ELSEVIER

Improved structural method for T-cell cross-reactivity prediction

Marcus F.A. Mendes [a,b,1], Dinler A. Antunes [a,b,1], Maurício M. Rigo [a,b],
Marialva Sinigaglia [a,b], Gustavo F. Vieira [a,b,*]

[a] 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, Brazil
[b] Programa de Pós-Graduação em Genética e Biologia Molecular (PPGBM), Universidade Federal do Rio Grande do Sul (UFRGS), Rio Grande do Sul, Porto Alegre, Brazil

ARTICLE INFO

ABSTRACT

Cytotoxic T-lymphocytes (CTLs) are the key players of adaptive cellular immunity, being able to identify and eliminate infected cells through the interaction with peptide-loaded major histocompatibility complexes class I (pMHC-I). Despite the high specificity of this interaction, a given lymphocyte is actually able to recognize more than just one pMHC-I complex, a phenomenon referred as cross-reactivity. In the present work we describe the use of pMHC-I structural features as input for multivariate statistical methods, to perform standardized structure-based predictions of cross-reactivity among viral epitopes. Our improved approach was able to successfully identify cross-reactive targets among 28 naturally occurring hepatitis C virus (HCV) variants and among eight epitopes from the four dengue virus serotypes. In both cases, our results were supported by multiscale bootstrap resampling and by data from previously published in vitro experiments. The combined use of data from charges and accessible surface area (ASA) of selected residues over the pMHC-I surface provided a powerful way of assessing the structural features involved in triggering cross-reactive responses. Moreover, the use of an R package (pvclust) for assessing the uncertainty in the hierarchical cluster analysis provided a statistical support for the interpretation of results. Taken together, these methods can be applied to vaccine design, both for the selection of candidates capable of inducing immunity against different targets, or to identify epitopes that could trigger undesired immunological responses.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

Cellular immunity is one of the two main branches of the adaptive immunologic response, focused on specific functions of the cytotoxic T-lymphocytes (CTLs). Although both cellular and humoral immunity are desired for an ideal and longstanding immunization, CTL response plays a central role in regard to antiviral immunity (Brehm et al., 2004). After infecting a host cell, the virus will use the host molecular machinery to replicate its genome and produce new virions. In addition to all the mechanisms that allow virus escape from circulating neutralizing antibodies, during its intracellular replication cycle the virus is virtually hidden from the action of humoral immunity. However, some viral proteins will unavoidably be marked to enter the endogenous antigen presentation pathway. Through this route, virus-derived peptides will be presented at the cell-surface in the context of major histocompatibility complex (MHC) class I molecules, forming stable peptide:MHC-I (pMHC-I) complexes. Each CTL produced by the host has one specific T-cell receptor (TCR), which is able to recognize pMHC-I complexes presenting nonself peptides. Therefore, through the interaction between pMHC-I complexes and TCRs, CTLs are able to identify and eliminate infected cells.

The TCR/pMHC-I interaction is highly specific, which allows the development of memory T-cells that will be once again triggered in future challenges with the same target. However, a given lymphocyte is able to recognize more than just one pMHC-I complex. This capacity of a CTL to recognize non-related peptides derived from the same virus, or even peptides from heterologous

viruses, was defined as cross-reactivity (Vieira and Chies, 2005). As expected, cross-reactivity has direct implications over vaccine development, autoimmunity and heterologous immunity, a process by which the immunization with one pathogen confers protection against another (Cornberg et al., 2010; Selin et al., 1994; Welsh and Fujinami, 2007; Welsh and Selin, 2002). Understanding of the molecular features driving these cross-reactivities became a major goal for several immunologists, but the system's complexity has delayed progress in the field. Wedemeyer et al. (2001) have proposed that cross-recognition of two heterologous epitopes could be triggered by the high amino acid sequence similarity between them. Similarity in terms of biochemical properties was also proposed as being the key for cross-recognition (Vieira and Chies, 2005), and was even applied with some success to predict cross-reactivity (Frankild et al., 2008; Moise et al., 2013). However, structural studies have shown that even epitopes with low sequence and biochemical similarity might present quite identical pMHC-I surfaces (Antunes et al., 2011; Sandalova et al., 2005), indicating that this structural similarity should account for the cross-stimulation of a given T-cell population.

