ANALYSES OF V_HH DOMAIN DYNAMICS

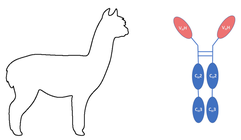
Akhila Melarkode Vattekatte[1], Julien Diharce[2], Catherine Etchebest[3], Alexandre G. de Brevern[4].

[1]DSIMB Laboratory, INSERM UMRS_1134, Biochemistry, University of Reunion Island, Saint Denis, France, [2]DSIMB, INSERM UMRS_1134, Univ Paris Cité, Paris, France, [3]DSIMB, INSERM UMRS_1134 Paris-Cité, Paris, France, [4]DSIMB, INSERM UMRS_1134 Univ Paris Cité, Paris, France.

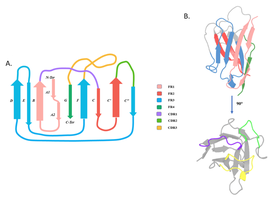
Abstract: Protein conformational flexibility is crucial for its structural stability and function. The concerted displacements of residues in an antigen-antibody complex facilitate and determine their interactions' strength, making their study indispensable to modulating their function. Members of the family *Camelidae* express a unique subset of Immunoglobulin Gamma called the Heavy Chain only Antibody (HCAbs), consisting of one Variable domain (V_HH) at the N-terminus of each heavy chain. Each V_HH domain comprises two types of amino acid regions varying in sequence identity arranged alternatingly called the Framework Regions (FRs) and Complementarity Determining Regions (CDRs). Even when expressed independently *in-vitro*, V_HH domains exhibit excellent solubility and thermostability compared to the V_H-V_L complexes, so they present a valuable opportunity to exploit their biophysical and biochemical properties to generate the next generation of therapeutic and diagnostic molecules. Recent studies have reported sequence and structural features of V_HH domains contributing to these abilities in comparison to classical V_H-V_L complexes. In this study, we performed large-scale classical molecular dynamics simulations for a dataset of unrelated V_HH structures to understand the local and global differences in their dynamics. We used classical metrics such as the Normalised B-factors of Cα atoms, RMSF of Cα atoms and an in-house method called the Protein Blocks (PBs) to investigate flexibility in V_HH domains and trajectories. We have classified the trajectories based on Cα Root Mean Squared Fluctuation, which revealed four main clusters of the V_HH trajectories. We observed various local changes in CDRs but within different ranges in trajectories within the same cluster as well as from other clusters. The FR-CDR boundary regions showed distinct local backbone conformational diversity when assessed using PBs. This study sheds light on region-wise changes in flexibility during dynamics which could aid in improving the design and function of V_HH domains.

Introduction

Heavy Chain only antibodies from Camelids and their Variable domains (V_HH).



2D and 3D organisation of a V_HH domain



- Are the FRs as conformationally similar as they have been previously observed?
- Are the CDRs as conformationally diverse as they have been previously observed?
- Can we have other perspectives on flexibility in the V_HH domains?

Methods:

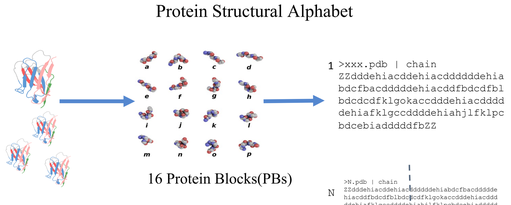
We used three different flexibility metrics to dissect conformational changes in VHH

1: Normalised B-factors

2: Cα RMSF

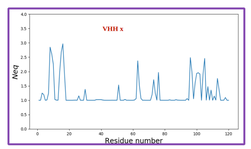
3: Protein Structural alphabet called as Protein Blocks(PBs) (de Brevern et al. 2000). PB assignment for a given protein using PBxplore(https://github.com/pierrepo/PBxplore).

