Application of a statistical inference model to the prediction of antibody affinity from sequence analysis

David Luna[1], Dr. Pierpaolo Bruscolini[1], and Dr. Sergio Pérez Gaviro[1]
[1]Universidad de Zaragoza

The interest in therapeutic antibodies has grown quickly due to their huge potential application in a large number of diseases; however they are still difficult to design. That is because they cannot be computationally designed with the necessary accuracy to reach the required affinity to the antigen and their equilibrium properties cannot be predicted precisely. Furthermore, it is not known with sufficient precision the mechanisms of antibody maturation and selection, that transform a naive antibody, generated by the random combination of copies of the V, (D), and J genes, into a specialized tool for recognizing and binding a particular antigen, without causing damages to the self.

In this work, we describe the evolution in the sequence space as a Ornstein-Uhlenbeck (OU) process, that is, a stochastic Markov process, where the sequences undergo a drift towards an attractor of high affinity, while they also diffuse randomly in sequence space. Introducing some approximation and simplifying hypotheses, we are able to perform a bayesian inference of the attractor and final variance. The objective is to model the maturation process using this OU process. In this approach, the clusters of observed sequences are considered as multivariate gaussian snapshots of the evolution. In this way, the ordinary OU process is replaced by a system of temporally linked gaussian distributions, as shown in Fig. 1, which are inferred from a sequence of sub-clusters identified in a experimental data set [1]. The attractors of this OU process are assumed to characterize the fully matured sequences cluster. It can be checked, using the MG score if the sequences of the hypermutated cluster are fully maturated or they were just maturated until their virus affinity was high enough.

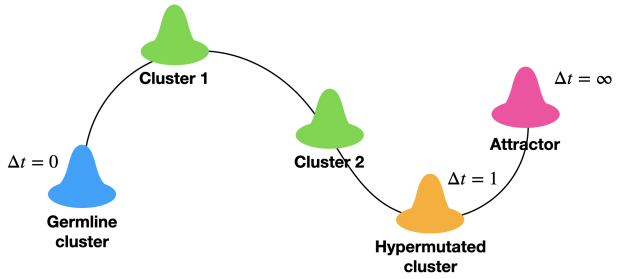

Fig. 1. Modified Ornstein-Uhlenbeck representation.

[1] Asti L, Uguzzoni G, Marcatili P, Pagnani A. Maximum-Entropy Models of Sequenced Immune Repertoires Predict Antigen-Antibody Affinity. PLoS Comput Biol 12(4): e1004870. (2016)doi:10.1371/journal.pcbi.1004870

[2] Marcou, Q., Mora, T. Walczak, A.M. High-throughput immune repertoire analysis with IGoR. Nat Commun 9, 561 (2018). doi:10.1038/s41467-018-02832-w

[3] Pilar Calvo Príncipe, Refinement of a statistical model for antibody humanization, TFM.(September 2019)

[4] Clavero-Alvarez, A., Di Mambro, T., Perez-Gaviro, S. et al. Humanization of Antibodies using a Statistical Inference Approach. Sci Rep 8, 14820. (2018) doi.10.1038/s41598-018-32986-y