Positive charge in the complementarity-determining regions of synthetic nanobody-Fc fusion prevents aggregation

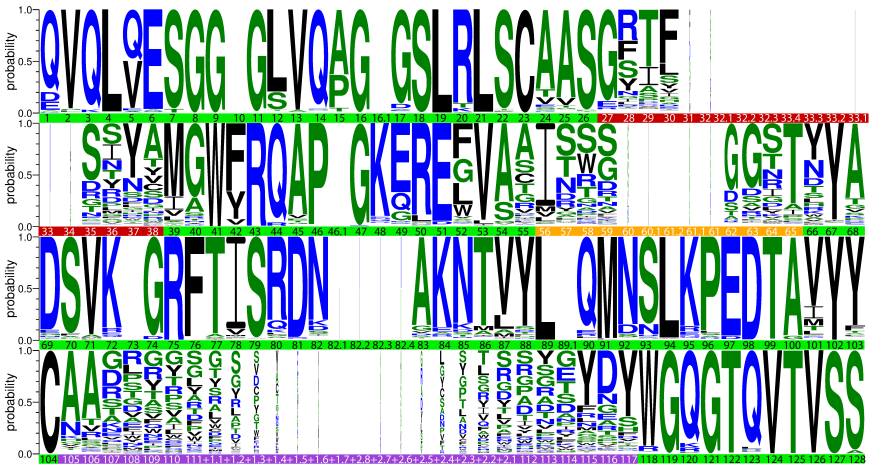
Zhenwei Zhong[1]

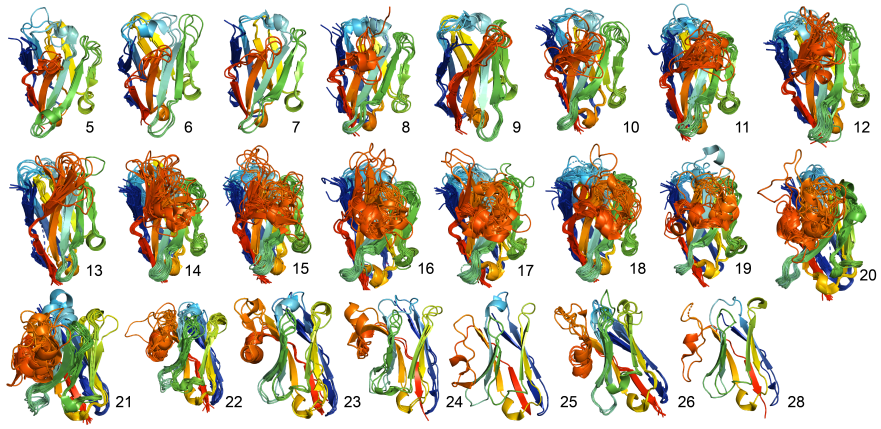
[1]Origincell Science and Technology Group Limited

March 07, 2024

Abstract

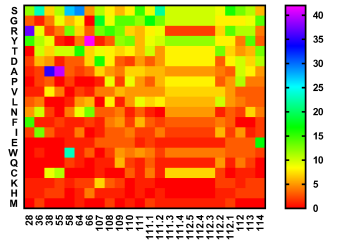
In the past, specificity and affinity were the priority for synthetic antibody library. However, therapeutic antibodies need good stability for medical use. Through carefully adjust the chemical diversity in CDRs, one hopes to design a synthetic antibody library with good developability. Here we thoroughly analyzed 296 nanobody sequences and structures, constructed a fully-functional synthetic nanobody library, evaluated the relationship between aggregation and isoelectric point, and found that high-pI nanobodies were more resistant to aggregation than low-pIs. As we used the same framework for constructing the library, CDRs charge played a crucial role in mediating nanobody aggregation. We also analyzed the theoretical pI of 296 nanobodies from PDB, about 75% had basic pI, only 25% were acidic. Those results provided useful guidelines for designing next-generation synthetic nanobody libraries and for identifying potent and safe nanobody therapeutics.



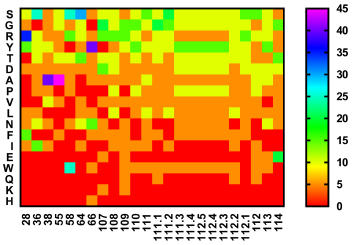


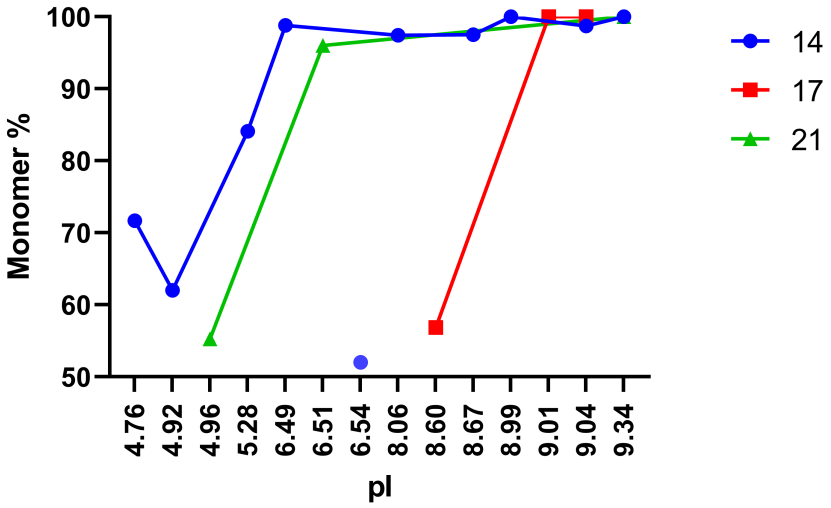
a.

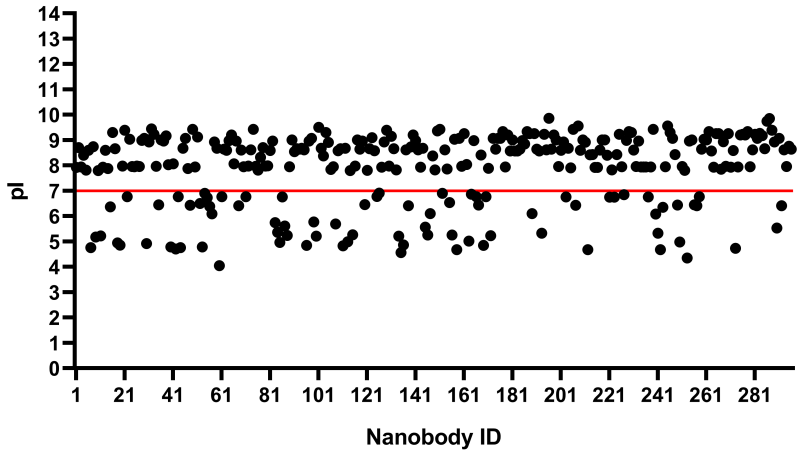
b.



c.







