Molecular Immunology 48 (2011) 1461–1467

Contents lists available at ScienceDirect

Molecular Immunology

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

Structural *in silico* analysis of cross-genotype-reactivity among naturally occurring HCV NS3-1073-variants in the context of HLA-A*02:01 allele

Dinler A. Antunes, Maurício M. Rigo, Jader P. Silva, Samuel P. Cibulski, Marialva Sinigaglia, José A.B. Chies, Gustavo F. Vieira*

NBLI - Núcleo de Bioinformática do Laboratório de Imunogenética, Department of Genetics, Federal University of Rio Grande do Sul, Porto Alegre, Brazil

ARTICLE INFO

Article history:
Received 19 January 2011
Received in revised form 15 February 2011
Accepted 28 March 2011
Available online 21 April 2011

Keywords:
Cross-reactivity
Immunogenicity
Structural bioinformatics
Hepatitis C virus

ABSTRACT

Cellular immune response plays a central role in outcome of Hepatitis C Virus (HCV) infection. While specific T-cell responses are related to viral clearance, impaired responses can lead to chronic infection, turning HCV variability into a major obstacle for vaccine development. In a recent work, Fytili et al. (2008) studied the cross reactive potential of HCV specific CD8+ T-cells and observed a large variation in immunogenicity among 28 naturally occurring $NS3_{1073}$ variants. In this work, we intend to evaluate this immunogenic variation at molecular level, through bioinformatics approaches. The *D1-EM-D2* strategy was used to build *in silico* MHC:peptide complexes (pMHC) of these HCV-derived peptides in the context of HLA-A*02:01 allele. The TCR-interacting surface of these complexes were evaluated using the GRASP2 program. Structural analysis indicated a sharing of topological and electrostatic features among complexes that induced strong response *in vitro*. Besides, complexes that induced low response presented an important positively charged spot in the center of TCR-interacting area. This spot was seen even in complexes with conservative amino acid changes and is consistent with the impairment of recognition by wild-type-specific T-cells, observed *in vitro*. Furthermore, the most remarkable difference in electrostatic potential was seen precisely in the only complex unable to induce *in vitro* stimulation. All these observations were confirmed by Principal Component Analysis (PCA) and this approach was also applied to a set of 45 non-related immunogenic viral epitopes, indicating possible new targets for cross-reactivity studies. Our results suggest structural *in silico* analysis of pMHC complexes as a reliable tool for vaccine development, affording to predict the impact of viral escape mutations and selection of epitopes with potential to induce cross-reactive immune responses.

1. Introduction

Hepatitis C Virus (HCV) infection is persistent in most of the infected individuals, being the major cause of chronic liver disease, cirrhosis, and hepatocellular carcinoma. Cytotoxic T Lymphocyte (CTL) response seems to play a central role in the outcome of this infection, associated with both viral clearance and immunopathogenesis of the infection (Thimme et al., 2002; Vertuani et al., 2002). Strong cellular immune response, especially in the acute phase of infection, has been related to viral control and eradication, nourishing expectations regarding vaccine development (Cucchiarini et al., 2000; Hiroishi et al., 2010). In this context, the HLA-A2-restricted epitope $NS3_{1073}$ comes to light as a promising target, being the most frequently recognized HCV-specific epitope (Engler et al., 2004; Schlaphoff et al., 2007). $NS3_{1073}$-specific CTL response has been related to clearance of acute HCV-infection (Lechner et al., 2000; Rehermann and Nascimbeni, 2005; Thimme et al., 2002) and, probably due to its importance for viral fitness, this region presents a very limited variation among the six HCV genotypes (Soderholm et al., 2006). However, even a limited variation within an immunodominant target for host CTL response can be a huge obstacle for vaccine development, since defective responses can lead to chronic infection (Wedemeyer et al., 2002). This idea is also supported by experimental HCV infection in chimpanzees, where animals rechallenged with heterologous HCV genotypes seem to be viremic for a longer time and develop chronic infection in a rate similar to primary infections (Prince et al., 2005).

In a recent work, Fytili et al. (2008) aimed to study the sequence variability of the $NS3_{1073}$ epitope across all HCV-genotypes, to

Abbreviations: HCV, Hepatitis C Virus; MHC, major histocompatibility complex; pMHC, MHC-peptide complex; HLA, human leukocyte antigen; TCR, T-cell receptor; PCA, principal component analysis.

* Corresponding author at: Av. Bento Gonçalves 9500 Building 43323 Room 225, Bairro Agronomia, P.O. Box 15053, 91501-970 - Porto Alegre, RS - Brazil. Tel.: +51 33089938; fax: +51 33087311.
E-mail addresses: dinler@gmail.com (D.A. Antunes), mauriciomr1985@gmail.com (M.M. Rigo), jader0001@hotmail.com (J.P. Silva), spcibulski@hotmail.com (S.P. Cibulski), msinigaglia@gmail.com (M. Sinigaglia), jabchies@terra.com.br (J.A.B. Chies), gusfioravanti@yahoo.com.br (G.F. Vieira).

0161-5890/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.molimm.2011.03.019

determine HLA-A2 binding affinities of all variants identified and to investigate cross-reactivity of $NS3_{1073}$ wild-type-specific CD8+ T-cells. In one of the performed assays, the authors used peripheral blood mononuclear cells (PBMCs) from a healthy individual who was vaccinated with the peptide vaccine IC41, which includes the wild-type epitope "CINGVCWTV", and observed a large variation of immunogenicity against 28 $NS3_{1073}$ peptides from naturally occurring variants, in the context of HLA-A2, through ELISPOT assay. Although some mutations in anchor positions partially affected the HLA binding affinity, these mutations did not prevent TCR stimulation *in vitro*. Similarly, immunogenic variation was not restricted to non-conservative amino acid exchanges, since even single conservative exchanges may abolish recognition of wild-type-specific T-cells. Therefore, the amino acid sequence analysis *per se* was not able to fully explain all the immunogenic variability observed *in vitro*, and this unpredictability remains as an obstacle for vaccine development.

