HIERARCHICAL CLUSTERING OF pMHC COMPLEXES BASED ON THE ELECTROSTATIC POTENTIAL OF THE TCR-INTERACTING SURFACE

Antunes, D. A., Figueiredo, D. F., Rigo, M.M., Silva, J.P., Chies, J.A.B., Sinigaglia, M., Vieira, G.F.

NBLI – Núcleo de Bioinformática do Laboratório de Imunogenética. Departamento de Genética, Universidade Federal do Rio Grande do Sul. Av. Bento Gonçalves 9500, Building 43323, room 225.

The vertebrates immune system possess several defense mechanisms, some of which are shared with less complex organisms (e.g. the existence of phagocytic cells). However, with the risen of vertebrates a new set of defense mechanisms has also emerged, being referred as adaptive immunity. This innovative "arsenal" has allowed not only to respond against typical pathogenic targets, but also to develop specific responses against any new and diverse antigen that infect our system. One of the key players of adaptive immunity is the Major Histocompatibility Complex class I (MHC-I), the final step of an important cellular pathway. The endogenous antigen-presenting pathway, or MHC-I pathway, is present in all nucleated cells, acting as a "quality control" of the proteins produced inside the cell. Through this route, a sample of the cytosolic proteins are cleaved by an enzymatic complex and the derived peptides (or epitopes) are transported to the Endoplasmic Reticulum (ER), were they can be bound to the MHC-I molecule. The MHC class I molecule is a heterodimer with a constant light chain (β2-microglobulin) and a variable heavy chain (α chain). Two domains of the heavy chain (α1 and α2) interact with each other to form a cleft, able of accommodating peptides with 8 to 11 amino acids long. This peptide:MHC complex (or pMHC) will then be transported to the cell surface, to be presented to Cytotoxic T Lymphocytes (CTLs) . Since CTLs are capable of identifying non-self peptides presented in the cleft of self MHCs, the MHC-I pathway has a central role in the adaptive immunity against intracellular pathogens and other cellular disorders (e.g. cancer). Once a virus infects a nucleated cell, for instance, it will start to produce a great number of viral proteins, in order to replicate itself. Some of these virus-derived proteins will unavoidable be directed to the MHC-I pathway, what will end up with virus-derived epitopes being presented at the cell surface. CTLs will interact with these pMHC complexes through the T Cell Receptor (TCR) and identify the non-self nature of the presented peptide. This recognition triggers the cytotoxic mechanisms of the CTL, destroying the infected cell [1].

In humans, the MHC-I heavy chain is encoded by three genes at the short arm of chromosome 6. These genes are located at the most variable region of the human genome and together have already more than 5,000 described alleles. The sequence variants of a given gene are called alleles and the proteins encoded by different alleles of the same gene are called allotypes. Each MHC-I allotype has its own specificity, being able to present some peptides, but not others. The variability of this interaction is enormous and, despite the variability in the binding energies and in the pMHC complex stability, a given MHC-I can present thousands of different peptides (self-peptides, pathogen-derived epitopes, tumoral antigens, etc).

Each CTL carries a specific TCR, which can recognize a given pool of virus-derived peptides in the context of a given MHC-I. The CTL recognizes not only the peptide, but the surface of the pMHC complex, which is also determined by specific characteristics of each MHC-I allotype. Therefore, CTLs are referred as being both specific and poly-specific: each CTL can recognize only a small fraction of the universe of possible pMHC complexes (specific recognition), but this specific group of targets can include completely unrelated pMHCs (poly-specific recognition). This mechanism, also referred as cross-reactivity, plays a central role in heterologous immunity – when the infection by a given pathogen protects against another pathogen – and, therefore, in vaccine development [2]. Cross-reactivity was initially described as being triggered by the great similarity between amino acid sequences of the presented peptides. However, immunologists have already

described events of cross-reactivity between epitopes sharing less than 50% of the linear amino acid sequence , and this phenomenon has proven to be far more difficult to predict. The CTL stimulation will be triggered by the specific recognition of a pMHC by a given TCR. The combined surface of MHC and peptide is referred as TCR-interacting surface and, according to an existent hypothesis, structural aspects of this surface can be the molecular players driving CTL stimulation and cross-reactivity.