Positive charge in the complementarity-determining regions of synthetic nanobody-Fc fusion prevents aggregation

Zhenwei Zhong[a], Yue Yang[a], Xiaorui Chen[a], Zhen Han[a], Jincai Zhou[a], Bohua Li[a], Xiaowen He[a]

[a]Department of Antibody, Origincell Therapeutics, Origincell Industrial Park, No. 1227 Zhangheng Road, Pudong New District, Shanghai, China

Correspondence to: Zhenwei Zhong; Email: zhongzhenwei@origincell.com; Xiaowen He; Email: peterhe@origincell.com

Abstract

In the past, specificity and affinity were the priority for synthetic antibody library. However, therapeutic antibodies need good stability for medical use. Through carefully adjust the chemical diversity in CDRs, one hopes to design a synthetic antibody library with good developability. Here we thoroughly analyzed 296 nanobody sequences and structures, constructed a fully-functional synthetic nanobody library, evaluated the relationship between aggregation and isoelectric point, and found that high-pI nanobodies were more resistant to aggregation than low-pIs. As we used the same framework for constructing the library, CDRs charge played a crucial role in mediating nanobody aggregation. We also analyzed the theoretical pI of 296 nanobodies from PDB, about 75% had basic pI, only 25% were acidic. Those results provided useful guidelines for designing next-generation synthetic nanobody libraries and for identifying potent and safe nanobody therapeutics.

Keywords: positive charge, synthetic nanobody, aggregation, sequence and structure alignment, isoelectric point, consensus protein design

Introduction

Nanobody derives from the variable domain of heavy-chain-only antibody in sera of camelids (Hamers-Casterman et al., 1993). From the first identification in 1993 to the first approval of Caplacizumab by FDA in 2019, nanobody begins to get a boost (Hamers-Casterman et al., 1993; Morrison, 2019). There are over ten companies that have nanobody pipelines, and most of them are bispecific antibody. At the same time, synthetic antibody library becomes an important resource for antibody discovery, Guselkumab, the first synthetic antibody from MorphoSys's HuCAL® phage library, was approved by FDA in 2017. Over the last two decades, synthetic antibody libraries have continuously evolved (Fellouse et al., 2007; Fellouse, Wiesmann, & Sidhu, 2004; Knappik et al., 2000; Persson et al., 2013; Prassler et al., 2011; Rothe et al., 2008; Tiller et al., 2013). With the advances in DNA synthesis technology, such as Trinucleotide phosphoramidites (Virnekas et al., 1994) and Slonomics technology (Tiller et al., 2013), we can precisely control the chemical diversity at each position of CDRs, which means desired characteristics, like high expression yield, nanomolar-range affinity, risky sites free, may be introduced into a mini-library. However, one primary disadvantage of synthetic libraries is that the man-made diversity in CDRs may also lead to antibody misfolding and aggregation (Sidhu & Fellouse, 2006). Aggregation not only inactivates the antibody's biological function but also induces immunogenicity (Hermeling, Crommelin, Schellekens, & Jiskoot, 2004; Jefferis, 2011; Singh, 2011). Many researchers had identified that CDRs commonly contained aggregation hotspots (Dudgeon, Famm, & Christ, 2009; Lee, Perchiacca, & Tessier, 2013), sometimes even one single mutation could inhibit aggregation, arginine, or aspartate in CDR1 of single domain antibody had been proven to be effective (Famm, Hansen, Christ, & Winter, 2008; Perchiacca,

Bhattacharya, & Tessier, 2011).

Functionally, synthetic nanobody had been used as intrabody (Moutel et al., 2016) and conformationally selective GPCR stabilizers (McMahon et al., 2018), while others focused on improving nanobody's stability. Efforts involving adding disulfide bond by introducing cysteine at positions 54 and 78 (Hussack, Hirama, Ding, Mackenzie, & Tanha, 2011), or putting negatively charged amino acids at CDR1 and near the edges of CDR3 of the human variable domain (Dudgeon et al., 2012; Perchiacca et al., 2011; Perchiacca, Ladiwala, Bhattacharya, & Tessier, 2012), or introducing arginine at position 29 had improved aggregation-prone single-domain antibody stability (Famm et al., 2008). However, our understanding of the stability of synthetic nanobody is limited. The enormous potential of fully human-designed nanobody in clinical benefit motivated our study of the aggregation of synthetic nanobody.

We analyzed 296 non-redundant nanobody sequences and structures downloaded from PDB, according to the principle of "Consensus Protein Design" (Porebski & Buckle, 2016), we made a CDR mutation strategy and constructed a phage-displayed synthetic nanobody library by using the framework of Caplacizumab. After solid-phase and cell-based panning against two antigens which have high and low pI respectively, we identified several high-affinity, different-pI nanobodies, the results of the SEC showed that with the same length of CDR3, high-pI nanobodies were more aggregation-resistant than low pIs. What's more, above 75% of nanobodies had basic pI in PDB, which further highlighted the preference of positively charged residues.

Results

Sequence logo

All nanobody sequences were aligned by using the online tool ANARCI (Dunbar & Deane, 2016), according to IMGT unique numbering (Lefranc et al., 2003). CDR1 is from position 27 to 38, CDR2 is from position 56 to 65, CDR3 is from position 105 to 117 as shown in figure 1. Position 28, 36, and 38 were less conserved than position 27, 29, 30, 35, and 37, the cysteines in position 38 and 55 were responsible for forming a disulfide bond with another cysteine in CDR3 (Figure S1). As a previous report, CDR2 preferred small residues, such as glycine and serine, to form the tight backbone turn in nanobody (Mitchell & Colwell, 2018b), however, at the edges of CDR2, position 55 and 66 were also diversified. CDR3 had the most diversity, exclude position 105, 106, 115, 116, and 117. Compared with conventional VH, Nanobody has four mutations within FR2, including F42, E49, R50, and G52, render VHH soluble (Muyldermans, 2013). And we found that Y42 was more frequent when CDR3 was short than 14, otherwise, VHH favored F42 (Figure S2, S3). With different frameworks, a unique library design should be taken.

Analysis of CDRs length and CDR3 repertoire content

We also investigated the length variation of CDRs, almost all CDR1 was 8, mean was 8.03 (Figure 2); over 65% CDR2 length was 8, about 30% was 7, mean was 7.77 (Figure 2); the range of CDR3 length was extremely broad, from 5 to 28, mean was 15.87 (Figure 2). To increase our understanding of CDR3 repertoire content, we analyzed the total amino acid usage of CDR3, excluding 105, 106, 115, 116, and 117, since those positions were conservative. As we can see, glycine and serine were at the top of two (Figure 2), providing CDR3 flexibility. It was not

surprising to see arginine took up 8.94%, many studies revealed that arginine, one of the hotspot residues, was not only capable of mediating a wide array of intermolecular interactions but also may aid the folding or structural stability of the heavy chain (Birtalan et al., 2008; Bogan & Thorn, 1998; Chakrabarti & Janin, 2002; Reichmann, Rahat, Cohen, Neuvirth, & Schreiber, 2007; Villar & Kauvar, 1994), Another important amino acid was tyrosine, 7.63%, researchers once constructed phage-displayed antigen-binding fragments or monobody library by using only tyrosine and serine, gained high-affinity binders (Fellouse et al., 2005; Koide, Gilbreth, Esaki, Tereshko, & Koide, 2007). Aspartate is negatively charged, which played a vital role in mediating antibody specificity (Rabia, Zhang, Ludwig, Julian, & Tessier, 2018), whose percentage was 6.86%. Free cysteine, histidine, and methionine were not welcomed in CDR3, duo to oxidation risk (Manning, Chou, Murphy, Payne, & Katayama, 2010), all those residues were lower than 2%.

