BMC Research Notes

RESEARCH NOTE

Open Access

Structure-based design and construction of a synthetic phage display nanobody library

Ernesto Moreno[1*], Mario S. Valdés-Tresanco[1], Andrea Molina-Zapata[1,2] and Oliberto Sánchez-Ramos[3]

Abstract

Objective: To design and construct a new synthetic nanobody library using a structure-based approach that seeks to maintain high protein stability and increase the number of functional variants within the combinatorial space of mutations.

Results: Synthetic nanobody (Nb) libraries are emerging as an attractive alternative to animal immunization for the selection of stable, high affinity Nbs. Two key features define a synthetic Nb library: framework selection and CDR design. We selected the universal VHH framework from the cAbBCII10 Nb. CDR1 and CDR2 were designed with the same fixed length as in cAbBCII10, while for CDR3 we chose a 14-long loop, which creates a convex binding site topology. Based on the analysis of the cAbBCII10 crystal structure, we carefully selected the positions to be randomized and tailored the codon usage at each position, keeping at particular places amino acids that guarantee stability, favoring properties like polarity at solvent-exposed positions and avoiding destabilizing amino acids. Gene synthesis and library construction were carried out by GenScript, using our own phagemid vector. The constructed library has an estimated size of 1.75×10^8. NGS showed that the amino acid diversity and frequency at each randomized position are the expected from the codon usage.

Keywords: Nanobody, Synthetic nanobody library, Phage display, Structure-based design

Introduction

Nanobodies (Nbs) are single domain antibody fragments derived from heavy-chain antibodies found in members of the Camelidae family. They have several important advantages such as their small size, high solubility and stability. Remarkably, in spite of their monomeric binding region composed only of three hypervariable loops or CDRs (complementarity determining regions), Nbs can achieve affinities in the nanomolar and subnanomolar orders, just as clasical antibodies. Their modularity allows the easy generation of multivalent constructs for many different applications [1].

Antigen-specific Nbs are obtained mostly from inmune libraries, which requires animal immunization. In recent years, however, synthetic libraries are emerging as an attractive alternative for the selection of stable, high affinity Nbs, as recently reviewed by our group [2], while offering important advantages in terms of cost and speed. Two key features define a synthetic Nb library: framework selection and CDR design. A few recent works have relied on both sequence and structural data to define the CDR positions to be randomized, as well as the sets of amino acids (aa) to be introduced at those positions [3–9].

Here we describe the design and construction of a new synthetic nanobody library with a 14 amino acid-long CDR3, in which the three hypervariable loops were subjected to position-specific randomization schemes. The design follows a structure-based approach that seeks to maintain the high stability shown by the original

*Correspondence: emoreno@udemedellin.edu.co; ernesto_moreno_frias@yahoo.com
[1] Faculty of Basic Sciences, University of Medellin, Medellín, Colombia
Full list of author information is available at the end of the article

© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

framework-donor nanobody and increase the number of functional variants within the combinatorial space of mutations. To this aim we analyzed the structural role played by individual residues in defining CDR conformation or exposing their side chains for antigen binding.

Main text

Library design and construction

Framework selection

The synthetic library should be built on a scaffold shown to be highly stable, in particular to reducing conditions (as in the cytoplasm). Also, the framework should be capable of accepting many different CDRs while keeping high stability. Since current algorithms to predict protein stability still have important limitations [10], the safest option is to choose a camelid heavy-chain variable domain (VHH) framework already proven to possess the desired characteristics. We selected the "universal VHH framework" from the cAbBCII10 nanobody [11], which has been used previously as scaffold for the construction of Nb libraries [12–14] and not only its sequence is available, but also a few crystal structures containing CDR-grafted, the original and humanized versions of its framework (PDB entries 1ZMY [15], 3DWT and 3EAK [16], respectively.

CDR1 and CDR2 design

A main issue in our design was to keep the correct immunoglobulin domain folding and guarantee its stability, for which we thought important to keep the original cAbBCII10 internal packing as well as CDR1/CDR2 conformations. In consequence, both CDR1 and CDR2 were designed with a fixed length—the same as in cAbBCII10. Analysis of the available crystal structures allowed to assess the structural role played by different CDR residues and so guide the design. Buried CDR amino acids were kept as in the original nanobody since they are most likely important in maintaining the high framework stability. Therefore, only selected residues with surface-exposed side chains were subjected to randomization. Sequence variability was included by means of degenerate codons, represented in the sequence using the IUPAC degenerate base symbols [17]. Codon usage was tailored to limit the presence of hydrophobic amino acids at these positions. Cysteines were not allowed at any position. A few CDR aa found to be highly or relatively conserved in nanobody sequences, or thought to be important in holding CDR conformation in cAbBCII10, were also kept unchanged. It should be noted that our CDR definition follows that of Chothia and coworkers, which for CDR1 includes positions outside the classical Kabat CDR definition [18].