HCV represents a difficult challenge for conventional vaccinology and some alternatives, such as reverse vaccinology, have already been suggested (Rappuoli, 2000). In this scenario, bioinformatics tools are used to analyse the pathogen genome and predict the antigens that are most likely to be vaccine candidates. In the past few years, many other applications of computational power to address immunological problems have been proposed, and immunoinformatic has emerged as a promising field (Korber et al., 2006; Tong and Ren, 2009). One of the goals in this field is to predict the three-dimensional structure of complexes formed by hosts major histocompatibility complex (MHC) and peptides derived from a given pathogen. These MHC:peptide complexes (pMHC) play a central role in cellular immune response, once it allows the host lymphocytes to identify infected cells through the interaction with T-Cell Receptor (TCR).

Our group has used structural bioinformatics tools, such as molecular docking and molecular dynamics, to build *in silico* models of pMHC complexes that can be further analysed regarding its topology and electrostatic potential, among other features (Antunes et al., 2010; Rigo et al., 2009). Our approach, previously described as *D1-EM-D2* (*Docking 1 - Energy Minimization - Docking 2*), uses a crystal structure of a given MHC as a macromolecule for cross-docking with the structure of each peptide. One intermediate step of energy minimization is also performed to accommodate the MHC side chains to each new ligand. In the present work, this strategy was applied to build *in silico* pMHC complexes of the 28 HCV derived peptides previously tested *in vitro* by Fytili et al. (2008), in the context of human MHC allele HLA-A*02:01. Aiming to evaluate the immunogenic variation of these complexes at molecular level, images of the TCR-interacting area of these complexes were obtained and used as input to multivariate statistical methods. In addition, we submitted these 28 complexes to a short molecular dynamics and reevaluated their TCR-interacting surfaces.

2. Material and methods

2.1. Construction of pMHC complexes

Complexes were built through combined use of molecular docking and energy minimization, previously described as *D1-EM-D2* approach (Antunes et al., 2010). Briefly, a crystal structure of HLA-A*02:01, Protein Data Bank (PDB) access code 2V2W, was used as "MHC donor" (macromolecule) for cross-docking with the structure of each peptide. Epitopes were built from its linear amino acid sequence using the crystal structure of an HLA-A2-restricted peptide, PDB access code 1T1Z, as a reference for backbone three-dimensional conformation. Molecular dockings were performed with AutoDock Vina 1.0.2 (Trott and Olson, 2009) using default values (e.g. exhaustiveness = 8). All torsions in the macromolecule were kept rigid, as torsions in ligand backbone, while ligand side chains were kept completely flexible. For each epitope, the same input files were used to run AutoDock Vina 20 times, generating a final population with up to thousand different conformations (20 outputs with up to 50 structures each). Binding energy (BE) values of the best conformation of each output were used to establish a threshold, and the conformations above this threshold were used to calculate the root mean square deviation (RMSD) among the selected conformations. These steps were performed using an automated script developed by our team, which uses the g_confrms software, GROMACS 4.0.7 package (Van Der Spoel et al., 2005), to calculate the RMSDs. The conformation with the lowest deviation among the selected ones was chosen as the final docking result (side chains in most frequent positions) and saved in a pdb file, along with the "MHC donor". This complex is then submitted to an energy minimization (EM) step, aiming to adjust the "MHC donor" to its new ligand, and the resulting complex is used into a new round of molecular docking. EMs were also performed with GROMACS 4.0.7 package (Van Der Spoel et al., 2005).

All complexes built for these work, as well as GRASP2 images of these complexes, are being compiled to be released as part of the CrossTope Structural Data Bank for Cross-Reactivity Assessment (www. crosstope.com.br). This database also stores the structures (PDB files) and GRASP2 images for other immunogenic viral peptides in the context of four different MHC alleles (HLA-A*02:01, HLA-B*27:05, H-2D^b and H-2K^b), allowing on-line comparison of these complexes data (Sinigaglia M, personal communication).

2.2. Analyses of topology and charges distribution

The MHC surface analyses were performed with the GRASP2 program (Petrey and Honig, 2003), on Windows 7 platform. Electrostatic potential of the 28 NS3-1073-variants was calculated with a scale from −10 to +10 kT. A narrower range, from −5 to +5 kT, was used to generate images of selected immunogenic viral epitopes. Images of complexes after molecular dynamics were depicted with a range from −3 to +3 kT. All images were obtained with complexes in the same position.

2.3. Images acquisition and data extraction

Top view images of pMHCs were generated with GRASP2 program (Petrey and Honig, 2003) and seven selected regions were defined over the TCR-interacting surface (Appendix A). Color histograms (RGB) of these selected areas were obtained from all GRASP2 images using the ImageJ 1.43u software (National Institute of Health, USA, http://rsb.info.nih.gov/ij). In total, 63 values were obtained from the seven histograms of each image, such as color mean, mode and standard deviation. Figures included in the article were edited with Adobe Photoshop CS2 v.9.0. program (Adobe, San Jose, CA).

