Molecular Immunology 67 (2015) 303–310

Contents lists available at ScienceDirect

Molecular Immunology

journal homepage: www.elsevier.com/locate/molimm

Improved structural method for T-cell cross-reactivity prediction

Marcus F.A. Mendes[a,b,1], Dinler A. Antunes[a,b,c,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
[c] Department of Computer Science, Rice University, Houston, Texas, 77005, USA

ARTICLE INFO

Article history:
Received 9 April 2015
Received in revised form 3 June 2015
Accepted 16 June 2015
Available online 2 July 2015

Keywords:
Cross-reactivity
pMHC-I
HCA
ASA
Pvclust
Vaccine development

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.

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

Abbreviations: CTLs, cytotoxic T-lymphocytes; MHC, major histocompatibility complex; pMHC-I, peptide: major histocompatibility complex class I; TCR, T-cell receptor; D1–EM–D2, docking 1–energy minimization–docking 2; HCV, hepatitis C virus; ASA, accessible surface area.

* Corresponding author at: 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. Tel.: +55 51 33089938.
E-mail addresses: marcus.famendes@gmail.com (M.F.A. Mendes), dinler@gmail.com (D.A. Antunes), mauriciomr985@gmail.com (M.M. Rigo), msinigaglia@gmail.com (M. Sinigaglia), gusfioravanti@yahoo.com.br (G.F. Vieira).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.molimm.2015.06.017
0161-5890/

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).

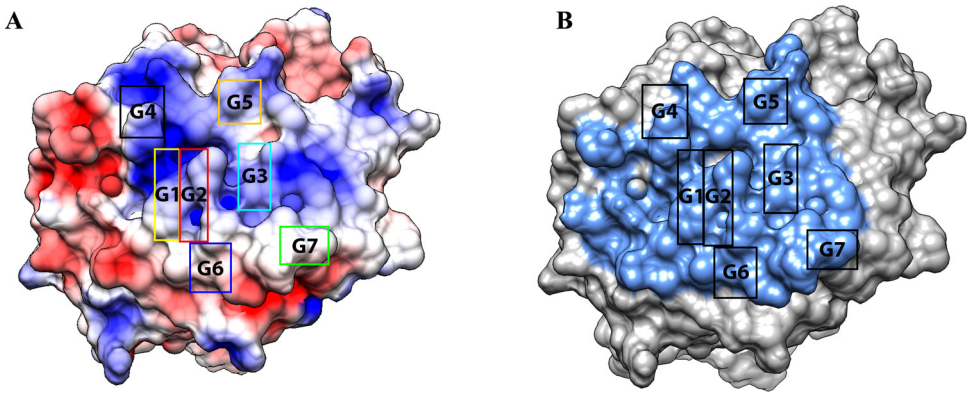

Fig. 1. Seven gates defined to obtain color histograms and selected residues for ASA assessment. Top view of a pMHC-I complex presenting a dengue-derived epitope in the cleft of HLA-A*02:01, obtained with the UCSF Chimera package (Trott et al., 2010). In (A), electrostatic potential over the surface was computed with the Delphi program and represented as red (negative charges) and blue (positive charges) spots, with a range from −3 to +3 kT. The seven gates (G1–G7) relate to conserved contacts with different TCRs, as observed in the crystal structures available, and were selected for the RGB analysis with ImageJ. In (B), the complex surface is depicted in grey while the surface of all residues selected for ASA assessment are indicated in blue. Black rectangles indicate the seven gates (from G1–G7) used in the RGB analysis. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Comparing results with and without ASA, we observed a better definition in the clusters, making the results more consistent with *in vitro* data. To exemplify this improvement, G6_26 (not including ASA values) appears in other branch, outside of the cross reactive cluster, being now included in the correct cross reactive cluster. For a full comparison, an image of clusterization analysis without ASA values can be viewed in Fig. A.1.

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

2.3. Method validation with a previously studied subset

Twenty-eight variants, covering all six HCV genotypes, were tested *in vitro* against the same T-cell population, which was

