Distinct types of VHHs in Alpaca

Xinhao Wang[1*], Lu Zhang[2], Yao Zhang[2], Jiaguo Li[2], Wenfeng Xu[1], and Weimin Zhu[2]
1. Chantibody Therapeutics Inc.; 2. Shanghai Cell Therapy Group Co. Ltd

VHHs (VH of heavy-chain-only antibodies) represent a unique alternative to conventional antibodies because of their smaller size, comparable binding affinity and biophysical properties. In this study, we systematically analyzed VHH NGS sequences from 22 Alpacas and structure data from public database, demonstrating that VHHs in Alpaca can be grouped into five main types with multiple distinct sequence and structure features. Based on the existence of hallmark residues in FR2 region, VHHs can be classified into two groups: non-classical VHHs (without hallmark residues) and classical VHHs (with hallmark residues). Based on VHH hallmark residues at 42 position (IMGT numbering, FR2 region) and number of cysteines, we found that Alpaca classical VHHs can be further separated into three main types: F_C2 VHHs with F (phenylalanine) at position 42 and having 2 cysteines within sequences, Y_C2 VHHs with Y(tyrosine) at position 42 and having 2 cysteines, and F_C4 with F at position 42 and having 4 cysteines. Non-classical VHHs can be further separated into 2 types based on germlines mapped: N_V3 for VHHs mapped to V3 germlines and N_V4 for V4 germlines. Based on whether FR2 residues are involved in binding, two kinds of paratopes can be identified. Different types of VHHs showed distinct associations with these two paratopes and displayed significant differences in paratope size, residue usage and other structure features. Such results will have significant implications in VHH discovery, engineering, and design for innovative therapeutics.

1. 5 distinct types of VHHs in Alpaca with unique sequence features

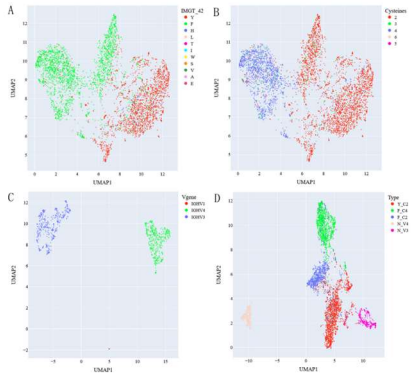

FIGURE 1 UMAP graph of top 5000 classical VHHs labeled with residue at IMGT 42 position (A) and number of cysteines (B) within sequences, top 1000 nonclassical VHHs labeled with types of mapped V genes (C). UMAP of classical and non-classical VHHs after removing non-assigned ones labeled with assigned types (D). AntiBERTy is used to generate sequence embeddings.

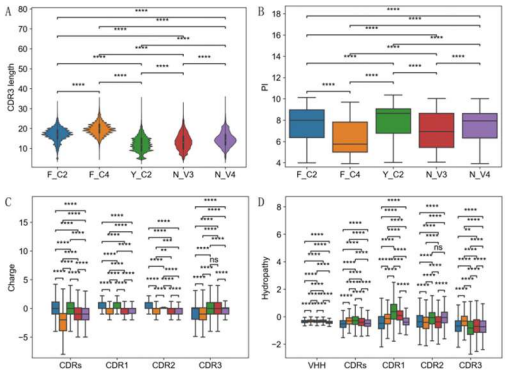

FIGURE 2 Sequence feature differences among 5 types of VHHs. Significant differences in all comparisons among 5 types were found for CDR3 length (A), PI (B). For charge (C), significant differences in CDR1/2/3 and CDR regions among 5 types were found, except CDR3 charge in Y_C2 and N_V3 comparison. For hydropathy (D), significant differences in CDR1/2/3/CDRs/VHH were found among 5 types, except CDR2 hydropathy in Y_C2 and N_V4 comparison.

2. Structure feature differences among different types of VHHs

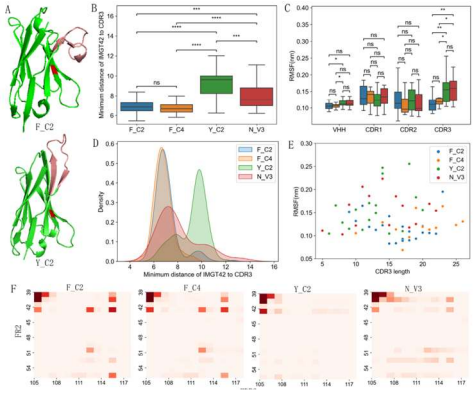

FIGURE 3 Structure feature differences among 4 types of VHHs. (A) Examples of VHHs with bent down (PDB ID: 6HJY) and extended (PDB ID: 6GKD) CDR3 (pink colored) conformation. Residue at IMGT 42 position is red colored. Y_C2 VHHs showed largest minimum distance between residue at 42 position and CDR3 among 4 types of VHHs based on both boxplot (B) and density map (D). Y_C2 and N_V3 showed larger flexibility in CDR3 as compared to other two types (C). No significant correlation was observed between flexibility and CDR3 length (E). Heat map showing the probability of interaction between FR2 and CDR3 residues (F). F_C2 and F_C4 VHHs showed more interactions between FR2 and CDR3 residues as compared to Y_C2 and N_V3.

