Taylor & Francis
Taylor & Francis Group

REPORT

🔓 OPEN ACCESS ✓ Check for updates

A library approach for the *de novo* high-throughput isolation of humanized VHH domains with favorable developability properties following camelid immunization

Paul Arras[a,b], Han Byul Yoo[a,c], Lukas Pekar[a], Christian Schröter[d], Thomas Clarke[e], Simon Krah[a], Daniel Klewinghaus[c], Vanessa Siegmund[c], Andreas Evers[a], and Stefan Zielonka [iD][a,b]

aAntibody Discovery & Protein Engineering, Merck Healthcare KGaA, Darmstadt, Germany; bInstitute for Organic Chemistry and Biochemistry, Technical University of Darmstadt, Darmstadt, Germany; cEarly Protein Supply & Characterization, Merck Healthcare KGaA, Darmstadt, Germany; dADCs & Targeted NBE Therapeutics, Merck KGaA, Darmstadt, Germany; eBioinformatics, EMD Serono, Billerica, MA, USA

ABSTRACT

In this study, we generated a novel library approach for high throughput *de novo* identification of humanized single-domain antibodies following camelid immunization. To achieve this, VHH-derived complementarity-determining regions-3 (CDR3s) obtained from an immunized llama (*Lama glama*) were grafted onto humanized VHH backbones comprising moderately sequence-diversified CDR1 and CDR2 regions similar to natural immunized and naïve antibody repertoires. Importantly, these CDRs were tailored toward favorable *in silico* developability properties, by considering human-likeness as well as excluding potential sequence liabilities and predicted immunogenic motifs. Target-specific humanized single-domain antibodies (sdAbs) were readily obtained by yeast surface display. We demonstrate that, by exploiting this approach, high affinity sdAbs with an optimized *in silico* developability profile can be generated. These sdAbs display favorable biophysical, biochemical, and functional attributes and do not require any further sequence optimization. This approach is generally applicable to any antigen upon camelid immunization and has the potential to significantly accelerate candidate selection and reduce risks and attrition rates in sdAb development.

Introduction

Antibody therapeutics have proven to be of utmost relevance for disease treatment.[1] This is exemplified by the fact that in 2022 antibody-based derivatives accounted for 30% of all entities that were granted marketing access by the US Food and Drug Administration.[2] Furthermore, within the biopharmaceutical sector, monoclonal antibody (mAb) therapeutics represent the dominant modality with respect to approval numbers and revenues.[3]

Along with canonical heterotetrameric antibodies composed of heavy chains and light chains, the adaptive immune system of camelids comprises homodimeric antibodies consisting of heavy chains only and devoid of light chains.[4,5] Intriguingly, these heavy chain-only antibodies (HcAbs) exploit paratopes composed of a single domain, referred to as VHH (variable domain of the heavy chain of a heavy chain-only antibody). Although only three complementarity-determining regions (CDRs) contribute to antigen binding compared with six CDRs of canonical antibody paratopes, VHH domains displaying affinities in the sub-nanomolar range and high specificities for a cognate antigen can readily be generated.[6–9] Additional inherent beneficial attributes include their small size, which might be beneficial for tissue penetration,[10] as well as generally good physicochemical stability.[11,12] Moreover, the simple molecular architecture of the VHH domain enables a plethora of engineering options with respect to the generation of bi- and multispecific antibody designs involving different paratope valences and spatial orientations of individual domains within a molecule.[13–16] Accordingly, camelid-derived single-domain antibodies (sdAbs) emerged as promising modalities for therapeutic applications.[17] As of mid-2023, three VHH-based antibody therapeutics, caplacizumab,[18] envafolimab,[19] and ozoralizumab,[20] have been approved by different health authorities.[21] However, the foreign nature of camelid-derived VHH domains poses an obstacle for therapeutic utilization due to the risk of immunogenicity and anti-drug antibody (ADA) development. While humanized VHHs can be generated using synthetic library approaches,[22,23] the humanization of VHHs from immunized camelids remains a laborious procedure involving the isolation of target-specific paratopes followed by sequence modulation (resurfacing),[24] CDR grafting,[25] or other recently described *in silico* workflows.[26] Often, such sequences require multiple cycles of optimization toward a favorable early developability profile, considering aspects like chemical liabilities, post-translational modifications, immunogenicity, or aggregation tendency. In some cases, it might not be possible to optimize such hits toward a favorable overall profile.[27]

Recently, Teixeira et al. implemented a semi-synthetic antibody library approach that explicitly takes into account developability and functional compatibility of antibody framework and CDR regions.[28] In their library design, CDR-H3 regions were directly amplified from B cells of 10 healthy adult human donors and incorporated into four paired human frameworks. These frameworks were selected from a diverse panel of well-behaved antibodies known for their favorable biophysical characteristics[29] and at the same time cover different germline families, thereby assuring structural and sequence diversity in the library to improve the ability to select binders against different antigens. Finally, to optimize for developable sequences within these frameworks, human CDR-L1-3s and CDR-H1-2s as found in human next-generation sequencing (NGS) data were purged from defined sequence motifs related to chemical instability, PTMs, polyreactivity, and surface hydrophobic patches.

In this work, we generated a generic high-throughput approach for the *de novo* isolation of humanized and *in silico* optimized sdAb panels following camelid immunization. To this end, VHH-derived CDR3 regions were amplified from the peripheral blood mononuclear cell (PBMC) repertoire of a recombinant human (rh) NKp46-immunized llama (*Lama glama*) and grafted onto two different humanized VHH backbone libraries harboring different hallmark residues in framework region 2 (FR2). Both backbone libraries were diversified in CDR1 and CDR2 to potentially compensate for a loss of affinity contributed by amino acid modification introduced through somatic hypermutation during the course of immunization (Figure 1a). To this end, sequence distributions observed in NGS datasets of non-immunized as well as immunized camelids and naïve human antibody repertoires served as starting point (Figure 1b, Supplementary Figure S1). By applying different *in silico* filters, this sequence diversity was
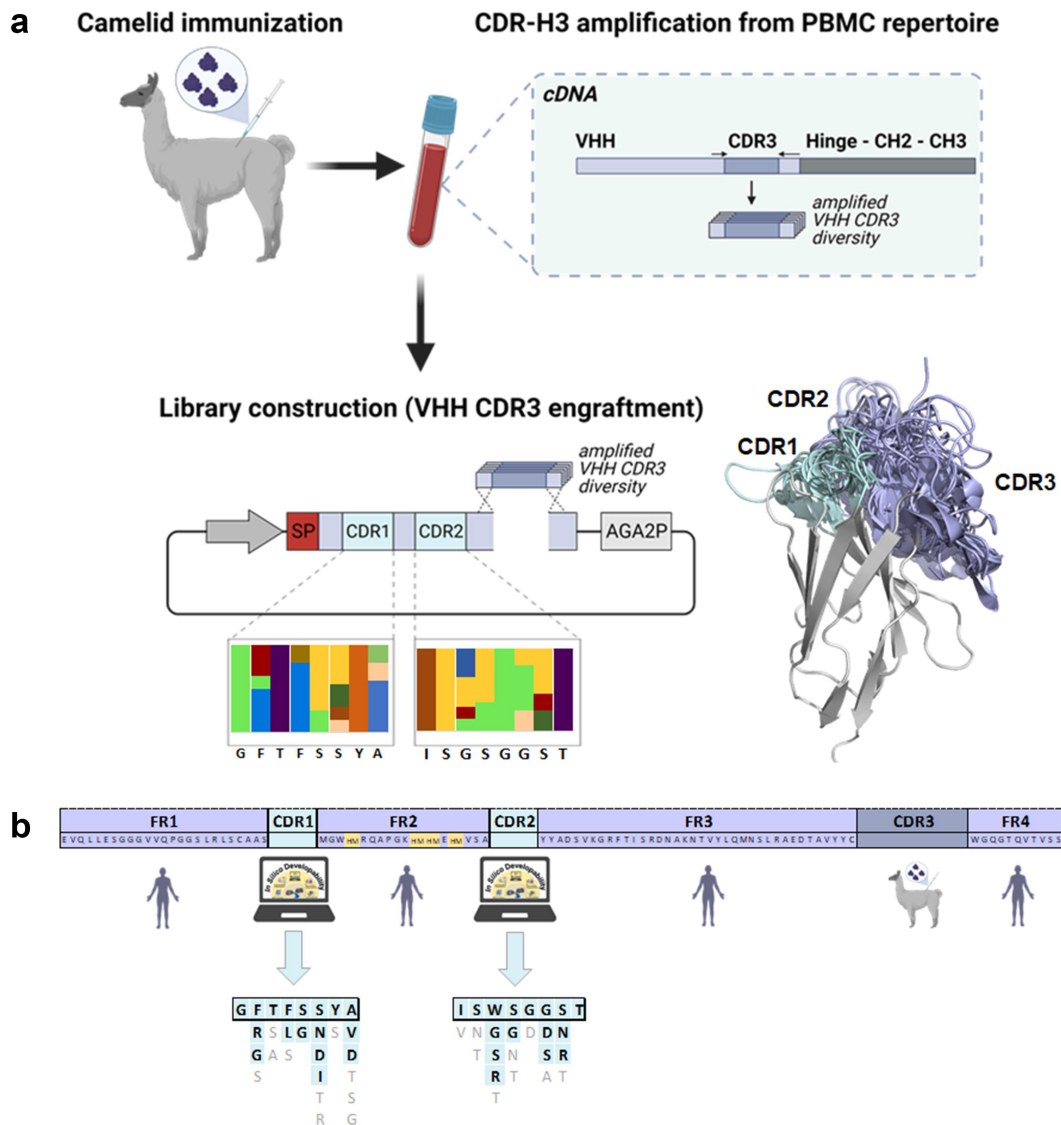
Figure 1. Overview of the library construction process. (a) The VHH CDR3 diversity was amplified from cDNA derived from PBMCs of an immunized llama and grafted onto humanized and sequence optimized sdAb backbones with artificially diversified CDR1 and CDR2 regions. (b) Library design for the humanized backbone libraries. The framework (FR) regions were derived from human IGHV3-23 × 1. Two humanized libraries were generated with different hallmark (HM) signatures, referred to as FERF and VGLW libraries. Residues used in combinatorial diversification of CDR1 and CDR2 are given in cyan and bold. Amino acids observed with frequencies of more than 4% in NGS data sets of WT llama repertoires, which were eliminated from the final design due to *in silico* developability and diversity aspects are indicated in gray. Figure generated using www.biorender.com and PyMOL software version 2.3.0.

tailored toward favorable developability properties, considering aspects such as human-likeness, predicted chemical liabilities and major histocompatibility complex II (MHC-II) binding. Screening those libraries *via* yeast surface display (YSD)[30,31] in direct comparison to the wild-type (WT) VHH library revealed a similar enrichment and comparable sequence diversity of the selection output, while exhibiting a significantly increased human-likeness and computed physical and chemical stability profile[32]. Biophysical, biochemical, and functional characterization of the resulting molecules showed further favorable inherent attributes of *de novo* humanized sdAbs. Ultimately, we demonstrate that this novel approach of *de novo* high-throughput humanization enables the one-step isolation of humanized and sequence-optimized sdAbs displaying comparable affinities to corresponding WT VHHs as well as favorable early signs of developability.

Results

Library design strategy for the de novo isolation of humanized sdAbs from camelid immune repertoires

Within this work, we aimed at generating a novel approach for the high throughput *de novo* isolation of humanized sdAbs after immunization of camelids. To this end, we developed a semi-immune/semi-synthetic strategy that relies on grafting the PBMC-amplified VHH CDR3 repertoire of llamas following immunization onto two humanized backbone libraries (Figure 1a). Based on sequence similarities between camelid VHH domains and the corresponding human VH germline sequences, both humanized backbone libraries were derived from human IGHV3–23 × 1 (Figure 1b, Supplementary Figure S1). As shown in Supplementary Table S1, IGHV3–23 × 1 shares high overall sequence similarity with the llama or alpaca V gene repertoire (https://www.imgt.org), and therefore represents a suitable general template for VHH humanization[26]. In detail, sequence identities of IGHV3–23 × 1 are in the range of 81.2–92.5% in the framework region (with most differences not located in Vernier positions) to the most prevalent V-genes observed in alpaca and llama repertoires. Leu11 (according to Kabat numbering) was replaced by valine aiming to reduce the reactivity to preexisting ADAs.[33,34] Essentially, the two libraries differed in their framework region 2 (FR2) hallmark residue signatures. It is known that camelid VHH hallmark residues in FR2 (37F, 44E, 45 R, and 47F) encoded in dedicated V germline genes contribute to the overall stability of the VHH domain[35,36]. These hallmark residues reshape the classical VL interface and are considered to largely contribute to the high water solubility of this domain.[37] In line with this, the VHH hallmark motif, as well as subtle modified versions thereof, was also predominant in VHHs from non-immunized llamas, as observed from NGS datasets of naïve camelids (Supplementary Table S2). Intriguingly, we found a subset (7.2%) of variable domains belonging to the HcAb repertoire (here, the V domain is directly fused to the hinge region without domain CH1) without this hallmark signature (comprising 37 V, 44 G, 45 L, and 47W; Supplementary Table S2). Previously, it was shown that such autonomous V domains can display acceptable biophysical attributes and are capable of

binding their cognate antigens with high affinities.[13] Importantly, residues at these hallmark positions are identical to the closest corresponding human germline.

Consequently, we designed a humanized backbone library comprising hallmark signatures of camelid VHH domains (37F, 44E, 45 R, and 47F, referred to as *FERF* library) and a second humanized library harboring the classical V signature (V37, G44, L45, and W47 termed *VGLW* library). To compensate for a potential loss of affinities caused by mutations of residues in CDR1 and CDR2 as introduced by somatic hypermutation during the course of immunization, we moderately diversified both CDRs (Figure 1). Assuming that the final yeast surface display library is limited to a size of $10^8 - 10^9$ different clones,[30] we aimed to diversify the CDR1 and CDR2 regions to a final number of 10^4 different sequence variations in the combinatorial library. Diversification of specific residues in those regions was 'inspired' by the amino acid distribution found in naïve and immunized llamas, considering only residues that appeared with a relative frequency of ≥4% at each specific position. This list was further reduced by eliminating residues that would result in 1) highly susceptible chemical liability motifs (NG, DG, methionine, and unpaired cysteine),[38] 2) N-glycosylation motifs (N-X-S/T), 3) strong predicted MHC-II binding peptide motifs, 4) with respect to diversity in terms of amino acids charge, size, and hydrophobicity, and 5) occurrence in the equivalent positions in NGS data of human antibody repertoires (Supplementary Figure S1). This procedure resulted in a theoretical combinatorial backbone library space of 1.04×10^4 variants (in CDR1 and CDR2). A stuffer region was exploited instead of CDR3, encoding for multiple stop codons in every possible reading frame that harbored multiple *Bsa* I restriction sites for CDR3 diversity engraftments by homologous recombination. For the amplification of CDR3 repertoires from immunized llamas, primers were designed that convert camelid FR3 and FR4 positions into human counterparts during polymerase chain reaction (Supplementary Table S3). Both humanized backbone libraries were generated for yeast surface display and importantly, NGS analysis revealed that the constructed libraries matched fairly well with the initial designs with respect to the amino acid distribution in CDR1 and CDR2 (Figure 2).