2.4. Principal component analysis (PCA)

Values of mean and standard deviation of each color (red, green and blue) were taken as correlated variables for a factor analysis using principal component method, using the SPSS software (PASW Statistics 18, IBM, Chicago IL. USA). For the 28 NS3-1073-variants, PCA was based on data from selected region 1 (S1) and two principal components explained 94.5% of total variance. PCA for selected immunogenic viral epitopes was based on data from three selected regions (S1, S2 and S5) and three principal components explained just 68.2% of total variance. In this specific case, values regarding green color were not included, given its poor contribution for the principal components.

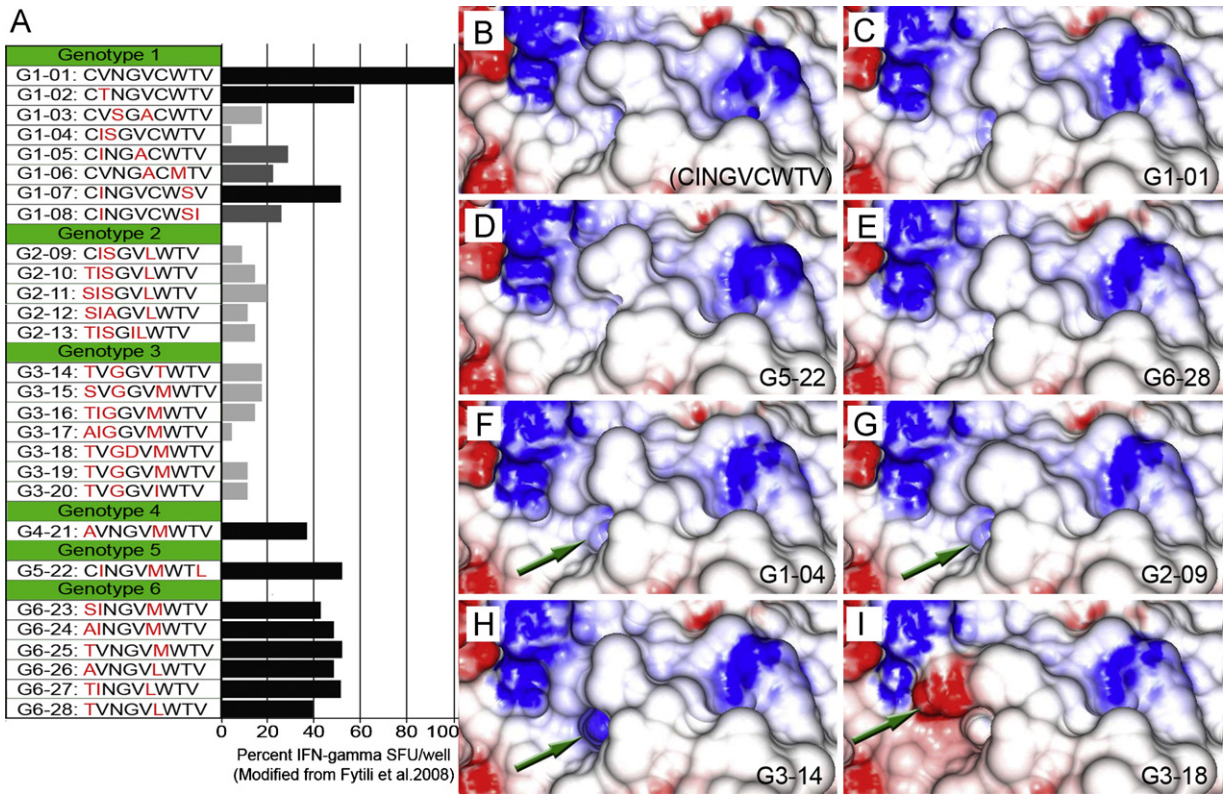

Fig. 1. Comparison between in vitro and in silico results for cross-reactivity potential among HCV- NS31073-variants. (A) Results of an *in vitro* assay performed by Fytili et al. (2008), in which $NS3_{1073}$ wild-type-specific CD8+ T-cells were tested for IFN-gamma production after stimulation with 28 $NS3_{1073}$ naturally occurring variants, in the context of HLA-A*02:01 allele (Fytili et al., 2008). (B–I) TCR-interacting area of pMHC complexes presenting some of these peptides, after construction by *D1-EM-D2* approach (Antunes et al., 2010). Charged areas over the complexes surface were computed with GRASP2 program and represented as red (negative charges) and blue (positive charges) spots, with a range from −10 to +10 kT. Complexes with strong *in vitro* response (B–E) presented similar topology and charge distribution, while those with poor *in vitro* response (F–I) presented charged areas that differentiate them from the wild-type complex (green arrows). GRASP2 images for the entire set of HLA-A2-$NS3_{1073}$ complexes are shown in Appendix C. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

2.5. Hierarchical cluster analysis (HCA)

Hierarchical cluster analysis was performed assisted by the SPSS software (PASW Statistics 18, IBM, Chicago IL. USA). The centroid clustering method was used with squared Euclidean distance interval and results were plotted as a dendrogram. HCA of the 28 NS3-1073-variants was based on data from selected region 1 (S1). HCA for selected immunogenic viral epitopes was performed with data from all seven selected regions over the TCR-interacting area of the pMHC complexes.

2.6. Molecular dynamics

Complexes presenting 28 NS3-1073-variants were submitted to a short molecular dynamics with GROMACS v4.0.7 package (Van Der Spoel et al., 2005), on Linux platform (Ubuntu 9.10). An energy minimization step was performed using steepest-descent integrator with no restrains (nsteps = 10000), allowing relaxation of the entire system. Then, the system was submitted to an equilibration stage for 100 ps, with position of non-hydrogen atoms restrained to allow the formation of solvation layers. Finally, the system was gradually heated from 50 to 310 K preserving 100 kJ^{-1} nm^{-1} of positional restraints for all heavy atoms, over a period of 100 ps.