In the present work, we analyzed 60 unrelated pMHC complexes presenting virus-derived peptides, in the context of the most frequent human MHC allele (HLA-A*02:01). These complexes were obtained from the CrossTope Data Bank for Cross-Reactivity Assessment, being 5 crystal structures and 55 *in silico* predicted structures (obtained with the previously described approach *D1-EM-D2*) [3, 4]. Images of the TCR-interacting surface of these complexes, presenting the electrostatic potential distribution, were used to extract the color histograms (RGB) of seven selected areas, corresponding to the points of interaction with the TCR. Values of mean and standard deviation of the three RGB components, for each one of these selected areas, were used as input for multivariate statistical methods, aiming to predict possible targets of cross-reactivity.

Our dataset included some peptides with already known cross-reactivity, and the result of a Hierarchical Cluster Analysis (HCA) of these data agreed with this experimental background. For instance, we included 10 variants of the wild-type epitope HCV-$NS3_{1073}$ (CV/INGVCWTV). The wild-type and all the cross-reactive variants (genotypes 4, 5 and 6) were placed in the same group, while the noncross-reactive variant (genotype 3) was placed in a completely unrelated group. The HCA dendrogram also clustered two cross-reactive variants of the epitope $A26L_{6-14}$, one from Vaccinia and other from Variola Virus. Other well-known case of cross-recognition, between the epitopes IV-$M1_{58-66}$ (GILGFVFTL) and HIV-GAG_{77-85} (SLYNTVATL), was also represented, with both complexes placed in the same group. Interestingly, these last two epitopes were placed exactly in the same group of the cross-reactive variants of HCV-$NS3_{1073}$ epitope. Cross-reactivity between this HCV immunodominant epitope and the HIV-GAG_{77-85} peptide has not been described so far. However, this same HIV epitope was already described as being cross-reactive with the HCV-NS5b peptide, in HIV/HCV co-infected individuals. There is yet another complex included in the same cluster, presenting the “LLWTLVVLL” peptide, from the Human herpesvirus 4 (LMP_{2329}). There is yet other two complexes included in the same cluster, presenting the “LLWTLVVLL” and the "NLVPMVATV" peptides, from the Human herpesvirus 4 (LMP_{2329}) and 5 ($pp65_{485}$), respectively. It is important to note that the former peptide does not share even a single amino acid with the target peptide (CV/INGVCWTV) and, nevertheless, presented almost the same topology and charge distribution when presented in the context of HLA-A*02:01. This result stresses the power of this structural approach in prospecting new cross-reactive targets.

Supported by: CNPq, CAPES and a grant from *Bill & Melinda Gates Foundation* through the Grand Challenges Exploration Initiative.

1. Blanchard N, Shastri N: **Topological journey of parasite-derived antigens for presentation by MHC class I molecules**. *Trends in Immunology* 2010, **31**:414-421.
2. Welsh RM, Fujinami RS: **Pathogenic epitopes, heterologous immunity and vaccine design**. *Nat Rev Microbiol* 2007, **5**(7):555-563.
3. Antunes DA, Vieira GF, Rigo MM, Cibulski SP, Sinigaglia M, Chies JA: **Structural allele-specific patterns adopted by epitopes in the MHC-I cleft and reconstruction of MHC:peptide complexes to cross-reactivity assessment**. *PLoS One* 2010, **5**(4):e10353.
4. Antunes DA, Rigo MM, Silva JP, Cibulski SP, Sinigaglia M, Chies JA, Vieira GF: **Structural in silico analysis of cross-genotype-reactivity among naturally occurring HCV NS3-1073-variants in the context of HLA-A*02:01 allele**. *Mol Immunol* 2011, **48**(12-13):1461-1467.