CDR3 engraftments onto humanized libraries enable the enrichment of sequence-diverse panels of antigen-specific sdAbs by YSD

To investigate whether this semi-synthetic, semi-immune library approach allows for the *de novo* isolation of humanized camelid-derived sdAbs similar to the conventional VHH isolation procedure by YSD,[7] we opted for PBMC-derived total RNA of a recombinant human (rh) NKp46-immunized llama as template for the generation of both humanized libraries. For direct comparison, we also constructed a conventional VHH library based on the same starting material.[7] NKp46 is an activating receptor on natural killer (NK) cells belonging to the group of natural cytotoxicity receptors.[39] It was shown by Vivier and coworkers as well as by our group that NKp46 can be harnessed for efficient NK cell redirection.[40–43] In order to generate both humanized libraries (*FERF* and *VGLW*,
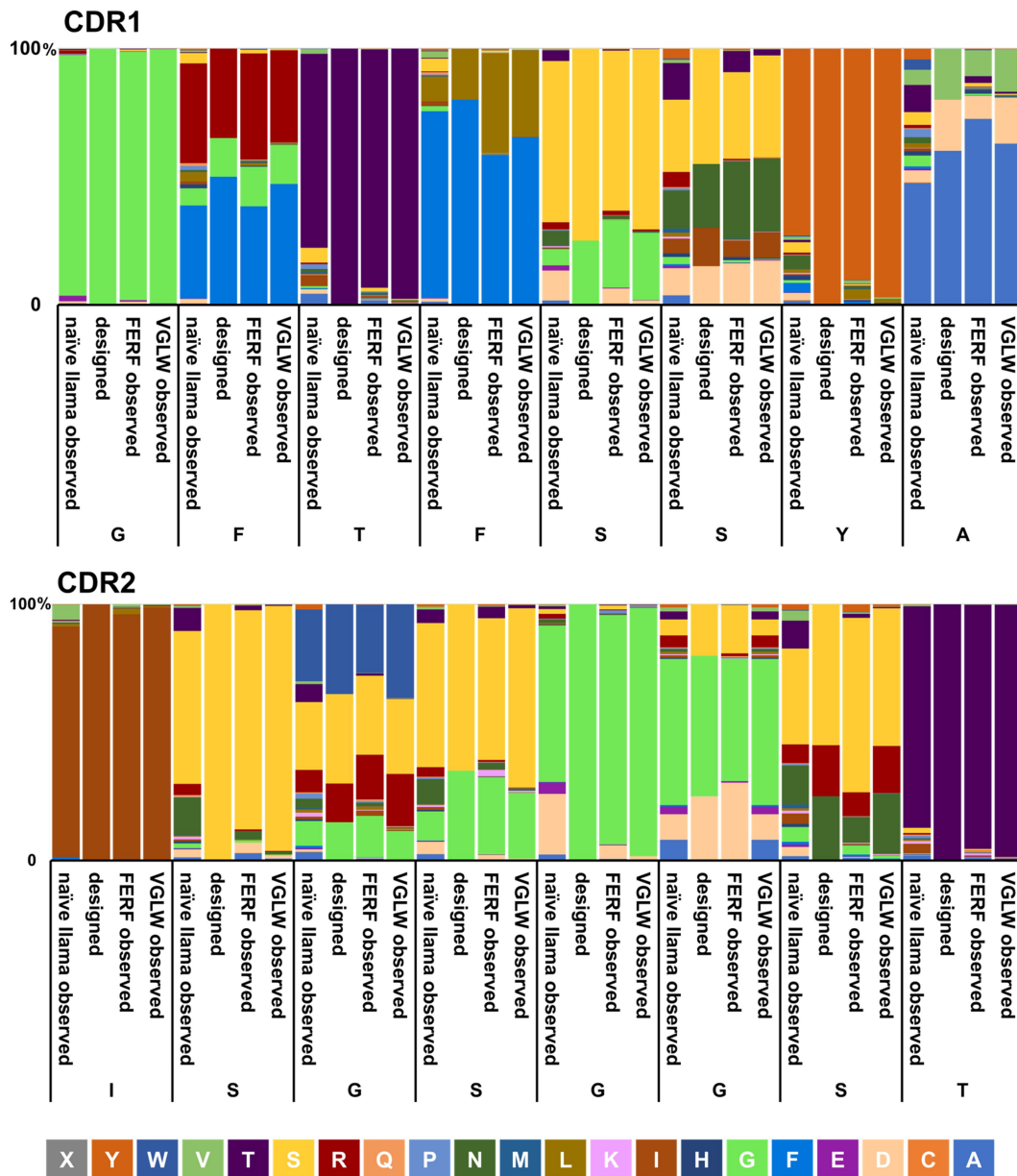
Figure 2. Amino acid distribution of CDR1 and CDR2 of non-immunized llamas compared with artificially designed diversities and observed compositions in humanized sdAb libraries. Corresponding CDR residues of human germline IGHV3-23 × 1 given below. Amino acid distributions were observed by NGS.

respectively), the VHH-derived CDR3 repertoire was amplified in a one-step PCR with oligonucleotides annealing to FR3 and FR4 (Supplementary Table S3). The CDR3 stuffer regions of both backbone library plasmids were digested using *Bsa*I and individual CDR3-engrafted humanization libraries were constructed in a homologous recombination-based process referred to as gap-repair cloning.[44] All three resulting libraries were subjected to library selection by fluorescence-activated cell sorting (FACS). To this end, a two-dimensional labeling strategy was exploited to simultaneously select for (rh) NKp46 binding and full-length sdAb display.[45] To potentially also select for antibody-derivatives displaying lower affinities for (rh) NKp46, an antigen concentration of 1 µM was used. Within the first round of selection, we already observed a distinct antigen-binding population for both humanized libraries similar to the WT VHH library, and after two sorting

rounds in total, we were able to significantly enrich for antigen-binding populations from all three libraries (Figure 3a). From the sorting output, 96 clones from each library were sent for sequencing. We merged the set of (3 × 96=) 288 sequences and applied a clustering strategy based on CDR3 length and 50%, 85%, or 100% sequence identity of the CDR3 sequence, assuming that sequences within the same clusters will bind to the same epitope. Encouragingly, inspection of the clusters revealed sequences with identical or similar CDR3s that originated from the different libraries (*WT*, *FERF* or *VGLW*, Figure 3b, Supplementary Table S4). However, there were also a few CDR3 clusters harboring only sequences from one of the three different libraries. As expected, within identical CDR3 clusters, we observed a certain diversity of the CDR1 and CDR2 sequences. Indeed, there was no occasion where WT or designed library sequences had identical CDR1-CDR2-
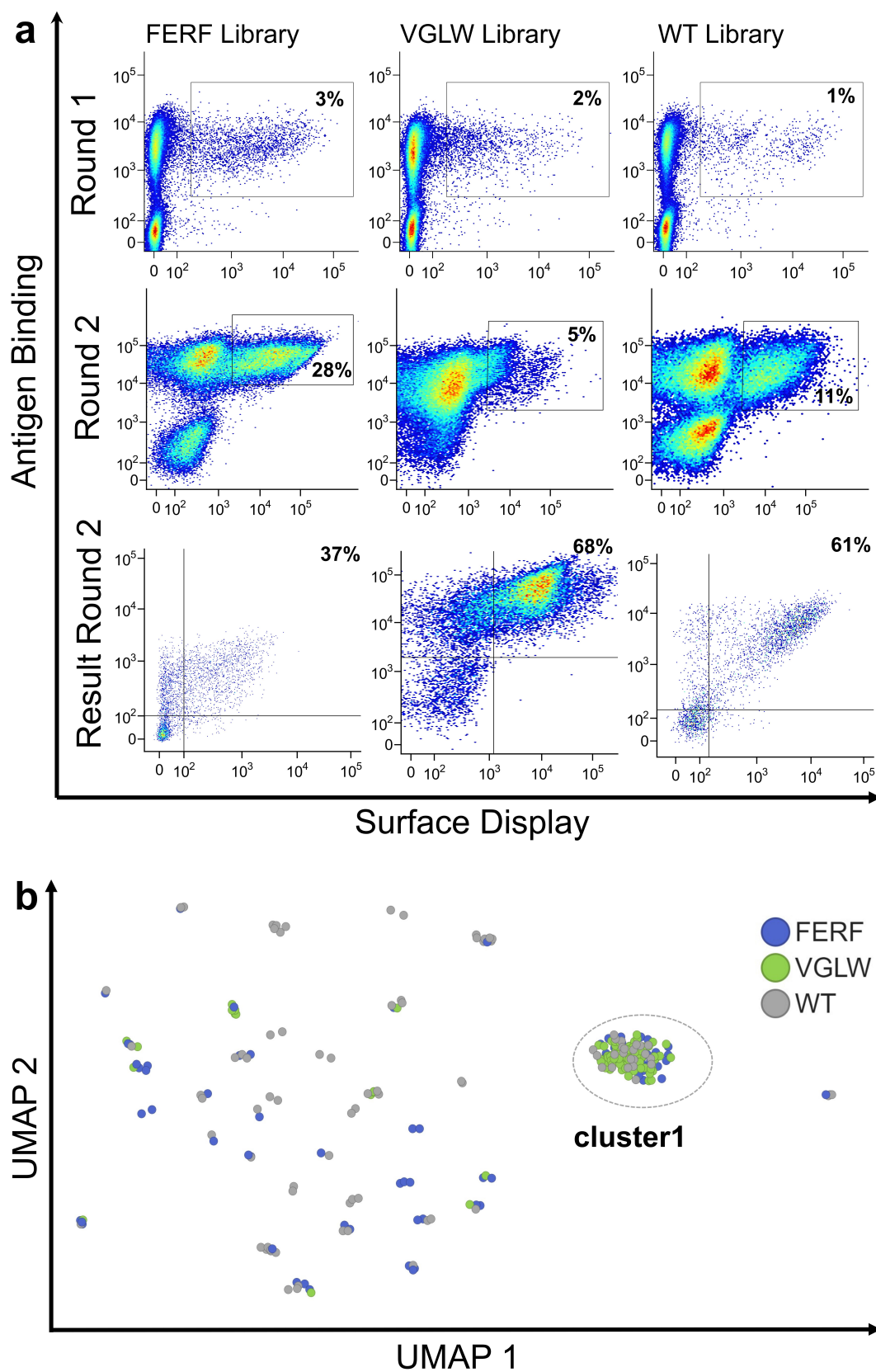
Figure 3. Yeast surface display enables the isolation of (rh) NKp46-targeting sdAbs from both humanized libraries. (a) FACS-based enrichment of the FERF, VGLW, and the WT libraries by applying a two-dimensional sorting strategy for simultaneous detection of antigen binding and full-length sdAb display. Two consecutive rounds of selection (Round 1 and Round 2) were conducted for each library. (b) Similarity of CDR3 sequences as analyzed by UMAP dimensionality reduction. Each dot represents the CDR3 of an individual VHH sequence. Dots are colored based on the library origin (FERF shown in blue, VGLW in green, and WT given in gray).

Table 1. CDR1–3 and hallmark (HM) sequences of immunized llama WT or humanized VHH libraries (FERF or VGLW) that were selected for production and experimental profiling (shown in Table 3). Residues that are different from the most potent sequences within each CDR3 sequence cluster are shown in orange; susceptible chemical liability and post-translational modification motifs (DG, NG, M, unpaired C, and NXS/T) are indicated as red boxes.