Structural analysis of pMHC-I complexes can provide a level of information much closer to that presented *in vivo* for the interaction with the TCR. On the other hand, structural approaches are frequently limited by the number of pMHC-I structures already produced by experimental methods, such as X-ray crystallography and NMR (nuclear magnetic resonance). Our group has used structural bioinformatics tools to build *in silico* models of pMHC-I complexes that were not yet determined by experimental methods. This approach, referred as *D1–EM–D2* (*docking 1–energy minimization–docking 2*), was previously validated through the successful reproduction of several crystal structures (Antunes et al., 2010; Sinigaglia et al., 2013) and has been used to provide novel complexes for the CrossTope Data Bank for cross-reactivity assessment (Sinigaglia et al., 2013). Our group has also combined this approach with the use of multivariate statistical methods to make structural-based cross-reactivity predictions (Antunes et al., 2011). In a previous study, we used images of the electrostatic potential distribution over the pMHC-I surface to predict the cross-reactivity pattern among 28 naturally occurring hepatitis C virus (HCV) variants, in the context of HLA-A*02:01 (Antunes et al., 2011). Hierarchical clustering of proteins based on electrostatic potential over the entire surface has being previously used to protein functional assignment and protein classification, as performed by the webPIPSA server (Richter et al., 2008). This approach, however, is not suitable for cross-reactive prediction since most of the pMHC surface will not be contacted by the TCR and only few residues from the TCR-interacting face will play a key role in triggering a T cell response. The innovative image-based clustering of pMHC-I complexes here described has been shown to be a fast and efficient way to predict cross-reactivity using structural information, being able to identify cross-reactive targets even between epitopes which shared no amino acids in sequence (Zhang et al., 2015).

In a previous study, one region over the pMHC-I surface was defined, based on the observation of the main spots of variation among the 28 complexes analyzed. Based on the extracted information from the pMHC-I structures, we were able to predict the same clusters of cross-reactivity observed *in vitro* (Antunes et al., 2011). Despite the success of this approach, the same parameters could not be applied to other subsets, since different regions of the pMHC-I surface might have diverse influence over the TCR recognition. In this context, we presented here an improved and standardized structural-based method for T-cell cross-reactivity prediction of HLA-A*02:01-restricted epitopes. In the present work, we aimed to provide a generic set of "gates" that could be applied to any subset of epitopes restricted to HLA-A*02:01. These gates were defined considering the key TCR interactions regions, which could be involved in cross-reactive responses.

Another improvement we implemented in this work was the inclusion of topography prediction. There are experimental evidences suggesting that charge similarity is more important than subtle topographic differences between the cross-reactive complexes (Jorgensen et al., 1992; Kessels et al., 2004). However, pMHC-I complexes are 3D structures and, hence, topography variation certainly has some influence over the TCR recognition. The accessible surface area (ASA) of a residue can provide a quantitative measure of how exposed or buried its side chain is, which will have impact over the pMHC-I topography. ASA values of the epitope residues, for instance, were previously related to immunogenicity (Meijers et al., 2005) and were also able to identify non-cross-reactive complexes (Antunes et al., 2010).

The predictive capacity of our method was enhanced by the inclusion of these new features such as mapping interaction zones in TCR/p:MHC complexes that are responsible for cytotoxic response, topography prediction, and a bootstrap-based statistical method to validate the hierarchal clusters. Our results with the analysis of hepatitis C virus and dengue virus epitopes support its use as an important guidance tool for rational vaccine development.

2. Results and discussion

2.1. Identification of conserved contacts among TCR-HLA-A*02:01 crystal structures

The human HLA-A*02:01 is largely studied for being the most frequent MHC-I allele in human populations (http://www.allelefrequencies.net/) (Fernandez-Vina et al., 1992). For this reason, the protein encoded by this specific allele (called allotype) also presents the larger number of crystal structures available at the Protein Data Bank (PDB). Aiming to identify the residues involved in the recognition of this allotype by different TCRs, we performed an extensive search for all available crystal structures of TCR/HLA-A*02:01complexes. This search returned 31 complexes (Table A.1), presenting 16 different TCRs and 20 different epitopes. Despite this variability, five epitope positions (p4–p8 – gates 1–3) and four MHC-I residues were consistently indicated as involved with TCR interactions, being present in more than 85% of these complexes. The P4–P6 positions of the epitope had already been observed as being directly involved in the stimulation of immunogenicity (Calis et al., 2012, 2013; Frankild et al., 2008; Hoof et al., 2010; Rudolph et al., 2006; Wucherpfennig et al., 2009). Several residues over the pMHC-I surface might participate in the interaction with the TCR, influencing the specific level of T-cell stimulation that will be triggered by each pMHC-I. However, we here postulate that changes in these nine conserved contacts might have greater impact over the T-cell recognition, therefore influencing cross-reactivity.

Supplementary material related to this article can be found, in the online version, at http://dx.doi.org/10.1016/j.molimm.2015.06.017

2.2. Inclusion of ASA values

We decided to include ASA values together with electrostatic potential information to improve our prediction method. It is important to note that the epitope amino acids composition will affect not only the charges and the ASA values of the epitope itself, but also of surrounding MHC-I residues. For that reason, in addition to the ASA values for the nine epitope residues, we also included ASA values from 28 frequently TCR-interacting MHC-I residues in our approach (Fig. 1B).