CDR3 conformation

CDR3 contributes the key contact with an antigen, but it does not have canonical structures (Al-Lazikani, Lesk, & Chothia, 1997). To get a deep view of the relationship between CDR3 length and conformation, we first grouped nanobodies according to CDR3 length, then aligned structures with the “alignment” function in PyMOL. When the length was lower than 14, FR2 was solvent-exposed, Y42 was more common than F42, CDR3 possessed a small “loop” structure, held by FR3, as the length elongating, the loop grew close to CDR1, we called this structure as “loop-stick” (Figure 3). Length 14 had the most conformational diversity, some still had “loop-stick”, others began to bend to FR2. As CDR3 became longer, CDR3 gradually covered FR2, like a “palm fan”, F42 may be incorporated with other hydrophobic amino acids in CDR3 to form a

hydrophobic core to stabilize nanobody.

Synthetic nanobody library: NanoOri_1.0

The success of the HuCAL antibody library let us adopt the same strategy to try to mimic the relative amino acid frequencies at X positions (Figure 4), for other less diversified positions, we used degenerate codon to cover residues that were prominent in PDB. For CDR1, length 8 was chosen, position 28 to 38 were mutated, G27 was maintained; for CDR2, the length was 8, position 57 to 64 were introduced mutations, I56, T65 unchanged, although position 55 and 66 were not included in CDR2 by IMGT, they were also taken as X position; For CDR3, three kinds of length were introduced: 14, 17, and 21, because 14 had the most diverse conformation (Figure 3), 17 was the most among all lengths (Figure 2), Caplacizumab has 21 amino acids in CDR3, so 21 was chosen. The relative amino acid frequencies at X position were calculated in Excel, except for cysteine and methionine, and the percentage of amino acids was adjusted to the multiple of 5 % (Figure 4). According to the length of CDR3, the whole library was divided into three sub-libraries. We introduced mutations by Kunkel mutagenesis in the HP153 phage display vector, then the plasmids were electroporated into SS320, 2.4×10^{10} independent clones were obtained. Combined with the 288-clone sequencing result, the total diversity of NanoOri_1.0 was 1.44×10^{10}.

SEC of synthetic nanobodies with different pIs

By coincidence, we did panning against two antigens with different pI using NanoOri_1.0, expressed with VHH-Fc format, and obtained several high-affinity nanobodies with variable pI

(Table 1). To know the aggregation property, all nanobodies were analyzed by SEC (Figure S4). To our surprise, we found that high-pI nanobodies recognized acidic antigen with a remarkable affinity (KD<10nM, data not shown) and vice versa (Table 1). Then we further analyzed the relationship between SEC and pI. Considering CDR3 could significantly influence nanobody stability, all nanobodies were grouped into three classes according to CDR3 length. 4A4, VHH1, and VHH13 had the same CDR3 length: 21, which pI was 9.34, 6.51, 4.96 respectively, showed variable degree of aggregation. 4A4 was purely 100% monomer, and 96.02% for VHH1, 55.26% for VHH13. The same relationship could be observed for 3G4, 4A2, and 3E9, whose CDR3 length was 17, so did 4F10, 3F5, 4B3, 3B5, VHH5, VHH8, VHH16, VHH17, and VHH18 (Figure 5). One exception was VHH3, other factors that help increasing its aggregation, maybe duo to the double phenylalanines in the middle of CDR3. Interestingly, CDR3 length 17 started to aggregate at a higher pI in comparison to lengths 14 and 21, and we need more data to explain this observation.

Theoretical pI of 296 nanobodies from PDB

Since the NanoOri_1.0 library was based on the data of PDB, we wanted to know the theoretical pI distribution of 296 nanobodies, found that basic-pIs took up 75%, and only 73 among 296 were acidic, the mean was 7.29 (Figure 6). As disulfide bond is a main factor that influences the stability of nanobody, we picked out 46 nanobodies which had disulfide bond between CDR3 and CDR1 or FR2, or within CDR3 calculated the theoretical pI. To our surprise, basic and acidic pI took up about 50% respectively (Figure S5b). Another disulfide bond between CDR3 and CDR1 or FR2, or within CDR3 may allow nanobody to use more diverse chemistry in CDRs.

Discussion

In this work, we successfully constructed a fully functional synthetic nanobody library which was easy to gain high-affinity nanobodies with low nM range and uncovered a piece of hidden information that high-pI nanobodies were more resistant to aggregation than low-pIs. First, we analyzed 296 nanobody sequences and structures, uncovered the conformations of CDR3 with variable lengths, short CDR3 had “loop-stick” conformation, long CDR3 was like a palm fan. For synthetic antibody library, both sequence and structure diversity are required. Another interesting finding was position 42. When CDR3 was short than 14, Y42 might help to increase the solubility of nanobody, as CDR3 elongating, F42 was suitable for forming a hydrophobic core with a bent CDR3. Besides, we found that low-pI antigen easily got high-pI nanobodies and vice versa. One explanation was that antigen-nanobody interaction depends heavily on static electricity, especially arginine (Mitchell & Colwell, 2018a).

Arginine was the most abounded amino acid among hypervariable sites of CDR3, apart from glycine and serine (Figure 2). Researchers once “supercharged” green fluorescent protein or the anti-MS2 scFV antibody fragment with arginine and lysine, improved aggregation resistance and thermal stability remarkably (Lawrence, Phillips, & Liu, 2007; Miklos et al., 2012). Our results also provided evidence that synthetic nanobodies having more arginine in CDRs (high-PIs), were more aggregation-resistant (Figure 5). Also, as a small molecular additive, arginine effectively prevents the aggregating of various kinds of proteins(Shiraki, Kudou, Fujiwara, Imanaka, & Takagi, 2002). It is reasonable to say arginine, in CDRs, functions as a stabilizer and contactor.

However, one report showed that aggregation-resistant VHs tend to have acidic isoelectric points(Arbabi-Ghahroudi et al., 2009). Researchers carried out a theoretical pI distribution

analysis of rearranged and germline VHH and VH sequences from camelids and humans, found that the combined pool of camelid VHHs (646 sequences) mostly has low pI, and the VH pool consists mainly of VH segments with high pI. There were some differences between our library and theirs, they kept the disulfide bond within CDR3 which may become an obstacle for industrial manufacturing. And they used the heat denaturation approach for selecting binders which the affinity was low (KD was from μM to $100 nM$), while we did conventional panning and got high-affinity nanobodies (KD<10nM). All the sequences which we analyzed were from PDB, had known structures, and arginine was more abundant than any other amino acids in CDR3, except glycine and serine (figure 2d).

We also analyzed the theoretical pI distribution of the pool of 296 nanobodies, revealed that 75% had basic pI, if exclude nanobodies with another disulfide bond, the percentage was up to 80% (Figure S5a), the controversial results told us that pI was only one of factors that affect the stability of nanobody, it could be a predictor only when other factors are controlled, like disulfide bond, CDR length, hydrophobic residue patch, and framework. Besides, another disulfide bond between CDR3 and CDR1 or FR2, or within CDR3 may help nanobody to use more diverse chemistry since their pI distribution was more impartial.

It also should be kept in mind that positively charged patch of antibody can interact with anionic sites of cell surfaces, increasing tissue uptake and fast clearance from circulation (Liu, 2018). There were reports demonstrated that an antibody with a lower isoelectric point (pI) had a longer half-life (Igawa et al., 2010; Li et al., 2014). Whether those rules could be applied to nanobody is unknown. How to balance the charge, pI, physicochemical characteristics, and PK is still a challenge for synthetic nanobody.