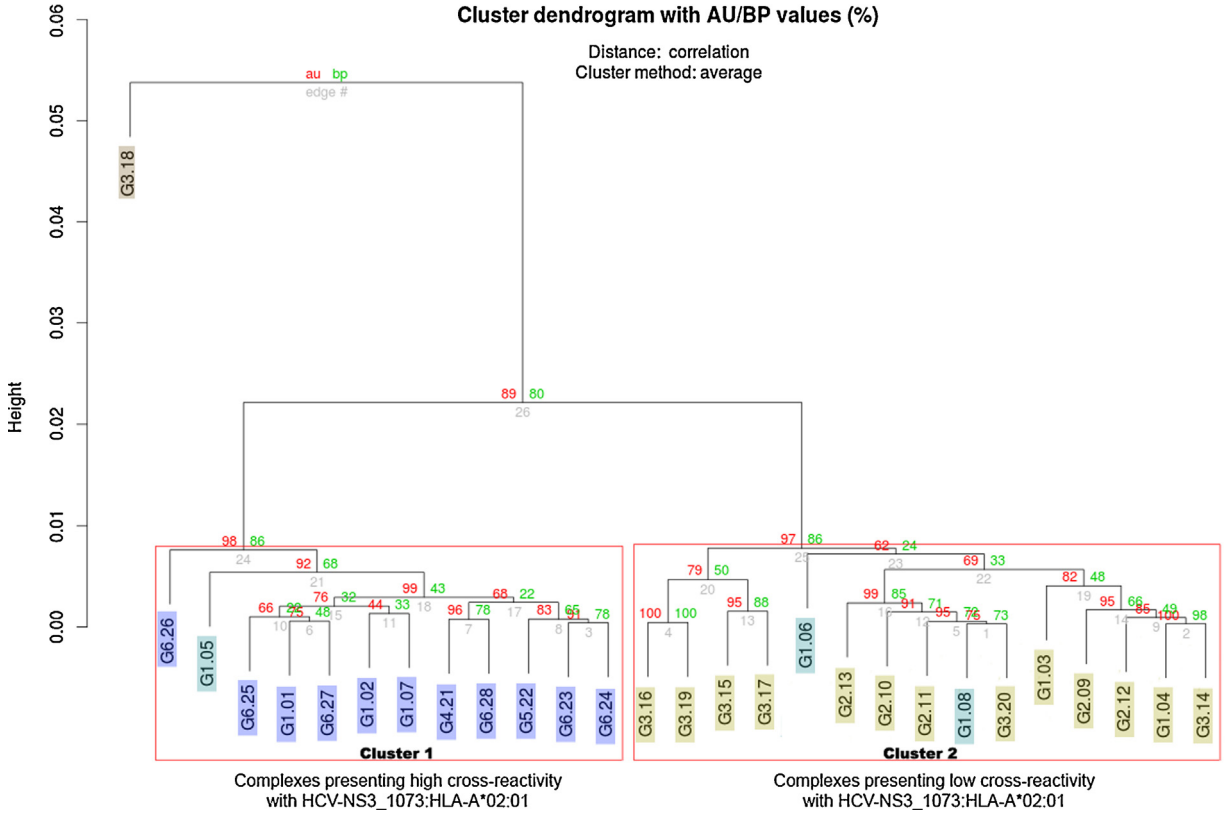

Fig. 2. HCA of 28 HCV natural occurring variants. Dendrogram representing the hierarchical cluster analysis (HCA) of 28 pMHC-I complexes loaded with HCV-derived epitopes covering all six HCV genotypes (from G1–G6). The input data was accessible surface area values and color histograms (RGB) for each pMHC-I, which provided information on topography and charges distribution over the surface. Red boxes indicate the main clusters identified (*alpha* = 0.95). High (Cluster 1) and low (Cluster 2) G1_01 cross-reactive complexes fell in independent main clusters. The only complex that presented no response *in vitro*, G3-18, fell alone in an independent branch. The strong (dark blue), intermediate (light blue), low (yellow) and without (brown) cross-reactive targets in respect to G1_01 are represented inside individual boxes. Information on the specific response presented by each complex in cross-reactivity tests (*in vitro*) is provided in Additional file 3. G, HCV genotype; AU, approximately unbiased; BP, bootstrap probability. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

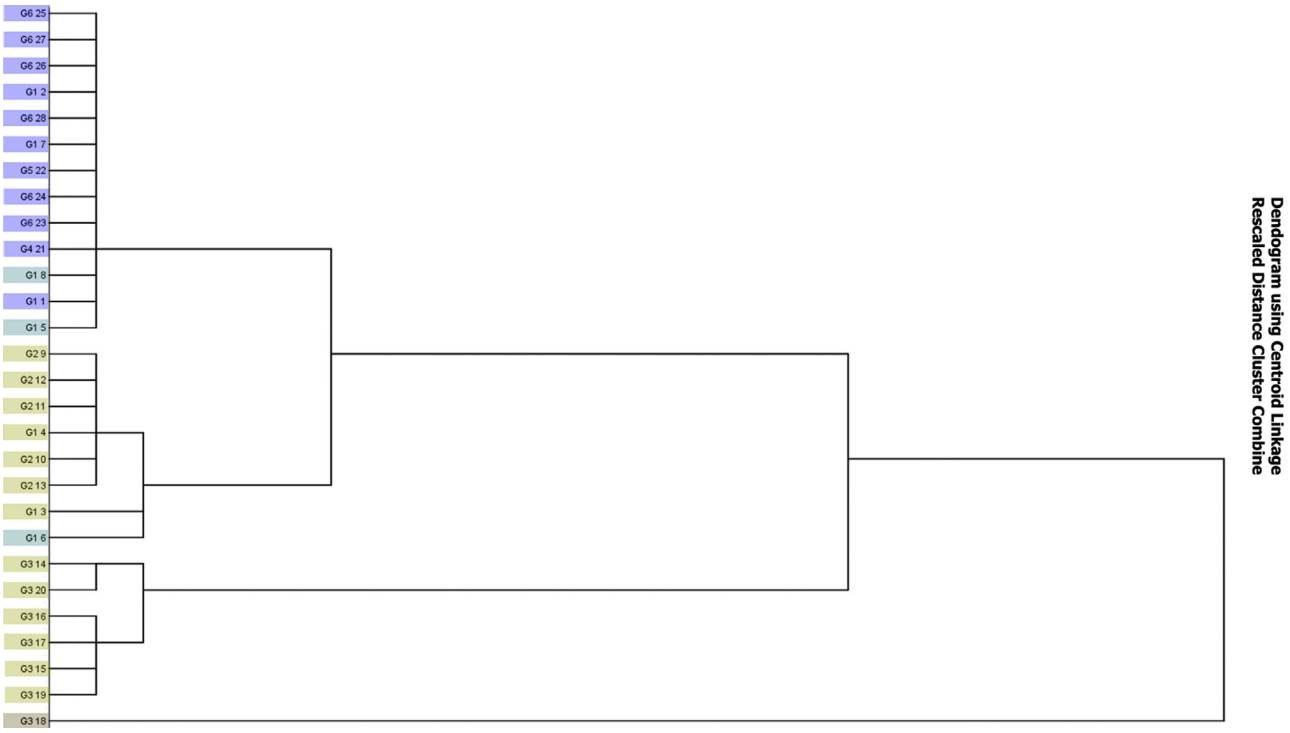

Fig. 3. HCA of 28 HCV naturally occurring variants from previous article. A modified figure from our previous article (Antunes et al., 2011), representing a hierarchical cluster analysis (HCA) of 28 pMHC-I complexes loaded with HCV-derived epitopes covering all six HCV genotypes (from G1–G6). The input data was extracted from a single spot in the surface, and provided information on charges distribution using color histograms (RGB) values. The dark blue boxes indicate the G1_01 cross-reactive complexes, light blue boxes depict the intermediate targets, yellow boxes indicate targets with low cross-reactives and brown boxes indicate the target with no cross-reactives. The dendogram was generated by the SPSS software, using hierarchal clustering, with centroid method and squared euclised. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

obtained from an individual vaccinated with the wild-type epitope HCV-$NS3_{1073}$ (CINGVCWTV) (Fytili et al., 2008). The level of IFN-gamma production stimulated against a highly cross-reactive variant from genotype 1 (G1-01: CVNGVCWTV) was defined as a reference of high response, which was used to classify the other variants into high, intermediate or low cross-reactive complexes (Table A.2).

