Volume 1 | Number 1 | Jan 2013 | Pages 1–100

NJC

New Journal of Chemistry
www.rsc.org/njc

A journal for new directions in chemistry

ROYAL SOCIETY OF CHEMISTRY

CNRS

This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the **Information for Authors**.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard **Terms & Conditions** and the **Ethical guidelines** still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/njc

Dynamical Insights into *Mycobacterium avium* subsp. *paratuberculosis* peptide binding differential characteristics to HLA-DRB1 proteins associated to Multiple Sclerosis

Amit Kumar[a,b,*], Leonardo A Sechi[c], Pierluigi Caboni[d], Maria Giovanna Marrosu[e], Luigi Atzori[b], Enrico Pieroni [a,*]

[a] CRS4 Science and Technology Park Polaris, Biomedicine Dept., Piscina Manna, 09010, Pula (CA) Italy.
[b] Department of Biomedical Sciences, Oncology and Molecular Pathology Unit, University of Cagliari, Cagliari, 09100, Italy.
[c] Department of Biomedical Sciences, Microbiology and Virology Unit, University of Sassari, Sassari, Italy.
[d] Department of Life and Environmental Sciences, University of Cagliari, Cagliari, Italy.
[e] Multiple Sclerosis Center, Department of Public Health and Clinical and Molecular Medicine, University of Cagliari, Cagliari, 09100, Italy.

*Corresponding authors

Dr. Amit Kumar
Visiting Scientist, Biomedicine Dept.
CRS4 Science and Technology Park Polaris,
Piscina Manna, Pula (CA). 09010.
Italy.
Tel. +39-070-9250355
Email. amit369@gmail.com, akumar@crs4.it

Abstract

Mycobacterium avium subsp. *paratuberculosis* (MAP) infections have been recently linked to multiple sclerosis in Sardinian population, with amino acid sequence 301-309 of MAP2694 protein possessing a high antigenic potential. Peptide presentation by HLA protein is an important step in adaptive immune response to pathogens, which is linked to stable peptide-HLA complex. In this study we focus on two predisposing (*15:01, *03:01) and two protective (*16:01, *15:02) HLA-DRB1 proteins in Sardinian population. We investigate in detail their binding characteristics to the MAP2694-derived peptide, with the aim to link disease susceptibility to protein structural-dynamical features and ultimately to molecular mechanisms contributing to the disease. Using dihedral angle principal components analysis, we observe distinct chemical configurations and dynamics of the HLA peptide binding, between the predisposing and protective proteins upon binding. The difference persists in the collective motion of residues near the amino-terminal region of the peptide-binding groove involving polymorphic HLA-DRB1 residue 86. Buried surface area and binding energy estimations also support the protein distinction, suggesting a preferential peptide binding towards the predisposing proteins. Finally, water dynamics analysis in the binding groove highlighted the role of slow water molecules in bridging H-bond interactions between the protein and peptide residues. Our work demonstrates the ability of molecular simulations to characterize MAP peptide binding to the proteins associated with multiple sclerosis in Sardinia. Overall, this distinction between the predisposing and protective proteins can be associated to a biological mechanism that functionally contributes to disease onset in predisposed individuals.

Keywords: Molecular Recognition, Molecular Dynamics, Multiple Sclerosis, Antigen presentation, Dihedral angle principal component analysis, Water dynamics.

1.   Introduction

1.1   Multiple Sclerosis

Multiple Sclerosis (MS) is a chronic inflammatory autoimmune disease[1], wherein nerve myelin sheets in the central nervous system get damaged or destroyed, possibly leading to permanent disability. No definite cause of MS and of the wrongly driven immune response are known, although it is widely believed that the disease onset is due to the combination of a predisposing genetic pattern and a largely unknown triggering environmental agents.[2, 3] Genome wide association studies identified specific genetic regions associated to MS.[4, 5] In particular, these studies indicated Human Leukocyte Antigen (HLA) genetic region class II DRB1 gene as the most relevant candidate for direct disease promotion.[6-8] The HLA regions codes for the Major Histocompatibility Complex (MHC) proteins,[9] that are hetero-dimer glycoproteins expressed on the surface of antigen presenting cells. In HLA class II proteins, the binding groove is shaped by two non-covalently associated polypeptide chains α1 and β1 that bind and present peptides to T cell receptors (TCR) for recognition[10] and activation of the adaptive immune system's response to foreign pathogens.[11]