Although we removed cysteine and methionine in CDR3, other potential post-translational modification sites, such as glycosylation sites: N-X-S/T, deamidation sites: NN, NS, NG, NH; isomerization sites: DS, DG, DD; protease cleavage sites: DP, DQ (Manning et al., 2010; Tiller et al., 2013) were maintained. Some nanobodies need to be further engineered for pharmaceutical development. Knowing the importance of arginine and nanobody sequence and structure characteristics should be helpful for next-generation library design. As the data in PDB grows, some hidden information may appear in the future, synthetic nanobody library will keep evolving.

Materials and Methods

Sequence alignment

Searched and downloaded nanobody sequences on the PDB website (https://www.rcsb.org/) by using keywords: nanobody, VHH, single-domain antibody, heavy-chain-only antibody. Extracted nanobody sequences into Excel, and removed repeating sequences. All sequences were imported into the ANARCI website (Dunbar & Deane, 2016) (http://opig.stats.ox.ac.uk/webapps/newsabdab/sabpred/anarci/) and aligned using the IMGT numbering scheme (Lefranc et al., 2003). According to the length of CDR3, sequences were grouped into 23 classes. In the end, the sequence logos were generated on the WebLogo 3 website (http://weblogo.threeplusone.com/create.cgi).

Structure alignment

After gaining nanobody sequences, structures were downloaded and removed non-nanobody

structures. Nanobody structures were aligned using the “Alignment” function in PyMOL.

Synthetic nanobody library construction

The nucleotide of Caplacizuamab and oligos were synthesized by Genewiz Corporation. Inserted Caplacizumab into HP153 phagemid vector (Persson et al., 2013), transformed the vector into CJ236. Isolate dU-ssDNA template as previously reported (Tonikian, Zhang, Boone, & Sidhu, 2007). After in vitro synthesis of heteroduplex CCC-dsDNA, the DNA was electroporated into SS320. The supernatant containing the phage-displayed nanobody library was collected and precipitated as described (Huang et al., 2018).

Panning

Phages from library NanoOri_1.0 were selected through four rounds of panning with antigen coated on 96-well Immunoplates as described (Tonikian et al., 2007). Cell-based panning followed the protocol as described (Eisenhardt, Schwarz, Bassler, & Peter, 2007). 400 phage clones were picked into 96-well format for each antigen, used the culture supernatants to do phage ELISA. Clones that bound antigen or antigen-expressing 293T but not BSA or 293T were sequenced.

Nanobodies expression and purification

All nanobodies were cloned into pcDNA3.0 plasmid, and transfected into EXPI293F cells with PEI, expressed with VHH-hFc format, purified by protein A affinity chromatography. Nanobodies were bound to a protein A column (10 mL MabSelect SuRe LX, GE Healthcare) in 1×PBS (Hyclone), eluted by 50 mM Gly-HCl (pH3.0) and neutralized with 1 M Tris (pH8.0).

SEC

Size-exclusion chromatography (SEC) was performed using an Agilent 1200 high-performance liquid chromatography system and a Sepax SRT-C SEC-300 column (7.8×300 mm). Nanobodies were prepared at 1 mg/ml and injected (10 µl) at a flow rate of 1 ml/min. The elution of each nanobody was monitored at 280 nm using an Agilent 1200 series UV Absorbance detector.

Supplementary Material

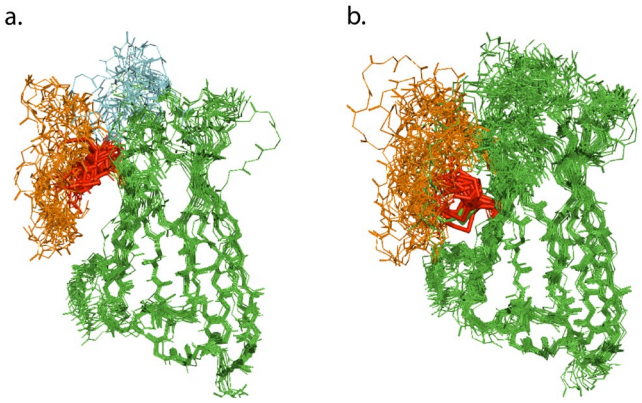


Figure S1. (a) CDR3 formed a disulfide bond with CDR1. (b) CDR3 formed a disulfide bond with FR2. Red lines represent disulfide bonds.

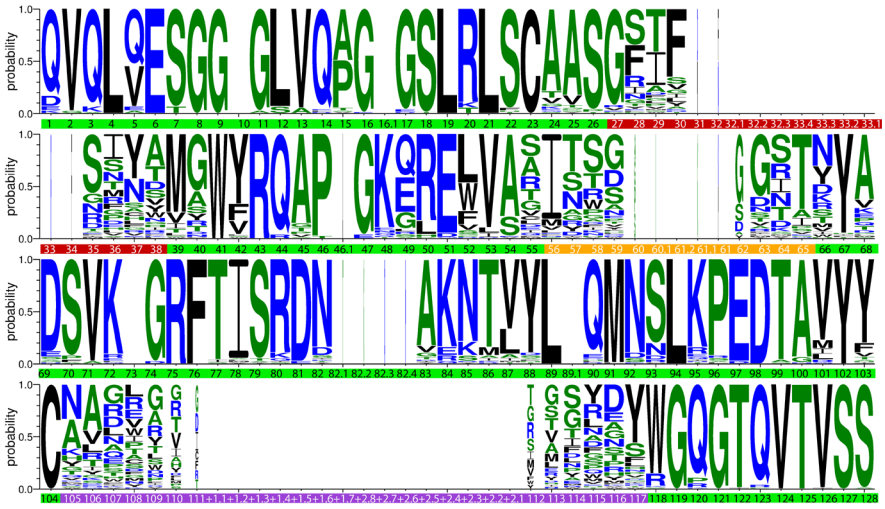


Figure S2. Sequence logo of nanobodies, CDR3 length was from 5-13. Position 42 favored tyrosine.

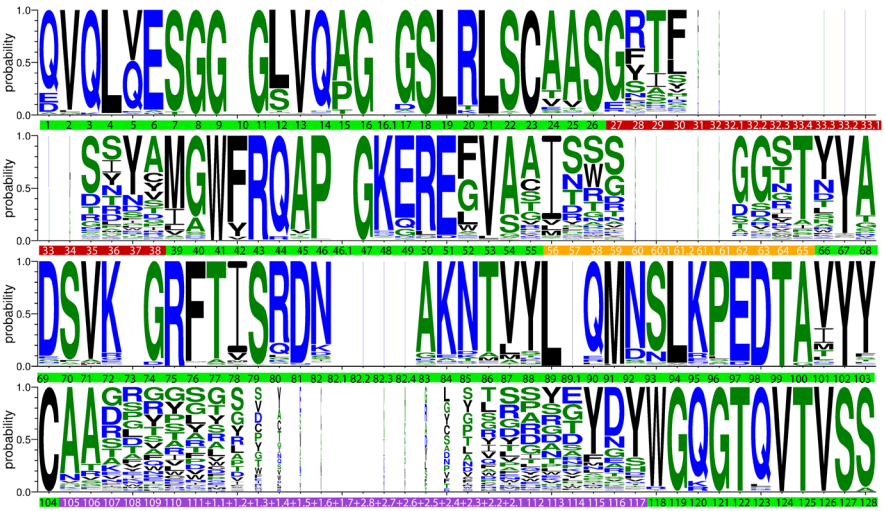


Figure S3. Sequence logo of nanobodies, CDR3 length was from 14-28. Position 42 preferred phenylalanine.