Protein Structural Alphabet

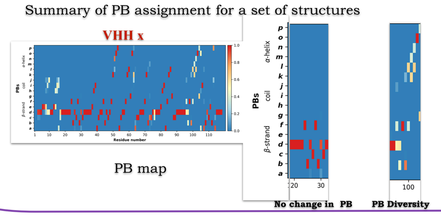


Protein Block entropy and PB frequency change

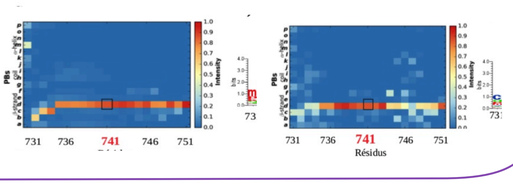
$$Neq = exp(-\sum_{x=1}^{16} f_x \ln f_x)$$



Summary of PB assignment for a set of structures



$$\Delta PB = \sum_{x}^{16} \left| (f_x^1 - f_x^2) \right|$$



Dataset

Sequence and structural features of 88 non redundant V_HH domains.

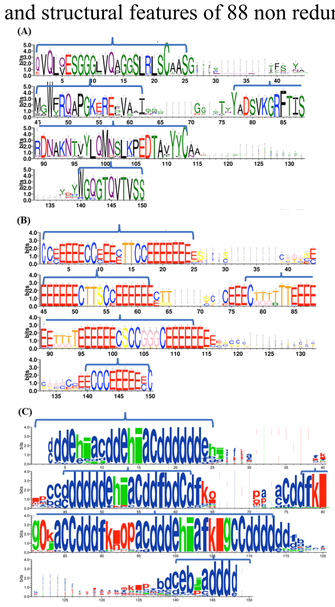


Figure 1: Sequence and Structure characteristics in V_HH *dataset*. Conservation of (A) Amino acid residues, (B) Secondary structures, and (C) Protein Blocks. The four Framework regions are delineated in each figure

Results:

- Flexibility in 88 V_HH PDB structures and corresponding 1μs trajectories.

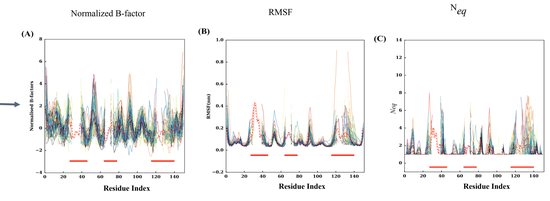
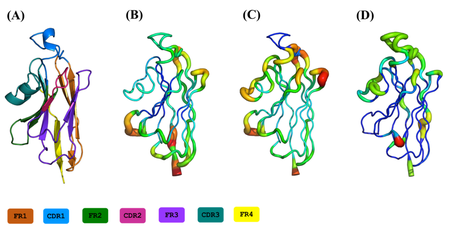


Figure 2: *Flexibility metrics at each residue position.* (A) Representation of normalised B-factors of Cα atoms, the x-axis is the residue position in the MSA and the y-axis the Normalised B-factor values, (B) Representation of all RMSF values Cα atoms, (C) Representation of N_{eq} values. The three CDR regions are highlighted using three red-coloured regions at the bottom of the plots. The average values of each metric are shown in dotted red lines.



- Flexibility metrics are not always correlated.
- Loops other than CDRs also show high RMSF.

Figure 3: *:Representation of mean values of flexibility metrics onto a 3D structure of a V_HH.* (A) FRs and CDRs coloured

- Classification of Flexibility in CDRs

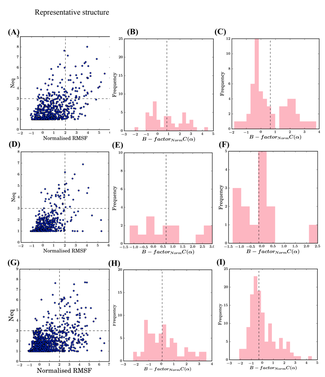


Figure 4: *CDR flexibility.* (A) to (C) is CDR1, (D) to F is CDR2 and (G) to (I) is CDR3. (A), (D) and (G) show the normalised RMSF vs N_{eq} while (B), (E) and (H) are the normalized B-factors of flexible quadrant (nRMSF > 2 and N_{eq} > 3), and (C), (F) and (I) for mobile quadrant (nRMSF > 2 and N_{eq} < 3). The dotted line in each B-factor distribution represents the mean value.