ID	CDR3 Cluster ID	CDR1	HM	CDR2	CDR3
WT.cluster1.1	1	GGPTRTYA	FERF	ISGDGATT	AAIRTPADSQ--VIVTLHWYRY
FERF.cluster1.2	1	GFTLSNYA	FERF	ISRSGGST	AAGRTPAESQ--IIVTLDWYRY
WT.cluster1.3	1	GGPTNVYA	FERF	ISGDGATT	AAIRTPAESQ--VIVTLDWYRY
FERF.cluster1.4	1	GRTFSIYA	FERF	ISRSGGST	AAIRTPAESQ--VIVTLDWYRY
FERF.cluster1.5	1	GFTFSSYA	FERF	ISSSGGST	AAIRTPAESQ--VIVTLDWYRY
WT.cluster1.6	1	GGPTRTYA	FERF	ISGDGATT	AAIRTPAESQ--VIVTLDWYRY
FERF.cluster1.7	1	GRTFSSYA	FERF	ISSSGGST	AAIRTPAESQ--VIVTLDWYRY
WT.cluster1.8	1	GDNFSSYI	FERF	ISGSSENT	AADKTRSQSI--VIVTSEWYDY
VGLW.cluster1.9	1	GRTFSSYA	VGLW	ISRSGGST	AAIRTPAESQ--VIVTLDWYRY
FERF.cluster2.10	2	GRTFSSYV	FERF	ISWGGDNT	AAALAPSGTT--VVVSPLGYDN
FERF.cluster2.11	2	GFTFGSYA	FERF	ISSSGGST	AAAALVSGTT--VIVSPVRYDY
FERF.cluster2.12	2	GRTLGSYA	FERF	ISSSGGNT	AATHHQPSTI--VVVMPIGYGY
WT.cluster2.13	2	TNIFSINT	YERF	ISWSGDST	AASSAGGNTI--VIVSPRGYGY
VGLW.cluster2.14	2	GRTLSSYV	VGLW	ISWSGGST	AAALAPSGTL--VVVSPLGYTY
FERF.cluster2.15	2	GFTFGNYA	FERF	ISWSGGRT	AASSAGGNTI--VIVSPRGYGY
FERF.cluster3.16	3	GGTLGSYA	FERF	ISSSGDNT	AAAGGMGSSI--VVVSTIPYKY
FERF.cluster3.17	3	GRTLSDYV	FERF	ISSSGSST	GAAETYGESA--VYISPHRYAY
FERF.cluster3.18	3	GRTFGSYA	FERF	ISSSGDST	AAAGGIGSST--VVVSPIPYAY
FERF.cluster3.19	3	GGTFSNYA	FERF	ISSGGGST	AAAGGMGSTT--VVVSTIPYKY
VGLW.cluster3.20	3	GFTLGDYA	VGLW	ISSSGDNT	GAAETAGESA--VYISPHRYAY
FERF.cluster4.21	4	GFTFSSYA	FERF	ISSSGGST	AAVQQPPSVA---VVSYRGYNY
FERF.cluster4.22	4	GFTFGSYA	FERF	ISWSGGST	AAVQQPPSVA---VVAYRGYNY
VGLW.cluster4.23	4	GFTLSNYA	VGLW	ISWGGSNT	AADQQPPSVA---VVAARGYRY
VGLW.cluster4.24	4	GGTFGIYA	VGLW	ISWGGSST	AAVQQPPSVA---VVSYRGYNY
WT.cluster5.25	5	GRTFSSYA	FERF	ISRSDDNT	AAVKVPVSST--IYTDQRIYTN
FERF.cluster5.26	5	GFTFSSYA	FERF	ISWSGGNT	AAVKVPVSST--IYTDQRIYTN
FERF.cluster5.27	5	GFTLGNYV	FERF	ISGSGDST	AAVKVPVSST--IYTDQRIYTN
FERF.cluster6.28	6	GRTLSSYA	FERF	ISSSGGRT	ATSLTYDQTT--VYVSPLPYDD
FERF.cluster6.29	6	GRTFSDYA	FERF	ISSSGGST	ATSLTYDQTT--VYVSPLPYND
WT.cluster6.30	6	GRTTSSYS	FERF	ISWTGTTI	ATSLTYDQTT--VYVSPLAYGD
WT.cluster7.31	7	RSTFGNYA	FERF	ISRSGGNT	AAAQGGSTTV---FITPQVYEY
FERF.cluster7.32	7	GFTFGDYA	FERF	ISSSGGST	AAAQGGSTTV---FITPQVYEY
WT.cluster8.33	8	GRTFSSYA	FERF	ISWSGEST	ASNPATSTVL---IVRDLGYAY
FERF.cluster8.34	8	GGTLGSYA	FERF	ISGGGGST	ASNPATSTVL---IVRDLGYAY
WT.cluster9.35	9	GFTFSSYA	VGPW	INSNGDRT	AQGGY-------------SDSP
WT.cluster10.36	10	GRTFSRYA	FERF	ISWSGDNT	AASYTMSAYQ--TLISSRAYAY
WT.cluster11.37	11	GFTFSSYY	VGLW	IYSEDSAT	ASKIIENGSSW-YPHRFHEYDY
WT.cluster12.38	12	GRTFSSYA	FERF	ISYSGDST	AAASVGSTSTV-VAVSDLLYNY
WT.cluster13.39	13	GRTFSSNA	FERF	ISSDGSA	AAKRTDQSSV--VWTSDLAYDY
WT.cluster14.40	14	GRTFANYA	FERF	INWNDGGT	AARESYSTTT--YYVNHMAYNY
FERF.cluster15.41	15	GFTFGNYA	FERF	ISSSGDST	AAVRSNYGEN--IYIHGMAYNV
VGLW.cluster16.42	16	GFTFSSYA	VGLW	ISSSGDNT	AASLIGPSTV--IVTNSRGYYY
WT.cluster17.43	17	GFTFSTYP	VGLW	INSGGGRT	AIGGYP----------RSRDS
WT.cluster18.44	18	GRTFSSYA	FERF	FGWSGGTT	AARTVFSDTDSAVYISEHMYRY
FERF.cluster19.45	19	GRTFSNYD	FERF	ISWSGGST	AAVGYPTC-----PLTANYDY
VGLW.cluster20.46	20	GRTFSSYA	VGLW	ISWSGSST	AAALWGDSG-----CLTGKPNF
VGLW.cluster21.47	21	GFTLGDYD	VGLW	ISWSGGNT	AAEWAAQAS----LDDAGDYDY
WT.cluster22.48	22	GFTFSSYA	VGLW	ISSGGSRT	AMGGI-------------TTAP
WT.cluster23.49	23	GSIFSFNV	YPRL	ITSG-YST	NFRHIF----------RQDEY

CDR3 sequences. Strikingly, more than 50% of all sequences from the three different libraries had an identical CDR3 sequence. Among all CDR3 sequence clusters, cluster 1 (Table 1 and Figure 3b) was significantly most populated within the sequences obtained from Sanger sequencing. This is in agreement with the frequency distributions observed from NGS analysis for the screening pools obtained after the second sorting round: CDR3 sequence cluster 1 was observed in 27.8% (Sanger) or 31.5% (NGS) of all sequences from the WT library, in 49.0% (Sanger) or 74.2% (NGS) in the *FERF* library and in 78.0% (Sanger) or 88.0% (NGS) in the *VGLW* library. All other CDR3 sequence clusters were observed with a significantly lower relative frequency.

From the set of sequences obtained by both approaches (Sanger sequencing and NGS), we selected 49 clones for in-depth characterization. Selection of these clones was motivated by the following questions: 1) do sequences from the designed libraries show significant differences in binding affinity or (*in silico*) developability properties compared to CDR3 analogues from the WT library, and 2) how far do sequence differences in CDR1/CDR2 within the same CDR3 clusters contribute to differences in binding affinity or developability properties? Furthermore, we selected sequences from several CDR3 clusters that contained only members of the WT, *FERF* or *VGLW* library. Nominated sequences are shown in Table 1.

De novo *humanized sdAbs display a favorable* in silico *developability profile*

In silico developability profiles were computed using our SUMO (Sequence Assessment Using Multiple Optimization Parameters) pipeline.[32] This procedure computes 3D models based on the provided VHH sequences, reports their human-likeness by sequence identity to the most similar human germline sequence, determines structure-based surface-exposed chemical liability motifs (unpaired cysteines, methionines, asparagine deamidation motifs, and aspartate deamidation sites) and sites susceptible to post-translational modification (N-linked glycosylation). In addition, the isoelectric point (pI), Schrodingers AggScore as estimate for hydrophobicity and aggregation tendency (for the entire Fv and for the CDRs only), and the calculated positive patch energy of the CDRs are reported. These scores are complemented with a green to yellow to red color coding to indicate deviations from the mean scores over a benchmarking dataset of marketed antibodies. The SUMO overview for the 49 sequences that were selected for production and experimental profiling is shown in Table 2. Due to the design strategy, the human-likeness was significantly higher for the sequences obtained from the humanized libraries with an average sequence identity to the closest germline sequence within the framework of 92.4% (range 90.0%–96.3%) and 80.0% (range 68.8%–92.5%) only for the WT sequences.

With regard to the computed physicochemical properties (pI, AggScore, charge patches), the VHHs showed a certain degree of diversity. All sequences covered a broad range of computed hydrophobicity/aggregation scores and positively charged patches, without showing substantial deviations from standard antibody drugs. We intentionally selected

sequences with large diversity in predicted pI, since the pI can strongly influence several developability and manufacturability properties, such as solubility and aggregation during purification or virus inactivation, colloidal stability or viscosity in formulation, nonspecific binding, or clearance. Consequently, we considered it beneficial to have different pI variants of a lead sequence available that could represent potential backups in a project.

Due to the design strategy, none of the sequences derived from the humanized library carried any N-glycosylation or highly susceptible deamidation (NG) or isomerization (DG) sites in CDR1 or CDR2. Other potential deamidation or isomerization motifs (e.g., NS, NN, NT; DS, DD, DT) might still result in chemical instability, depending on their solvent accessibility, conformational flexibility, or environmental effects such as the formulation.[38] However, since degradation of these motifs occurs significantly less frequent based on internal and literature data,[38] we did not exclude them from the design, but usually assess their relevance case-by-case, either by post-filtering based on more rigorous *in silico* liability assessments or by experimental profiling. In very few cases, we observed enrichment of sequences harboring an unpaired cysteine residue within CDR3, apparently due to the fact that this residue naturally forms a disulfide bond with another cysteine situated in FR2 of the WT repertoire. This cysteine counterpart was not designed to be present in the humanized library. In realistic project applications, such sequences with unpaired cysteine residues would be disregarded for follow-up studies. Notwithstanding, since such non-canonical disulfide bonds were observed in more than 15% of NGS data of llama WT sequences (a cysteine in CDR3 as well as a cysteine located in the last position of FR2) (see Supplementary Table S4), an alternative option might be the design of an additional humanized VHH library with a cysteine residue in this last position of FR2.

De novo *humanized sdAbs display high-affinity antigen binding and favorable early developability properties*

For antibody production, nominated sdAbs were reformatted as one-armed, monovalent sdAb-Fc fusions (Supplementary Figure S2A). To this end, we utilized the strand-exchanged engineered domain (SEED) technology for Fc heterodimerization.[46] This technique is based on beta strand exchanges between the CH3 domains of IgG and IgA, preferentially resulting in the formation of heavy-chain heterodimers. VHH domains were grafted onto the hinge region of the AG chain of the SEED molecule, while the GA chain was expressed in a paratope-less fashion. In addition, we implemented the RF mutation into the GA chain of the SEEDbody to prevent binding of GA:GA homodimers to protein A,[47] since this misassembled side product would affect downstream analysis. Production was performed in ExpiCHO™ cells at a scale of 5 ml. Of note, clones we have not been able to express were derived from all three different libraries (two *WT*: *WT. cluster6.30, WT.cluster23.49*; four *FERF*: *FERF.cluster3.19, FERF.cluster7.32, FERF.cluster8.34, FERF.cluster19.45;* and two *VGLW*: *VGLW.cluster3.20, VGLW.cluster4.24* members).

Table 2. *In silico* developability assessment of VHHs obtained from different library approaches. sdAbs were analyzed for their sequence identity compared to the most similar human germline (MOST SIMILAR GERMLINE) either based on the entire variable chain region (SEQ-ID) or the framework region only (SEQ-ID FR), as well as for their total number of specific chemical liabilities and PTMs, i.e., non-canonical cysteines, methionine oxidations, asparagine deamidations or aspartate isomerizations, and N-glycosylations, in structurally exposed CDR residues as derived from automatically generated models. As calculated physicochemical developability descriptors (IN SILICO PHYSCHEM), structure-based pl values (plfv 3D), the AggScores of the entire variable regions, and the AggScores of CDR regions only (CDR AggScore), as well as the positive patch energy of the CDRs (CDR positive patch energy), are shown. The complementing color coding indicates scores within one standard deviation from a benchmark mean (dataset of 77 biotherapeutics approved for human application) as green, scores above one standard deviation as yellow, and scores above two standard deviations as red. For the AggScores, this classification was slightly adjusted based on correlation analyses to internal experimental HIC data.