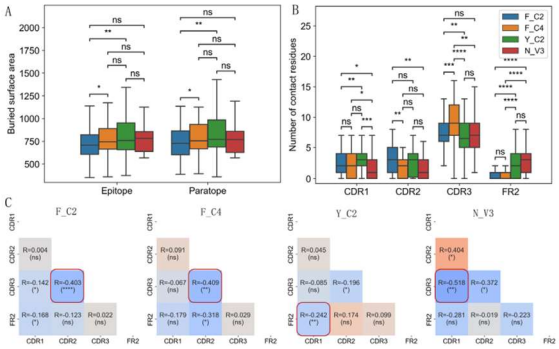

FIGURE 4 Interaction interface differences among 4 types of VHHs. Y_C2 VHHs showed slightly larger epitope and paratope size as measured by buried surface area (A). Number of contact residues in CDR1/2/3 and FR2 regions showed significant differences among 4 types of VHHs (B). The correlations of number of contact residues among 4 regions (CDR1, CDR2, CDR3, FR2) (C) showed distinct dependency of contact residues among 4 types of VHHs. Most significant correlation for each type is highlighted with red box. P-values are marked as followings: ns: P > 0.05; **: 0.001 <= P < 0.01; ***: 0.0001<= P < 0.001; ****: P <= 0.0001.

3. Two types of binding paratopes and their correlation with different types of VHHs

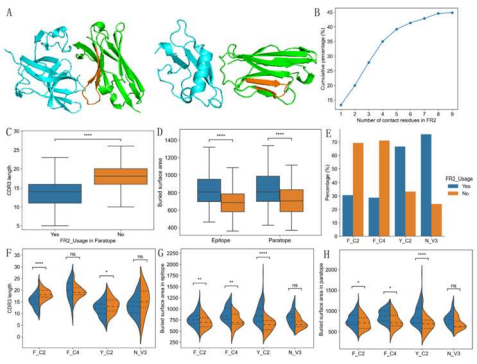

FIGURE 5 Two types of binding paratopes and their differences among different types of VHHs. Example structures with contact residues from FR2 (PDB ID: 7B5G, FR2 region orange colored) vs not (PDB ID: 5IP4, FR2 region orange colored) (A). Cumulative percentage of structures with different number of contact residues from FR2 (B). VHHs with contact residues in FR2 have significantly shorter CDR3 (C) and larger epitope and paratope sizes (D). Y_C2 and N_V3 are more likely to have FR2 contact residues than F_C2 and F_C4 VHHs (E). F_C2 and Y_C2 VHHs using FR2 in binding have shorter CDR3 (F). VHHs using FR2 in binding have larger epitope (G) and paratope (H) sizes.

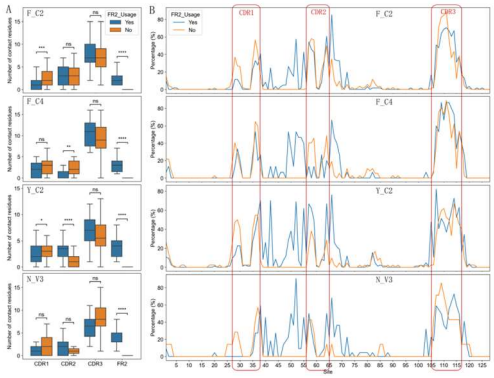

FIGURE 6 Comparison of number of contact residues in CDR1/2/3 and FR2 regions (A) and contact residue distribution (B) between two types of paratopes for each type of VHHs. The Y-axis (B) represents the percentage of VHHs involved in binding at a specified position.

4. Conclusions

In this study, we systematically analyzed different types of VHHs in Alpaca, identifiable using simple sequence features and showed distinct sequence and structure feature differences among them. Furthermore, we compared two kinds of paratopes in VHHs and their usage in different types of VHHs. We found that paratopes involving FR2 residues are used mostly by VHHs with short CDR3. This type of VHH may enable us to design novel therapies with distinct binding modalities.

Correspondance to: Xinhao.wang@chantibody.com
https://www.frontiersin.org/journals/immunology/articles/10.3389/fimmu.2024.1447212/full