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

A HCA based in our improved approach was able to divide the complexes into two main clusters (Cluster 1 and Cluster 2) and one out-group represented by G3-18 (Fig. 2). A threshold was defined with the *pvrect* function to highlight these groups (*alpha* = 0.95), which are corroborated by AU *p*-values with low standard errors (Fig. 2). The variant G3-18 (from genotype 3) fell in a completely independent branch. This result is in agreement with our previous analysis and with the experimental data, since G3-18 was the only among the 28 complexes that presented no detectable response *in vitro* (Fytili et al., 2008). All the high G1_01 (HCV-$NS3_{1073}$) cross-reactive complexes fell in Cluster 1 (*AU* = 98). Of note, in the *in vitro* assay, the complexes with the higher IFN-γ levels within the cross-reactive complexes were G1-02, G1-07, G5-22, G6-25 and G6-27 (Antunes et al., 2010, 2011; Fytili et al., 2008; Sinigaglia et al., 2013). With the exception of G5-22, all other complexes fell in the same sub-cluster of the reference variant G1-01 (*AU* = 76). It is important to note that this level of information was not contemplated by our previous work (Fig. 3). The high responder variant G6-26 and the intermediate responder G1-05 fell in separate branches, but still within the main cluster of the cross-reactive complexes (*AU* = 98). It is also important to mention that our previous analysis of these complexes presented the intermediate responder G1-05 as the closest related complex to the reference complex G1-01 (Antunes et al., 2011). We explained this unexpected result by suggesting that despite the surface charges distribution other issues might account for the lower response presented by G1-05. Our improved approach was able to identify neglected structural differences between G1-01 and G1-05, and correctly placed G1-05 outside the sub-clusters of high responders.

All low cross-reactive complexes fell in Cluster 2 (*AU* = 97). The low responders from genotype 1, G1-03 and G1-04, fell correctly into this main low responders cluster, as well as the intermediate responders G1-06 and G1-08. The complex G1-06 was also placed within the low responders in the original analysis (Fig. 3). Of note, a trend to the separation of the variants according to their genotypes is also observed, since we have a sub-cluster only with G3 complexes (*AU* = 79) and a sub-cluster with the majority of G2 complexes (*AU* = 99). Our HCA results also provide other suggestions, such as that G1-08 is more closely related to G2-11 and G3-20 (*AU* = 95) than to G1-06. However, to these new cross reactive suggested targets, there is no experimental background in Fytili's paper to support this level of speculation (Fytili et al., 2008). Note that the *in vitro* assay with these 28 HCV variants was performed to verify the cross-reactivity against the wild-type HCV-$NS3_{1073}$. Cross-reactivity also depends on the T-cell population involved, so to evaluate the cross-reactivity against G1-08, an assay with a G1-08-specific T-cell population would be needed.

2.4. Cross-reactivity prediction among dengue virus serotypes

Dengue virus (DV) represents a major challenge for vaccine development (Halstead, 2013). Despite effective immunization against one serotype is easy to achieve, and protective T-cell response is observed, challenge of an immunized individual with an heterologous serotype often leads to severe symptoms, such as dengue hemorrhagic fever and dengue shock syndrome (DHF/DSS).