1.2   MAP-derived peptide in Sardinia

Sardinia is an Italian island in the Mediterranean sea, which has a high incidence of MS, commonly attributed to the genetic peculiarity of its population.[12] To better understand the role of HLA-DRB1 variation in MS, four susceptibility proteins in Sardinian population were analyzed, namely *03:01, *15:01, *16:01 *15:02 (Fig. 1).[13] A notable difference with respect to

North Europe population is noted: the *15:01 protein is scarcely represented in Sardinia (1.5%), while proteins *16:01 (19.2%) and *03:01 (21.9%) are very common.[14] In this study HLA binding to *Mycobacterium avium* subp. *paratuberculosis* (MAP) derived antigenic peptide is analyzed. MAP is an intracellular pathogen which causes chronic inflammation of the intestine in animals commonly referred as Johne's disease in ruminants.[15] MAP infection has been associated to human diseases[16], including Crohn's disease[17], type 1 diabetes[18] and recently to MS.[19, 20] In particular, the presence of MAP specific DNA in 42% of the MS patients in Sardinia was reported.[19] Quite interestingly, the specific protein MAP2694 was also found to share high sequence similarity with chain C of a specific TCR. The same authors also identified a MAP2694 fragment with a high antigenic index, corresponding to amino acids 301-309.[19, 21] The immunological assays on MAP2694 peptides[21] are based on humoral immunity response (B-cell driven), and can be related to T-cell mediated immunity, which is the rationale of the modeling work proposed here.

1.3   Past and present modeling on antigen presentation

In our recent studies[22, 23], we have addressed the structural and dynamical comparison of two HLA proteins (the predisposing *15:01 and the protective *16:01) relevant for MS disease, in complex with one myelin-derived and one virus-derived peptide. The detailed analysis of their physicochemical properties evidenced the different binding patterns. In particular, we found the predisposing protein to be able to bind promiscuously both self and foreign peptides with similar binding characteristics, while the protective protein displayed a preferential binding towards the self peptide. These facts suggested that the predisposing protein is able to lead to a functional and molecular similarity (mimicry) between self and non-self peptides, as presented by HLA and recognized by TCR, which could erroneously trigger an immune reaction against the wrong

target.[24] Here, we extend our investigation analyzing two predisposing (*03:01, *15:01) and two

protective HLA proteins (*16:01, *15:02), in complex with MAP2694 peptide 301-309 (Fig. 2).

In our previous studies[22, 23, 25] we had already investigated myelin-derived 14-mer peptide bound

to three of these four proteins, therefore we focus here in simulating the same MAP-derived

peptide evidenced by immunological assays,[21] although 9-mer. This study represents a first step

towards understanding the molecular basis of MAP-linked MS onset. In this work we present an

extended molecular picture of peptide-protein interactions, including: (i) dihedral angle principal

component analysis (dPCA), (ii) residue collective motions, (iii) global and local flexibility, (iv)

binding groove buried surface area, (v) binding free energies estimation and (vi) water molecule

exchange dynamics bound to the peptide.