2.7. Selection of complexes presenting immunogenic viral epitopes

A set of 45 pMHC complexes presenting immunogenic viral epitopes was selected from CrossTope Data Bank (http://www.crosstope.com.br/). These epitopes were curated from two well-known epitope databases: SYFPEITHI (http://www.syfpeithi.de/) (Rammensee et al., 1999) and IEDB (http://www.immuneepitope.org) (Vita et al., 2010). Information about protein source and epitope position was manually verified (Appendix B), as well as results of *in vitro* experiments performed to identify these epitopes as immunogenic. Selected epitopes were complexed to HLA-A*02:01 structure for further analysis of these pMHC complexes and inclusion in the CrossTope Data Bank.

3. Results/discussion

3.1. In silico study of cross-reactivity potential among the wild-type complex HLA-A2-$NS3_{1073}$ and 28 pMHCs presenting naturally occurring $NS3_{1073}$-variants

Two sequences are considered as wild-type for the $NS3_{1073}$ epitope, CVNGVCWTV and CINGVCWTV. In one of their experiments, Fytili et al. (2008) used PBMCs from a healthy individual vaccinated with the peptide vaccine IC41 (which includes "CINGVCWTV") in an ELISPOT assay. These cells were tested for IFN-gamma production after stimulation with 28 naturally occurring variants, covering all six HCV genotypes and including the wild-type CVNGVCWTV (these results were summarized in Fig. 1A).

In a previous work, our group have already analysed Accessible Surface Area (ASA) of this set of $NS3_{1073}$ peptides in the context of HLA-A*02:01 allele (Antunes et al., 2010). Although this approach alone was not able to explain all variation observed *in vitro*, ASA values from genotype 6 presented the same pattern of the wild type

peptide, while peptides with higher deviation from this "ASA pattern" were those with the lowest levels of IFN-gamma production in all ELISPOT assays.

In this work, TCR-interacting surface analysis of these 28 pMHC complexes confirmed a sharing of topological and electrostatic aspects among complexes that induced strong response *in vitro* (depicted with black bars in Fig. 1A), especially among the wild-type (G1-1) and the variants from genotypes 4, 5 and 6 (Fig. 1C–E). Differences in electrostatic potential over the pMHC surface are referred to as being more crucial for TCR recognition than subtle changes in topology (Jorgensen et al., 1992; Kessels et al., 2004). Therefore, despite some minimal topological variation, these complexes can be recognized by the same specific T-cell population. On the other hand, complexes that induced lower responses, especially variants from genotypes 2 and 3, presented an important positively charged spot in the center of the TCR-interacting area (Fig. 1F–H). Furthermore, the most remarkable difference in electrostatic potential was seen precisely in G3-18, the only complex unable to induce an *in vitro* stimulation of wild-type-specific T-cells (Fig. 1I and Appendix C). This is not the only pMHC subset in which our approach agreed with *in vitro* data regarding cross-reactive potential against a wild-type-specific T-cell population. The basis of our *in silico* approach for cross-reactivity assessment was used in a previously work to explain the immunogenic variation within a group of HCV-derived alanine exchanged peptides in the context of HLA-A*02:01 allele (Antunes et al., 2010).

Fytili et al. (2008) have also called attention for an observation that even single conservative amino acid changes impaired recognition of wild-type-specific T-cells, highlighting the risk for escape in this epitope. This feature cannot be predicted from the linear amino acid sequence, and could limit the potential efficacy of vaccines containing the $NS3_{1073}$-wild-type peptide. However, our results suggest that this feature could be predicted by TCR-interacting surface analysis, since the complexes with affected CTL recognition due to conservative amino acid changes (G1-3, G1-4 and G1-6) also presented a positively charged spot in the center of TCR-interacting area (Fig. 1F and Appendix C).

3.2. Confirmation of in silico observation by principal component analysis (PCA)

Principal component Analysis has been applied to a wide range of problems, including structural bioinformatics and immunoinformatics (Andrusier et al., 2008; Bremel and Homan, 2010a,b; Tian et al., 2008). In this work, PCA was performed with a view to evaluating the contribution of a positively charged spot, seen in some complexes surface, to the recognition by wild-type-specific T-cells. This analysis was based on color histograms extracted from GRASP2 images (see Section 2) and confirmed our previous observations, grouping peptides from genotypes 4, 5 and 6 with the optimal responders from genotype 1 (Fig. 2). Peptides from genotypes 2 and 3, which presented poor *in vitro* stimulation, formed separate clusters. Also in agreement with *in vitro* data, complex G3-18, the only non-inducer complex, was placed in a completely separated position in the scatter plot.

Complexes G1-3, G1-4 and G1-6, in which conservative amino acid changes affected CTL recognition, were distributed in positions far from genotype 6. In this analysis, complexes G1-5 and G1-8 were placed among the optimal responders, although these complexes induced an intermediate stimulation *in vitro* (depicted with dark grey bars in Fig. 1A). It is important to note that G1-5 complex, along with complexes G1-3 and G1-6, presented the most remarkable topological differences among all the studied complexes (Appendix C). This difference is given by the presence of an alanine at the fifth position (p5), which occurs only in these complexes. Topological aspects, though less important than electrostatic potential, can also