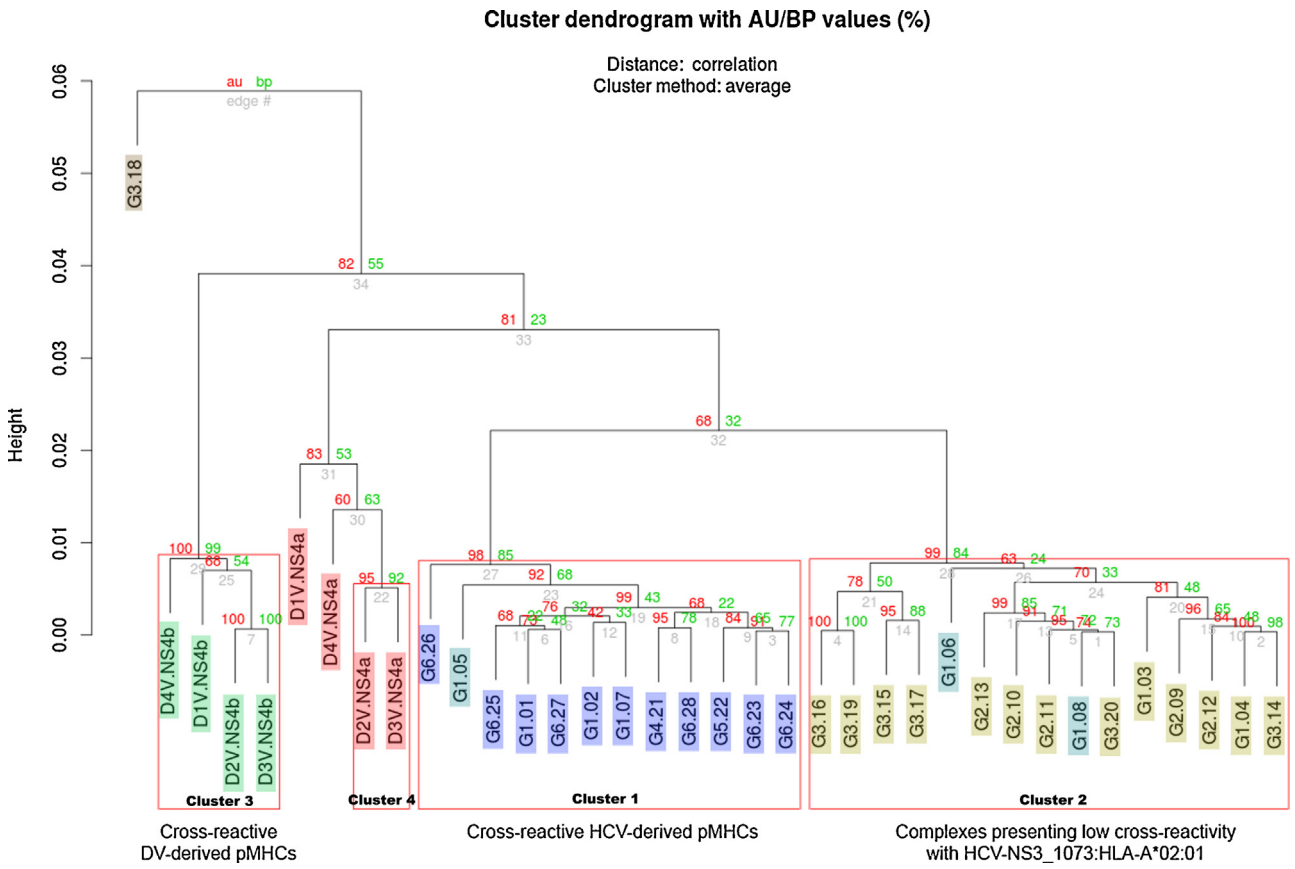

Fig. 4. Structure-based hierarchical clustering of pMHC-I complexes. Dendrogram of 36 pMHC-I complexes representing the hierarchical cluster analysis performed with the Pvclust R package. The input data was accessible surface area values and color histograms (RGB) for each pMHC-I, which provided information on topography and charges distribution over the surface. Red boxes indicate the main clusters identified (*alpha* = 0.95). Cross-reactive and non-cross-reactive complexes of both subsets (HCV and DV) fell in independent clusters. AU, approximately unbiased; BP, bootstrap probability. Dark blue box indicates the G1.01 cross-reactive complexes, light blue box indicates the intermediate targets, yellow box indicates targets with low cross-reactives and brown box indicate the target with no cross-reactives. Green box indicates NS4b targets and red box indicates NS4a targets. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

In this context, cross-reactive T-cells are believed to mediate the immunopathogenesis of DHF/DSS during secondary heterologous challenge (Duan et al., 2012). Therefore, the identification of non-cross-reactive immunogenic targets, specific for each DV serotype, is one way to develop a combined tetravalent vaccine. In a recent publication, Duan et al. (2012) identified HLA-A*02:01-restricted peptides from the four DV serotypes, and examined their immunogenicity and cross-reactivity. From their data, we extracted the epitope sequence of two groups of targets, one being identified as (i) cross-reactive variants, and the other as (ii) non-cross-reactive variants (Fig. A.2).

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

We performed new predictions with the combined data from both subsets (HCV and DV), totaling 36 pMHC-I complexes. The HCV and DV variants fell in independent main clusters, HCV maintaining the same complexes in Clusters 1 and 2 and defining two more groups (Clusters 3 and 4). The same threshold (*alpha* = 0.95) was able to identify cross-reactive and non-cross-reactive complexes within these groups (Fig. 4). All four NS4b variants fell in the same cluster (Cluster 3) (*AU* = 100). This was expected, since cross-reactive *in vitro* response was indeed observed for these four variants. The same level of clustering was not observed for the NS4a variants (*AU* = 83), a group that did not present cross-reactivity in the study of Duan et al. (2012).

The variants D1V-NS4a$_{140}$ and D4V-NS4a$_{140}$ fell in independent branches, while the other two (D2V-NS4a$_{140}$ and D3V-NS4a$_{140}$) fell in the same cluster (*AU* = 95). Our HCA, therefore, indicates a possible cross-reactivity between D2V-NS4a$_{140}$ and D3V-NS4a$_{140}$, which could be understood as a false positive result. However, it is important to highlight that cross-reactivity is also dependent on the specific T-cell population involved, and normally produces responses with lower intensity when compared to the challenge with the cognate peptide. Of note, the D2V-NS4a$_{140}$ presented really low levels of response even upon challenge with the cognate epitope (Fig. A.1) (Duan et al., 2012). Despite of a possible structural similarity (Fig. 5), a cross-reactive response would be probably undetectable with this T cell population. However, our approach relies exclusively on structural features of the pMHC-I surface, such as charges distribution and ASA values, and therefore is capable of identifying the closer related complexes. Also, other features in antigen processing might prevent the T cell stimulation process.

Finally, the combined HCA (HCV and DV) was able to reproduce the same results observed in the independent HCV analysis. This combined approach corroborates the consistency of our method, even with a greater number of complexes, suggesting its possible use in a larger scale as a virtual screening method. In this sense, we also explored an alternative way to present our HCA results. Instead of a dendrogram, this data can be used as input for relational networks, which can provide more intuitive information about the