2. Methodology

2.1 Peptide-HLA model preparation

The starting structures for the predisposing proteins: DRB1*03:01 the X-ray structure of HLA-

DR3 in complex with CLIP peptide (PDB id: 1A6A), and DRB1*15:01 from HLA-DR2 in

complex with MBP peptide (PDB id:1BX2), were obtained from protein data bank. In cases of

protective protein, where no X-ray crystallographic or NMR structures were available, we

performed Homology modeling, using MODELLER software[26], for the proteins DRB1*1601,

DRB1*15:02 using as template the available X-ray structure of DRB1*01:01 in complex with

CLIP peptide (PDB id: 3PDO), which shares a high sequence similarity 94 % with the target

protective proteins. We generated 50 models of our target structure using MODELLER software

and choose the best structure based on lowest DOPE score. The quality of the modeled protein

structure was analyzed using PROCHECK program[27], which assess the stereochemical quality of

the given structure. This analysis was done using Swiss model web interface.[28] For DRB1*16:01

model structure, Ramachandran plot displayed 92.1 % of residues in the most favored region, 7.3

% in the additionally allowed region and 0.6 % in the generously allowed region. For

DRB1*15:02 model structure, Ramachandran plot displayed 92.8 % of residues in the most

favored region, 6.0 % in the additionally allowed region and 1.2 % in the generously allowed

region. The sequence of MAP peptide 301-309 was taken from MAP2694 protein, and the

structure was modeled using the same backbone conformation of CLIP peptide (PDB id: 1A6A)

and side chain was built and optimized using SCWRL program[29], which is an advanced

algorithm for predicting sidechain conformation using rotamer library approach.. Subsequently,

we used AUTODOCK software[30] to dock MAP peptide in the respective binding cleft of the four

proteins under investigation. At first, we build grid of dimension [60 x 60 x 60], with grid point

spacing of 0.575 covering the HLA peptide binding groove. We used genetic algorithm (GA) and

Lamarckian genetic algorithm (LGA) parameters for conformational sampling of MAP peptide.

In each docking experiment performed, we employed (i) 50 runs with (ii) 100 random

individuals in the first population, (iii) 25000000 energy evaluations and (iv) 27000000 as

maximum number of generation. The generated conformations of the peptide were subjected to

cluster analysis with root mean square tolerance of 2.0 Å and ranked based on docked energy

scores and population. The peptide conformer with lowest docking energy score and higher

population number was selected for performing molecular dynamics (MD) simulation.

2.2 Molecular Dynamics Simulations of MAP-HLA complexes

The missing hydrogen atoms for the MAP-HLA complexes were built using psfgen package of

VMD software[31] and using segment command to add hydrogen atoms for the residues based on

the residue definitions from the topology file. The formed MAP-HLA complex system was then

immersed in rectangular water box, followed by addition of counter-ions in order to have a neutral system. Further details about the simulation box size and the number of water molecules for the four protein systems, see supplementary material, Table S1. The TIP3P parameters[32] was used for water molecules and Charmm 27 force-field parameter[33] was used for the protein and the peptide. We used propka software[34] to assign the protonation states to all the residues of the proteins, with exception for α-Asp 66 and β-Glu 9 residues which were protonated, in accordance to a previous MD study.[35] For His residue, neutral protonation state was assigned.[22, 23, 35] The simulations were performed both with/without MAP peptide for the four protein systems. Each molecular system was energy minimized and slowly heated to 310 K in steps of 30 K with initial positional constraints of 50 kcal/(mol Å^2) on C-alpha atoms. Subsequently, positional constrains was slowly released in steps of 10 kcal/(mol Å^2). Molecular dynamic simulation of 100 ns was performed in NPT ensemble with T=310 K, and 1 atm pressure.[36] Further simulation protocol details have been described in our previous works.[22, 23] All-atom molecular dynamics (MD) simulations were performed employing NAMD software package[37] on 64 processors cluster.

2.3   MD simulation analysis

We employ backbone dihedral angles as coordinates to perform principal component analysis[38] for the HLA class II binding site residues (α chain:5-76 and β-chain 5-90) both in presence and absence of MAP peptide on MD simulation trajectory using CARMA software.[39] Subsequently cluster analysis was performed to identify grouping of clusters with the most prominent configuration. The most populated cluster (Cluster 1, see supplementary material Table S2) was then analyzed in detail. All possible binding groove conformations present in the Cluster 1 was used to calculate the average binding groove structure for the cluster. Subsequently, the binding

groove conformation with the smallest root mean square deviation value from the average structure was chosen as a representative structure of the Cluster1.