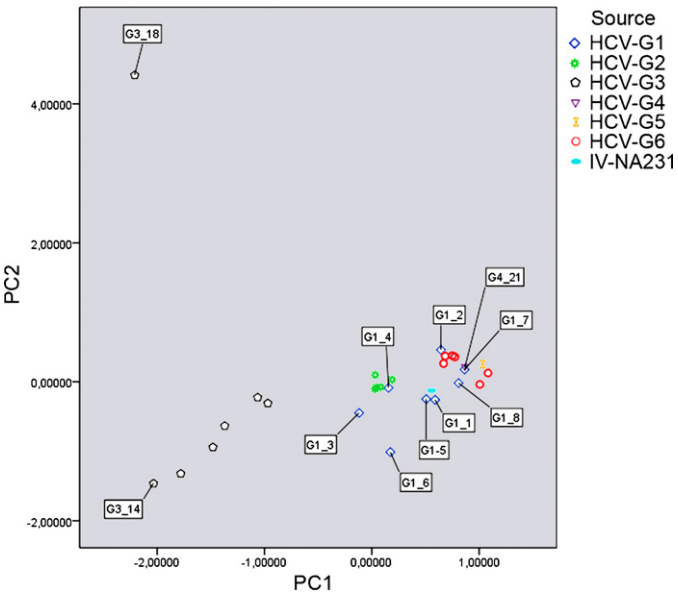

Fig. 2. Simple scatter plot of a principal component analysis (PCA). GRASP2 images were analysed using the ImageJ software and the color histogram (RGB) of a selected area was used to a PCA assisted by the PASW Statistics 18 software (for more information, see Section 2). Complexes presenting peptides from genotypes 4, 5 and 6 were grouped with the optimal responders from genotype 1. Complex G3-18, the only non-inducer complex, was placed in a completely separated position in the scatter plot. Moreover, the already described cross-reactive target IV-NA_{231} was placed near the $NS3_{1073}$ wild-type epitope.

influence the TCR/pMHC interaction. Since this PCA was based only on color information, it does not reflect topological issues.

Although according to *in vitro* assay the complex G1-8 induced intermediate levels of IFN-gamma production, this complex does not present any important structural difference as compared to the wild type. However, it is important to stress that we are using as reference just one of the experiments carried out by Fytili et al. (2008). We choose to focus our discussion in the results of the assay performed with PBMCs recovered from the vaccinated individual, aiming to avoid any external issues that could enhance the variability in our system and bias our results, but the authors presented in their work data from 7 ELISPOT assays (representative of a larger group). These experiments were performed with PBMCs recovered from seven individuals with different immunological backgrounds: healthy volunteers (anti-HCV negative), as well as acute and chronically HCV infected patients. It was observed some variation in the levels of IFN-gamma production among different experiments, even comparing the values of the same HCV variant. Considering the results of all these seven independent assays, the three complexes that we grouped as intermediate responders (G1-5, G1-6 and G1-8) presented distinct patterns. IFN-gamma production by complex G1-6 was around 20% in all assays and complex G1-5 presented similar levels, except for one assay in which this complex stimulated 40% of IFN-gamma production. On the other hand, complex G1-8 presented more than 40% of IFN-gamma production in three assays performed both with CTLs from healthy volunteers, as well as from recovered or chronically HCV infected patients. Therefore, our structural analysis is in agreement with *in vitro* data, since it was able to separate complex G1-8 from complexes G1-5 and G1-6, identifying in the former the potential to be a cross-reactive target for $NS3_{1073}$-specific T-cells.

Apart from being involved in cross-genotype-reactivity, the HCV-derived $NS3_{1073}$ epitope is also implicated in cross-reactivity with epitopes from non-related viruses. For instance, Wedemeyer et al. (2001) were able to expand specific-$NS3_{1073}$-CTLs from the blood of HCV-negative blood donors. The authors identified the

neuraminidase-derived NA_{231} epitope, from Influenza Virus (IV), as the target for these cross-reactive memory T-cells. This IV-derived epitope was included in principal component analysis and, in agreement with *in vitro* data, was placed near the $NS3_{1073}$ wild-type epitope (Fig. 2). This result indicates that structural-based approaches for cross-reactivity assessment can be used both to study determinants of related viruses, as well as to identify possible targets for cross-reactivity among any other epitopes of interest.

In addition to PCA, we used the same data from GRASP2 images to perform a hierarchical cluster analysis (HCA). The results agreed with PCA, and were plotted as a dendrogram using centroid linkage (Appendix D). All complexes from genotypes 4–6 were placed in the same cluster, along with G1-1, G1-2, G1-7, G1-8 and IV-NA_{231}. Complexes G1-3, G1-4 and G1-6 were placed in a group close to complexes from genotype 2. Finally, complexes from genotype 3 formed a separated cluster and the G3-18 complex remained as the most distant one.

3.3. Molecular dynamics with position restrains

All structural analyses described so far were taken over "static" complexes that were obtained by *D1-EM-D2* approach. Despite general similarity, some topological differences were seen among these 28 complexes, especially in complexes G1-3, G1-5 and G1-6. In order to verify if these topological differences would remain even in solution, all complexes were submitted to an energy minimization (EM) followed by a short molecular dynamics (MD). Once the main goal was to observe subtle variation in topology and electrostatic potential, a mild position restrain for heavy atoms was preserved during the simulations. These restrains allowed to accommodate side chains with unfavorable steric interactions, while avoiding a greater variability of the solvent exposed side chains.

Structural analysis taken after MD presented results in agreement with our first observations (Appendix E). Complexes G1-3, G1-5 and G1-6 preserved their topological differences even after MD, with the absence of a neutral crest which was present in all other complexes. All pMHCs from genotypes 2 and 3 preserved their positively charged spot in the TCR-interacting area and, in all cases, this spot was represented as a small cavity in the complexes surface. On the other hand, none of the complexes with good response *in vitro* presented this cavity or the positively charged spot.