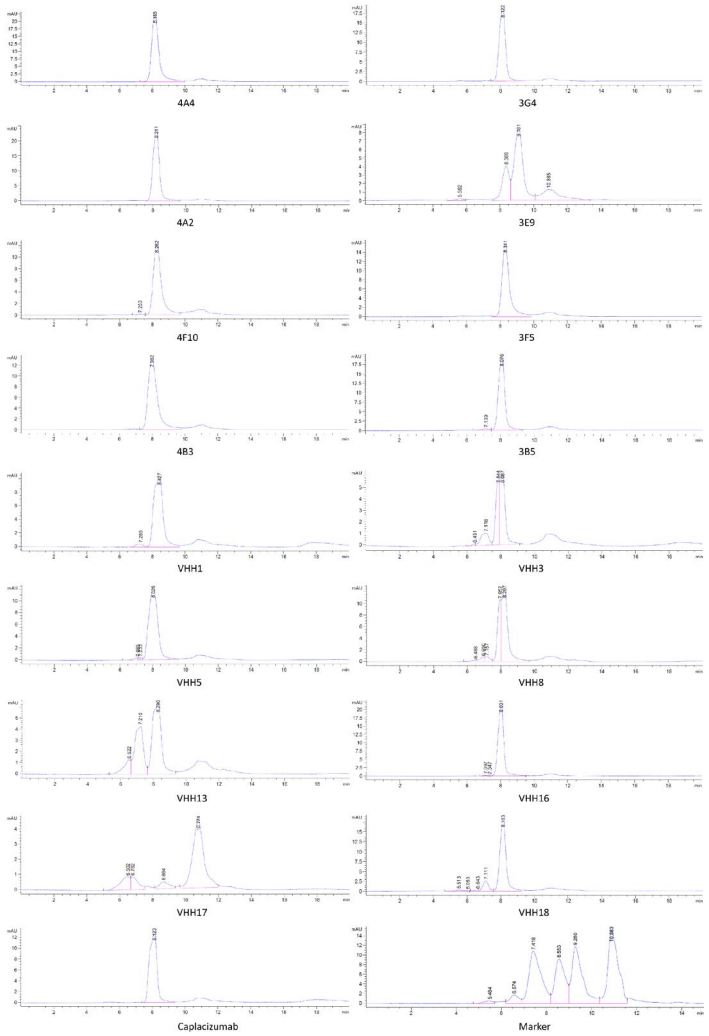


Figure S4. SEC of nanobodies.

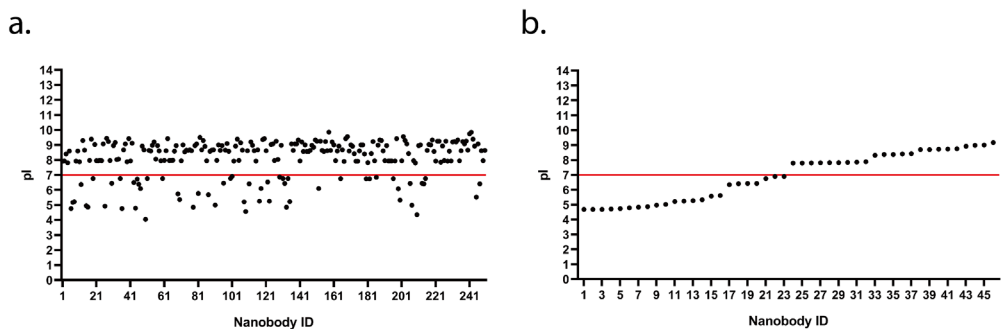


Figure S5. (a) Theoretical pI distribution of 250 nanobodies without disulfide band, basic pIs was

77.2% (193/250), acidic pIs was 22.8% (57/250), the mean was 7.47. (b) 46 nanobodies with disulfide bond between CDR3 and CDR1 or FR2, or within CDR3, nearly one half had acidic pI, another half was basic. The mean was 6.27.

Acknowledgments

No.

Disclosure of potential conflicts of interest

All authors are current employees of Origincell Meditech, which develops and commercializes antibody and CAR-T therapeutics.

References

Al-Lazikani, B., Lesk, A. M., & Chothia, C. (1997). Standard conformations for the canonical structures of immunoglobulins. *J Mol Biol, 273*(4), 927-948. doi:10.1006/jmbi.1997.1354

Arbabi-Ghahroudi, M., To, R., Gaudette, N., Hirama, T., Ding, W., MacKenzie, R., & Tanha, J. (2009). Aggregation-resistant VHs selected by in vitro evolution tend to have disulfide-bonded loops and acidic isoelectric points. *Protein Eng Des Sel, 22*(2), 59-66. doi:10.1093/protein/gzn071

Birtalan, S., Zhang, Y., Fellouse, F. A., Shao, L., Schaefer, G., & Sidhu, S. S. (2008). The intrinsic contributions of tyrosine, serine, glycine and arginine to the affinity and specificity of antibodies. *J Mol Biol, 377*(5), 1518-1528. doi:10.1016/j.jmb.2008.01.093

Bogan, A. A., & Thorn, K. S. (1998). Anatomy of hot spots in protein interfaces. *J Mol Biol, 280*(1), 1-9. doi:10.1006/jmbi.1998.1843

Chakrabarti, P., & Janin, J. (2002). Dissecting protein-protein recognition sites. *Proteins, 47*(3), 334-343. doi:10.1002/prot.10085

Dudgeon, K., Famm, K., & Christ, D. (2009). Sequence determinants of protein aggregation in human VH domains. *Protein Eng Des Sel, 22*(3), 217-220. doi:10.1093/protein/gzn059

Dudgeon, K., Rouet, R., Kokmeijer, I., Schofield, P., Stolp, J., Langley, D., . . . Christ, D. (2012). General strategy for the generation of human antibody variable domains with increased aggregation resistance. *Proc Natl Acad Sci U S A, 109*(27), 10879-10884. doi:10.1073/pnas.1202866109

Dunbar, J., & Deane, C. M. (2016). ANARCI: antigen receptor numbering and receptor classification. *Bioinformatics, 32*(2), 298-300. doi:10.1093/bioinformatics/btv552

Eisenhardt, S. U., Schwarz, M., Bassler, N., & Peter, K. (2007). Subtractive single-chain antibody (scFv) phage-display: tailoring phage-display for high specificity against function-specific conformations of cell membrane molecules. *Nat Protoc, 2*(12), 3063-3073. doi:10.1038/nprot.2007.455

Famm, K., Hansen, L., Christ, D., & Winter, G. (2008). Thermodynamically stable aggregation-resistant antibody domains through directed evolution. *J Mol Biol, 376*(4), 926-931. doi:10.1016/j.jmb.2007.10.075

Fellouse, F. A., Esaki, K., Birtalan, S., Raptis, D., Cancasci, V. J., Koide, A., . . . Sidhu, S. S. (2007). High-throughput generation of synthetic antibodies from highly functional minimalist phage-displayed libraries. *J Mol Biol, 373*(4), 924-940. doi:10.1016/j.jmb.2007.08.005

