. 1982; 129: 596–602. PMID: 6979579

17. Sasso EH, Silverman GJ, Mannik M. Human IgA and IgG F(ab')$_2$ that bind to staphylococcal protein A belong to the VHIII subgroup. J Immunol. 1991; 147: 1877–1883. PMID: 1909733

18. Sasano M, Burton DR, Silverman GJ. Molecular selection of human antibodies with an unconventional bacterial B cell antigen. J Immunol. 1993; 151: 5822–5839. PMID: 8228264

19. Sasso EH, Silverman GJ, Mannik M. Human IgM molecules that bind staphylococcal protein A contain VHIII H chains. J Immunol. 1989; 142: 2778–2783. PMID: 2495325

20. Roben PW, Salem AN, Silverman GJ. VH3 family antibodies bind domain D of staphylococcal protein A. J Immunol. 1995; 154: 6437–6445. PMID: 7759880

21. Seppala I, Kaartinen M, Ibrahim S, Makela O. Mouse Ig coded by VH families S107 or J606 bind to protein A. J Immunol. 1990; 145: 2989–2993. PMID: 2120342

22. Cary S, Krishnan M, Marion TN, Silverman GJ. The murine clan V(H)III related 7183, J606 and S107 and DNA4 families commonly encode for binding to a bacterial B cell superantigen. Mol Immunol. 1999; 36: 769–776.

23. Hillson JL, Karr NS, Oppliger IR, Mannik M, Sasso EH. The structural basis of germline-encoded VH3 immunoglobulin binding to staphylococcal protein A. J Exp Med. 1993; 178: 331–336. PMID: 8315388

24. Potter KN, Li Y, Capra JD. Staphylococcal protein A simultaneously interacts with framework region 1, complementarity-determining region 2, and framework region 3 on human VH3-encoded Igs. J Immunol. 1996; 157: 2982–2988. PMID: 8816406

25. Graille M, Stura EA, Corper AL, Sutton BJ, Taussig MJ, Charbonnier JB, et al. Crystal structure of a Staphylococcus aureus protein A domain complexed with the Fab fragment of a human IgM antibody: structural basis for recognition of B-cell receptors and superantigen activity. Proc Natl Acad Sci U S A. 2000; 97: 5399–5404. PMID: 10805799

26. Nguyen VK, Hamers R, Wyns L, Muyldermans S. Camel heavy-chain antibodies: diverse germline V (H)H and specific mechanisms enlarge the antigen-binding repertoire. EMBO J. 2000; 19: 921–930. PMID: 10698934

27. Harmsen MM, Ruuls RC, Nijman IJ, Niewold TA, Frenken LG, de Geus B. Llama heavy-chain V regions consist of at least four distinct subfamilies revealing novel sequence features. Mol Immunol. 2000; 37:579–590. PMID: 11163394

28. 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. J Immunol. 2010; 184: 5696–5704. doi: 10.4049/jimmunol.0903722 PMID: 20404276

29. De Genst E, Saerens D, Muyldermans S, Conrath K. Antibody repertoire development in camelids. Dev Comp Immunol. 2006; 30: 187–198. PMID: 16051357

30. Muyldermans S. Single domain camel antibodies: current status. J Biotechnol. 2001; 74: 277–302. PMID: 11526908

31. Arbabi-Ghahroudi M, To R, Gaudette N, Hirama T, Ding W, MacKenzie R, et al. Aggregation-resistant VHs selected by in vitro evolution tend to have disulfide-bonded loops and acidic isoelectric points. Protein Eng Des Sel. 2009; 22: 59–66. doi: 10.1093/protein/gzn071 PMID: 19033278

32. Arbabi-Ghahroudi M, Tanha J, MacKenzie R. Isolation of monoclonal antibody fragments from phage display libraries. Methods Mol Biol. 2009; 502: 341–364. doi: 10.1007/978-1-60327-565-1_20 PMID: 19082566

33. Baral TN, MacKenzie R, Arbabi Ghahroudi M. Single-domain antibodies and their utility. Curr Protoc Immunol. 2013; 103: Unit 2.17. doi: 10.1002/0471142735.im0217s103 PMID: 24510545

34. Kumaran J, MacKenzie CR, Arbabi-Ghahroudi M. Semiautomated panning of naive camelidae libraries and selection of single-domain antibodies against peptide antigens. Methods Mol Biol. 2012; 911: 105–124. doi: 10.1007/978-1-61779-968-6_7 PMID: 22886248

35. Zhang J, Li Q, Nguyen TD, Tremblay TL, Stone E, To R, et al. A pentavalent single-domain antibody approach to tumor antigen discovery and the development of novel proteomics reagents. J Mol Biol. 2004; 341: 161–169. PMID: 15312770

36. Bell A, Wang ZJ, Arbabi-Ghahroudi M, Chang TA, Durocher Y, Trojahn U, et al. Differential tumor-targeting abilities of three single-domain antibody formats. Cancer Lett. 2010; 289: 81–90. doi: 10.1016/j.canlet.2009.08.003 PMID: 19716651

37. Hussack G, Arbabi-Ghahroudi M, van Faassen H, Songer JG, Ng KK, MacKenzie R, et al. Neutralization of Clostridium difficile toxin A with single-domain antibodies targeting the cell receptor binding domain. J Biol Chem. 2011; 286: 8961–8976. doi: 10.1074/jbc.M110.198754 PMID: 21216961

38. Richard G, Meyers AJ, McLean MD, Arbabi-Ghahroudi M, MacKenzie R, Hall JC. *In vivo* neutralization of alpha-cobratoxin with high-affinity llama single-domain antibodies (VHHs) and a VHH-Fc antibody. PLoS One. 2013; 8: e69495. doi: 10.1371/journal.pone.0069495 PMID: 23894495

39. Abulrob A, Arbabi-Ghahroudi M, Stanimirovic D, inventors; National Research Council of Canada, assignee. Anti-ICAM-1 single domain antibody and uses thereof. United States patent US8623369. 2011 Apr 27.