Correlated motion analysis on the representative protein structure was performed using quasi-rigid domain decomposition, PiSQRD web server.[40] The collective nature of low-energy modes in proteins suggests that their internal dynamics can be described in terms of the relative motion of a limited number of rigid subunits.[41] The dynamical decomposition implemented in the PiSQRD web server is done by calculating ten lowest energy modes via elastic network model.[42] The best assignment of amino acids into five rigid domains was done automatically by the server on choosing a desired threshold value (>80%) for the protein internal essential dynamics.

Solvent accessible surface area (SASA) analysis of the peptide-HLA binding groove structures was calculated using NACESS program.[43] Denoting by pMHC the complex composed by MAP peptide p and HLA protein MHC, then the buried surface area (BSA) is calculated as follows: BSA = SASA (p) + SASA (MHC) - SASA (pMHC). SASA calculations were performed with 200 ps time step, on 100 ns MD trajectory. The H-bonds between peptide-HLA residues were calculated using a geometrical criteria, with a donor-acceptor cutoff distance of 3.1 Å and donor-hydrogen-acceptor-cutoff angle 130 degree. H-bond that are present for at least 20% of trajectory time length are reported.

Configurational entropy of MAP peptide inside the binding groove was estimated based on quasi-harmonic approximation developed by Schlitter[44] and using CARMA software package.[39]

The binding energy of antigenic MAP peptide in the HLA complex was estimated using solvated interaction energy (SIE) approach,[45] using SIETRAJ software package.[46] We performed additional MD simulations using AMBER-99 force-field parameters, which is a prerequisite for SIETRAJ calculations, and utilized 1000 structures of the MAP-HLA complex for the free

energy calculations. In the SIE method, the peptide–HLA binding energy (ΔG_{bind}) in aqueous solution is approximated by (i) an interaction energy contribution (E_{inter}) and (ii) a desolvation free energy contribution (ΔG_{desolv}), which resemble the formalism used in MM-PBSA. Even though entropy is not included explicitly, calibration of the obtained SIE free energy is done using an empirically determined parameter, obtained by fitting a training set of 99 protein–ligand complexes. Thus, allowing a crude but effective treatment of entropy–enthalpy compensation. The calculated binding energy values are only rough estimates of true peptide-HLA binding affinity.

To understand the importance of water mediated interaction, we investigated dynamics of water molecules bound to the MAP peptide by computing survival probability N_w (t) for water-peptide bonds, as described by Sterpone et. al.[47] Furthermore, to distinguish contributions from very different time scales to water-peptide bond lifetimes we used the following function

$F (t) = n_{fast} \exp [- (t/\tau_{fast})^c] + n_{medium} \exp [- (t/\tau_{medium})] + n_{slow} \exp [- (t/\tau_{slow})]$ to fit the survival probability and extract the different temporal scales (the three half life τ).[48] In this equation, the first term corresponds to fast regime ($\tau_{fast} \sim 15$ ps in our simulation), the second term to medium regime ($\tau_{medium} \sim 100$ ps) and the last term to slow regime ($\tau_{slow} > 500$ ps). The variables n_{fast}, n_{medium} and n_{slow} are the average number of water molecules in the three time regimes. In addition, we also performed 100 ns simulation of MAP peptide in bulk water alone and calculated survival probability of water bound to MAP residues.

To check the stability of peptide-HLA complexes, root mean square deviation (RMSD) values for the C-alpha atoms of the binding groove was monitored during MD simulations (see supplementary material Fig. S1). To provide a reasonable estimate of convergence in our simulations, we calculated probability distribution of RMSD values (see supplementary material

supplementary Fig. S2) and also adapted the Good-Turing statistical approach, recently proposed.[49] The latter method allows to estimate the probability distribution of unobserved configurations as a function of RMSD distance between unobserved and observed molecular configurations in our MD simulations. We observed no significant change in shape of the probability distribution curves (see supplementary material Fig. S3).

