Structural Analysis of Antibody Repertoire in Response to Yellow Fever Virus Vaccine

Luke Marques
Franca Fraternali

Abstract

Antibodies are immunoglobulin proteins produced by B-cells as part of an immune response. Understanding their structure is vital for designing antibody therapeutics.

This project models, using the web server ABodyBuilder [Leem et al. 2016], antibody variable-region sequences from a Yellow Fever vaccination trial, and aims to classify these sequences based on their 3-dimensional structure and physicochemical properties.

Initial results demonstrate the biophysical constraints of mutations within antibodies and that sequentially different antibodies can converge to similar conformations. These results illustrate the importance of structural analysis to understand the dynamics of antibodies in response to vaccines.

Antibody Structure

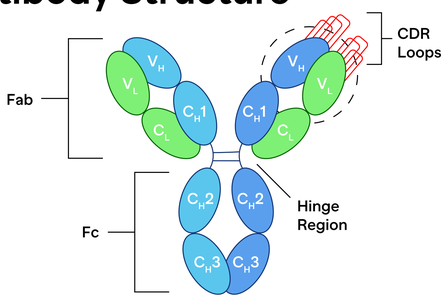


Figure 1. Diagram of typical antibody structure, showing the fragment antigen-binding region (Fab), the constant region (Fc), the variable region (FV), and the antigen-binding site / complementary determining region (CDR) loops.

- Variability in the variable region originates from **VDJ gene recombination** and **mutations**.
- The antigen-binding site, composed of the **CDR loops**, is where **most of this variability manifests**.

Repertoire Data

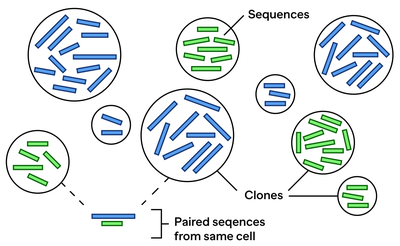


Figure 2. Schematic representation of how antibody nucleotide sequences, from immunoglobulin repertoires, are clustered to clones and identified by parent B-cell.

Project Overview

Research Questions:

1. **What information can we gain from modeling antibody sequences from large immune repertoires sampled from individuals?**
2. **How can we classify large amounts of antibody structural data efficiently and what determines these classifications?**
3. **Where are mutations populated within antibody structures?**

Clean repertoire data and select paired sequences

↓

Submit sequences to ABodyBuilder

↓

Analyse results:
Cluster models on structure
Get SASA of models
Investigate mutations trends

Structural Alphabet

The PDBencode package [Pandini et al. 2010] allows fast determination of protein conformations by storing structural information as an alphabetical code.

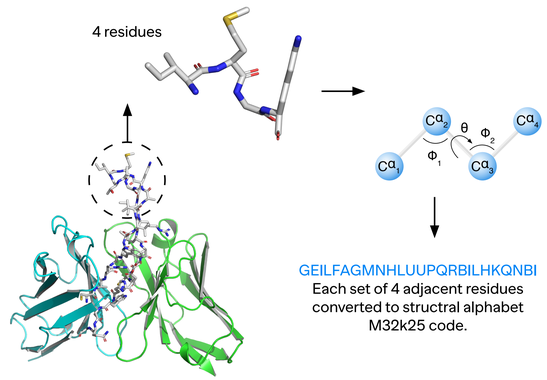


Figure 3. Schematic representation of the process of converting the 3-dimensional structural data of an antibody model to a conformational code, using PDBencode.

Structural Clustering

Conformational convergence of CDRH3 loop is not entirely determined by clinical data or by V, D, or J-genes.

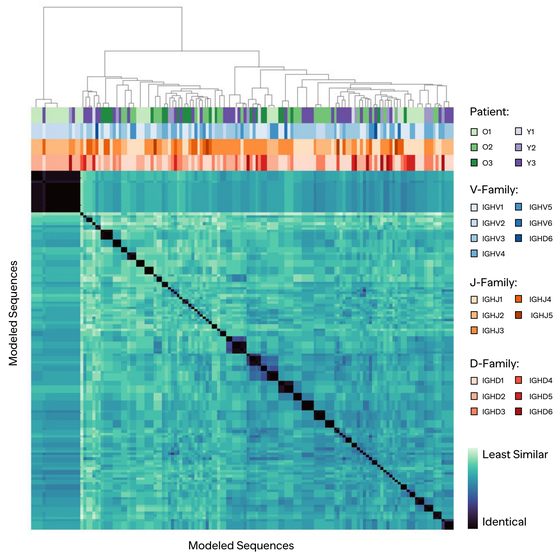
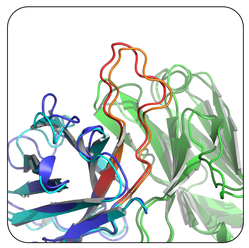


Figure 4. Modelled heavy chain antibody sequences, with varying CDRH3 lengths (short, medium, long), clustered hierarchically by the Levenshtein distance between their alphabet encoded conformations, gathered using PDBencode.

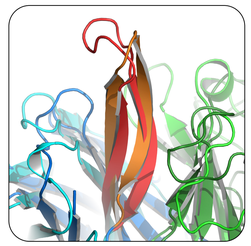


Clustering sequences based on PDBencode data results in structures with very high conformational similarity, across patients and gene usage.

Figure 5. Two CDRH3 models of the same conformational cluster from different patients with different VDJ-gene combinations superimposed.

Clustering models based on their root mean squared deviation (RMSD) will be performed, similar to that of the conformational clustering.

Figure 6. Example of two CDHR3 models from the same patient, superimposed and RMSD of C-alpha atoms minimised.



SASA Trends

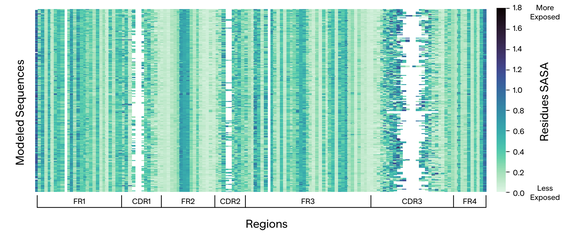


Figure 7. SASA values of each residue for all modelled antibody sequences, across all patients, calculated using POPScomp. [Kleinjung et al. 2005] Darker colours indicate a more exposed residue, and gaps in the sequences are due to length variability.

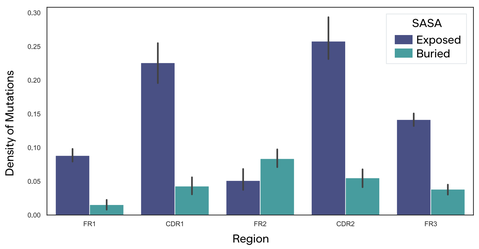


Figure 8. Density of mutations at exposed or buried positions within each region, relative to the number of exposed or buried positions for each region.

- Calculated **SASA values** for residues within sequences are consistent **across models**.
- Allows assignment of exposed/buried regions to non-modelled sequences.
- Mutation data for largest clones in the repertoire show an **overexpression of mutations within exposed regions**, particularly in CDR loops.

Conclusions

Antibodies with different sequences can have highly similar conformations.

Solvent accessibility dictates the density of mutations on antibody structures.

Future work
Expand structural clustering methods to all CDR loops. Molecular dynamics simulations to study conformational diversity and impact of mutations.

References: Leem et al. (2016), mAbs, 8:1259-1268 // Pandini et al. (2010), BMC Bioinformatics, 11:97 // Kleinjung, J. and Fraternali, F. (2005), Nucleic Acids Research, 33:342-346