Fellouse, F. A., Li, B., Compaan, D. M., Peden, A. A., Hymowitz, S. G., & Sidhu, S. S. (2005). Molecular recognition by a binary code. *J Mol Biol, 348*(5), 1153-1162. doi:10.1016/j.jmb.2005.03.041

Fellouse, F. A., Wiesmann, C., & Sidhu, S. S. (2004). Synthetic antibodies from a four-amino-acid code: a dominant role for tyrosine in antigen recognition. *Proc Natl Acad Sci U S A, 101*(34), 12467-12472. doi:10.1073/pnas.0401786101

Hamers-Casterman, C., Atarhouch, T., Muyldermans, S., Robinson, G., Hamers, C., Songa, E. B., . . . Hamers, R. (1993). Naturally occurring antibodies devoid of light chains. *Nature, 363*(6428), 446-448. doi:10.1038/363446a0

Hermeling, S., Crommelin, D. J., Schellekens, H., & Jiskoot, W. (2004). Structure-immunogenicity relationships of therapeutic proteins. *Pharm Res, 21*(6), 897-903. doi:10.1023/b:pham.0000029275.41323.a6

Huang, G., Zhong, Z., Miersch, S., Sidhu, S. S., Hou, S. C., & Wu, D. (2018). Construction of Synthetic Phage Displayed Fab Library with Tailored Diversity. *J Vis Exp*(135). doi:10.3791/57357

Hussack, G., Hirama, T., Ding, W., Mackenzie, R., & Tanha, J. (2011). Engineered single-domain antibodies with high protease resistance and thermal stability. *PLoS One, 6*(11), e28218. doi:10.1371/journal.pone.0028218

Igawa, T., Tsunoda, H., Tachibana, T., Maeda, A., Mimoto, F., Moriyama, C., . . . Hattori, K. (2010). Reduced elimination of IgG antibodies by engineering the variable region. *Protein Eng Des Sel, 23*(5), 385-392. doi:10.1093/protein/gzq009

Jefferis, R. (2011). Aggregation, immune complexes and immunogenicity. *MAbs, 3*(6), 503-504. doi:10.4161/mabs.3.6.17611

Knappik, A., Ge, L., Honegger, A., Pack, P., Fischer, M., Wellnhofer, G., . . . Virnekas, B. (2000). Fully synthetic human combinatorial antibody libraries (HuCAL) based on modular consensus frameworks and CDRs randomized with trinucleotides. *J Mol Biol, 296*(1), 57-86. doi:10.1006/jmbi.1999.3444

Koide, A., Gilbreth, R. N., Esaki, K., Tereshko, V., & Koide, S. (2007). High-affinity single-domain binding proteins with a binary-code interface. *Proc Natl Acad Sci U S A, 104*(16), 6632-6637. doi:10.1073/pnas.0700149104

Lawrence, M. S., Phillips, K. J., & Liu, D. R. (2007). Supercharging proteins can impart unusual resilience. *J Am Chem Soc, 129*(33), 10110-10112. doi:10.1021/ja071641y

Lee, C. C., Perchiacca, J. M., & Tessier, P. M. (2013). Toward aggregation-resistant antibodies by design. *Trends Biotechnol, 31*(11), 612-620. doi:10.1016/j.tibtech.2013.07.002

Lefranc, M. P., Pommie, C., Ruiz, M., Giudicelli, V., Foulquier, E., Truong, L., . . . Lefranc, G. (2003). IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. *Dev Comp Immunol, 27*(1), 55-77. doi:10.1016/s0145-305x(02)00039-3

Li, B., Tesar, D., Boswell, C. A., Cahaya, H. S., Wong, A., Zhang, J., . . . Kelley, R. F. (2014). Framework selection can influence pharmacokinetics of a humanized therapeutic antibody through differences in molecule charge. *MAbs, 6*(5), 1255-1264. doi:10.4161/mabs.29809

Liu, L. (2018). Pharmacokinetics of monoclonal antibodies and Fc-fusion proteins. *Protein Cell, 9*(1), 15-32. doi:10.1007/s13238-017-0408-4

Manning, M. C., Chou, D. K., Murphy, B. M., Payne, R. W., & Katayama, D. S. (2010). Stability of protein pharmaceuticals: an update. *Pharm Res, 27*(4), 544-575. doi:10.1007/s11095-009-0045-6

McMahon, C., Baier, A. S., Pascolutti, R., Wegrecki, M., Zheng, S., Ong, J. X., . . . Kruse, A. C. (2018). Yeast surface display platform for rapid discovery of conformationally selective nanobodies. *Nat Struct Mol Biol, 25*(3), 289-296. doi:10.1038/s41594-018-0028-6

Miklos, A. E., Kluwe, C., Der, B. S., Pai, S., Sircar, A., Hughes, R. A., . . . Ellington, A. D. (2012). Structure-based design of supercharged, highly thermoresistant antibodies. *Chem Biol, 19*(4), 449-455. doi:10.1016/j.chembiol.2012.01.018

Mitchell, L. S., & Colwell, L. J. (2018a). Analysis of nanobody paratopes reveals greater diversity than classical antibodies. *Protein Eng Des Sel, 31*(7-8), 267-275. doi:10.1093/protein/gzy017

Mitchell, L. S., & Colwell, L. J. (2018b). Comparative analysis of nanobody sequence and structure data. *Proteins, 86*(7), 697-706. doi:10.1002/prot.25497

Morrison, C. (2019). Nanobody approval gives domain antibodies a boost. *Nat Rev Drug Discov, 18*(7), 485-487. doi:10.1038/d41573-019-00104-w

Moutel, S., Bery, N., Bernard, V., Keller, L., Lemesre, E., de Marco, A., . . . Perez, F. (2016). NaLi-H1: A universal synthetic library of humanized nanobodies providing highly functional antibodies

and intrabodies. *Elife, 5*. doi:10.7554/eLife.16228

Muyldermans, S. (2013). Nanobodies: natural single-domain antibodies. *Annu Rev Biochem, 82*, 775-797. doi:10.1146/annurev-biochem-063011-092449

Perchiacca, J. M., Bhattacharya, M., & Tessier, P. M. (2011). Mutational analysis of domain antibodies reveals aggregation hotspots within and near the complementarity determining regions. *Proteins, 79*(9), 2637-2647. doi:10.1002/prot.23085

Perchiacca, J. M., Ladiwala, A. R., Bhattacharya, M., & Tessier, P. M. (2012). Aggregation-resistant domain antibodies engineered with charged mutations near the edges of the complementarity-determining regions. *Protein Eng Des Sel, 25*(10), 591-601. doi:10.1093/protein/gzs042

Persson, H., Ye, W., Wernimont, A., Adams, J. J., Koide, A., Koide, S., . . . Sidhu, S. S. (2013). CDR-H3 diversity is not required for antigen recognition by synthetic antibodies. *J Mol Biol, 425*(4), 803-811. doi:10.1016/j.jmb.2012.11.037

Porebski, B. T., & Buckle, A. M. (2016). Consensus protein design. *Protein Eng Des Sel, 29*(7), 245-251. doi:10.1093/protein/gzw015