ID	CDR3 Cluster ID	human likeness			CDR LIABILITIES 3D (*in silico*)				CDR PTMs (in silico)	IN SILICO PHYSCHEM			
		MOST SIMILAR GERMLINE	SEQ-ID	SEQ-ID FW	NON-CANONICAL CYS	MET OXIDATION	ASN DEAMIDATION	ASP ISOMERIZATION	N-GLYCOSYDATION	plFv 3D	AggScore	CDR AggScore	CDR Positive Patch Energy
WT.cluster1.1	1	IGHV3-23*01	64,0	68,8	0	0	0	3	0	6,9	74,4	51,0	442,1
FERF.cluster1.2	1	IGHV3-23*01	86,0	91,3	0	0	0	1	0	8,5	85,0	42,0	412,2
WT.cluster1.3	1	IGHV3-23*01	65,5	70,9	0	0	0	2	0	5,9	94,1	61,0	369,4
FERF.cluster1.4	1	IGHV3-23*01	86,0	91,3	0	0	0	1	0	8,7	111,9	63,6	617,8
FERF.cluster1.5	1	IGHV3-23*01	87,7	91,3	0	0	0	1	0	7,9	105,3	75,6	369,9
WT.cluster1.6	1	IGHV3-23*01	64,0	68,8	0	0	0	2	0	6,2	58,3	33,6	398,7
FERF.cluster1.7	1	IGHV3-23*01	86,8	91,3	0	0	0	1	0	8,4	79,1	42,7	433,5
WT.cluster1.8	1	IGHV3-23*01	76,8	81,0	0	0	1	1	1	4,4	106,9	67,7	301,1
VGLW.cluster1.9	1	IGHV3-23*01	90,4	96,3	0	0	0	1	0	9,0	114,9	57,9	658,3
FERF.cluster2.10	2	IGHV3-23*01	83,3	91,3	0	0	1	1	0	6,3	140,1	87,5	262,5
FERF.cluster2.11	2	IGHV3-23*01	88,5	91,3	0	0	0	1	0	7,9	99,2	59,9	329,6
FERF.cluster2.12	2	IGHV3-23*01	85,0	91,3	0	1	1	1	0	8,5	180,2	121,4	416,6
WT.cluster2.13	2	IGHV3-23*01	73,2	77,2	0	0	1	2	0	4,5	127,7	103,1	427,9
VGLW.cluster2.14	2	IGHV3-23*01	89,4	96,3	0	0	0	1	0	8,3	180,2	133,3	472,8
FERF.cluster2.15	2	IGHV3-23*01	85,8	91,3	0	0	1	1	0	9,0	81,9	50,1	585,2
FERF.cluster3.16	3	IGHV3-23*01	84,1	91,3	0	1	1	1	0	7,9	123,0	71,1	307,6
FERF.cluster3.17	3	IGHV3-23*05	83,5	91,3	0	0	0	1	0	7,0	78,5	45,5	341,9
FERF.cluster3.18	3	IGHV3-23*01	85,8	91,3	0	0	0	2	0	7,8	109,7	74,8	395,9
FERF.cluster3.19	3	IGHV3-23*01	85,8	91,3	0	1	0	1	0	8,3	93,3	55,9	448,9
VGLW.cluster3.20	3	IGHV3-23*01	86,1	96,3	0	0	1	1	0	4,3	111,2	38,8	270,8
FERF.cluster4.21	4	IGHV3-23*01	88,5	91,3	0	0	0	1	0	8,5	70,1	24,9	385,4
FERF.cluster4.22	4	IGHV3-23*01	87,6	91,3	0	0	0	1	0	8,3	93,0	43,2	351,8
VGLW.cluster4.23	4	IGHV3-23*01	86,1	96,3	0	0	1	1	0	8,3	76,7	46,0	478,9
VGLW.cluster4.24	4	IGHV3-23*01	87,6	96,3	0	0	0	1	0	8,5	176,2	101,5	411,5
WT.cluster5.25	5	IGHV3-23*01	74,8	82,5	0	0	1	2	0	8,4	30,9	13,4	596,9
FERF.cluster5.26	5	IGHV3-23*01	84,3	90,0	0	0	1	1	0	8,3	70,2	52,7	368,5
FERF.cluster5.27	5	IGHV3-23*01	81,7	90,0	0	0	0	2	0	7,9	58,1	40,2	363,6
FERF.cluster6.28	6	IGHV3-23*01	85,1	91,3	0	0	0	2	0	6,4	101,8	75,1	344,4
FERF.cluster6.29	6	IGHV3-23*01	86,7	91,3	0	0	1	1	0	4,9	120,6	99,1	319,5
WT.cluster6.30	6	IGHV3-48*03	71,1	77,5	0	0	0	0	0	4,3	144,0	96,4	293,8
WT.cluster7.31	7	IGHV3-23*01	68,4	73,4	0	0	1	1	0	6,6	109,3	50,9	486,6
FERF.cluster7.32	7	IGHV3-23*01	84,3	90,0	0	0	0	1	0	6,3	127,5	86,1	232,1
WT.cluster8.33	8	IGHV3-23*01	77,7	79,7	0	0	0	1	0	8,4	99,2	67,1	404,0
FERF.cluster8.34	8	IGHV3-23*01	85,8	91,3	0	0	0	1	0	7,8	103,5	66,0	265,1
WT.cluster9.35	9	IGHV3-23*05	83,3	88,8	0	0	1	2	0	3,8	51,9	3,8	269,7
WT.cluster10.36	10	IGHV3-23*01	77,0	82,5	0	1	2	0	0	8,8	99,4	90,6	605,8
WT.cluster11.37	11	IGHV3-23*05	86,7	92,5	0	0	1	2	1	5,4	185,6	98,8	403,4
WT.cluster12.38	12	IGHV3-23*01	79,6	82,5	0	0	0	2	0	6,8	83,1	65,0	384,7
WT.cluster13.39	13	IGHV3-23*05	67,3	71,3	0	0	0	1	0	3,8	107,0	70,9	293,5
WT.cluster14.40	14	IGHV3-64*04	70,1	77,2	0	1	1	2	0	5,5	89,4	63,6	475,3
FERF.cluster15.41	15	IGHV3-23*01	83,9	90,0	0	1	0	2	0	7,1	117,4	71,1	431,4
VGLW.cluster16.42	16	IGHV3-23*01	90,3	96,3	0	0	2	1	0	7,7	104,4	68,6	358,6
WT.cluster17.43	17	IGHV3-23*01	81,4	91,3	0	0	1	2	0	7,0	53,7	18,7	591,3
WT.cluster18.44	18	IGHV3-23*01	73,9	78,8	0	0	0	2	0	5,7	130,0	92,3	504,2
FERF.cluster19.45	19	IGHV3-23*01	86,7	91,3	1	0	0	1	0	6,3	66,1	30,7	336,9
VGLW.cluster20.46	20	IGHV3-23*01	91,2	96,3	1	0	0	2	0	8,4	236,6	105,8	511,0
VGLW.cluster21.47	21	IGHV3-23*01	87,7	96,3	0	0	1	3	0	3,2	69,1	30,6	67,3
WT.cluster22.48	22	IGHV3-23*03	85,8	91,1	0	0	0	1	0	9,4	36,3	6,0	375,6
WT.cluster23.49	23	IGHV3-66*01	73,9	84,8	0	0	0	0	1	8,2	86,5	48,4	366,3

To assess early developability properties, we exploited analytical size-exclusion chromatography (SEC) after protein A purification as a first filter. In general, purities above 85% target peak are considered as adequate attributes for transient antibody expression, while purities of more than 90% indicate favorable properties. Intriguingly, besides one VHH SEEDbody obtained from the *VGLW* library sort (*VGLW. cluster20.46*) harboring an unpaired cysteine residue in CDR3, purities for all other constructs were above 90% target monomer peak (Table 3). Importantly, aggregation properties

for sdAb-based entities obtained from the independent libraries were not appreciably different. As an additional layer of biophysical attributes in order to monitor early developability, we also scrutinized all VHH SEEDbodies using analytical hydrophobic interaction chromatography (HIC). For this, we utilized two therapeutic antibodies as reference, cetuximab and avelumab, with HIC retention times of 5.7 min and 7.2 min, respectively. In general, HIC retention times of the vast majority of sdAb-based antibodies were favorable. In this respect, most molecules displayed even shorter retention times than cetuximab, clearly highlighting a beneficial, i.e., low hydrophobicity. While retention times were quite similar between sdAbs derived from the humanized *FERF* library design and WT VHHs (HIC retention times for all constructs ≤5.7 min), we noticed a trend toward slightly higher hydrophobic properties for sdAbs obtained from the humanized *VGLW* designs (with 5/6 humanized VHHs displaying retention times >6 min).

We also looked at the thermal stability of sdAbs obtained from both humanization approaches in comparison to the stability of WT VHHs. As read out we determined the Tonset that is defined as the lowest temperature at which a given protein starts to unfold. Overall, Tonsets varied between 51.6°C (*WT.cluster17.43*) and 59.5°C (*WT.cluster7.31*). Basically, differences in thermal stability between WT VHHs (mean Tonset for all *WT* VHHs = 57.2°C) as well as sdAbs retrieved from the different designed humanization libraries were negligible (mean Tonset *FERF* set of sdAbs = 57.9°C and mean Tonset *VGLW* = 57.8°C), again highlighting quite similar biophysical attributes of *de novo* humanized and sequence-optimized library-derived sdAbs to WT VHHs isolated from immunized llamas.

Aside from early developability assessment, we set out to investigate binding functionalities of engineered humanized sdAbs. Therefore, we determined affinities against (rh) NKp46 (Figure 4, Table 3) using biolayer interferometry (BLI). Interestingly, we observed substantial differences in binding behavior between sdAbs derived from the *FERF* library design and the designed *VGLW* library. Overall, affinities of *de novo* humanized *FERF*-based sdAbs were quite similar to the WT VHH counterparts. For instance, affinities for WT VHHs within cluster 1 varied from 1.5 nM to 45 nM, with a mean KD of 16.6 nM. Binding to (rh) NKp46 for *de novo* humanized sdAbs from the *FERF* design ranged from 4.1 nM to 19.8 nM, with an average affinity of 10.9 nM. In cluster 2, *WT* VHH *WT.cluster2.13* displayed an affinity of 7.4 nM, whereas affinities for *FERF*-derived clones varied between 2.9 nM and 12 nM (average KD = 6.4 nM). In cluster 5, *WT.cluster5.25* showed affinities for binding to (rh) NKp46 of 0.5 nM and *FERF.cluster5.26* and *FERF.cluster5.27* displayed affinities of 3.6 nM and 9.3 nM, respectively. Essentially, similar to anti-NKp46 VHHs obtained from WT library sorting, we were also able to generate *de novo* humanized sdAbs from the *FERF* library design with high-affinity binding to (rh) NKp46 in the sub-nanomolar range. In contrast to this, out of all six humanized sdAbs originating from the *VGLW* design that showed protein expression, only two clones displayed binding to (rh) NKp46. For both clones, binding in terms of capacities (i.e., interference pattern shifts in BLI) and affinities was vastly

diminished. In this respect, sdAb *VGLW.cluster16.42* displayed affinities for (rh) NKp46 in the high triple digit nanomolar range and while humanized VHH *VGLW.cluster2.14* bound to (rh) NKp46 with affinities of 7.9 nM, maximum binding capacities (i.e., interference pattern shift in BLI) were substantially reduced (Figure 4 4d).

De novo *humanized sdAbs facilitate the construction of NK cell engagers triggering NK cell mediated lysis of EGFR-overexpressing tumor cells*

Since it was shown previously that NKp46 can be efficiently harnessed for effector cell redirection,[41,43] we tested whether *de novo* humanized sdAbs represent versatile building blocks for the construction of NK cell engagers (NKCEs). Thus, we focused on WT VHHs and *de novo* humanized sdAbs belonging to identical sequence clusters. Because *de novo* humanized sdAb *VGLW.cluster.2.14* was the only clone derived from the *VGLW* humanization approach displaying (diminished) antigen binding and belonging to a sequence cluster that also comprised derivatives from both other approaches (*WT* and *FERF* design strategy), we nominated this single *VGLW*-derived sdAb in addition to WT VHHs and sdAbs from the *FERF* library design. At first, we wanted to get a deeper understanding about early developability by implementing two additional parameters, expression yields post protein A purification from antibody production in a volume of 25 mL and affinity-capture self-interaction nanoparticle spectroscopy (AC-SINS).[48,49] This technique allows the assessment of the self-interaction propensities of generated sdAbs. To this end, selected sdAbs were again expressed as one-armed, monovalent VHH SEEDbodies (Supplementary Figure S2A) using ExpiCHO™ cells as production host. Except for the nominated sdAb derived from the *VGLW* approach (*VGLW.cluster2.14*, expression yield = 98.4 mg/L), production yields after protein A chromatography for *de novo* humanized sdAbs were in the triple digit mg per L scale, generally indicating adequate expression profiles for transient antibody production (Table 4). In general, the production of *de novo* humanized sdAbs was fairly similar to expression yields of WT VHHs. Furthermore, aggregation propensities determined by analytical SEC unveiled favorable purities of more than 95% target species for all 10 sdAb-harboring monovalent SEEDbodies (Table 4). In addition to production characteristics, the antibody colloidal stability and the tendency to self-association and aggregation have been reported as key factors for selection of mAbs with favorable developability properties, which has been successfully correlated with the AC-SINS measure.[49,50] For this, sdAb SEEDbodies were captured onto particles by exploiting immobilized capture antibodies, and self-association was judged in phosphate-buffered saline (PBS) buffer at pH 7.4 by shifts in the plasmon wavelengths. Trastuzumab was used as control indicating favorable biophysical properties with mean plasmon wavelength shifts of approximately 0.2 nm after subtraction of buffer blanks. Final AC-SINS scores for sdAb SEEDbodies were calculated *via* subtraction of blank and trastuzumab scores. The calculated scores of all molecules including *de novo* humanized sdAbs were in the range of −0.46 and

Table 3. Analytical and early developability data of one-armed constructs organized by cluster and affinity, including SEC purity, HIC retention time, mean Tonset, and dissociation constant measured *via* BLI.

ID	SEC Purity [%]	HIC tR [min]	Mean Tonset [°C]	KD [nM]
WT.cluster1.1	96.3	5.8	58.8	1.5
FERF.cluster1.2	100	4.9	57.6	4.1
WT.cluster1.3	98.8	5.5	58.1	6.9
FERF.cluster1.4	98.8	5.5	58.4	7.7
FERF.cluster1.5	100	5.4	58.1	12.0
WT.cluster1.6	92.2	5.6	57.8	12.9
FERF.cluster1.7	99.0	5.3	57.2	19.8
WT.cluster1.8	97.9	5.0	57.9	45.0
VGLW.cluster1.9	96.7	6.6	57.4	non binding
FERF.cluster2.10	100	4.9	57.5	2.9
FERF.cluster2.11	94.5	4.8	57.7	4.4
FERF.cluster2.12	96.6	5.6	58.7	6.4
WT.cluster2.13	97.8	5.3	58.2	7.4
VGLW.cluster2.14	98.9	7.3	57.5	7.9
FERF.cluster2.15	99.2	4.8	58.0	12.0
FERF.cluster3.16	100	4.9	58.0	0.7
FERF.cluster3.17	100	4.8	58.1	2.2
FERF.cluster3.18	98.8	5.2	58.2	2.8
FERF.cluster4.21	98.4	4.9	57.5	5.0
FERF.cluster4.22	100	5.3	57.9	6.1
VGLW.cluster4.23	97.0	6.1	57.8	non binding
WT.cluster5.25	100	4.6	57.6	0.5
FERF.cluster5.26	96.2	4.9	57.4	3.6
FERF.cluster5.27	98.4	4.5	57.1	9.3
FERF.cluster6.28	96.9	5.3	58.6	3.4
FERF.cluster6.29	99.0	5.5	57.5	5.0
WT.cluster7.31	9.7	4.6	59.5	20.4
WT.cluster8.33	100	5.3	58.5	0.5
WT.cluster9.35	95.7	5.0	52.5	0.1
WT.cluster10.36	99.4	5.1	58.0	0.7
WT.cluster11.37	95.6	5.4	58.5	1.7
WT.cluster12.38	97.1	5.1	56.6	1.9
WT.cluster13.39	99.1	5.0	57.5	2.4
WT.cluster14.40	99.3	5.1	58.0	3.1
FERF.cluster15.41	95.6	5.2	58.4	2.2
VGLW.cluster16.42	96.8	7.2	58.8	722.8
WT.cluster17.43	97.8	5.3	51.6	non binding
WT.cluster18.44	100	5.5	57.9	3.8
VGLW.cluster20.46	84.2	5.5	57.6	non binding
VGLW.cluster21.47	92.9	6.6	53.0	non binding
WT.cluster22.48	100	5.1	56.3	non binding
WT.cluster27.50	100	4.6	56.0	non binding

0.06, indicating favorable developability properties very similar to those of trastuzumab (Table 4).

To investigate the capacities of *de novo* humanized sdAbs to elicit NK cell-mediated lysis of EGFR-overexpressing tumor cells *via* triggering NKp46 (Figure 5a), nominated sequences were produced as bispecific Fc effector silenced SEEDbodies harboring the Fab arm of a humanized version of cetuximab that was expressed on the GA chain (Supplementary Figure S2B). After expression in ExpiCHO™ cells, aggregation properties were assessed by analytical SEC. Besides, WT VHH-derived bispecific SEEDbody *WT.cluster5.25* (purity = 88.8%) SEC profiles of all other sdAb-based NKCEs were above 90% target monomer species, indicating adequate purities for all generated bispecifics. EGFR-strongly overexpressing tumor cell line A431 was exploited as target cells, whereas NK cells, freshly isolated from PBMCs of healthy human donors were used as effector cells (Figure 5b–d). As negative control we utilized a one-armed, monovalent Fc effector-silenced SEEDbody incorporating the Fab arm of a humanized version of cetuximab that was also used for the construction of sdAb-based bispecifics.