Figure 1A and B shows the sequence positions selected for randomization in each of the CDRs, marked on the sequence of the original cAbBCII10 and on its crystal structure. The rationale followed to define whether or not a CDR position is randomized, as well as the implemented codon variability for each position, is presented in Table 1. It is worth noting that, as for other libraries described in the literature, our design is based on premises involving a degree of subjectivity.

CDR3 design

For this library we chose a 14-long CDR3, which creates a "convex" binding site topology by bending over the flank that in classical antibodies is buried at the VH-VL interface [19, 20]. This represents an important difference as compared with a previously constructed library, carrying a 10 aa-long CDR3, as we analyze in the following section.

Three degenerate codons: VRN (coding for 8 polar or charged aa, and Gly), WMY (coding for Asn, Ser, Thr and Tyr in equal proportions) and the highly variable VNN (coding for 16 aa, excluding Cys, Phe, Trp and Tyr) alternate along most of CDR3, starting from its N-terminus (Table 1). The presence of five WMY codons places small aa (Asn, Ser, Thr) and Tyr every two positions, conferring conformational variability, while the high representation of Arg + Lys within the VRN and VNN codons increments their frequency in the alternating positions.

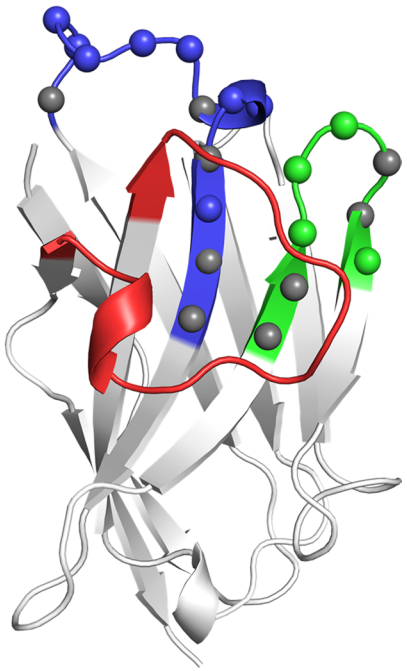
For the three C-terminal CDR3 residues ("n-2", "n-1" and "n") we chose particular degenerate codons to account for the aa frequency observed at these positions in nanobodies with long CDR3 loops, as well as for the structural role played by the particular amino acids found in the original cAbBCII10 Nb: Phe, Arg and Tyr (positions 115, 116 and 117, respectively).

For position "n" (117) we chose TMY, coding in equal proportions for Ser and Tyr—the two aa most commonly found at this position, with Tyr being the most frequent. In cAbBCII10, Tyr remains mostly packed with its OH group exposed. The side chain at the "n-1" position (116) is exposed to the solvent, so we chose the VRN codon. This residue holds also a backbone turn that bends down CDR3 on its C-terminal side. Finally, at the "n-2" position (115) in cAbBCII10, Phe is buried against a hydrophobic patch a the Nb flank, most likely playing an important structural role in the folding of CDR3. For this position we limited the repertoire to only two aa—Phe and Tyr (codon TWY). In fact, Tyr is the most frequent aa at this position. The final nucleotide sequence for the whole nanobody is shown in Fig. 1C.

A

```
Sequential          22      30    36
                     |       |     |
QVQLVESGGGSVQAGGSLRLSCTASGGSEYSYSTFSLGWFRQAPGQEREA
   55   60
    |    |
VAAIASMGGLTYYADSVKGRFTISRDNAKNTVTLQMNNLKPEDTAIY
99                 118
 |                  |
YCAAVRGYFMRLPSSHNFRYWGQGTQVTVSS
```