Prassler, J., Thiel, S., Pracht, C., Polzer, A., Peters, S., Bauer, M., . . . Enzelberger, M. (2011). HuCAL PLATINUM, a synthetic Fab library optimized for sequence diversity and superior performance in mammalian expression systems. *J Mol Biol, 413*(1), 261-278. doi:10.1016/j.jmb.2011.08.012

Rabia, L. A., Zhang, Y., Ludwig, S. D., Julian, M. C., & Tessier, P. M. (2018). Net charge of antibody complementarity-determining regions is a key predictor of specificity. *Protein Eng Des Sel, 31*(11), 409-418. doi:10.1093/protein/gzz002

Reichmann, D., Rahat, O., Cohen, M., Neuvirth, H., & Schreiber, G. (2007). The molecular architecture of protein-protein binding sites. *Curr Opin Struct Biol, 17*(1), 67-76. doi:10.1016/j.sbi.2007.01.004

Rothe, C., Urlinger, S., Lohning, C., Prassler, J., Stark, Y., Jager, U., . . . Urban, M. (2008). The human combinatorial antibody library HuCAL GOLD combines diversification of all six CDRs according to the natural immune system with a novel display method for efficient selection of high-affinity antibodies. *J Mol Biol, 376*(4), 1182-1200. doi:10.1016/j.jmb.2007.12.018

Shiraki, K., Kudou, M., Fujiwara, S., Imanaka, T., & Takagi, M. (2002). Biophysical effect of amino acids on the prevention of protein aggregation. *J Biochem, 132*(4), 591-595. doi:10.1093/oxfordjournals.jbchem.a003261

Sidhu, S. S., & Fellouse, F. A. (2006). Synthetic therapeutic antibodies. *Nat Chem Biol, 2*(12), 682-688. doi:10.1038/nchembio843

Singh, S. K. (2011). Impact of product-related factors on immunogenicity of biotherapeutics. *J Pharm Sci, 100*(2), 354-387. doi:10.1002/jps.22276

Tiller, T., Schuster, I., Deppe, D., Siegers, K., Strohner, R., Herrmann, T., . . . Urlinger, S. (2013). A fully synthetic human Fab antibody library based on fixed VH/VL framework pairings with favorable biophysical properties. *MAbs, 5*(3), 445-470. doi:10.4161/mabs.24218

Tonikian, R., Zhang, Y., Boone, C., & Sidhu, S. S. (2007). Identifying specificity profiles for peptide recognition modules from phage-displayed peptide libraries. *Nat Protoc, 2*(6), 1368-1386. doi:10.1038/nprot.2007.151

Villar, H. O., & Kauvar, L. M. (1994). Amino acid preferences at protein binding sites. *FEBS Lett, 349*(1), 125-130. doi:10.1016/0014-5793(94)00648-2

Virnekas, B., Ge, L., Pluckthun, A., Schneider, K. C., Wellnhofer, G., & Moroney, S. E. (1994). Trinucleotide phosphoramidites: ideal reagents for the synthesis of mixed oligonucleotides for random mutagenesis. *Nucleic Acids Res, 22*(25), 5600-5607. doi:10.1093/nar/22.25.5600

Figure legends

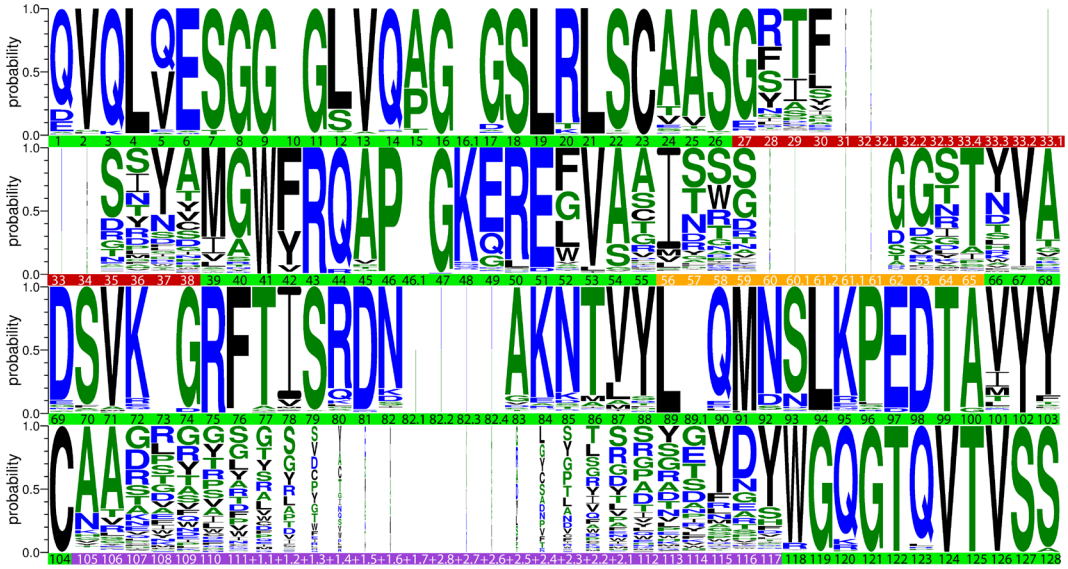


Figure 1. The logo of 296 nanobody sequences. IMGT numbering was applied. FR1 is from position 1 to 26, which is green background; CDR1 is from 27 to 38, wine red; FR2 is from 39 to 55, green; CDR2 is from 56 to 65, brown; FR3 is from 66 to 104, green; CDR3 is from 105 to 117, purple; FR4 is from 118 to 128, green.

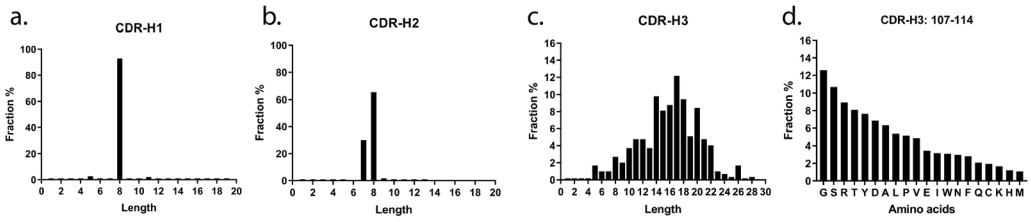


Figure 2. CDRs length distribution and CDR3 amino acid usage. (a) The length distribution of CDR1, over 90% was 8, mean was 8.03. (b) CDR2 length profile, around 65% was 8, about 30% was 7, mean was 7.77. (c) CDR3 length was broad, from 5 to 28, the distribution fitted into a bell curve, mean was 15.84. (d) Amino acid percentage of CDR3 from 107 to 114, glysine and serine were on the top of two, followed by arginine, threonine, tyrosine, and aspartate. Cysteine, lysine, histidine, and methionine took the least percentage.