For NKCEs constructed from sdAbs belonging to the first sequence cluster, we have not been able to determine significant effector cell redirection resulting in tumor cell lysis appreciably higher than the negative control at concentrations of up to 10 nM. At higher, unphysiological concentrations, marginal killing was observed. This is in line with previous observations that not every paratope specific to a natural cytotoxicity receptor is capable of triggering NK cell activation upon reformatting into a NK cell engager architecture.[9,51] In this respect, capacities in terms of NK cell redirection were comparable between NKCEs constructed from *de novo* humanized sdAbs of the *FERF* design and the bsAb harboring a WT representative of this particular sequence cluster. The group of NKCEs harboring sdAbs belonging to sequence cluster 2 comprised members of all three different approaches, a WT representative (*WT.cluster2.13*), one of the few clones derived from the *VGLW* library strategy that showed binding to (rh) NKp46 (*VGLW.cluster2.14*) as well as three candidates obtained from the *FERF de novo* humanization design (*FERF*.cluster*2.10*, *FERF*.cluster*2.15*, and *FERF*.cluster*2.11*). In accordance with diminished binding capacities to (rh) NKp46 in BLI, the
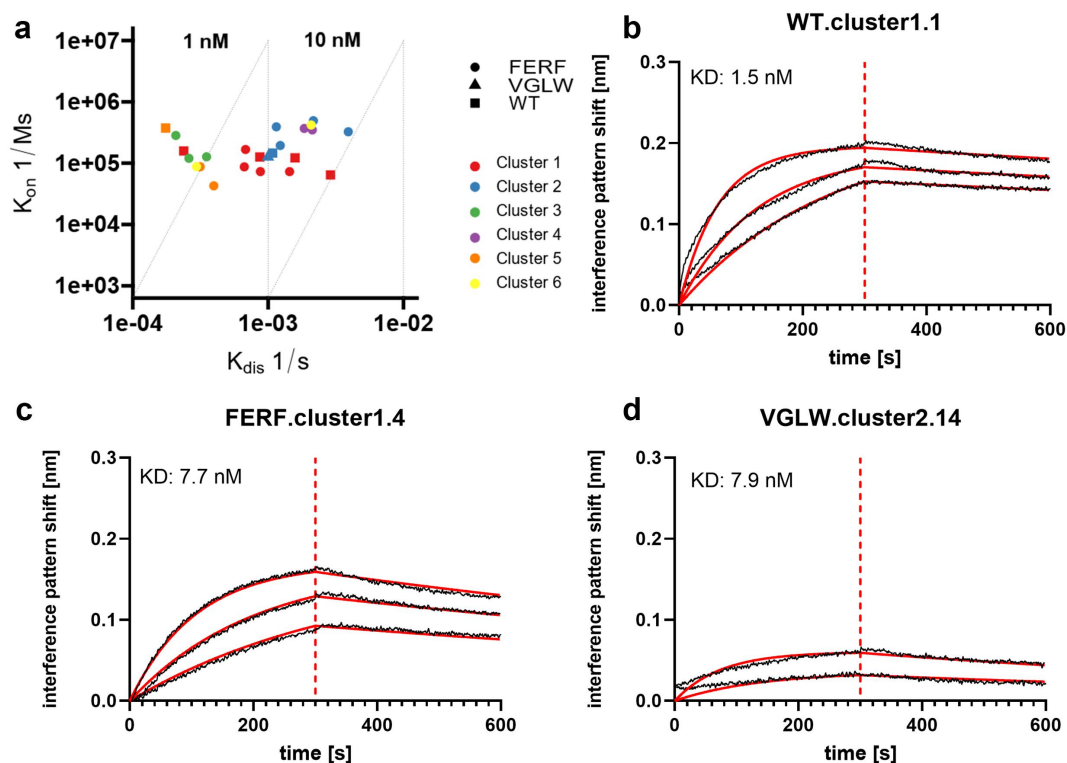
Figure 4. Binding capacities of humanized sdAbs and WT VHHs reformatted as one-armed SEEDbodies as determined by BLI. (a) Binding kinetics of reformatted sdAbs belonging to different CDR3 sequence clusters. sdAbs originating from humanized FERF library design are represented as dots, clones derived from humanized VGLW approach are shown as triangles, and the VHHs obtained from WT library are given as squares. (b)–(d) Affinity determination by BLI is exemplarily shown for individual clones derived from the different library approaches (WT, FERF, and VGLW). sdAb-derived SEEDbodies were loaded onto sensor tips. After sensor rinsing, antigen binding was conducted at different concentrations for 300 s, followed by a dissociation step in KB buffer for 300 s.

Table 4. Production yields and analytical purity of nominated sequences and their ability to function as NKCE with humanized cetuximab (ctx) on the GA chain, as indicated by the EC_{50} for killing of target cells. Evaluation of AC-SINC is included for the one-armed constructs.

one armed (hu)VHH-SEEDbodies				VHH x Ctx SEEDbodies	
ID	Yield o.a. [mg/L]	SEC Purity [%]	AC-SINS	SEC Purity [%]	EC_{50} [nM]
FERF.cluster1.4	193.2	100	0.1	93.0	no killing
FERF.cluster1.7	147.9	100	−0.2	95.7	no killing
WT.cluster1.3	122.3	100	−0.4	96.1	no killing
FERF.cluster2.15	183.2	100	0.0	97.6	0.05
WT.cluster2.13	146.8	100	−0.5	97.9	4.07
VGLW.cluster2.14	98.4	100	−0.3	97.1	no killing
FERF.cluster2.10	131.3	100	−0.5	9.6	0.06
FERF.cluster2.11	133.4	100	−0.4	94.6	0.05
WT.cluster5.25	193.0	100	−0.4	88.8	0.03
FERF.cluster5.26	221.1	97.6	−0.5	9.8	4.65

NKCE incorporating *VGLW* clone *VGLW.cluster2.14* provoked negligible NK cell-mediated killing of A431 cells. Only at very high concentrations (500 nM) moderate killing was observed. This is in contrast to NKCEs harboring either a WT VHH or sdAbs isolated from the *FERF* design strategy. NKCE based on WT representative *WT.cluster2.13* triggered significant target-dependent NK cell-mediated lysis of A431 cells with potencies in the single digit nanomolar range (EC_{50} killing = 4.07 nM, Table 4).

Surprisingly, all NKCEs incorporating representatives from the *FERF* approach elicited lysis of EGFR-overexpressing A431 cells with significantly enhanced potencies (EC_{50} killing ranging from 0.05 nM to 0.06 nM). This effect was not strictly dependent on binding affinities to (rh) NKp46, since within

this particular sequence cluster clone *FERF.cluster2.15* displayed affinities of 12 nM, whereas all other clones including the WT VHH *WT.cluster2.13* showed high-affinity binding to (rh) NKp46 in the single digit nanomolar range. However, it is important to note that all *FERF*-derived clones within this cluster were neither CDR3 nor CDR1–2 sequence identical with *WT.cluster2.13*, and it is appreciated that subtle changes in the sequence might have substantial consequences in terms of effector cell redirection[52,53]. Furthermore, besides residues within the different CDRs, framework residues might also contribute to antigen binding, and consequently to redirection capacities.[54]

Interestingly, for NKCEs based on clones from sequence cluster 5, we observed the opposite phenomenon. Here, *WT.*
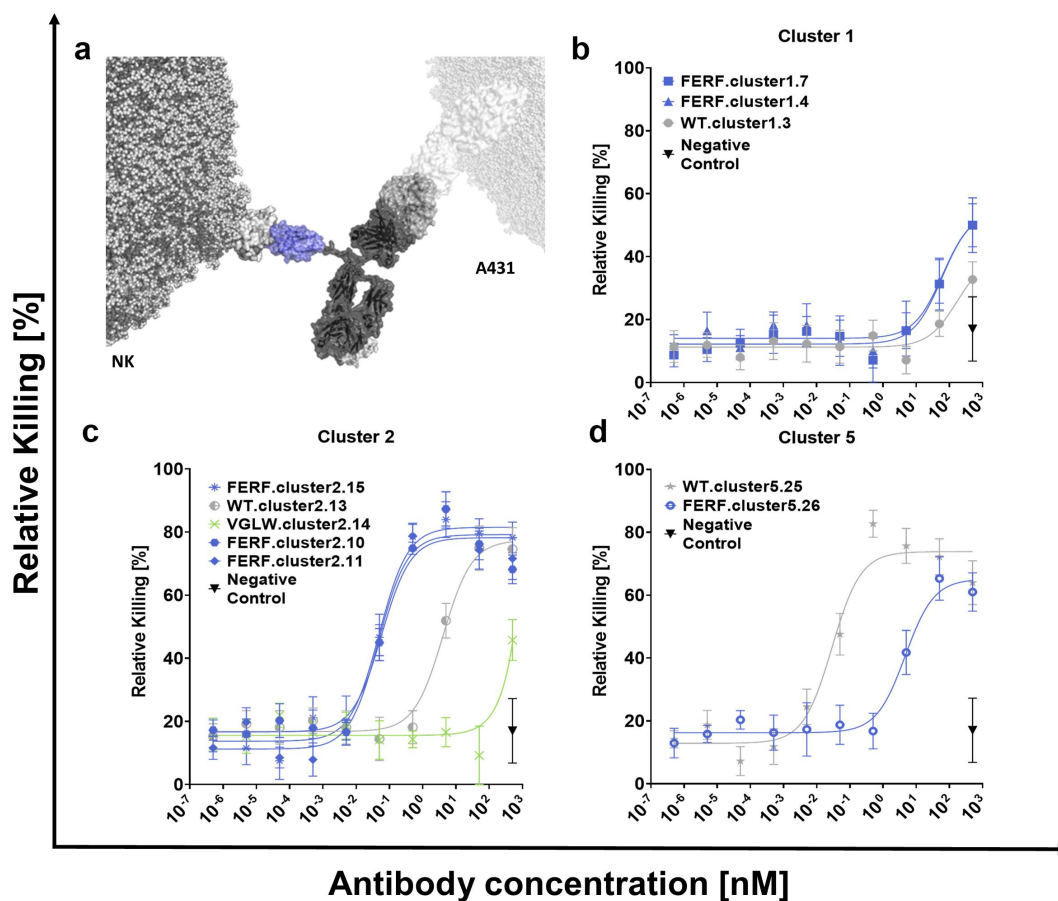
Figure 5. NK cell engagers (NKCEs) harboring *de novo* humanized sdAbs from the FERF design trigger significant NK cell mediated killing of EGFR overexpressing tumor cells. (a) Schematic depiction of NK cell redirection by exploiting an Fc-silenced SEEDbody harboring a *de novo* humanized sdAb targeting NKp46 on the NK cell and a Fab arm derived from cetuximab (ctx) for binding to EGFR on A431 cells. (b)–(d) Fluorescence-based NK cell killing assay of A431 target cells using freshly isolated NK cells from PBMCs of human healthy donors at an E:T ratio of 5:1. Respective Fc-silenced SEEDbodies were tested at increasing concentrations. Tumor cell lysis was normalized to Cetuximab at 50 nM. Killing capacities shown for NKCEs harboring different sdAbs derived from the FERF (blue) and VGLW (green) *de novo* library designs and obtained from the WT library (gray). An EGFR-targeting one-armed SEEDbody with an effector-negative Fc region was exploited as negative control. Mean values ± SEM of eight independent experiments with duplicates are shown.

cluster5.25 was significantly more potent compared with *FERF. cluster5.26* (*WT.cluster5.25* EC$_{50}$ killing = 0.03 nM *vs. FERF. cluster5.26* EC$_{50}$ killing = 4.65 nM; Table 4). Despite being CDR3 identical, *WT.cluster5.25* displayed an approximately sevenfold higher affinity for binding to (rh) NKp46 than *FERF.cluster5.26*, which might be attributed to differences in CDR1 and CDR2 (Table 1, Table 3). Overall, the data gives clear evidence that the herein presented *de novo* humanization approach enables the generation of humanized and sequence optimized sdAbs that are versatile for effector cell redirection. Of note, minor sequence variations introduced during this novel humanization approach might have major implications in terms of functionalities in one way or another.

Epitope binning of NKp46-specific sdAb-based NKCEs

To further characterize the humanized sdAbs and most importantly to investigate whether epitope drifts of humanized sdAbs explain differences in killing capacities, competition for binding to NKp46 was performed for selected pairs of SEEDbodies (Supplementary Figure S3). These included all sequences that had been evaluated for tumor cell lysis and an additional set of WT and humanized sequences from different

CDR3 clusters and several singletons. BLI experiments were conducted, where the first antibody sample was immobilized on the sensor tip followed by quenching of the sensor tip with unrelated IgG. Subsequently, an association of (rh) NKp46 was conducted, followed by a second association step exploiting a second NKp46 SEEDbody. The results revealed that from the set of 29 tested sequences, 28 bind to the same epitope. Only one singleton (WT.cluster 9.35) binds to a distinct epitope. This observation is in agreement with results from our previous study,[43] where the major fraction of VHHs that had been obtained from immunization of three camelids were bound to one single epitope. Hence, it is tempting to speculate that there is an immune-dominant epitope on the surface of human NKp46.

