Reduced structural flexibility of eplet amino acids in HLA proteins

Diego Amaya-Ramirez, Romain Lhotte, Magali Devriese, Constantin Hays, Jean-Luc Taupin, Marie-Dominique Devignes

Reduced structural flexibility of eplet amino acids in HLA proteins

Amaya-Ramirez Diego[1], Romain Lhotte[2], Magali Devriese[2], Constantin Hays[2], Taupin Jean-Luc[2] and Devignes Marie-Dominique[1]

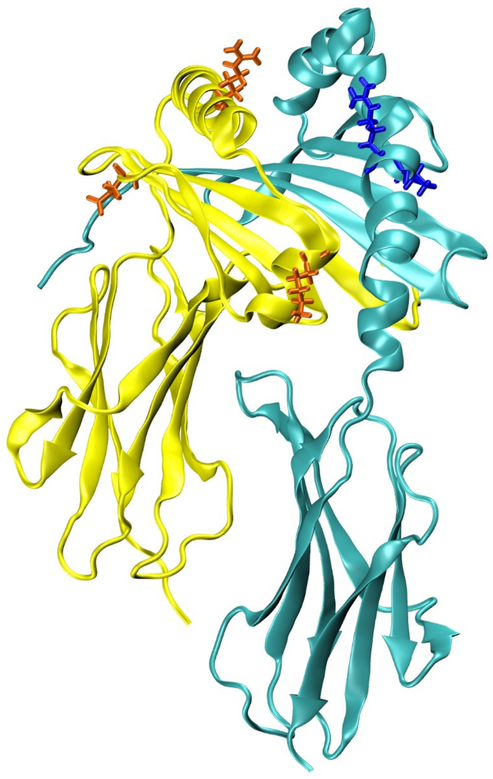
[1] Université de Lorraine, CNRS, INRIA, LORIA, Nancy, F-54000, France; [2] INSERM, Hôpital Saint Louis, F-75010 Paris

Contact : diego.amaya-ramirez@inria.fr

Introduction

The proteins encoded in the **HLA (Human Leukocyte Antigen) system** are largely responsible for the compatibility in organ transplants. To date, the molecular determinants involved in the recognition of HLA antigens by recipient antibodies are unknown. Here we explore flexibility as a potential determinant. For this purpose we compare in terms of **N-RMSF (Normalized Root Mean Square Fluctuation)** amino acids labeled as confirmed **eplets (regions defined around polymorphic amino acids)** against amino acids that have not been reported as eplets. We found that eplet amino acids tend to be less flexible than non-eplet amino acids, which would indicate that the antibodies would have a preference for binding with less mobile regions.

Materials and Methods



The HLA proteins are heterodimers. Here the DQA1*01:01-DQB1*05:01 protein is presented. In yellow the chain α with some eplets in orange. In cyan the chain β with some eplets in dark blue.

Dataset :

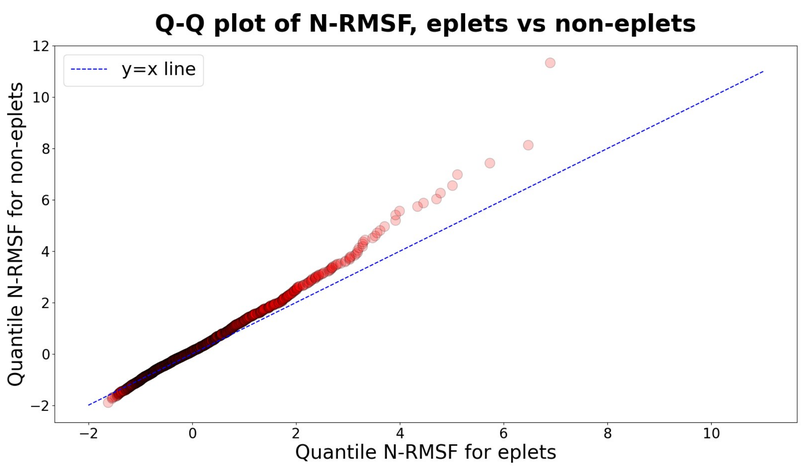
- 207 HLA proteins (the most prevalent in the European population, Luminex kit) were modeled.
- Initial templates for **Molecular Dynamics (MD) simulations** were obtained from **PDB**[1] or **Alphafold**[2].
- 10ns equilibration MD simulations were performed for each HLA protein.
- **NAMD3**[3], **CHARMM36**[4] force field and **TIP3P** explicit solvent model were used for MD runs.
- For the **RMSF** calculations, 500 frames extracted from the last 5ns of the trajectory were considered.
- List of eplet amino acids were obtained from **EpRegistry**[5].
- **RSASA**[6] **(Relative Solvent Accessible Surface Area)** was used to identify solvent-accessible amino acids.
- Only solvent-accessible amino acids (RSASA > 20%) were considered for the N-RMSF calculations : eplet amino acids = 2087, non-eplet amino acids = 38364.
- **Z-score normalization** method were applied to all the RMSF values (μ=0, σ=1). $\frac{RMSF - \mu}{\sigma}$

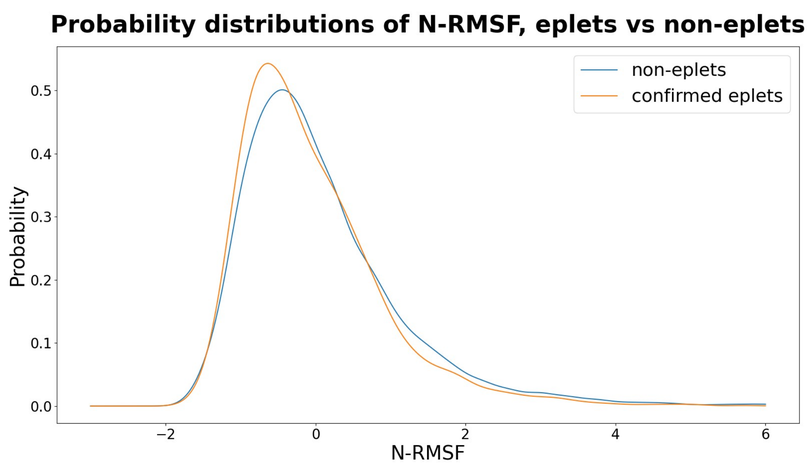
# HLA proteins modeled per gene					
Class I			Class II		
A	B	C	DP	DQ	DR
34	59	17	31	30	36

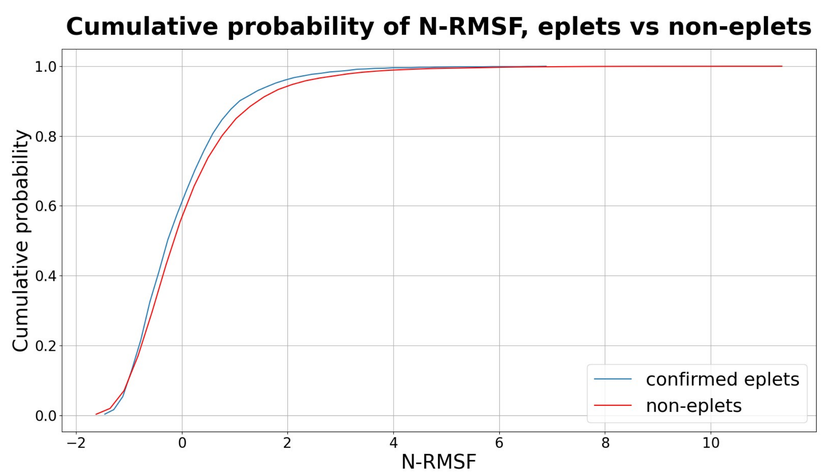
$$\mathrm{RMSF}_i = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left\langle\left(r_i'(t) - r_i\left(t_{\mathrm{ref}}\right)\right)^2\right\rangle}$$

i indicates a given amino acid
$r_i(t)$ represents the set of atoms coordinates for a given amino acid for frame t.

Results







The **Quantile-Quantile (Q–Q) plot** has a steeper slope than the random line y=x implying that the N-RMSF distribution of the non-eplet is more widely dispersed, which indicates higher structural flexibility.

Kolmogorov–Smirnov test (K–S test) is a statistical test sensitive to differences in both location and shape of the cumulative distribution functions of two samples. To verify whether the difference observed in the Q-Q plot is significant, K–S test was conducted for the two N-RMSF distributions and confirmed statistically significant differences between the two distributions (**p-value = 1,06x10^{-6}**).

Conclusion

Here we present an unprecedented set of molecular dynamics simulations of the HLA system. These simulations revealed that the key amino acids for the recognition of an HLA antigen by an antibody (termed eplets) are less flexible than the rest of the surface amino acids. These findings demonstrate the relationship between structural flexibility and HLA-antibody binding preference. It opens the door to the use of structural flexibility to identify antibody binding sites on HLA proteins.

References

[1] H. M. Berman et al., « The Protein Data Bank », Nucleic Acids Research, 2000.

[2] J. Jumper, R. Evans, et al., « Highly accurate protein structure prediction with AlphaFold », Nature, 2021.

[3] Phillips, J. *et al.* Scalable molecular dynamics on CPU and GPU architectures with NAMD. Journal of Chemical Physics, 2020.

[4] Huang, J. *et al.* CHARMM36m: an improved force field for folded and intrinsically disordered proteins. Nat Methods, 2017.

[5] R. J. Duquesnoy et al., « 16th IHIW: A Website for Antibody-Defined HLA Epitope Registry », International Journal of Immunogenetics, 2017.

[6] [1] M. Z. Tien, et al, « Maximum Allowed Solvent Accessibilites of Residues in Proteins », PLOS ONE, 2013, doi: 10.1371/journal.pone.0080635.

Acknowledgements

Diego Amaya-Ramirez is recipient of an Inria-Inserm doctoral fellowship.

Experiments presented in this poster were carried out using :

- The Grid'5000 testedbed, supported by a scientific interest group hosted by Inria and including CNRS, RENATER and several Universities as well as other organizations (see https://www.grid5000.fr).
- The MBI-DS4H platform hosted by Inria/Loria funded by CPER IT2MP (Contrat Plan État Région, Innovations, Technologiques, Modélisation & Médicine Personalisée, including a FEDER co-funding).