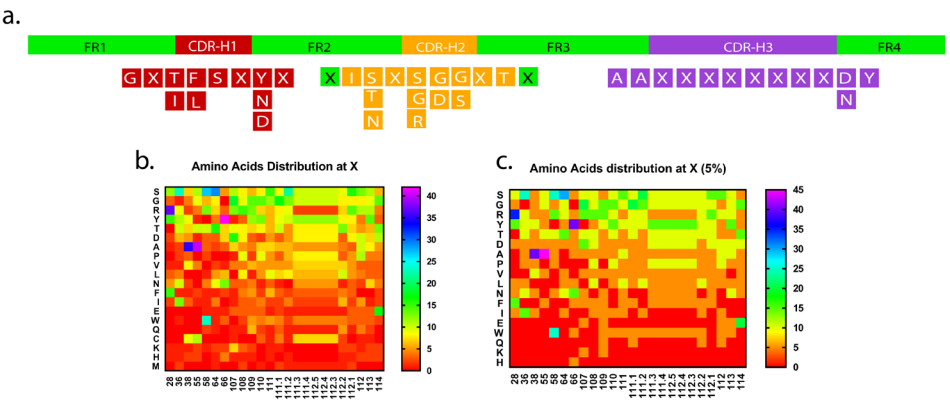


Figure 3. Structure alignment. To get a deep view of CDR3 structure characteristics, all nanobodies are grouped according to the CDR3 length, as length elongating, from 5 to 28, CDR3 gradually bent to FR2, in the end, covered it to form a hydrophobic core.

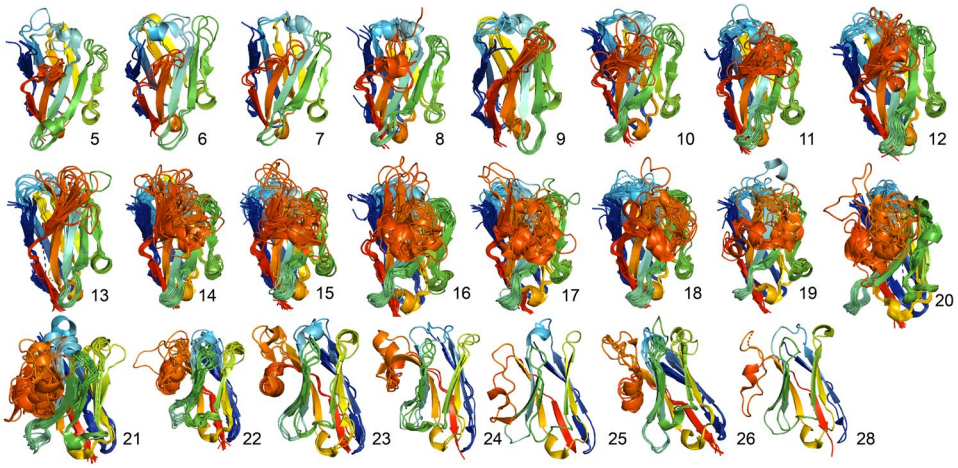


Figure 4. NanoOri_1.0 design. (a) CDR1, CDR2, and CDR3 were mutated by degenerate codon or by fixed-percentage amino acids. (b) The amino acid distribution at X position in PDB. (c) The percentage of amino acids was adjusted to the multiple of 5%.

No.	Target	pI (Target)	CDR1	CDR2	CDR3	Length	Net charge	pI (VHH)	Monomer
Caplacizumab	vWf	5.36	GRTFSYNP	AISRTGGSTY	AAAGVRAEDGRVRTLPSEYTF	21	+2	9.07	100.00%
4A4	A	3.71	GFIFSRNA	AINWSGSSTF	AAGSRTNTGTDWSRTRYTYNY	21	+3	9.34	100.00%
3G4			GSIFSTYR	GISPGGSDTY	AATPGRFALLSPRGYNY	17	+2	9.04	100.00%
4A2			GFIFSRYD	VITWGGGSTY	AARSASAYRTGRLEYNY	17	+2	9.01	100.00%
3E9			GFIFSRYD	AINWGDSSTY	AARRYYSYYPSSGGYNY	17	+1	8.60	56.87%
4F10			GFIFSLNA	AISPGDGTTY	AARRRSLSGYGYNY	14	+2	9.04	98.73%
3F5			GFILSRYV	FITWSDGSTY	AAKSRTYRGGLYDY	14	+2	8.99	100.00%
4B3			GLIFSRYV	FINSSGGSTN	AASSRTSTTRGYNY	14	+3	9.34	100.00%
3B5			GFIFSRYG	AINWSDGTTY	AADRRLRSQTGYDY	14	+1	8.67	97.53%
VHH1	B	9.04	GRTFSYNP	AISRTGGSTY	AAGDVSWFLPWYDTSSSEYNY	21	-1	6.51	96.02%
VHH3			GRTFSPYA	AISWRGSNTY	AAFREDFFEADYNY	14	-2	6.54	52.04%
VHH5			GSTFSSYD	AISARGGNTF	AAFGRSFTAPPYDY	14	0	8.06	97.45%
VHH8			GITLSSYV	AISAGGGSTW	AADDISSFLSGYDY	14	-3	4.92	62.04%
VHH13			GRTFSYDA	YISAGGGSTY	AATDIPFGVWLTYDGYDTYNY	21	-3	4.96	55.26%
VHH16			GFIFSNYV	AISRTGGSTY	AADDSSFFGRAYDY	14	-1	6.49	98.81%
VHH17			GSTFSADY	AISSSDGYTW	AAGYFVSYWTEYDY	14	-4	4.76	71.69%
VHH18			GSIFSIYT	AISRTGGSTY	AADDFWSSDAGYNY	14	-2	5.28	84.10%

Table 1 The relationship between pI and aggregation of nanobodies. pI was theoretically computed on the website (https://web.expasy.org/compute_pi/). The net charge was calculated by assuming the charges of glutamate (-1), aspartate (-1), arginine (+1), lysine (+1)(Rabia et al., 2018), there was no Histidine in CDRs.

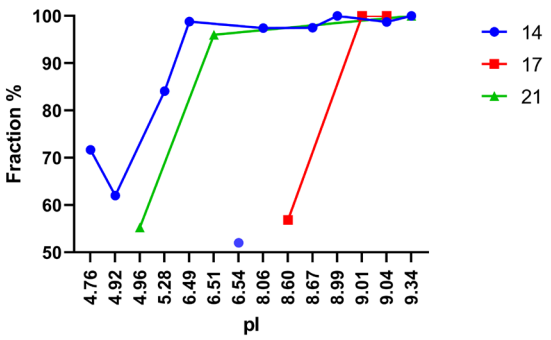


Figure 5. SEC analysis of nanobodies at 1 mg/ml in PBS buffer. To analyze the relationship between pI and aggregation, all nanobodies were classified by CDR3 length. For length 14, 17, and 21, high-pIs showed more resistant to aggregation than low-pIs.

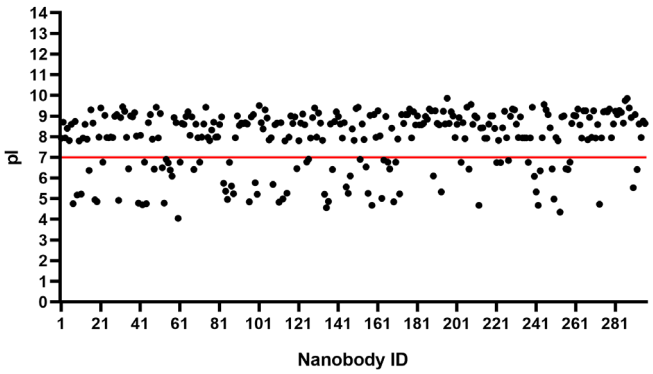


Figure 6. Theoretical pI distribution of 296 nanobodies from PDB. High-pI nanobody outnumbered low-pIs, which pI>7 took up about 75% (223/296), pI<7 was only 25% (73/296).