Discussion

In this work, we generated a novel one-step *de novo* high-throughput approach for the generation of humanized and sequence-optimized sdAbs from immunized llamas. Camelid-derived sdAbs exhibit many desirable properties in terms of biomedical applications, such as a small size that might be beneficial for tissue penetration,[10] adequate stability,

and simple architecture that enables facile reformatting into a plethora of different bi- and multispecific antibody formats.[13–16,55] Notwithstanding, the non-human origin of camelid VHHs might pose a risk of immunogenicity when injected into patients. Hence, VHHs are typically humanized for therapeutic purposes. Although, no strict correlation between the degree of human-likeness of conventional IgG-like antibodies and the incidence of ADA has been reported, a trend toward less immunogenicity when molecules have higher human-likeness has been noted.[56] Moreover, the percentage of human-likeness of marketed therapeutic antibodies significantly increased in the past, which correlated quite well with the observation that in recent years, antibody therapeutics with lower incidences of ADAs were approved. Investigations of humanized camelid-derived sdAbs have revealed minimal immunogenicity in general, similar to fully human IgGs.[24]

As discussed above, humanization and sequence optimization of VHHs can be tedious and cumbersome, requiring the design, production, and thorough characterization of multiple variants[24–26]. This procedure often requires multiple optimization cycles, and, under certain circumstances, it might not be possible to sufficiently humanize the sequence due to incompatibilities of human germline residues with regions that are structurally important for the bioactive VHH conformation. Our approach, composed of two main steps, library construction following camelid immunization and yeast surface display-based (humanized and sequence optimized) sdAb discovery, is characterized by its low complexity and speed similar to the generation of WT VHHs following immunization.[6,45,57]

The procedure itself, grafting a PCR-amplified and immunized CDR3 repertoire onto a moderately diversified, humanized, and sequence optimized scaffold library, was fairly similar to a method previously described by our group for the isolation of ultralong CDR-H3 antibodies from immunized cattle.[58] In that study, after the immunization of cattle, we specifically amplified ultralong CDR-H3 regions and grafted those paratopes onto a fixed Fab scaffold that was exploited for yeast surface display. Of note, in the herein presented approach, we introduced a slightly diversified CDR1 and CDR2 repertoire into the scaffold library that was inspired by the natural amino acid composition found in NGS datasets of naïve and of immunized llamas as well as in NGS datasets of human antibody VH repertoires. Due to limitations in library sizes for yeast surface display (i.e., 10^8 – 10^9 unique clones[30] and subsequent FACS-based hit discovery, we restricted the diversities within the humanized and sequence-optimized scaffold libraries (in regions CDR1 and CDR2) to approximately 10^4 different sequence variations. This does not represent the entire diversity that can theoretically be generated by the immune system of a camelid. Consequently, several camelid-derived CDR3 sequences that show high binding affinity in a WT VHH environment may not be structurally compatible with the herein generated humanized sdAb scaffolds. Since such WT VHHs could pose considerable difficulties for classical humanization and multi-parameter optimization, it might be considered a benefit that in our library approach combined with FACS-based sdAb discovery, only CDR3 regions that are compatible with

the internally humanized and sequence-optimized sdAb scaffold library are identified. This increases the likelihood of obtaining potent and readily humanized sdAbs directly from a screen that might need further *in silico* assessment or sequence optimization only of the CDR3 region.

The generated humanized and sequence optimized *FERF* library resulted in the facile discovery of various sdAbs displaying high affinity binding to (rh) NKp46 and efficient NK cell redirection. Intrigued by the fact that within the HcAb repertoire of naïve llamas we found 7.2% of sequences comprising the 37 V, 44 G, 45 L and 47W (*VGLW*) hallmark motif, we also designed a corresponding humanized and sequence-optimized sdAb library. Importantly, sdAbs obtained from this strategy displayed either no or only significantly compromised binding to (rh) NKp46. Furthermore, redirection capacities of NK cells were also negligible. This is in line with NGS datasets and YSD-based WT library sampling, which both suppose that from the (rh) NKp46-immunized llama primarily VHHs were enriched comprising the *FERF* hallmark motif (37F, 44E, 45 R, and 47F). These findings support the notion that antigen-specific CDR3 sequences obtained from a *FERF* hallmark environment might not necessarily be sufficiently compatible within a *VGLW* background and *vice versa*.

In this study, we generated two separate libraries to account for only two specific hallmark signatures (*FERF* and *VGLW*). As evident from Supplementary Table S2, these signatures cover only 35.5% of hallmark motifs observed in the llama wild type VHH repertoire. Since the hallmark residues are known to be important for the bioactive conformation of CDR3 and often mediate interactions with antigens,[54] this might explain why we observed a stronger bias for a given sequence cluster after sorting in the humanized and sequence optimized libraries. Furthermore, we noticed that 15.9% of VHH sequences in the NGS dataset of the llama wild type VHH repertoire carry a non-canonical disulfide bond between cysteines in CDR3 and the last position of framework region 2. Consequently, by taking into consideration the different hallmark motifs and non-canonical disulfide bonds, follow-up libraries might be constructed that enable the *de novo* generation of humanized and sequence optimized sdAbs displaying an even broader diversity output.

Another angle of optimization might rely on a more adequate exploitation of NGS for sdAb discovery. Within the present investigations, NGS was primarily used to gather information about diversities, amino acid distributions of CDRs and hallmark motifs and to scrutinize CDR cluster frequencies after YSD-based sdAb discovery. Several studies have demonstrated that NGS analysis of sequence pools before and after library selection can offer new ways of finding rare but highly enriched potent binders.[59–62] Moreover, such sequence pools can be used as input for artificial intelligence and machine learning approaches to design new sequences with even further improved potency or developability properties.[63–68]

In addition, for CDR1 and CDR2 diversification, we only considered amino acid compositions at specific positions as inclusion criterion that was further refined by removal of specific sequence liability motifs. This procedure, however, ignores how residue types interact to form stabilizing

interactions, potentially resulting in library members that fold and express poorly. This issue might be overcome using oligonucleotide array-based synthesis of sequences with CDR1-CDR2 combinations that are observed in NGS data of wild type or immunized subjects, as recently applied in a library approach described by Teixeira et al.[28] As a further advantage, this procedure significantly reduces the combinatorial diversity of artificial CDR1-CDR2 combinations and thereby leaves design space for other CDR1-CDR2 combinations that are observed *in vivo* in camelids and humans or to regions of the VHH framework region that are known from structural analysis to mediate interactions with antigens[54]. The design of a follow-up humanized scaffold library using specific sequences instead of a combinatorial approach would allow for consideration of a higher diversity of hallmark signatures as suggested above. Within the present libraries, we focused on two classical hallmark signatures (*FERF* and *VGLW*). Whereas the *FERF* motif is most predominant in llama WT sequences (28.2%; see Supplementary Table S2), the *VGLW* motif that is also present in both different sdAb-based paratopes of ozoralizumab[20] is observed at frequencies of only 7.2%. A follow-up humanized library might include further specific hallmark motifs mirroring their prevalence in camelid NGS data.

An aspect that might need further investigation is the question of how far the observed preservation of CDR3 diversity between WT and humanized library also translates to the preservation of epitope diversity. In this and our previous study on NKp46,[43] the epitope diversity observed from WT sequences is rather low, possibly due to the existence of an immune-dominant epitope on the surface of human NKp46.

In conclusion, we presented a novel approach that combines the generation of antigen-specific CDR3s from camelid immunization with a synthetic VHH scaffold library that has been optimized toward human-likeness and developability. This approach yields high-affinity binders with good developability properties from initial selections, substantially reducing or even eliminating the need for further sequence optimization. It is generally applicable for any antigen upon camelid immunization and has the potential to significantly accelerate candidate selection and reduce risks and attrition rates in sdAb development.

Materials and methods

Generation of the synthetic VHH scaffold libraries

For the design of the randomized scaffold libraries, NGS data sets of naïve human and llama V sequences and data sets from immunized camelids were used. In more detail, only amino acids with ≥4% frequency at each specific position were considered. Residues that would result in susceptible chemical liability motifs, N-glycosylation motifs, or strong predicted MHC-II binding peptide motifs, were excluded. Diversity with respect to the amino acids charge, size, and hydrophobicity and the occurrence in the equivalent positions in human antibody repertoires were taken into account. This resulted in a theoretical backbone library space of 1.04×10^4 variants (in CDR1 and CDR2). These randomized CDR sequences were

embedded in a synthetically synthesized human IGHV3–23 × 1 germline framework sequence. Moreover, two different FR2 designs were generated, harboring either a VHH dedicated AA motif (37F, 44E, 45 R, and 47F, referred to as FERF library) or the classical VH signature (V37, G44, L45, and W47, termed VGLW library). Both *in silico* designed scaffold libraries were generated at Geneart (Thermo Fisher Scientific) using TRIM technology. To enable subsequent camelid CDR3 engraftment in the randomized scaffold libraries, sequences were designed to harbor a stuffer sequence at CDR3 position containing multiple stop codons in every reading frame and BsaI recognition sites, allowing for stuffer digestion and replacement by camelid CDR3. The gene strings were also designed to already contain *N*- and *C*-terminal nucleotide additions enabling a seamless incorporation of the framework libraries into the pYD-derived display plasmid backbone (pDest) via yeast-mediated homologous recombination cloning as described elsewhere.[45]

Llama immunization

All procedures involving animals were conducted at preclinics GmbH, Germany. Procedures and animal care were in accordance with local regulations and animal welfare protection laws. Of note, animals remained alive after immunization process and final blood collection. For the immunization, one llama (*Lama glama*) was vaccinated with recombinant human (rh) NKp46 extracellular domain (ECD; Acro). Four antigen administrations of 300 μg rh NKp46 ECD, each conducted as subcutaneous injections at three sites, were performed over a period of 42 days in total (at d0, d14, d28, and d35). For this, the antigen was diluted to a stock concentration of 1 mg/mL in PBS and emulsified for initial immunization with Complete Freund's Adjuvant, or with Incomplete Freund's Adjuvant for subsequent immunizations. Seven days after final administration (d42), a volume of 100 mL blood was collected followed by RNA extraction.

Yeast display library generation

cDNA was generated by reverse transcription using the RNA material derived from immunized camelid as template. SuperScript III First-Strand Synthesis Kit (Invitrogen) was used according to the manufacturer's instructions. A total of 10 reactions were performed with 8 μl of template RNA in each reaction. Next, PCR was exploited to specifically amplify complete VHH DNA (for WT library) or CDR3 regions only (for FERF and VGLW libraries) using 2×Q5-Polymerase Mastermix (New England Biolabs) and 0.5 μM of the respective primers per reaction. The primers were designed to insert homologous nucleotides matching the sequence in pDest (Supplementary Table S3), hence allowing for subsequent gap repair cloning in yeast. The PCR was conducted under the following conditions: initial denaturation at 98°C for 30 s, followed by 35 cycles of denaturation (98°C, 10 s) and combined annealing and elongation (72°C, 60 s) steps, finalized by an extension step at 72°C for 5 min. A total of 96×50 μl reactions were conducted in parallel per library to yield adequate amount of PCR amplicons. The resulting PCR product

was purified using the Wizard SV Gel and PCR Clean-Up System (Promega) according to the manufacturer's instructions. In parallel, pDest containing the VHH scaffold libraries (for FERF and VGLW libraries) and pDest without inserted scaffold (for WT library) were linearized with specific restriction enzyme *Bsa*I (New England Biolabs). A total of 200 U *Bsa*I was used to digest 150 μg pDest in CutSmart buffer (New England Biolabs) in a total volume of 300 μl. The reaction was carried out at room temperature overnight prior to purification via the Wizard SV Gel and PCR Clean-Up System (Promega). For the final assembly of the display vector, gap repair cloning in EBY100 *Saccharomyces cerevisiae* cells, adapted from Bernatuil and coworkers;[44] was exploited and explained in detail elsewhere.[45] In brief, digestion of the stuffer sequence enables the genetic fusion of library candidates in frame to Aga2p, resulting in the presentation of (humanized) sdAb variants on the yeast cell surface. An additional insertion of a HA tag *C*-terminally linked to Aga2p on the plasmid backbone allows for the detection of proper full-length sdAb presentation on the yeast surface.

Library sorting

For FACS purposes, the transformed cells were cultivated in an SD medium with dropout mix lacking tryptophan (−Trp) at 30°C and 120 rpm overnight. sdAb expression was induced subsequently by cell transfer (10^7 cells/ml) to SG-Trp medium and incubation at 20°C and 120 rpm for 48 h. Full-length sdAb surface expression was monitored via application of an anti-HA mouse antibody (Alexa Fluor 488, R&D systems, catalog number IC6875G, diluted 1:20), while antigen binding was detected by indirect immunofluorescence using 1 μM rh his-tagged NKp46 ECD (Acro Biosystems, catalog number NC1-H52H6) and an anti-his mouse monoclonal detection antibody (Alexa Fluor 647, Qiagen, catalog number 35,370, or Allophycocyanin, BioLegend, catalog number J095G46, both diluted 1:20), thereby allowing for a two-dimensional sorting strategy. For the FACS procedure, a BD FACSAria™ (BD Biosciences) device was used for the initial sorting rounds, while subsequent cell selection was performed on a Sony LE-SH800 sorter. For optimal sorting gate adjustments, control samples, i.e., cells incubated with secondary labeling reagents, and an unrelated antigen, as well as cells incubated with only secondary labeling reagents, were used in every experiment. Sorting gates were designed to maximize the enrichment of displaying cells while minimizing unspecific binding (Figure 3a).

NGS sequencing

To prepare RNA material for NGS analysis, two defined antisense primer sequences were used that specifically aligned with nucleotides in the upper hinge regions of camelid IgG2 and IgG3 antibody isotypes, facilitating directed cDNA synthesis. Within a subsequent PCR utilizing index primers for Illumina sequencing, the sdAb sequences were amplified and tagged. For the samples derived from the sdAbs diversities embedded in the plasmid vector system, the sequences are processed accordingly, but lacking the cDNA synthesis step. During the DNA amplification process, the AMPure system (Beckman Coulter) was used to purify the sdAb amplicons, while for the purification of the final sequencing library a Pippin Prep (Sage Science) was used. For sequencing purposes, a MiSeq (Illumina) device with the v3 600 cycle kit according to the manufacturer's protocol was used. Resulting FASTQ files were uploaded to Geneious Biologics (www.geneious.com/biopharma) for analysis. Reads were overlapped, filtered for length, and the VHH sequences were annotated using the *Lama glama* reference library. Normalized counts for each CDR3 were used to identify sequences that were enriched in the sorted samples relative to the baseline diversity.

Protein expression and analysis

Selected sdAb variants were cloned into pTT5 mammalian expression vector (Thermo Fisher Scientific) as *N*-terminal fusion to the hinge region of Fc immune effector-silenced (eff-) SEED AG chains, consequently enabling the production of either eff-one-armed (oa) SEEDbodies using a paratope-less SEED-GA-chain or the production of bispecific SEEDbodies in combination with eff- humanized cetuximab Fab fused to the SEED GA chain.