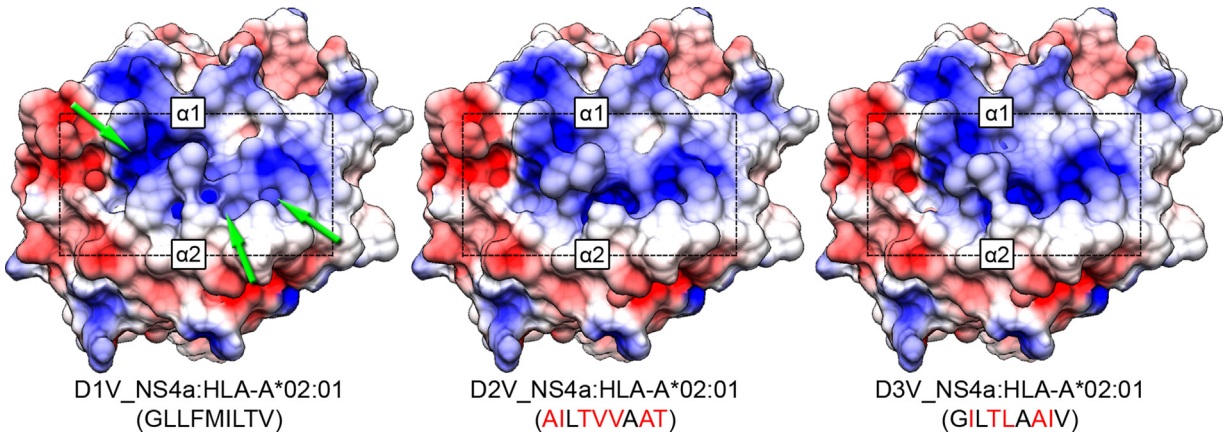

Fig. 5. Topography and electrostatic potential comparison among pMHCs presenting dengue-derived epitopes. "TCR-interacting surfaces" of three pMHC-I complexes presenting epitopes derived from three different Dengue Virus serotypes are depicted. Regions with positive (blue) and negative (red) charges are represented with a scale from −3 to +3 kT. Sequences of presented peptides are depicted below each complex, with mutations in relation to "D1V" indicated in red. Alpha-1 and Alpha-2 MHC domains are also shown. TCR-interacting surfaces of complexes "D2V" and "D3V" share greater similarity in terms of electrostatic potential, while "D1V" presents some differences in three positively charged spots (green arrows). Images were obtained with the UCSF Chimera package (Trott et al., 2010). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

cross-reactive-networks studied (Fig. 6), indicating, however, the same relationships.

2.5. Applicability to vaccine development

Several immunogenic targets were identified and successful immunization can be achieved, but HCV diversity remains a major challenge. The identification of targets capable of triggering cross-genotype responses could drive the efforts to develop a new generation of vaccines, improving vaccination coverage.

On the other hand, cross-reactivity is an issue to be avoided in a DV vaccine development, since it is involved in the immunopathogenesis of DHF/DSS. Once again, our improved structurally based prediction could be applied as a virtual screening method to identify undesirable cross-reactive responses that are unknown, and must be tested before the use of predicted targets in an anti-DV vaccine.

Traditional methods of vaccine development provided some successful results, but have been unable to overcome some of the major challenges for global health, such as the control of HIV and HCV. In that context, a new generation of rationalized vaccines is

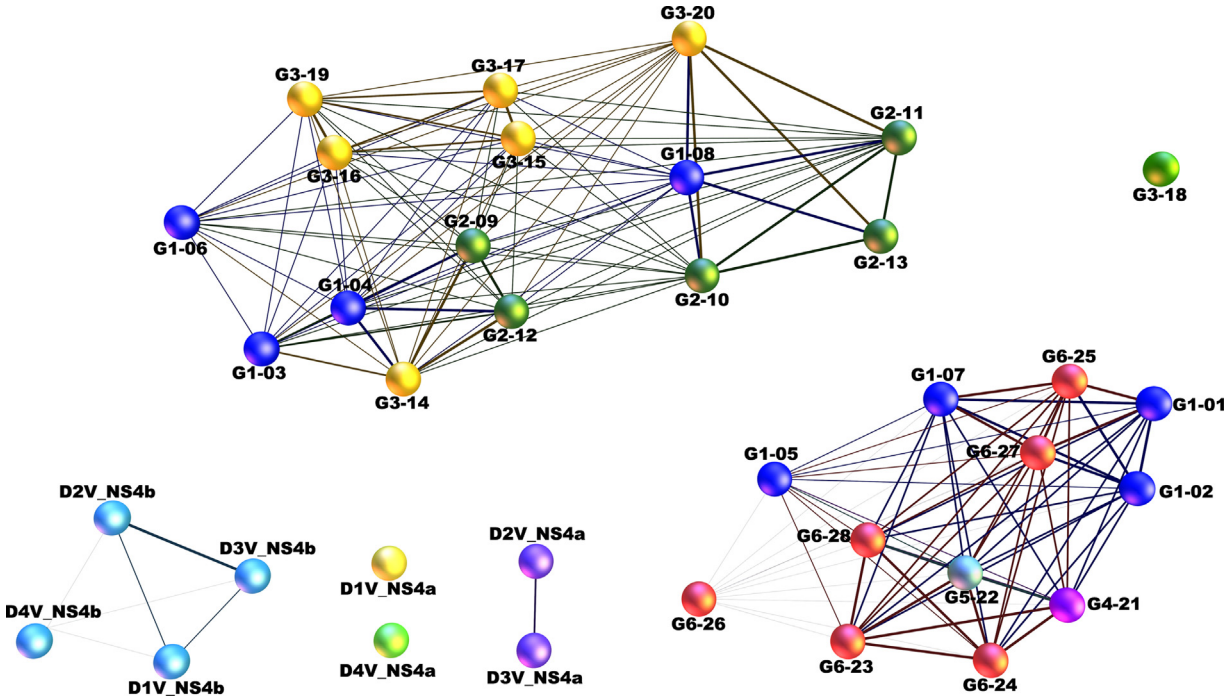

Fig. 6. Relational network of 36 pMHC-I complexes. Relational network generated with the Gephi program, based on the dendrogram of 36 pMHC-I complexes (Fig. 4). Each sphere represents a given pMHC-I and different colors indicate different HCV genotypes or DV serotypes. For instance, red spheres indicate pMHC-I complexes loaded with HCV genotype six epitopes. Lines (*edges*) indicate cross-reactivity between the connected complexes (*nodes*), complexes without connections are considered non-cross-reactive. The strength of each line indicates the similarity between the connected complexes, being a structure-based indicative of the strength of the cross-reactivity between them. The distribution of the clusters is merely representative, and distance between *nodes* in the picture has no meaning. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

starting to be planned, and bioinformatics tools are playing a major role in this process (Donati and Rappuoli, 2013; Dormitzer et al., 2012). Combined *in silico* approaches can save time and money, identifying the candidates more likely to stimulate the desired immune response, which can then be tested with *in vitro* and *in vivo* experiments to confirm its safety and efficacy for the use in a new vaccine.