B



C

```
ATGCGAAGACTccatggGCCAGGTGCAACTGGTCGAGAGTGGCGGCGGCTCCGT
CCAAGCTGGTGGCTCGCTTCGCTTGTCCTGCACGGCTTCGGGT DMY VRN
VRN HWY VRN TAC DMY VNN TTC KMY TTAGGCTGGTTTCGCCAAGCAC
CAGGACAGGAACGTGAGGCTGTAGCCGCAATC KMY WBG VRN GGTGGA
ADM ACTTACTACGCAGATTCCGTAAAAGGCCGCTTTACAATTTCGCGTGATAA
CGCCAAGAATACCGTGACATTGCAAATGAATAACTTGAAACCGGTAGACACTGC
CATTTACTATTGCGCGGCC VNN VRN WMY VRN WMY VNN WMY VRN
WMY VNN WMY TWY VRN TMY TGGGGACAGGGAACCCAGGTCACCGTGTCT
TCTgcggccgcGTCT TCGATC
```

Fig. 1 Sequence and structural basis for synthetic library design. **A** cAbBCII10 amino acid sequence showing the CDRs (underlined) and randomized positions (highlighted in gray). **B** cAbBCII10 structure (PDB: 3DWT). CDRs 1, 2 and 3 are colored in blue, green, and red, respectively. Colored spheres in CDRs 1 and 2 represent the randomized positions, while gray spheres represent CDR positions that were kept fixed. **C** The final nucleotide sequence for the whole nanobody, including the NcoI and NotI cloning sites

Library construction and verification

Nanobody genes were cloned into our *ad-hoc* designed pMAC phagemid vector, whose map is shown in Additional file 1. This vector includes a pelB leader whose coding sequence contains a NcoI restriction site at the 3′ end. Other 3 unique restriction sites (EcoRI, BamHI and NotI) were added in tandem, followed by a sequence coding for a short linker (SGGGG) and a 6xHis tag, an amber stop codon and, finally, the M13 PIII protein. The synthesis of nanobody genes, library construction and next-generation sequencing (NGS) verification were carried out by GenScript (NJ, USA). Nb genes were cloned into the pMAC vector using the NcoI and NotI restriction sites, followed by transformation by electroporation of SS320 E. coli cells. The estimated library size was 1.75×10^8.

Figure 2A shows the obtained aa frequencies, as assessed by NGS, for all the randomized positions, together with the theoretically designed variability. At every position, the aa diversity is the expected from the codon usage, while the experimental aa frequencies are close to the theoretical ones.

Structural analysis of CDR3 conformations

In a previous work we constructed and tested against a panel of antigens a synthetic Nb library based also on the cAbBCII10 framework, but having an important structural difference with the one reported here: a shorter, 10 aa-long CDR3 (*manuscript in preparation*). It has been shown that different CDR3 lenghts create different binding site topologies [5]. Here we assessed these structural differences for the two implemented CDR3 lenghts (10 and 14 aa), using the nanobody structural data available in the Protein Data Bank (PDB) [21, 22], which has been rapidly growing in the last few years.

Currently, the PDB contains >600 entries including a nanobody structure, either alone or in complex with an antigen [23]. Figure 2 shows the superposition of 33 and 37 nanobody crystal structures we found with either 10 (CDR3-10) or 14 aa-long (CDR3-14) CDR3 loops. CDR3-10 loops display a variety of conformations, going from the "upright" geometry adopted by most CDR H3 loops in classical antibodies to slightly or entirely bent loops (Fig. 2B). CDR-14 loops, on the other hand, adopt mostly bent conformations, with relatively few exceptions (Fig. 2C). Additionally, we used the NanoNet program [24] to construct structural models of 10,000 sequences of the CDR3-14 library, randomly generated based on the aa frequencies yielded by the degenerate codons used at each randomized position. Figure 2D shows the superposition of 100 randomly selected models, all of them