All proteins were expressed in the ExpiCHO™ Expression System (Thermo Fisher Scientific) either in 5 or 25 ml scale according to the manufacturer's manual standard protocol with 2:1 ratio of AG to GA chain or a 2:2:1 ratio of light chain to AG to GA chain. After 7 days, the protein containing supernatants were purified with MabSelect™ antibody purification chromatography resin (Cytiva). After sterile filtration, protein concentrations were determined by A_{280} absorption measurement. For the assessment of protein sample quality regarding monomer content [%], analytical SEC was applied using 7.5 μg protein per sample on a TSKgel UP-SW3000 column (2 μm, 4.6 × 300 mm, Tosoh Bioscience) on an Agilent HPLC 1260 Infinity system with a flow rate of 0.35 ml/min using 50 mM sodium phosphate, 0.4 M NaClO4 pH 6.3 as mobile phase. Signals were recorded at 214 nm. Moreover, hydrophobicity of the different molecules was determined by HIC using 20 μg protein per sample on a TSKgel Butyl-NPR column (2.5 μm, 4.6 × 100 mm, Tosoh Bioscience) on an Agilent HPLC 1260 Infinity system with a flow rate of 0.5 ml/min. Samples were premixed with 50% (v/v) 2 M ammonium sulfate solution prior to injection. A gradient running from mobile phase A (1.2 M ammonium sulfate in PBS) to mobile phase B (50% methanol in 0.1× PBS) over 15 min at 25°C was applied. Signals were recorded at 214 nm. Anti-PD-L1 avelumab and anti-EGFR cetuximab (both produced in house) were used as reference molecules. To assess thermal unfolding properties of the antibodies, differential scanning fluorimetry (DSF) on a Prometheus NT.PLEX nanoDSF instrument (NanoTemper) was utilized. Samples were measured in duplicates using nanoDSF Standard Capillary Chips. A temperature gradient from 20°C to 95°C at a slope of 1°C/min was used while recording fluorescence at 350 and 330 nm. Unfolding transition midpoints (Tm) and Tonset values were determined from the melting curves or from the first derivative of the fluorescence ratio 350 nm/330 nm.

Biolayer interferometry

Biophysical properties of the evaluated binders were assessed using an Octet Red BLI system (Sartorius). Binding experiments were conducted in KB buffer (PBS pH 7.4, 0.1% BSA, 0.02% Tween-20) using Anti-Human Fc Capture (AHC) biosensors. Biosensors were loaded with the one-armed antibody samples at a concentration of 3 µg/ml for 180 s. The samples were subjected to a twofold serial dilution of (rh) NKp46 (Acro Biosystems, catalog number NC1-H52H6), starting at a concentration of 100 nM using a measurement window of 300 s for association and dissociation each.

The data were aligned to the association step, and inter-step correction was applied at the dissociation step, as well as noise reduction using Savitzky-Golay filtering. The resulting data were analyzed using a 1:1 binding model.

Epitope binning

The BLI technology (Octet RH16 BLI, Sartorius) was used for epitope binning. Samples were loaded (200 s) on Anti-Human IgG Capture (AHC) biosensors in PBS. Subsequently, AHC sensor was quenched (300 s) using an unrelated IgG (20 µg/ml) in PBS. The sensor tip was then immersed into a solution of 100 nM recombinant human NKp46 spiked with the unrelated IgG (20 µg/ml) in KB-Buffer for the initial association phase (800 s).

Following NKp46 association, the biosensor was transferred into a well containing the second antibody sample and measured for 800 s. This sample was spiked with 100 nM NKp46 and unrelated IgG (20 µg/ml) to mitigate shifts arising from dissociation. An interference shift surpassing 0.1 nm beyond the level established by NKp46 association, subsequent to subtraction of the baseline, was considered indicative of binding to a different epitope than the first antibody.

AC-SINS

Molecules were captured onto particles via immobilized capture antibodies, and self-association was judged in PBS buffer at pH 7.4 by shifts in the plasmon wavelengths.[50] Trastuzumab (produced in house) was used as control indicating favorable biophysical properties with mean Δλmax values of ~0.2 nm after subtraction of buffer blanks. Final AC-SINS scores for molecules were calculated via subtraction of blank and trastuzumab scores.

Killing assay

The ability to engage NK cells and mediate lysis of tumor cells was assessed using a fluorescence-based A431-NK-cell killing assay. To this end, human PBMCs were isolated from whole blood of healthy donors (donors provided informed consent) via SepMate tubes (Stemcell) and Lymphoprep density gradient media (Stemcell) according to the supplier's manual prior subsequent automated negative-NK cell isolation (Stemcell). The isolated NK cells were rested in AIMV (Thermo Fisher Scientific) medium supplemented with low dose rh IL-2 (100 U/ml; R&D) overnight at 37°C and 5% CO_2. EGFR-positive A431 were cultivated in DMEM medium supplemented with 10% (v/v) heat inactivated fetal bovine serum, while EGFR-negative ExpiCHO^{TM} (Thermo Fisher Scientific) cells were cultivated using ExpiCHO^{TM} (Thermo Fisher Scientific) according to the manufacturer's protocol. For the assay, target cells were labeled with Cell Tracker DeepRed dye (Thermo Fisher Scientific) prior adjustment to 0.125×10^6 viable cells/ml in corresponding media and transfer of 2'500 viable cells per well into a Bio-One µClear 384-well plate (Greiner). After cell adherence, NK cells were added to an effector-to-target (E:T) ratio of 5:1, followed by 5 µl sample addition, pre-diluted in AIMV media. For EC_{50} evaluation, a 10-fold serial dilution, starting at 500 nM, was used. Controls included NK cells incubated with target cells only (basal killing), one-armed cetuximab eff- (negative control) and cetuximab (positive control). Addition of 0.03 µM SYTOX-green (Invitrogen) enabled the measurement of dead cells. Red and green fluorescence signals were recorded for 12 h using an Incucyte Live-Cell Analysis Instrument (Sartorius). Determination of the red and green fluorescence overlay signals consequently allowed for dead target cell-specific analysis.

In silico developability assessment

The in silico developability profile was computed using our internal pipeline termed "Sequence Assessment Using Multiple Optimization Parameters (SUMO)."[32] Briefly, this approach automatically generates VHH models based on the provided input sequences, calculates human-likeness to the most similar human germline sequence, and reports the following sequence motifs as potential liability sites: DG, DD, DS (isomerization), NG, ND, NN, NT, NA, NS (deamidation), M (oxidation), C (unpaired cysteines), and N-X-S/T (glycosylation) if these are >20% solvent exposed according to the structural model. Moreover, the pI of the variable domain, Schrodingers AggScore as predictor for hydrophobicity and aggregation tendency and the calculated positive patch energy of the CDRs are determined. These scores are complemented with a green to yellow to red color coding, indicating scores within one standard deviation from the mean over a benchmarking dataset of biotherapeutics approved for human application as green scores above one standard deviation as yellow and those above two standard deviations as red. For the AggScore values, these cutoffs were slightly adjusted based on correlation analyses to internal experimental HIC data.

MHC-II binding predictions

To eliminate residues from CDR1 and CDR2 library design that would result in strong predicted MHC-II binding peptides, all enumerated variants plus their N- and C-terminally flanked framework residues were scored against a reference set of MHC class II alleles using the IEDB software (version 3.1.6; http://tools.iedb.org/mhcii/) using the following settings: IEDB-recommended Prediction Method; Selected species/locus: Human, HLADR; the IEDB-recommended reference panel of 27 alleles; peptide length: 15.

Abbreviations

AC-SINCS	affinity-capture self-interaction nanoparticle spectroscopy
ADA	anti-drug antibody
BLI	biolayer interferometry
CDR	complementarity-determining region
DSF	differential scanning fluorimetry
FACS	fluorescence-activated cell sorting
FR	framework region
HcAb	heavy chain-only antibody
HIC	hydrophobic interaction chromatography
MHC	mahor histocompatibility complex
NGS	next-generation sequencing
NK cell	natural killer cell
NKCE	NK cell engager
oa	one-armed
PBMC	peripheral blood mononuclear cell
PBS	phosphate-buffered saline
rh	recombinant human
sdAb	single domain antibody
SEC	size exclusion chromatography
SEED	strand-exchanged engineered domain
VHH	variable domain of the heavy chain of a heavy chain-only antibody
WT	wild-type
YSD	yeast surface display

Disclosure statement

PA, HY, LP, SK, DK, VS, CS, TC, AE, and SZ are or were employees at Merck Healthcare KGaA or EMD Serono. Besides, this work was conducted in the absence of any further commercial interest.

ORCID

Stefan Zielonka 🔵 http://orcid.org/0000-0002-4649-2843

References

1. Kaplon H, Crescioli S, Chenoweth A, Visweswaraiah J, Reichert JM. Antibodies to watch in 2023. MAbs. 2023;15(1):2153410. [accessed 2023 Mar 10]. doi:10.1080/19420862.2022.2153410.
2. Mullard A. 2022 FDA approvals. Nat Rev Drug Discovery [Internet]. 2023 [accessed 2023 Jan 10]. https://www.nature.com/articles/d41573-023-00001-3.
3. Walsh G, Walsh E. Biopharmaceutical benchmarks 2022. Nat Biotechnol. 2022;40(12):1722–60. doi:10.1038/s41587-022-01582-x.
4. Conrath KE, Wernery U, Muyldermans S, Nguyen VK. Emergence and evolution of functional heavy-chain antibodies in Camelidae.

Dev Comp Immunol. 2003;27(2):87–103. doi:10.1016/S0145-305X(02)00071-X.
5. Hamers-Casterman C, Atarhouch T, Muyldermans S, Robinson G, Hammers C, Songa EB, Bendahman N, Hammers R. Naturally occurring antibodies devoid of light chains. Nature. 1993;363 (6428):446–48. doi:10.1038/363446a0.
6. Sellmann C, Pekar L, Bauer C, Ciesielski E, Krah S, Becker S, Toleikis L, Kügler J, Frenzel A, Valldorf B, et al., A one-step process for the construction of phage display scFv and VHH libraries. Mol Biotechnol. 2020;62(4):228–39. [accessed 2020 Jan 28]. doi:10.1007/s12033-020-00236-0.
7. Roth L, Krah S, Klemm J, Günther R, Toleikis L, Busch M, Becker S, Zielonka S. Isolation of antigen-specific VHH single-domain antibodies by combining animal immunization with yeast surface display. Methods Mol Biol. 2020;2070:173–89.
8. Pardon E, Laeremans T, Triest S, Rasmussen SGF, Wohlkönig A, Ruf A, Muyldermans S, Hol WGJ, Kobilka BK, Steyaert J. A general protocol for the generation of nanobodies for structural biology. Nat Protoc. 2014;9(3):674–93. doi:10.1038/nprot.2014.039.
9. Klausz K, Pekar L, Boje AS, Gehlert CL, Krohn S, Gupta T, Xiao Y, Krah S, Zaynagetdinov R, Lipinski B, et al. Multifunctional NK cell–engaging antibodies targeting EGFR and NKp30 elicit efficient tumor cell killing and proinflammatory cytokine release. J Immunol. 2022;209(9):1724–35. doi:10.4049/jimmunol.2100970.
10. Li Z, Krippendorff B-F, Sharma S, Walz AC, Lavé T, Shah DK. Influence of molecular size on tissue distribution of antibody fragments. MAbs. 2016;8(1):113–19. doi:10.1080/19420862.2015.1111497.
11. Könning D, Zielonka S, Grzeschik J, Empting M, Valldorf B, Krah S, Schröter C, Sellmann C, Hock B, Kolmar H. Camelid and shark single domain antibodies: structural features and therapeutic potential. Curr Opin Struct Biol. 2017;45:10–16. doi:10.1016/j.sbi.2016.10.019.
12. Krah S, Schröter C, Zielonka S, Empting M, Valldorf B, Kolmar H. Single-domain antibodies for biomedical applications. Immunopharmacol Immunotoxicol. 2016;38(1):21–28. doi:10.3109/08923973.2015.1102934.
13. Pekar L, Busch M, Valldorf B, Hinz SC, Toleikis L, Krah S, Zielonka S. Biophysical and biochemical characterization of a VHH-based IgG-like bi- and trispecific antibody platform. MAbs. 2020;12(1):1812210. doi:10.1080/19420862.2020.1812210.
14. Yanakieva D, Pekar L, Evers A, Fleischer M, Keller S, Mueller-Pompalla D, Toleikis L, Kolmar H, Zielonka S, Krah S. Beyond bispecificity: controlled Fab arm exchange for the generation of antibodies with multiple specificities. MAbs. 2022;14(1):2018960. [accessed 2022 Jul 1]. doi:10.1080/19420862.2021.2018960.
15. Chanier T, Chames P. Nanobody engineering: toward next generation immunotherapies and immunoimaging of cancer. Antibodies. 2019;8(1):13. doi:10.3390/antib8010013.
16. Bannas P, Hambach J, Koch-Nolte F. Nanobodies and nanobody-based human heavy chain antibodies as antitumor therapeutics. Front Immunol [Internet]. 2017;8:1603. [accessed 2019 Sep 24]. http://journal.frontiersin.org/article/10.3389/fimmu.2017.01603/full.10.3389/fimmu.2017.01603.
17. Jovčevska I, Muyldermans S. The therapeutic potential of nanobodies. BioDrugs [Internet]. 2019 [accessed 2019 Dec 3]. http://link.springer.com/10.1007/s40259-019-00392-z.
18. Duggan S. Caplacizumab: first global approval. Drugs. 2018;78 (15):1639–42. doi:10.1007/s40265-018-0989-0.
19. Markham A. Envafolimab: first approval. Drugs. 2022;82(2):235–40. doi:10.1007/s40265-022-01671-w.
20. Keam SJ. Ozoralizumab: first approval. Drugs. 2023;83(1):87–92. doi:10.1007/s40265-022-01821-0.
21. Lyu X, Zhao Q, Hui J, Wang T, Lin M, Wang K, Zhang J, Shentu J, Dalby PA, Zhang H, et al. The global landscape of approved antibody therapies. Antib Ther. 2022;5(4):233–57. doi:10.1093/abt/tbac021.
22. Murakami T, Kumachi S, Matsunaga Y, Sato M, Wakabayashi-Nakao K, Masaki H, Yonehara R, Motohashi M, Nemoto N,

Tsuchiya M. Construction of a humanized artificial VHH library reproducing structural features of camelid VHHs for therapeutics. Antibodies. 2022;11(1):10. doi:10.3390/antib11010010.

23. Moutel S, Bery N, Bernard V, Keller L, Lemesre E, de Marco A, Ligat L, Rain J-C, Favre G, Olichon A, et al. NaLi-H1: a universal synthetic library of humanized nanobodies providing highly functional antibodies and intrabodies. Elife [Internet]. 2016;5:e16228. [accessed 2023 Jan 12]. https://elifesciences.org/articles/16228.