3. Conclusions

The CD8+ T-cell cross-reactivity is a complex phenomenon triggered by the structural similarity between two different pMHC-I complexes that are recognized by the same TCR. Despite the enormous variability of TCRs and epitopes involved in these interactions, there are few conserved contacts that are shared by all TCR/pMHC-I crystal structures available, providing a map of the most important regions over the pMHC-I surface. Moreover, cross-reactivity between two pMHC-I complexes can be predicted based on the electrostatic potential over these selected regions. Although there are many studies about possible characteristics that trigger cross-reactivity, our method applying electrostatic potential (Antunes et al., 2011) and topology data to predict cross reactivity is a new one in this field. Our innovative approach showed that use of ASA values can improve this prediction, adding valuable information on the topography of these complexes. Finally, the use of an R package to assess the uncertainty of the hierarchical clustering provided a statistical validation of the results. Our method can be applied in rational vaccines construction, allowing to predict the impact of heterologous immunity and anticipate individual response to vaccination (Wlodarczyk et al., 2009; Zhang et al., 2015). It can also be used to predict unexpected off-target toxicity in T cell based immunotherapies for cancer, field in which cross-reactivity has become a major concern (Linette, 2013; Stone et al., 2015).

The presented results demonstrate that our technique is on the right track. The next steps to consolidate this approach will come with the increase on analyzed cross-reactive networks, through the recovery and inclusion of *in vivo* experimental data available in scientific literature. This increase in the number of networks will strengthen the specificity of the approach, decreasing the number of false positive results. Alternatively, we aim to implement a strategy using neural network or Suport Vector Machine algorithms to infer immunogenicity in pMHC complexes considering their charge distribution and topographic patterns. These different tools will became available in our immunoinformatics platform Crosstope – Structural Data Bank for Cross-Reactivity Assessment (http://www.crosstope.com.br).

4. Materials and methods

4.1. Identification of conserved contacts between TCRs and pMHCs

An extensive search for all available crystal structures of TCR/pMHC-I complexes restricted to HLA-A*02:01 with 9 residues epitopes was performed in the Protein Data Bank and IMGT/3D structure-DB (Kaas et al., 2004). Curated and calculated contacts between TCR and pMHC, for each complex, were obtained from IEDB-3D (Ponomarenko et al., 2011). An arbitrary cut-off of 85% and 60% was used to select TCR-interacting residues of the pMHC to retrieve electrostatic potential and ASA (Fig. 1) values, respectively. Information on included complexes is provided in Table A.1. Considering the nine key positions identified in crystal structures, we defined a group of seven regions over the pMHC-I surface (Fig. 1A). These regions, or "gates", were defined considering the specific contribution of each one of these residues to the pMHC-I surface. Three regions were defined covering the epitope surface. The contribution of epitope positions p4 and p5 were collected by two independent gates (G1 and G2). In the case of positions p6, p7 and p8, only one gate was defined, centered over p7 (G3). This was decided because p7 is much more exposed to the contact with the TCR, while p6 and p8 have a lower contribution to the pMHC-I surface. Other four gates were defined over selected MHC-I residues (G4, G5, G6 and G7). These seven key regions are in agreement with previously described "TCR footprints" for this allotype (Gras et al., 2009, 2012; Rudolph et al., 2006) and, therefore, will be probably involved in cross-reactive responses.

4.2. Construction of pMHC-I complexes

All our structural analysis were performed with pMHC-I complexes obtained through the previously described *D1–EM–D2* approach (Antunes et al., 2010). Briefly, only the FASTA sequence of the epitopes was recovered from the reference studies (Duan et al., 2012; Fytili et al., 2008) and used as input to produce 3D structures of these epitopes, with PyMOL scripts. A "donor" structure of an empty HLA-A*02:01 was obtained by removing the epitope from a reference PDB structure (Protein Data Bank code 2V2W). The new pMHC-I structure, harboring the epitope of interest in the context of HLA-A*02:01, was then obtained by a combined sequence of molecular docking and energy minimization steps. These steps were performed with AutodockVina (Trott et al., 2010) and GROMACS 4.5.1 (Pronk et al., 2013), respectively. The accuracy and reliability of this *D1–EM–D2* approach was tested in previous studies (Antunes et al., 2010; Sinigaglia et al., 2013).

4.3. Electrostatic potential and ASA calculations over the pMHC-I complexes

Electrostatic potential for each pMHC-I structure was calculated with Delphi (Li et al., 2012), with custom parameters (*e.g.*: *indi* = 1.0, *exdi* = 80.0, *prbrad* = 1.4, *salt* = 0.2). Accessible surface area (ASA) from each pMHC-I complex was calculated with NACCESS V2.1.1 (http://www.bioinf.manchester.ac.uk/naccess/), which in a simplified explanation calculates the atomic accessible surface by rolling a probe of specific size around a van der Waals surface, of the selected residues. In this work, we used a probe size 1.40 Å.

4.4. Image acquisition and data extraction

Images of the electrostatic potential distribution over the "TCR-interacting surface" of each pMHC-I were obtained with the UCSF Chimera package from the Resource for Biocomputing, Visualization, and Informatics of the University of California, San Francisco (Pettersen et al., 2004). The "Electrostatic surfacing coloring" option of Chimera was used to import and visualize the electrostatic potential calculated with Delphi, using a range from −3 to +3 kT. Selected regions over these images were defined, and color histograms (RGB) of these areas were obtained with ImageJ 1.43u software (National Institute of Health, USA, http://rsb.info.nih.gov/ij). In total, 42 values were obtained from the seven histograms of each image, such as color mean and standard deviation for each RGB component. Figures included in the article were edited with Adobe Photoshop CS2 v.9.0. program (Adobe, San Jose, CA).