Table 1 Rationale for CDR design

Sequence position	Original aa	Degenerate codon	Coded aa	Analysis
CDR1				
26	Gly	–	Gly	Conserved (our own analysis). An outlier in the Ramachandran plot. Keep fixed
27	Gly	DMY	Ala:2 Asn:2 Asp:2 Ser:2 Thr:2 Tyr:2	An outlier in the Ramachandran plot. Randomize, but keep it small. The degenerate codon includes also Tyr
28	Ser	VRN	Arg:6 Asn:2 Asp:2 Gln:2 Glu:2 Gly:4 His:2 Lys:2 Ser:2	Surface-exposed side chain. Set it variable, but polar
29	Glu			
30	Tyr	HWY	Asn:2 His:2 Ile:2 Leu:2 Phe:2 Tyr:2	It is packed against the extremes of CDR1. Keep it mostly big and hydrophobic/aromatic
31	Ser	VRN	*see above*	Surface-exposed side chain. Set it variable, but polar
32	Tyr	–	Tyr	Buried aa playing an important packing role. Keep fixed
33	Ser	DMY	*see above*	It's placed on an alpha turn, outside the binding site. Keep it mostly small, as the original aa
34	Thr	VNN	Ala:4 Arg:6 Asn:2 Asp:2 Gln:2 Glu:2 Gly:4 His:2 Ile:3 Leu:4 Lys:2 Met:1 Pro:4 Ser:2 Thr:4 Val:4	Points "upwards" at the base of CDR1. Set it very variable
35	Phe	–	Phe	Buried aa playing an important packing role. Keep fixed
36	Ser	KMY	Ala:2 Asp:2 Ser:2 Tyr:2	Support for CDR H3 (found also making a Cys bridge with CDR H3). Becomes buried by long CDRs 3. Small aa are commonly found, as well as Tyr
CDR2				
53	Ala	–	Ala	Within a β-sheet. Packed with the long CDR3. Keep fixed
54	Ile	–	Ile	Very conserved, buried aa playing an important packing role. Keep fixed
55	Ala	KMY	*see above*	Interacting with the bent CDR3 in the original Nb. Ser and Tyr are commonly found, as well as small aa
56	Ser	WBG	Arg:1 Leu:1 Met:1 Ser:1 Thr:1 Trp:1	Small aa side chains, like Ser and Thr, remain partially buried. The larger amino acids in the set would pack against CDR1 and the N-ter of CDR3
57	Met	VRN	*see above*	Surface-exposed side chain. Set it variable, but polar
58, 59	Gly	–	Gly	Conserved pair of residues, outliers in the Ramachandran plot. Keep fixed
60	Leu	ADM	Arg:1 Asn:1 Ile:2 Lys:1 Ser:1	Combines long charged aa. that can protrude upwards with small polar aa. Arg, Asn Ile and Ser are among the most frequently found aa at this position
CDR3				
VNN VRN WMY VRN WMY VNN WMY VRN WMY VNN WMY TWY VRN TMY				WMY: Asn:2 Ser:2 Thr:2 Tyr:2 TWY: Phe:2 Tyr:2 TMY: Ser:2 Tyr:2

IUPAC degenerate base symbols [14]:

B = C, G, T	H = A, C, T	M = A, C	R = A, G	V = A, C, G	Y = C, T	D = A, G, T	K = G, T	N = A, C, G, T	S = C, G	W = A, T

Numbers in bold correspond to CDR positions that were randomized

displaying a bent CDR3 conformation. Overall, these analyses indicate that the two libraries display different binding site topologies, which may account for different "preferences" regarding antigen shape.

The huge amino acid variability introduced in the library creates also a huge variability in interaction networks between CDR amino acids, as well as between CDR and framework amino acids. For the longer CDR3-14 loops, the alternation of different hydrophobic and