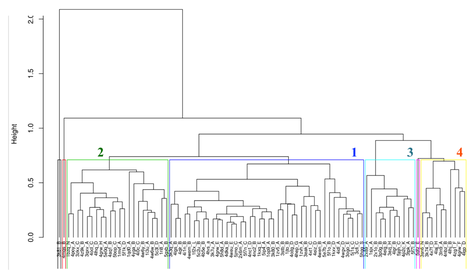


Figure 5: Hierarchical clustering of V_HH *trajectories using RMSF.* The different RMSF clusters are demarcated using coloured boxes, and their cluster number is marked accordingly.

- Hierarchical clustering of 88 1μs trajectories Distance between two RMSF vectors were calculated using .

$$d(v,w) = \sqrt{\frac{1}{n-m}\sum_{i=1}^{n}(v(i) - w(i))}$$

Four dense clusters and 3 outliers were obtained from the clustering.

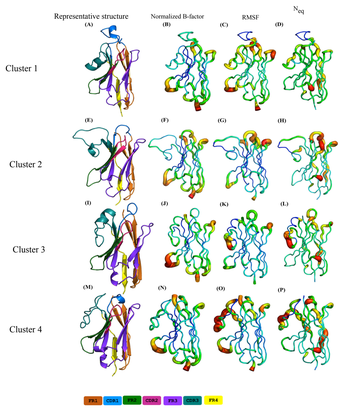


Figure 6: *:Representation of mean values of flexibility metrics onto a 3D structure of a V_HH from each cluster.*

- Mean flexibility metrics represented for each cluster.
- Not all the residue positions of CDRs are diverse.
- FRs show relatively less diversity when compared to CDRs except in the C-terminal loop regions and the 4th N-terminal loop.

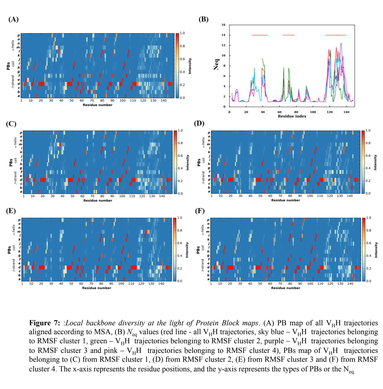


Figure 7: *:Local backbone diversity at the light of Protein Block maps.* (A) PB map of all V_HH trajectories aligned according to MSA, (B) N_{eq} values (red line – all V_HH trajectories, sky blue – V_HH trajectories belonging to RMSF cluster 1, green – V_HH trajectories belonging to RMSF cluster 2, purple – V_HH trajectories belonging to RMSF cluster 3 and pink – V_HH trajectories belonging to RMSF cluster 4), PBs map of V_HH trajectories belonging to (C) from RMSF cluster 1, (D) from RMSF cluster 2, (E) from RMSF cluster 3 and (F) from RMSF cluster 4. The x-axis represents the residue positions, and the y-axis represents the types of PBs or the N_{eq}

- Total conformational diversity observed in various V_HH trajectory groups quantified using Protein Blocks.
- The **β**-core is well conserved in the FRs represented using the PB 'd', except in some regions of FR2 and FR3.
- CDR1 also exhibits higher conformational diversity in terms of PBs.

Conclusions:

- Protein Blocks offer a unique perspective which is both qualitative and quantitative to characterize backbone diversity in structures and trajectories of V_HH domains.
- Higher RMSF values need not be always considered as flexible residues.
- There are few residue positions in CDR3 that do not show much diversity.
- The 4th N-terminal loop exhibits unexpected backbone diversity.

Acknowledgements

- This work is supported by the POE FEDER 2014-20 of the Conseil Régional de La Réunion (S3D VHH program, N° SYNERGIE RE0022962), EU-H2020 and Université de la Réunion.
- A.M.V. acknowledges access granted to high performance computing (HPC) resources at the French National Computing Centre CINES under grants no. A0040710426 and A0110713032 funded by the GENCI (Grand Equipement National de Calcul Intensif).