Molecular dynamics also seems to smooth out the electrostatic potential over the pMHC complex. Principal component analysis is affected by this, since PCA is based on charge distribution differences. PCA using data from MD was able to cluster complexes G1-7, G4-21, G5-22 and G6-23 as the most closely related with the wild-type (G1-01). However, this analysis was not able to clearly separate all G6 complexes from those from genotypes 2 and 3, even using images with a narrow electrostatic potential range.

3.4. In silico analysis of cross-reactivity among the HCV wild-type epitope CVNGVCWTV ($NS3_{1073}$) and a set of selected immunogenic viral epitopes

Aiming at applying this structural analysis to a larger set of non-related pMHC complexes, forty-five complexes presenting immunogenic viral epitopes in the context of HLA-A*02:01 were selected from the CrossTope Data Bank (see Section 2). Most of these complexes do not possess crystal structures available, and were built using the *D1-EM-D2* approach. The IV-derived epitope CVNGSCFTV (NA_{231}) was also included in this analysis, along with the HCV-$NS3_{1073}$ wild-type and all variants from genotypes 4 to 6. Relevant information about selected epitopes, totaling 55 sequences, is shown in Appendix B.

PCA results for these 55 non-related structures preserved the clusterization of the immunogenic HCV-$NS3_{1073}$ variants studied before, as well as the proximity between the epitopes HCV-$NS3_{1073}$ (wild-type) and IV-NA_{231} (Appendix F). In addition, other complexes were identified as having similar distribution according to three principal components. For this analysis, we used data obtained from three different regions selected over the TCR-interacting surface (see Section 2). The inclusion of these new areas, however, made it impossible to summarize the whole span of variation in just two or even three principal components. For instance, the three components used in this case explained just 68.2% of the variance.

Inclusion of new selected regions can increase the power to discriminate among complexes, and this information can be used directly to an hierarchical cluster analysis (HCA). For this alternative approach, GRASP2 images from the 55 selected complexes were used to collect data from seven regions over the TCR-interacting surface (Appendix A). Data from all these regions were used as input for a HCA and plotted as a dendrogram using centroid linkage (Fig. 3). As expected, all HCV-$NS3_{1073}$ variants were grouped together, along with IV-NA_{231}. Interestingly, five complexes presenting non-related viral epitopes – CrossTope IDs A0201_0014, A0201_0073, A0201_0076, A0201_0083 and A0201_0095 – were placed in the same group.

Further evaluation of these five complexes revealed that at least two of them (A0201_0073 and A0201_0095) presented little differences in both topology and electrostatic potential, and could be promising candidates for *in vitro* cross-reactivity assays (Appendix G). Two of the remaining complexes presented some topological differences (A0201_0014 and A0201_0083) and complex A0201_0076 presented a positively charged spot in a position that was not represented among the selected regions, also differing from the HLA-A2-$NS3_{1073}$ complex presenting the wild-type.

One of the two promising complexes presents the HIV-1-derived peptide "SLYNTVATL" (GAG_{77}), which has not been so far described as a possible target for cross-reactivity with the HCV-$NS3_{1073}$ epitope. Interestingly, this same HIV epitope was already described as being cross-reactive with the HCV-NS5b peptide, in HIV/HCV co-infected individuals (Vali et al., 2010). The other indicated complex present the "LLWTLVVLL" peptide, from the Human herpesvirus 4 ($LMP2_{329}$). It is important to note that this peptide does not share even a single amino acid with the target peptide (CV/INGVCWTV) and, nevertheless, presented almost the same structural pattern 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.

The present approach was able to reproduce *in vitro* data for the identification of cross-reactive targets and suggested new candidates for future *in vitro* experiments. The seven regions used to build the dendrogram were chosen considering the spots of variation in charge distribution among the selected complexes in the whole TCR-interacting surface. Interestingly, all these selected regions are placed within an area that corresponds to already described footprints of public TCRs (Gras et al., 2009; Rudolph et al., 2006; Turner et al., 2006). Therefore, the same variation that allowed us to discriminate among complexes is probably involved with *in vivo* differential stimulation of CTLs. However, it is important to consider that this structural approach must be refined before it can be safely and efficiently used in other applications, such as vaccine development. For instance, we could use information about the whole TCR-interacting surface, extracting data from several spots and allowing the PCA to infer which spots contribute more or less to group the complexes in a given analysis.

The main idea behind this work is that structural information about pMHC complexes, especially regarding the electrostatic