3. Results

3.1 Dihedral angle principal component analysis for the peptide binding groove

dPCA of the HLA binding groove for all the four peptide-HLA complexes, was performed. This allowed to identify in an immediate way the distinct binding groove conformations explored by system. In Fig. 3 are the first two principal components for each trajectory snapshot from MD simulations. Each point in the plane represents, with a good accuracy, a specific configuration explored by the binding groove during MD simulation. Point density (reported in a color scale) measures the occurrence of each configuration. We note a distinctive grouping of the binding groove conformational states (Fig. 3, see supplementary material Table S2). The most populated cluster in each system, denoted as Cluster 1, represents the most relevant state and corresponds to a specific chemical configuration of the binding groove, including small fluctuations, as sampled during MD simulation. Quite interestingly, distribution of HLA binding groove fluctuations in two molecular systems (*03:01, 15:01) are quite similar (Fig. 3A and Fig. 3B, see supplementary material Fig. S4), while the protective complexes show quite distinct features (Fig. 3C and Fig. 3D). In general, in protective proteins Cluster 1 and Cluster 2 are more populated than the predisposing proteins (see supplementary material Table S2), thus suggesting a broader phase space exploration by the predisposing proteins.

3.2 **Correlated motions of binding groove residues**

To provide more insights into the structure-function relationship, the collective motions on a representative structure (the closest structure to the average one, see Methodology section) in Cluster1 was analyzed. This calculations allowed to classify the binding groove residues into five different domains, that captured more than 80% of the whole collective motions (See Methodology section). In Fig. 4, we evidence the residues of the binding groove grouped in these five domains. A notable difference between the predisposing (Fig. 4A, 4B) and protective proteins (Fig. 4C, 4D) is observed. This difference is mainly localized in the residues belonging to chains α and β near the amino-terminal region of the peptide-binding groove. More specifically, residues β 85-90 are grouped to α 45-53 only for the two protective proteins (Fig. 4C, 4D). On the other hand, residues β 83-90 are found to be grouped to α 25-29 only for the two predisposing proteins (Fig. 4A, 4B). Finally, we observe residues β 9-11 and residues α 67-76 to be grouped together for the four proteins. Overall, the main difference between the predisposing and protective proteins is localized in the amino-terminal region of the peptide-binding groove, where the residues interact in a correlated manner only for protective proteins. This analysis is also supported by cross correlation matrices of C-alpha fluctuations calculated between all possible pairs of structures grouped in Cluster 1 (see supplementary material Fig. S5).

3.3 **Flexibility of HLA binding groove and binding energy calculation**

To understand the importance of flexibility in peptide binding, the HLA-peptide binding groove was divided into four regions, denoted as D1, D2, D3, D4, with D1 being the region where the N-terminal fragment of the peptide is located.[35] The center of mass distances variation between selected residues of α- and β-chain were calculated for each compartment, with and without MAP peptide (Fig. 5). Regarding the two predisposing proteins (*03:01,*15:01) in complex with

MAP peptide, we observe the distance profile distribution to be right shifted in the regions D1, D2 and D3, and no change is observed in region D4, with respect to their respective free HLA simulations (Fig. 5A, 5B). On the other hand, for the protective *16:01 protein in complex with MAP (Fig. 5C), region D1 displays a narrow unimodal distribution, and the width profile is observed to be right shifted in the regions D2, D3 and D4, with respect to free HLA. For the second protective protein, *15:02 (Fig. 5D) complexed to MAP, a slightly different behavior is observed, with a single broad unimodal width distribution profile for region D1, while the profile is right shifted in regions D2 and D3 and no change in region D4 was noted, with respect to free HLA. The most relevant pattern distinction between protective and predisposing proteins is thus noted in region D1, where pocket 1 is located.

3.4 Configurational Entropy and Binding Energy estimates

We calculated the configurational entropy of MAP peptide in the four protein complexes (Table 1). MAP displayed higher configurational entropy when it is bound to the two predisposing protein (Table 1). Binding free energy estimates were also performed on the four MAP-protein complexes (see column 2, Table 1). Overall, the predisposing proteins displayed a slightly preferential binding to MAP peptide than the protective ones, despite an higher configurational entropy estimates. In the following sections we will show that these evidences are coherent with the buried surface area and peptide hydration analysis.