4.5. Clustering analysis

As previously described, our prediction method was based on the use of pMHC-I structural features as input for multivariate statistical methods (Antunes et al., 2011). Originally, only information on electrostatic potential was used to define the clusters of putative

cross-reactive complexes. Now, we combined additional information on ASA values and improved our approach with the use of an R package (*pvclust*) to assess the uncertainty of the hierarchical cluster analysis (HCA) (Suzuki and Shimodaira, 2006). This package provides both bootstrap probability (BP) and approximately unbiased (AU) *p*-values, which are computed by multiscale bootstrap resampling, and has been shown to be less biased than other methods in typical cases of phylogenetic tree selection (Shimodaira, 2002). The "average" linkage method was used with "correlation" distance, and the number of bootstrap replications was set to 10,000. Results were plotted as dendrograms with bootstrap probabilities (BP) and approximately unbiased (AU) *p*-values. Main clusters were identified with *pvrect* (*alpha* = 0.95) and standard errors for AU *p*-values were obtained with *seplot*. Relational networks were plotted with the open-source platform Gephi (https://gephi.org).

This improvement adds a statistical validation to the dendrogram, enriching the discussion of the results, and avoiding unsubstantiated conclusions.

Acknowledgements

We thank Jader Peres da Silva, Artur Krumberg Schüller and Marina Roberta Scheid for collaboration in some steps of this work. This work was supported by Coordenação de Aperfeiçoamento de Pessoal de Nivel Superior (CAPES) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

References

Antunes, D.A., Rigo, M.M., Silva, J.P., Cibulski, S.P., Sinigaglia, M., Chies, J.A.B., Vieira, G.F., 2011. 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. 48, 1461–1467.

Antunes, D.A., Vieira, G.F., Rigo, M.M., Cibulski, S.P., Sinigaglia, M., Chies, J.A.B., 2010. Structural allele-specific patterns adopted by epitopes in the MHC-I cleft and reconstruction of MHC:peptide complexes to cross-reactivity assessment. PLoS ONE 5, e10353.

Brehm, M.A., Selin, L.K., Welsh, R.M., 2004. CD8 T cell responses to viral infections in sequence. Cell. Microbiol. 6, 411–421.

Calis, J.J.A., Boer, R.J., Kesmir, C., 2012. Degenerate T-cell recognition of peptides on MHC molecules creates large holes in the T-cell repertoire. PLoS Comput. Biol. 8, e1002412.

Calis, J.J.A., Maybeno, M., Greenbaum, J.A., Weiskopf, D., De Silva, A.D., 2013. Properties of MHC class I presented peptides that enhance immunogenicity. PLoS Comput. Biol. 9 (10), e1003266.

Cornberg, M., Clute, S.C., Watkin, L.B., Saccoccio, F.M., Kim, S.-k., Naumov, Y.N., Brehm, M.A., Aslan, N., Welsh, R.M., Selin, L.K., 2010. CD8 T cell cross-reactivity networks mediate heterologous immunity in human EBV and murine vaccinia virus infections. J. Immunol. 184, 2825–2838.

Donati, C., Rappuoli, R., 2013. Reverse vaccinology in the 21st century: improvements over the original design. Ann. N. Y. Acad. Sci. 1285, 115–132.

Dormitzer, P.R., Grandi, G., Rappuoli, R., 2012. Structural vaccinology starts to deliver. Na.t Rev. Microbiol. 10, 807–813.

Duan, Z.L., Li, Q., Wang, Z.B., Xia, K.D., Guo, J.L., Liu, W.Q., Wen, J.S., 2012. HLA-A*0201-restricted CD8+ T-cell epitopes identified in dengue viruses. Virol. J. 9, 259.

Fernandez-Vina, M.A., Falco, M., Sun, Y., Stastny, P., 1992. DNA typing for HLA class I alleles: I. Subsets of HLA-A2 and of -A28. Hum. Immunol. 33, 163–173.

Frankild, S., de Boer, R.J., Lund, O., Nielsen, M., Kesmir, C., 2008. Amino acid similarity accounts for T cell cross-reactivity and for "holes" in the T cell repertoire. PLoS ONE 3, e1831.

Fytili, P., Dalekos, G.N., Schlaphoff, V., Suneetha, P.V., Sarrazin, C., Zauner, W., Zachou, K., Berg, T., Manns, M.P., Klade, C.S., Cornberg, M., Wedemeyer, H., 2008. Cross-genotype-reactivity of the immunodominant HCV CD8 T-cell epitope NS3-1073. Vaccine 26, 3818–3826.

Gras, S., Burrows, S.R., Turner, S.J., Sewell, A.K., McCluskey, J., Rossjohn, J., 2012. A structural voyage toward an understanding of the MHC-I-restricted immune response: lessons learned and much to be learned. Immunol. Rev. 250, 61–81.

Gras, S., Saulquin, X., Reiser, J.-B., Debeaupuis, E., Echasserieau, K., Kissenpfennig, A., Legoux, F., Chouquet, A., Le Gorrec, M., Machillot, P., Neveu, B., Thielens, N., Malissen, B., Bonneville, M., Housset, D., Gorrec, M.L., Alerts, E., 2009. Structural bases for the affinity-driven selection of a public TCR against a dominant human cytomegalovirus epitope. J. Immunol. 183, 430–437.

Halstead, S.B., 2013. Identifying protective dengue vaccines: guide to mastering an empirical process. Vaccine 31, 4501–4507.

Hoof, I., Perez, C.L., Buggert, M., Gustafsson, R.K.L., Nielsen, M., 2010. Interdisciplinary analysis of HIV-specific CD8+ T cell responses against variant epitopes reveals restricted TCR promiscuity. J. Immunol. 184, 5383–5391.

Jorgensen, J.L., Esser, U., Fazekas de St Groth, B., Reay, P.A., Davis, M.M., 1992. Mapping T-cell receptor-peptide contacts by variant peptide immunization of single-chain transgenics. Nature 355, 224–230.