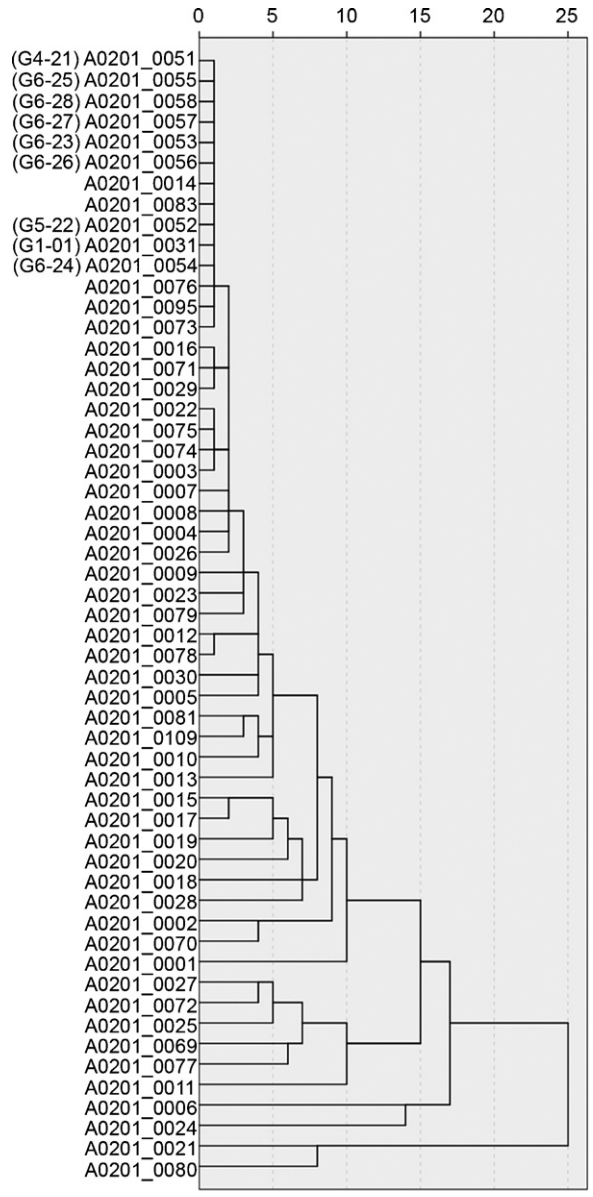

Fig. 3. Dendrogram of 55 complexes presenting selected immunogenic viral epitopes. GRASP2 images were analysed using the ImageJ software and the color histogram (RGB) of seven selected areas were used to a Hierarchical Cluster Analysis using the PASW Statistics 18 software (see Section 2). All HCV-$NS3_{1073}$ variants were grouped together, along with IV-NA_{231}. In addition, five complexes presenting non-related viral epitopes (CrossTope IDs A0201_0014, A0201_0061, A0201_0073, A0201_0076 and A0201_0083) were also placed in the same group. This complexes presented similar pattern of charges distribution over the TCR-interacting area and at least two of them (A0201_0073 and A0201_0095) could be promising candidates for *in vitro* cross-reactivity assays (Appendix G).

potential over the TCR-interacting area, can be used to predict targets of cross-reactivity for immunological applications. In addition, we postulate that future approaches for cross-reactivity prediction should incorporate even more information about the complexes, such as topology measures, ASA values, binding affinity and even data over pMHC stability, feeding one global analysis that can cover most of many details underlying TCR/pMHC interaction. This work has important implications especially to vaccine development against HCV, as it can be used to design new cross-reactive vaccines that can safely protect against a group of related viral genotypes (Hiroishi et al., 2010; Vieira and Chies, 2005; Wang et al., 2008; Yu and Chiang, 2010).

Acknowledgements

We thank the *Centro Nacional de Supercomputação* (CESUP-RS) for allowing access to its computational resources, and the collaboration of the Group of Structural Bioinformatics from the Center of Biotechnology (Cbiot/UFRGS). In particular, we thank Guilherme Menegon Giesel, for contributions about molecular dynamics parameters and standardization. We also thank Bruna Renata Silva Corrêa, from the The Laboratory for Biological Information Processing (LabPIB/USP), for suggesting the use of ImageJ software, and Sidia Maria Callegari Jacques, from the Department of Statistics (UFRGS), for contributions about the multivariate statistical methods.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.molimm.2011.03.019.

References

Andrusier, N., Mashiach, E., Nussinov, R., Wolfson, H.J., 2008. Principles of flexible protein-protein docking. Proteins 73, 271–289.
Antunes, D.A., Vieira, G.F., Rigo, M.M., Cibulski, S.P., Sinigaglia, M., Chies, J.A., 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.
Bremel, R.D., Homan, E.J., 2010a. An integrated approach to epitope analysis I: dimensional reduction, visualization and prediction of MHC binding using amino acid principal components and regression approaches. Immunome Res. 6, 7.
Bremel, R.D., Homan, E.J., 2010b. An integrated approach to epitope analysis II: A system for proteomic-scale prediction of immunological characteristics. Immunome Res. 6, 8.
Cucchiarini, M., Kammer, A.R., Grabscheid, B., Diepolder, H.M., Gerlach, T.J., Gruner, N., Santantonio, T., Reichen, J., Pape, G.R., Cerny, A., 2000. Vigorous peripheral blood cytotoxic T cell response during the acute phase of hepatitis C virus infection. Cell Immunol. 203, 111–123.
Engler, O.B., Schwendener, R.A., Dai, W.J., Wölk, B., Pichler, W., Moradpour, D., Brunner, T., Cerny, A., 2004. A liposomal peptide vaccine inducing CD8+ T cells in HLA-A2.1 transgenic mice, which recognise human cells encoding hepatitis C virus (HCV) proteins. Vaccine 23, 58–68.
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., 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., 2009. Structural bases for the affinity-driven selection of a public TCR against a dominant human cytomegalovirus epitope. J. Immunol. 183, 430–437 (Baltimore Md.: 1950).
Hiroishi, K., Eguchi, J., Ishii, S., Hiraide, A., Sakaki, M., Doi, H., Omori, R., Imawari, M., 2010. Immune response of cytotoxic T lymphocytes and possibility of vaccine development for hepatitis C virus infection. J. Biomed. Biotechnol., 263810, 2010.
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.
Kessels, H.W., de Visser, K.E., Tirion, F.H., Coccoris, M., Kruisbeek, A.M., Schumacher, T.N., 2004. The impact of self-tolerance on the polyclonal CD8+ T cell repertoire. J. Immunol. 172, 2324–2331.
Korber, B., LaBute, M., Yusim, K., 2006. Immunoinformatics comes of age. PLoS Comput. Biol. 2, e71.
Lechner, F., Gruener, N.H., Urbani, S., Uggeri, J., Santantonio, T., Kammer, A.R., Cerny, A., Phillips, R., Ferrari, C., Pape, G.R., Klenerman, P., 2000. CD8+ T lymphocyte responses are induced during acute hepatitis C virus infection but are not sustained. Eur. J. Immunol. 30, 2479–2487.
Petrey, D., Honig, B., 2003. GRASP2: visualization, surface properties, and electrostatics of macromolecular structures and sequences. Methods Enzymol. 374, 492–509.
Prince, A.M., Brotman, B., Lee, D.H., Pfahler, W., Tricoche, N., Andrus, L., Shata, M.T., 2005. Protection against chronic hepatitis C virus infection after rechallenge with homologous, but not heterologous, genotypes in a chimpanzee model. J. Infect. Dis. 192, 1701–1709.
Rammensee, H., Bachmann, J., Emmerich, N.P., Bachor, O.A., Stevanovic, S., 1999. SYFPEITHI: database for MHC ligands and peptide motifs. Immunogenetics 50, 213–219.
Rappuoli, R., 2000. Reverse vaccinology. Curr. Opin. Microbiol. 3, 445–450.
Rehermann, B., Nascimbeni, M., 2005. Immunology of hepatitis B virus and hepatitis C virus infection. Nat. Rev. Immunol. 5, 215–229.