3.5 Buried surface area

To confirm the stability of formed peptide:HLA complex in the four proteins, beyond standard RMSD analysis (see supplementary material Fig. S1, S2), we calculated the buried surface area (BSA) of the peptide HLA binding groove (see Methodology section) and monitored it over the simulation length of 100 ns (Fig. 6). The observed variation (~10%) in BSA over the simulation

length of 100 ns, indicates a stable peptide HLA binding groove in the four protein complexes.[50]

Quite interestingly, both the predisposing protein complexes result in having a higher BSA with

respect to the protective ones, that is coherent with binding energy estimates.

3.6   Hydrogen bond interaction

We evaluated the H-bond interaction between the MAP and the β-chain binding site residues. In

Table 2, we report the HLA binding site residues which form persistent H-bond interactions with

the MAP peptide. A detailed comparison with the initial starting structure is summarized in

supplementary material, Table S3. We observe a mixture of protein specific and promiscuous

interactions. For instance, we observe the interacting HLA-peptide pairs H81-D301, N82-T303,

T77/N77-T303 and Y60-D309 to be shared among the four proteins complexes. On the other

hand, interacting pairs T12-K307, D28-K307 are specific for the *03:01 predisposing protein,

while the Q64-D309 interacting pair is characteristic of the two predisposing protein complexes.

It is interesting to note that MAP residue N305 is involved in two different interactions, with R13

and D28, only in the protective protein *16:01, while it is engaged in a unique interaction with

Y26, only for the predisposing *03:01 protein, in agreement with antigenic index residue

scoring.[19] Finally, we observe MAP residue V306 to interact with binding sites residues R13,

Q70, for the predisposing *15:01 and protective *15:02 proteins, and with residue 71, for the

predisposing *03:01 and protective *16:01 proteins. The only relevant interaction distinguishing

predisposing and protective complexes is confined to pocket 7, where only predisposing proteins

are able to form an hydrogen bond between Q64 with peptide residue D309. This fact is coherent

with an higher binding affinity of the predisposing protein towards the peptide.

3.7   Water dynamics inside binding groove

We calculated the average number of water molecules bound to each peptide residue, and classified them into three time regimes: fast (τ ~15 ps), medium (~ 100 ps), slow (> 500 ps), based on their survival probability around the peptide during MD simulations, fitted as superposition of exponential decay terms (See Methodology section). The analysis was carried for all the four complexes and compared to a reference simulation with MAP alone in bulk water (Table 3). To extract spatial informations, we divided the peptide in three fragments : (i) N-terminal (N-ter), comprising residues 301-303; (ii) central, 304-306; and (iii) C-terminal (C-ter), 307-309 (Fig. 2). As expected, we observe 100% of water molecules in fast regime only for MAP peptide in bulk water, and for the N-ter section in the four complexes. Overall, the emerging patterns in water dynamics, although distinguishing all the proteins and providing an in-depth understanding of the water role in mediating interactions, do not give a direct insight in disease predisposition. For this latter point, the only relevant aspect is that the total peptide hydration inside the protective proteins is higher than inside predisposing ones, that is coherent with binding affinities and BSA estimates.

3.7.1 MAP N-ter section (D301, P302, T303)

In all the four complexes, N-ter section is characterized by fast exchange of water molecules, as observed for the whole MAP peptide in water simulations. Interestingly, we note a slight increase (~1 water molecule, Table 3), of the average number of water molecules, in the case of the two protective proteins with respect to the two predisposing ones.

3.7.2 MAP Central section of (R304, N305, V306)

In general, in the central section of MAP peptide we observe ~2-3 water molecules in excess for the protective proteins with respect to the predisposing ones. Additionally, a difference between the two predisposing proteins in water dynamics was observed: *03:01 protein is characterized

by only fast (70-80% of total water molecules) and medium (22%) water exchanges, while *15:01 protein shows only fast (83%) and slow (17%) water exchanges.