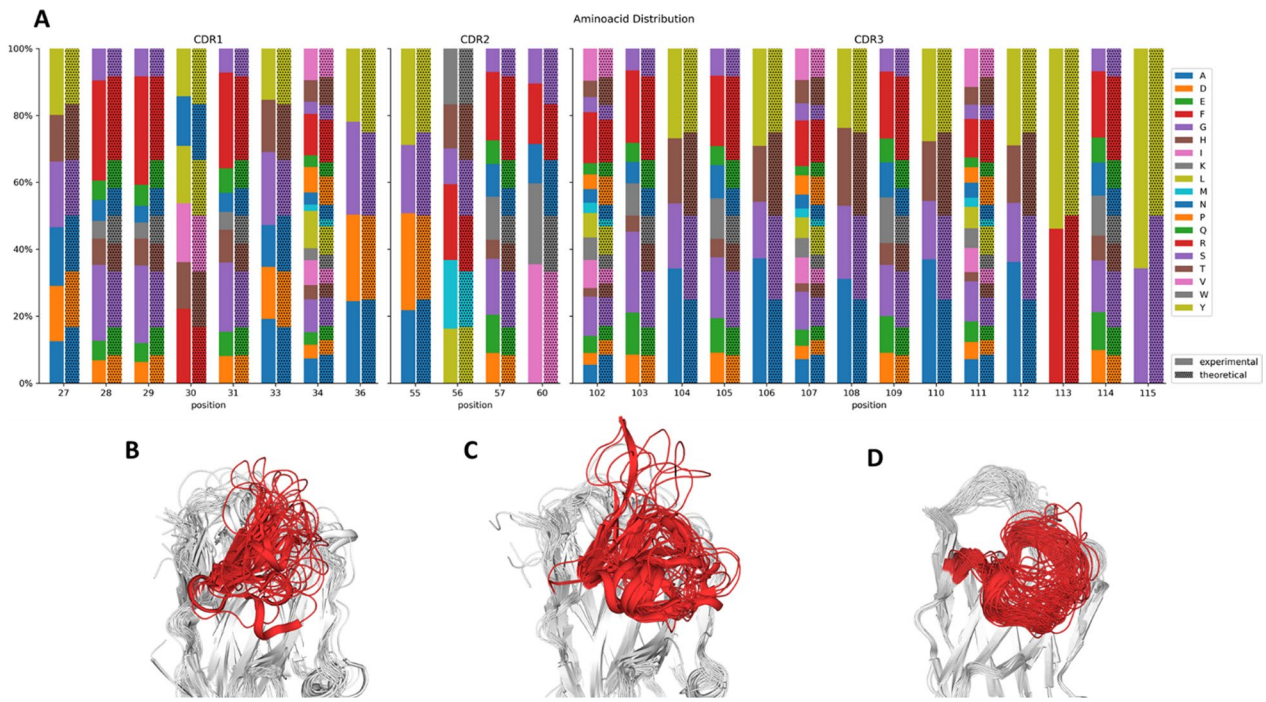


Fig. 2 **A** Amino acid frequencies per randomized position for the constructed library. Each pair of bars represents the experimental frequencies obtained by NGS and the corresponding theoretical (designed) frequencies. **B**, **C** Structural superposition of 33 and 37 Nb crystal structures showing 10 aa-long and 14 aa-long CDR3 loops, respectively. **D** Structural superposition of 100 randomly selected library Nbs modeled with NanoNet

hydrophilic amino acids may define its particular bended conformation and also exert different effects in protein stability, because of the interactions of this loop with the framework flank that in conventional antibodies interact with the light chain VL domain.

Concluding remarks

By carefully selecting the positions to be randomized and tailoring the codon usage for each position, we sought to increase the number of functional combinations within the huge theoretical combinatory, keeping at particular positions amino acids that guarantee stability, favoring properties like polarity at solvent-exposed positions and avoiding destabilizing aa at certain positions. The design strategy presented here may be used on other stable frameworks with different CDR1-3 lengths.

Limitations

We have yet to demonstrate that the constructed library can provide stable, high affinity nanobody binders for different antigens. This experimental work, though, has advanced for the CDR3-10 library, with very encouraging results (manuscript in preparation).

Abbreviations

Nb: Nanobody; aa: Amino acid(s); CDR: Complementarity determining region; NGS: Next-generation sequencing; VHH: Camelid heavy-chain variable domain.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13104-022-06001-7.

Additional file 1. Map of the designed pMAC phagemid vector.

Acknowledgements

Not applicable.

Authors' contributions

EM designed the library and wrote the manuscript. MSVT performed sequence and structural analyses and contributed to manuscript preparation. AMZ carried out experimental work to test the phagemid vector. OSR designed the pMAC vector. All authors read and approved the final manuscript.

Funding

Library construction and work by AMZ was supported by Minciencias (Colombia) (Grant No. 849-2017).

Availability of data and materials

The data generated or analyzed during this study were included in this article and its additional file.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Faculty of Basic Sciences, University of Medellin, Medellín, Colombia. [2]Grupo de Micología Médica y Experimental, Corporación para Investigaciones Biológicas (CIB), Medellín, Colombia. [3]Biological Sciences School, University of Concepción, Concepción, Chile.

Received: 15 December 2021 Accepted: 14 March 2022
Published online: 29 March 2022