Kaas, Q., Ruiz, M., Lefranc, M.-P., 2004. IMGT/3D structure-DB and IMGT/structural query, a database and a tool for immunoglobulin, T cell receptor and MHC structural data. Nucleic Acids Res. 32, D208–D210.

Kessels, H.W.H.G., de Visser, K.E., Tirion, F.H., Coccoris, M., Kruisbeek, A.M., Schumacher, T.N.M., 2004. The impact of self-tolerance on the polyclonal CD8+ T cell repertoire. J. Immunol. 172, 2324–2331.

Li, L., Li, C., Sarkar, S., Zhang, J., Witham, S., Zhang, Z., Wang, L., Smith, N., Petukh, M., Alexov, E., 2012. DelPhi: a comprehensive suite for DelPhi software and associated resources. BMC Biophys. 5, 9.

Linette, Gerald P., et al., 2013. Cardiovascular toxicity and titin cross-reactivity of affinity-enhanced T cells in myeloma and melanoma. Blood 122 (6), 863–871.

Meijers, R., Lai, C.-C.C., Yang, Y., Liu, J.-H.H., Zhong, W., Wang, J.-H.H., Reinherz, E.L., 2005. Crystal structures of murine MHC class I H-2 D(b) and K(b) molecules in complex with CTL epitopes from influenza A virus: implications for TCR repertoire selection and immunodominance. J. Mol. Biol. 345, 1099–1110.

Moise, L., Gutierrez, A.H., Bailey-Kellogg, C., Terry, F., Leng, Q., Abdel Hady, K.M., Verberkmoes, N.C., Sztein, M.B., Losikoff, P.T., Martin, W.D., Rothman, A.L., De Groot, A.S., 2013. The two-faced T cell epitope: examining the host–microbe interface with JanusMatrix. Hum. Vaccin. Immunother. 9, 1577–1586.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. UCSF chimera – a visualization system for exploratory research and analysis. J. Comput. Chem. 25, 1605–1612.

Ponomarenko, J., Papangelopoulos, N., Zajonc, D.M., Peters, B., Sette, A., Bourne, P.E., 2011. IEDB-3D: structural data within the immune epitope database. Nucleic Acids Res. 39, D1164–D1170.

Pronk, S., Pall, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M.R., Smith, J.C., Kasson, P.M., van der Spoel, D., Hess, B., Lindahl, E., 2013. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics 29, 845–854.

Richter, S., Wenzel, A., Stein, M., Gabdoulline, R.R., Wade, R.C., 2008. webPIPSA: a web server for the comparison of protein interaction properties. Nucleic Acids Res. 36 (Suppl. 2), W276–W280.

Rudolph, M.G., Stanfield, R.L., Wilson, I.a., 2006. How TCRs bind MHCs, peptides, and coreceptors. Annu. Rev. Immunol. 24, 419–466.

Sandalova, T., Michaelsson, J., Harris, R.A., Odeberg, J., Schneider, G., Karre, K., Achour, A., Michaëlsson, J., Kärre, K., 2005. A structural basis for CD8+ T cell-dependent recognition of non-homologous peptide ligands: implications for molecular mimicry in autoreactivity. J. Biol. Chem. 280, 27069–27075.

Selin, L.K., Nahill, S.R., Welsh, R.M., 1994. Cross-reactivities in memory cytotoxic T lymphocyte recognition of heterologous viruses. J. Exp. Med. 179, 1933–1943.

Shimodaira, H., 2002. An approximately unbiased test of phylogenetic tree selection. Syst. Biol. 51, 492–508.

Sinigaglia, M., Antunes, D.A., Rigo, M.M., Chies, J.A., Vieira, G.F., 2013. CrossTope: a curate repository of 3D structures of immunogenic peptide: MHC complexes. Database (Oxford), bat002.

Stone, J.D., Harris, D.T., Kranz, D.M., 2015. TCR affinity for p/MHC formed by tumor antigens that are selfproteins: impact on efficacy and toxicity. Curr. Opin. Immunol. 33, 16–22.

Suzuki, R., Shimodaira, H., 2006. Pvclust: an R package for assessing the uncertainty in hierarchical clustering. Bioinformatics 22, 1540–1542.

Trott, O., Olson, A.J., News, S., 2010. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J. Comput. Chem. 31, 455–461.

Vieira, G.F., Chies, J.A.B., 2005. Immunodominant viral peptides as determinants of cross-reactivity in the immune system – can we develop wide spectrum viral vaccines? Med. Hypotheses 65, 873–879.

Wedemeyer, H., Mizukoshi, E., Davis, A.R., Bennink, J.R., Rehermann, B., 2001. Cross-reactivity between hepatitis C virus and Influenza A virus determinant-specific cytotoxic T cells. J. Virol. 75, 11392–11400.

Welsh, R.M., Fujinami, R.S., 2007. Pathogenic epitopes, heterologous immunity and vaccine design. Nat. Rev. Microbiol. 5, 555–563.

Welsh, R.M., Selin, L.K., 2002. No one is naive: the significance of heterologous T-cell immunity. Nat. Rev. Immunol. 2, 417–426.

Wlodarczyk, M.F., Kraft, A., Chen, H., Selin, L.K., 2009. Protection or immunopathology upon heterologous virus infection: a decision of memory cells. J. Immunol. 182, 43.15 (meeting abstract supplement).

Wucherpfennig, K.W., Call, M.J., Deng, L., Mariuzza, R., 2009. Structural alterations in peptide–MHC recognition by self-reactive T cell receptors. Curr. Opin. Immunol. 21, 590–595.

Zhang, S., Bakshi, R., Suneetha, P., Fytili, P., Antunes, D., Vieira, G., Jacobs, R., Klade, C., Manns, M., Kraft, A., Wedemeyer, H., Schlaphoff, V., Cornberg, M., 2015. Frequency, privacy and cross-reactivity of pre-existing HCV-specific CD8+ T-cells in HCV seronegative individuals: implication for vaccine responses. J. Virol. (in press).