Rigo, M., Antunes, D., Vieira, G., Chies, J., 2009. MHC: Peptide Analysis: Implications on the Immunogenicity of Hantaviruses' N protein, 5676. Springer, Berlin/Heidelberg, 160–163.

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

Schlaphoff, V., Klade, C.S., Jilma, B., Jelovcan, S.B., Cornberg, M., Tauber, E., Manns, M.P., Wedemeyer, H., 2007. Functional and phenotypic characterization of peptide-vaccine-induced HCV-specific CD8+ T cells in healthy individuals and chronic hepatitis C patients. Vaccine 25, 6793–6806.

Soderholm, J., Ahlen, G., Kaul, A., Frelin, L., Alheim, M., Barnfield, C., Liljestrom, P., Weiland, O., Milich, D.R., Bartenschlager, R., Sallberg, M., 2006. Relation between viral fitness and immune escape within the hepatitis C virus protease. Gut 55, 266–274.

Thimme, R., Bukh, J., Spangenberg, H.C., Wieland, S., Pemberton, J., Steiger, C., Govindarajan, S., Purcell, R.H., Chisari, F.V., 2002. Viral and immunological determinants of hepatitis C virus clearance, persistence, and disease. Proc. Natl. Acad. Sci. U.S.A. 99, 15661–15668.

Tian, F., Lv, F., Zhou, P., Yang, Q., Jalbout, A.F., 2008. Toward prediction of binding affinities between the MHC protein and its peptide ligands using quantitative structure-affinity relationship approach. Prot. Pept. Lett. 15, 1033–1043.

Tong, J.C., Ren, E.C., 2009. Immunoinformatics: current trends and future directions. Drug Discov. Today 14, 684–689.

Trott, O., Olson, A.J., 2009. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J. Comput. Chem..

Turner, S.J., Doherty, P.C., McCluskey, J., Rossjohn, J., 2006. Structural determinants of T-cell receptor bias in immunity. Nat. Rev. Immunol. 6, 883–894.

Vali, B., Tohn, R., Cohen, M., Sakhdari, A., Sheth, P.M., Yue, F.Y., Wong, D., Kovacs, C., Kaul, R., Ostrowski, M.A., 2010. Characterization of cross-reactive CD8+ T-cell recognition of HLA-A2-restricted HIV-Gag: SLYNTVATL and HCV-NS5b: ALYDVVSKL epitopes in individuals infected with Human Immunodeficiency and Hepatitis C Viruses. J. Virol. (JVI. 01743-10-).

Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A.E., Berendsen, H.J., 2005. GROMACS: fast, flexible, and free. J. Comput. Chem. 26, 1701–1718.

Vertuani, S., Bazzaro, M., Gualandi, G., Micheletti, F., Marastoni, M., Fortini, C., Canella, A., Marino, M., Tomatis, R., Traniello, S., Gavioli, R., 2002. Effect of interferon-alpha therapy on epitope-specific cytotoxic T lymphocyte responses in hepatitis C virus-infected individuals. Eur. J. Immunol. 32, 144–154.

Vieira, G.F., Chies, J.A., 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.

Vita, R., Zarebski, L., Greenbaum, J., Emami, a., Hoof, H., Salimi, I., Damle, N., Sette, R., Peters, A.B., 2010. The immune epitope database 2.0. Nucleic acids Res. 38, D854–D862.

Wang, S., Kennedy, J.S., West, K., Montefiori, D.C., Coley, S., Lawrence, J., Shen, S., Green, S., Rothman, A.L., Ennis, F.A., Arthos, J., Pal, R., Markham, P., Lu, S., 2008. Cross-subtype antibody and cellular immune responses induced by a polyvalent DNA prime-protein boost HIV-1 vaccine in healthy human volunteers. Vaccine 26, 1098–1110.

Wedemeyer, H., He, X.-S., Nascimbeni, M., Davis, A.R., Greenberg, H.B., Hoofnagle, J.H., Liang, T.J., Alter, H., Rehermann, B., 2002. Impaired effector function of Hepatitis C Virus-Specific CD8+ T cells in chronic Hepatitis C Virus Infection. J. Immunol. 169, 3447–3458.

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.

Yu, C.I., Chiang, B.-L., 2010. A new insight into hepatitis C vaccine development. J. Biomed. Biotechnol. 2010, 548280.