24. Rossotti MA, Bélanger K, Henry KA, Tanha J. Immunogenicity and humanization of single□domain antibodies. FEBS J. 2022;289 (14):4304–27. doi:10.1111/febs.15809.

25. Vincke C, Loris R, Saerens D, Martinez-Rodriguez S, Muyldermans S, Conrath K. General strategy to humanize a camelid single-domain antibody and identification of a universal humanized nanobody scaffold. J Biol Chem. 2009;284(5):3273–84. doi:10.1074/jbc.M806889200.

26. Sulea T. Humanization of camelid single-domain antibodies [Internet]. In: Hussack G Henry K, editors. Single-domain antibodies. New York, NY: Springer US; 2022. pp. 299–312. [accessed 2023 Mar 10]. doi:10.1007/978-1-0716-2075-5_14.

27. Rabia LA, Desai AA, Jhajj HS, Tessier PM. Understanding and overcoming trade-offs between antibody affinity, specificity, stability and solubility. Biochem Eng J. 2018;137:365–74. doi:10.1016/j. bej.2018.06.003.

28. Teixeira AAR, Erasmus MF, D'Angelo S, Naranjo L, Ferrara F, Leal-Lopes C, Durrant O, Galmiche C, Morelli A, Scott-Tucker A, et al. Drug-like antibodies with high affinity, diversity and developability directly from next-generation antibody libraries. MAbs. 2021;13(1):1980942. [accessed 2023 May 1]. doi:10.1080/19420862.2021.1980942.

29. Jain T, Sun T, Durand S, Hall A, Houston NR, Nett JH, Sharkey B, Bobrowicz B, Caffry I, Yu Y, et al. Biophysical properties of the clinical-stage antibody landscape. Proc Natl Acad Sci U S A. 2017;114(5):944–49. doi:10.1073/pnas.1616408114.

30. Valldorf B, Hinz SC, Russo G, Pekar L, Mohr L, Klemm J, Doerner A, Krah S, Hust M, Zielonka S. Antibody display technologies: selecting the cream of the crop. Biol Chem. 2022;403(5–6):455–77. doi:10.1515/hsz-2020-0377.

31. Doerner A, Rhiel L, Zielonka S, Kolmar H. Therapeutic antibody engineering by high efficiency cell screening. FEBS Lett. 2014;588 (2):278–87. doi:10.1016/j.febslet.2013.11.025.

32. Evers A, Malhotra S, Bolick W-G, Najafian A, Borisovska M, Warszawski S, Fomekong Nanfack Y, Kuhn D, Rippmann F, Crespo A, et al. SUMO: in Silico sequence assessment using multiple optimization Parameters [Internet]. In: Zielonka S Krah S, editors. Genotype phenotype coupling: methods and protocols. New York, NY: Springer US; 2023. pp. 383–98. [accessed 2023 Jul 6]. doi:10.1007/978-1-0716-3279-6_22.

33. Lin J, Lee SL, Russell AM, Huang RF, Batt MA, Chang SS, Ferrante A, Verdino P, Henry KA. A structure-based engineering approach to abrogate pre-existing antibody binding to biotherapeutics. PloS One. 2021;16(7):e0254944. doi:10.1371/journal.pone.0254944.

34. Johansson MU, Weinert C, Reichardt DA, Mahler D, Diem D, Hess C, Feusi D, Carnal S, Tietz J, Giezendanner N, et al. Design of antibody variable fragments with reduced reactivity to preexisting anti-drug antibodies. MAbs. 2023;15(1):2215887. doi:10.1080/19420862.2023.2215887.

35. Soler MA, Medagli B, Wang J, Oloketuyi S, Bajc G, Huang H, Fortuna S, de Marco A. Effect of humanizing mutations on the stability of the llama single-domain variable region. Biomolecules. 2021;11(2):163. doi:10.3390/biom11020163.

36. Nguyen VK, Desmyter A, Muyldermans S. Functional heavy-chain antibodies in camelidae [Internet]. In: Advances in immunology. Elsevier; 2001. pp. 261–96. [accessed 2023 Jan 19]. https://linkinghub.elsevier.com/retrieve/pii/S0065277601790062.

37. 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(10):5696–704. doi:10.4049/jimmunol.0903722.

38. Lu X, Nobrega RP, Lynaugh H, Jain T, Barlow K, Boland T, Sivasubramanian A, Vásquez M, Xu Y. Deamidation and isomerization liability analysis of 131 clinical-stage antibodies. MAbs. 2019;11(1):45–57. doi:10.1080/19420862.2018.1548233.

39. Barrow AD, Martin CJ, Colonna M. The natural cytotoxicity receptors in health and disease. Front Immunol [Internet]. 2019;10. [accessed 2020 May 25]. https://www.frontiersin.org/article/10.3389/fimmu.2019.00909/full.

40. Demaria O, Gauthier L, Debroas G, Vivier E. Natural killer cell engagers in cancer immunotherapy: next generation of immuno□ oncology treatments. Eur J Immunol. 2021;51(8):1934–42. doi:10.1002/eji.202048953.

41. Gauthier L, Morel A, Anceriz N, Rossi B, Blanchard-Alvarez A, Grondin G, Trichard S, Cesari C, Sapet M, Bosco F, et al. Multifunctional natural killer cell engagers targeting NKp46 trigger protective tumor immunity. Cell. 2019;177(7):1701–13.e16. doi:10.1016/j.cell.2019.04.041.

42. Peipp M, Klausz K, Boje AS, Zeller T, Zielonka S, Kellner C. Immunotherapeutic targeting of activating natural killer cell receptors and their ligands in cancer. Clin Exp Immunol [Internet]. 2022;209(1):22–32. [accessed 2022 Jul 18]. doi:10.1093/cei/uxac028/6553894.

43. Lipinski B, Arras P, Pekar L, Klewinghaus D, Boje AS, Krah S, Zimmermann J, Klausz K, Peipp M, Siegmund V, et al. NKp46-specific single domain antibodies enable facile engineering of various potent NK cell engager formats. Protein Sci [Internet]. 2023;32(3): [accessed 2023 Mar 8]. doi:10.1002/pro.4593.

44. Benatuil L, Perez JM, Belk J, Hsieh C-M. An improved yeast transformation method for the generation of very large human antibody libraries. Protein Eng Des Sel. 2010;23(4):155–59. doi:10.1093/protein/gzq002.

45. Roth L, Krah S, Klemm J, Günther R, Toleikis L, Busch M, Becker S, Zielonka S. Isolation of antigen-specific VHH single-domain antibodies by Combining animal immunization with yeast surface display [Internet]. In: Zielonka S Krah S, editors. Genotype phenotype coupling. New York, NY: Springer US; 2020. pp. 173–89. [accessed 2020 Jul 6]. doi:10.1007/978-1-4939-9853-1_10.

46. Davis JH, Aperlo C, Li Y, Kurosawa E, Lan Y, Lo K-M, Huston JS. Seedbodies: fusion proteins based on strand-exchange engineered domain (SEED) CH3 heterodimers in an Fc analogue platform for asymmetric binders or immunofusions and bispecific antibodies†. Protein Eng Des Sel. 2010;23(4):195–202. doi:10.1093/protein/gzp094.

47. Tustian AD, Endicott C, Adams B, Mattila J, Bak H. Development of purification processes for fully human bispecific antibodies based upon modification of protein a binding avidity. MAbs. 2016;8(4):828–38. doi:10.1080/19420862.2016.1160192.

48. Sule SV, Sukumar M, Weiss WF, Marcelino-Cruz AM, Sample T, Tessier PM. High-throughput analysis of concentration-dependent antibody self-association. Biophys J. 2011;101(7):1749–57. doi:10.1016/j.bpj.2011.08.036.

49. Liu Y, Caffry I, Wu J, Geng SB, Jain T, Sun T, Reid F, Cao Y, Estep P, Yu Y, et al. High-throughput screening for developability during early-stage antibody discovery using self-interaction nanoparticle spectroscopy. MAbs. 2014;6(2):483–92. doi:10.4161/mabs.27431.

50. Makowski EK, Wu L, Gupta P, Tessier PM. Discovery-stage identification of drug-like antibodies using emerging experimental and computational methods. MAbs. 2021;13(1):1895540. [accessed 2023 Apr 15]. doi:10.1080/19420862.2021.1895540.

51. Klewinghaus D, Pekar L, Arras P, Krah S, Valldorf B, Kolmar H, Zielonka S. Grabbing the bull by both horns: bovine ultralong CDR-H3 paratopes enable engineering of 'almost natural' common light chain bispecific antibodies suitable for effector cell redirection. Front Immunol [Internet]. 2022;12:801368. [accessed 2022 Jul 1]. https://www.frontiersin.org/articles/10.3389/fimmu.2021.801368/full.

52. Pekar L, Klausz K, Busch M, Valldorf B, Kolmar H, Wesch D, Oberg H-H, Krohn S, Boje AS, Gehlert CL, et al. Affinity maturation of B7-H6 translates into enhanced NK cell–mediated tumor cell lysis and improved proinflammatory cytokine release of

bispecific immunoligands via NKp30 engagement. J Immunol. 2021;206(1):225–36. doi:10.4049/jimmunol.2001004.

53. Bortoletto N, Scotet E, Myamoto Y, D'Oro U, Lanzavecchia A. Optimizing anti-CD3 affinity for effective T cell targeting against tumor cells. Eur J Immunol. 2002;32(11):3102–07. doi:10.1002/1521-4141(200211)32:11<3102:AID-IMMU3102>3.0.CO;2-C.

54. Mitchell LS, Colwell LJ, Bonvin AMJJ. Analysis of nanobody paratopes reveals greater diversity than classical antibodies. Protein Eng Des Sel. 2018;31(7–8):267–75. doi:10.1093/protein/gzy017.

55. Lipinski B, Unmuth L, Arras P, Becker S, Bauer C, Toleikis L, Krah S, Doerner A, Yanakieva D, Boje AS, et al. Generation and engineering of potent single domain antibody-based bispecific IL-18 mimetics resistant to IL-18BP decoy receptor inhibition. MAbs. 2023;15(1):2236265. doi:10.1080/19420862.2023.2236265.

56. Martin KP, Grimaldi C, Grempler R, Hansel S, Kumar S. Trends in industrialization of biotherapeutics: a survey of product characteristics of 89 antibody-based biotherapeutics. MAbs. 2023;15(1): 2191301. [accessed 2023 Apr 19]. doi:10.1080/19420862.2023.2191301.

57. Uchański T, Zögg T, Yin J, Yuan D, Wohlkönig A, Fischer B, Rosenbaum DM, Kobilka BK, Pardon E, Steyaert J. An improved yeast surface display platform for the screening of nanobody immune libraries. Sci Rep. 2019;9(1). [accessed 2020 Jan 28]. doi:10.1038/s41598-018-37212-3.

58. Pekar L, Klewinghaus D, Arras P, Carrara SC, Harwardt J, Krah S, Yanakieva D, Toleikis L, Smider VV, Kolmar H, et al. Milking the cow: cattle-derived chimeric ultralong CDR-H3 antibodies and their engineered CDR-H3-Only knobbody counterparts targeting epidermal growth factor receptor elicit potent NK cell-mediated cytotoxicity. Front Immunol. 2021;12:4378. doi:10.3389/fimmu.2021.742418.

59. Hu D, Hu S, Wan W, Xu M, Du R, Zhao W, Gao X, Liu J, Liu H, Hong J, et al. Effective optimization of antibody affinity by phage display Integrated with high-throughput DNA synthesis and sequencing technologies. PloS One. 2015;10(6):e0129125. doi:10.1371/journal.pone.0129125.

60. Larman HB, Jing Xu G, Pavlova NN, Elledge SJ Construction of a rationally designed antibody platform for sequencing-assisted selection. Proceedings of the National Academy of Sciences 2012;109:18523–28. https://www.pnas.org/.

61. Mathonet P, Ullman CG. The application of next generation sequencing to the understanding of antibody repertoires. Front Immunol [Internet]. 2013;4:265. [accessed 2023 May 1]. http://journal.frontiersin.org/article/10.3389/fimmu.2013.00265/abstract.

62. Barreto K, Maruthachalam BV, Hill W, Hogan D, Sutherland AR, Kusalik A, Fonge H, DeCoteau JF, Geyer CR. Next-generation sequencing-guided identification and reconstruction of antibody CDR combinations from phage selection outputs. Nucleic Acids Res. 2019;47(9):e50–e50. doi:10.1093/nar/gkz131.

63. Hu R, Fu L, Chen Y, Chen J, Qiao Y, Si T. Protein engineering via Bayesian optimization-guided evolutionary algorithm and robotic experiments. Briefings Bioinf [Internet]. 2023;24(1):bbac570. [accessed 2023 May 1]. doi:10.1093/bib/bbac570/6958505.

64. Liu G, Zeng H, Mueller J, Carter B, Wang Z, Schilz J, Horny G, Birnbaum ME, Ewert S, Gifford DK, et al. Antibody complementarity determining region design using high-capacity machine learning. Bioinformatics. 2020;36(7):2126–33. doi:10.1093/bioinformatics/btz895.

65. Makowski EK, Kinnunen PC, Huang J, Wu L, Smith MD, Wang T, Desai AA, Streu CN, Zhang Y, Zupancic JM, et al. Co-optimization of therapeutic antibody affinity and specificity using machine learning models that generalize to novel mutational space. Nat Commun [Internet]. 2022;13(1):3788. [accessed 2023 May 1]. doi:10.1038/s41467-022-31457-3.

66. Mason DM, Friedensohn S, Weber CR, Jordi C, Wagner B, Meng SM, Ehling RA, Bonati L, Dahinden J, Gainza P, et al. Optimization of therapeutic antibodies by predicting antigen specificity from antibody sequence via deep learning. Nat Biomed Eng. 2021;5(6):600–12. doi:10.1038/s41551-021-00699-9.

67. Parkinson J, Hard R, Wang W. The RESP AI model accelerates the identification of tight-binding antibodies. Nat Commun. 2023;14(1):454. [accessed 2023 May 1]. https://www.nature.com/articles/s41467-023-36028-8.

68. Saka K, Kakuzaki T, Metsugi S, Kashiwagi D, Yoshida K, Wada M, Tsunoda H, Teramoto R. Antibody design using LSTM based deep generative model from phage display library for affinity maturation. Sci Rep [Internet]. 2021;11(1):5852. [accessed 2023 May 1]. https://www.nature.com/articles/s41598-021-85274-7.