References

1. Muyldermans S. A guide to: generation and design of nanobodies. FEBS J. 2021;288(7):2084.
2. Valdés-Tresanco MS, Molina-Zapata A, González Pose A, Moreno E. Structural insights into the design of synthetic nanobody libraries. Molecules. 2022 (*manuscript accepted*)
3. Moutel S, Bery N, Bernard V, Keller L, Lemesre E, de Marco A, et al. NaLi-H1: a universal synthetic library of humanized nanobodies providing highly functional antibodies and intrabodies. eLife. 2016;5(JULY):1–31.
4. McMahon C, Baier AS, Pascolutti R, Wegrecki M, Zheng S, Ong JX, et al. Yeast surface display platform for rapid discovery of conformationally selective nanobodies. Nat Struc Mol Biol. 2018. https://doi.org/10.1038/s41594-018-0028-6.
5. Zimmermann I, Egloff P, Hutter CAJ, Arnold FM, Stohler P, Bocquet N, et al. Synthetic single domain antibodies for the conformational trapping of membrane proteins. eLife. 2018. https://doi.org/10.7554/eLife.34317.
6. Sevy AM, Chen MT, Castor M, Sylvia T, Krishnamurthy H, Ishchenko A, Hsieh CM. Structure- and sequence-based design of synthetic single-domain antibody libraries. Protein Eng Des Sel. 2020;33:1–13. https://doi.org/10.1093/PROTEIN/GZAA028.
7. Zimmermann I, Egloff P, Hutter CAJ, Kuhn BT, Bräuer P, Newstead S, et al. Generation of synthetic nanobodies against delicate proteins. Nat Protoc. 2020;15(5):1707–41.
8. Zhao Y, Wang Y, Su W, Li S. construction of synthetic nanobody library in mammalian cells by DsDNA-based strategies. ChemBioChem. 2021;22:2957–65. https://doi.org/10.1002/CBIC.202100286.
9. Chen X, Gentili M, Hacohen N, Regev A. A cell-free nanobody engineering platform rapidly generates SARS-CoV-2 neutralizing nanobodies. Nat Commun. 2021;12(1):1–14.
10. Sanavia T, Birolo G, Montanucci L, Turina P, Capriotti E, Fariselli P. Limitations and challenges in protein stability prediction upon genome variations: towards future applications in precision medicine. Comput Struct Biotechnol J. 2020;18:1968–79.
11. Conrath KE, Lauwereys M, Galleni M, Matagne A, Frère J-M, Kinne J, et al. β-Lactamase inhibitors derived from single-domain antibody fragments elicited in the Camelidae. Antimicrob Agents Chemother. 2001;45(10):2807.
12. Wei G, Meng W, Guo H, Pan W, Liu J, Peng T, et al. Potent neutralization of influenza A virus by a single-domain antibody blocking M2 ion channel protein. PloS ONE. 2011;6(12). https://pubmed.ncbi.nlm.nih.gov/22164266/. Accessed 27 Nov 2021.
13. Yan J, Li G, Hu Y, Ou W, Wan Y. Construction of a synthetic phage-displayed Nanobody library with CDR3 regions randomized by trinucleotide cassettes for diagnostic applications. J Transl Med. 2014;12(1):1–12. https://doi.org/10.1186/s12967-014-0343-6.
14. Chi X, Liu X, Wang C, Zhang X, Li X, Hou J, et al. Humanized single domain antibodies neutralize SARS-CoV-2 by targeting the spike receptor binding domain. Nat Commun. 2020. https://doi.org/10.1038/s41467-020-18387-8.
15. Saerens D, Pellis M, Loris R, Pardon E, Dumoulin M, Matagne A, et al. Identification of a universal VHH framework to graft non-canonical antigen-binding loops of camel single-domain antibodies. J Mol Biol. 2005;352(3):597–607.
16. 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.
17. Cornish-Bowden A. Nomenclature for incompletely specified bases in nucleic acid sequences: recommendations 1984. Nucleic Acids Res. 1985;13(9):3021.
18. Al-Lazikani B, Lesk AM, Chothia C. Standard conformations for the canonical structures of immunoglobulins. J Mol Biol. 1997;273(4):927–48.
19. Mitchell LS, Colwell LJ. Comparative analysis of nanobody sequence and structure data. Proteins. 2018;86(7):697–706.
20. Mitchell LS, Colwell LJ. Analysis of nanobody paratopes reveals greater diversity than classical antibodies. Protein Eng Des Sel. 2018;31(7–8):267–75.
21. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The protein data bank. Nucleic Acids Res. 2000. http://www.rcsb.org/pdb/status.html. Accessed 11 May 2020.
22. Burley SK, Berman HM, Bhikadiya C, Bi C, Chen L, di Costanzo L, et al. RCSB protein data bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. Nucleic Acids Res. 2019;47(D1):D464–74.
23. Dunbar J, Krawczyk K, Leem J, Baker T, Fuchs A, Georges G, et al. SAbDab: the structural antibody database. Nucleic Acids Res. 2014;42(D1):D1140–6.
24. Cohen T, Halfon M, Schneidman-Duhovny D, Rachel T, Benin S. NanoNet: rapid end-to-end nanobody modeling by deep learning at sub angstrom resolution. bioRxiv. 2021. https://doi.org/10.1101/2021.08.03.454917v1.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions