Yoram Louzoun
Bar Ilan University
e-mail: louzouy@math.biu.ac.il
Tal Vider
Yaacov Maman
Alexandra Agaranovich
Lea Tsaban

Viruses selectively mutate their CD8+ T cell epitopes an optimization framework, a novel machine learning methodology and a large scale genetic analysis.

The relation between organisms and proteins complexity and between the rate of evolution has been discussed in the context of multiple generic models. The main robust claim from most such models is the negative relation between the organism complexity and the rate of mutation accumulation.

We here validate this conclusion, through the relation between viral gene length and their CD8 T cell epitope density. Viruses mutate their epitopes to avoid detection by CD8 T cells and the following destruction of their host cell. We propose a theoretical model to show that in viruses the epitope density is negatively correlated with the length of each protein and the number of proteins.

In order to validate this conclusion, we developed a novel machine learning methodology to combine multiple modalities of peptide-protein docking measurement. We use this methodology and large amount of genomic data to compute the epitope repertoire presented by over 1,300 viruses in many HLA alleles. We show that such a negative correlation is indeed observed.This negative correlation is specific to human viruses.

The optimization framework also predicts a difference between human and non-human viruses, and an effect of the viral life cycle on the epitope density. Proteins expressed early in the viral life cycle are expected to have a lower epitope density than late proteins.

We define the "Size of Immune Repertoire (SIR) score," which represents the ratio between the epitope density within a protein and the expected density. This score is applied to all sequenced viruses to validate the prediction of the optimization model.

The removal of early epitopes and the targeting of the cellular immune response to late viral proteins, allow the virus a time interval to propagate before its host cells are destroyed by T cells. Interestingly, such a selection is also observed in some bacterial proteins. We specifically discuss the cases of Herpesviruses, HIV and HBV showing interesting selection biases.