40. Stanimirovic D, Kemmerich K, Haqqani AS, Sulea T, Arbabi-Ghahroudi M, Massie B, et al., inventors; National Research Council of Canada, applicant. Insulin-like growth factor 1 receptor-specific antibodies and uses thereof. World Intellectual Property Organization patent application WO2015131256. 2014 Dec 4.

41. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins. 2006; 65: 712–725. PMID: 16981200

42. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, et al. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J Am Chem Soc. 1995; 117: 5179–5197.

43. Krivov GG, Shapovalov MV, Dunbrack RL Jr., Improved prediction of protein side-chain conformations with SCWRL4. Proteins. 2009; 77: 778–795. doi: 10.1002/prot.22488 PMID: 19603484

44. Naim M, Bhat S, Rankin KN, Dennis S, Chowdhury SF, Siddiqi I, et al. Solvated interaction energy (SIE) for scoring protein-ligand binding affinities. 1. Exploring the parameter space. J Chem Inf Model. 2007; 47: 122–133. PMID: 17238257

45. Sulea T, Purisima EO. The solvated interaction energy method for scoring binding affinities. Methods Mol Biol. 2012; 819: 295–303. doi: 10.1007/978-1-61779-465-0_19 PMID: 22183544

46. Guerois R, Nielsen JE, Serrano L. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. J Mol Biol. 2002; 320: 369–387. PMID: 12079393

47. Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L. The FoldX web server: an online force field. Nucleic Acids Res. 2005; 33: W382–388. PMID: 15980494

48. Rohl CA, Strauss CE, Misura KM, Baker D. Protein structure prediction using Rosetta. Methods Enzymol. 2004; 383: 66–93. PMID: 15063647

49. Ó Conchúir S, Barlow KA, Pache RA, Ollikainen N, Kundert K, O'Meara MJ, et al. A web resource for standardized benchmark datasets, metrics, and Rosetta protocols for macromolecular modeling and design. PLoS One. 2015; 10: e0130433. doi: 10.1371/journal.pone.0130433 PMID: 26335248

50. Sulea T, Vivcharuk V, Corbeil CR, Deprez C, Purisima EO. Assessment of solvated interaction energy function for ranking antibody-antigen binding affinities. J Chem Inf Model. 2016; 56: 1292–1303. doi: 10.1021/acs.jcim.6b00043 PMID: 27367467

51. Henry KA, Tanha J, Hussack G. Identification of cross-reactive single-domain antibodies against serum albumin using next-generation DNA sequencing. Protein Eng Des Sel. 2015; 28:379–383. doi: 10.1093/protein/gzv039 PMID: 26319004

52. Alturki N, Henry KA, MacKenzie R, Arbabi-Ghahroudi M. Isolation of camelid single-domain antibodies against native proteins using recombinant multivalent peptide ligands. Methods Mol Biol. 2015; 1348: 167–189. doi: 10.1007/978-1-4939-2999-3_16 PMID: 26424272

53. Magoc T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics. 2011; 27: 2957–2963. doi: 10.1093/bioinformatics/btr507 PMID: 21903629

54. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. Bioinformatics. 2011; 27: 863–864. doi: 10.1093/bioinformatics/btr026 PMID: 21278185

55. Murase T, Eugenio L, Schorr M, Hussack G, Tanha J, Kitova EN, et al. Structural basis for antibody recognition in the receptor-binding domains of toxins A and B from *Clostridium difficile*. J Biol Chem. 2014; 289: 2331–2343. doi: 10.1074/jbc.M113.505917 PMID: 24311789

56. To R, Hirama T, Arbabi-Ghahroudi M, MacKenzie R, Wang P, Xu P, et al. Isolation of monomeric human V(H)s by a phage selection. J Biol Chem. 2005; 280: 41395–41403. PMID: 16221664

57. Fridy PC, Thompson MK, Ketaren NE, Rout MP. Engineered high-affinity nanobodies recognizing staphylococcal Protein A and suitable for native isolation of protein complexes. Anal Biochem. 2015; 477: 92–94. doi: 10.1016/j.ab.2015.02.013 PMID: 25707320

58. Zhang J, MacKenzie R, Durocher Y. Production of chimeric heavy-chain antibodies. Methods Mol Biol. 2009; 525: 323–336. doi: 10.1007/978-1-59745-554-1_17 PMID: 19252853

59. De Groeve K, Deschacht N, De Koninck C, Caveliers V, Lahoutte T, Devoogdt N, et al. Nanobodies as tools for *in vivo* imaging of specific immune cell types. J Nucl Med. 2010; 51: 782–789. doi: 10.2967/jnumed.109.070078 PMID: 20395332

60. Pardon E, Laeremans T, Triest S, Rasmussen SG, Wohlkonig A, Ruf A, et al. A general protocol for the generation of nanobodies for structural biology. Nat Protoc. 2014; 9: 674–693. doi: 10.1038/nprot.2014.039 PMID: 24577359

61. Cortez-Retamozo V, Backmann N, Senter PD, Wernery U, De Baetselier P, Muyldermans S, et al. Efficient cancer therapy with a nanobody-based conjugate. Cancer Res. 2004; 64: 2853–2857. PMID: 15087403