Interestingly, the unique water molecule labeled w8194 (Table 4, Fig. 7B) is found to be involved in H-bond interaction between the peptide residue N305 and *15:01 residues Q70 and D28. We also observed differences in the water dynamics between the two protective proteins. Specifically, *15:02 protein is characterized by the absence of medium temporal water exchanges (Table 3). Moreover, we found two slow water molecules for *16:01 protein and one slow water molecule for *15:02 protein. These slow water molecules were found to bridge H-bond interactions between peptide residue N305 and HLA residues Y78, R13 in the *16:01 protein case (Fig. 7C, Table 4), and between R304 and HLA N82 in the *15:02 protein (Fig. 7D, Table 4).

3.7.3 MAP C-ter (K307, G308, D309)

This region comprises a glycine residue that is sandwiched between two charged amino acids. It is interesting to note that, unlike the N-ter section, the average number of water molecules bound to the C-ter section of MAP peptide is higher than in bulk water simulations (Table 4). Similar to the central section, in *03:01 and *16:01 proteins, we observed the simultaneous presence of all the three water molecule time regimes. However, we observe an higher percentage of slow water molecules for the protective *16:01 protein compared to the predisposing *03:01 protein. Regarding the other two proteins, the water dynamics is characterized by fast (78%) and slow (22%) water exchanges for *15:01, while we observe only fast water exchanges for *15:02. On the other hand, we found one slow water molecule for *03:01 protein (Fig. 7A), and two slow water molecules for *15:01 and *16:01 proteins, bridging interactions between peptide and HLA residues (Table 4).

4.   Discussion

The main objective of our study is to provide a dynamical insight into differences in MAP peptide molecular recognition operated by the predisposing and protective HLA class II proteins linked to multiple sclerosis in Sardinia. The four proteins investigated here share a high sequence identity (>90%, Fig. 1), yet *03:01 and *15:01 are susceptible while *16:01 and *15:02 are protective against multiple sclerosis, a fact that is not so unusual for autoimmune diseases.[25]

In our recent work, we had hypothesized that amino acid at position 86 could discriminate the susceptible and protective proteins belonging to the same HLA serotype subgroup, namely *15:01, *15:02 and *16:01.[13] However, the molecular mechanisms by which the protein confer susceptibility to multiple sclerosis is still debated and show little differences depending on HLA protein serotype classification.

In previous studies,[51-53] the authors have suggested the importance of MHC binding groove plasticity in the context of HLA peptide complex recognition process by T-lymphocytes via their receptor. In this work, we were guided by the hypothesis that the dynamics of the peptide is related to HLA binding groove plasticity, in particular close to the polymorphic residue β 86,[22] and is able to influence T cell receptor binding and thus differentiate predisposing and protective proteins. Therefore, a detailed molecular analysis was performed in order to identify physicochemical differences in MAP binding properties between the predisposing and protective HLA class II complexes. From dihedral angle principal component analysis, we were able to identify the most relevant molecular binding groove conformations sampled during MD simulations. Interestingly, for MAP complexed to the two predisposing proteins, the projection of the trajectory onto the first two principal component axis displayed similar cluster patterns (Fig.

3A, Fig. 3B, supplementary material Fig. S4). The analysis also pinpoints a higher ability of the binding groove to sample distinct conformations for the predisposing protein. Moreover, from correlated motion analysis, a notable difference emerged in the protein dynamics between the predisposing and protective proteins near the polymorphic residue β 86. In detail, only for MAP peptide complexed to the two protective proteins, the residues α 45-53 and residues β 85-90, were found to belong to the same domain (Fig. 4), thus dictating a correlated motion in the amino-terminal region of the peptide-binding groove. A clear difference between the predisposing and protective proteins thus emerges from the dynamic picture of binding groove plasticity near polymorphic residue β 86, that supports our hypothesis. This fact is coherent with the observation that polymorphism Gly/Val is able to characterize protective/predisposing proteins.[13, 23]

In addition, the correlated motion in the protective proteins results in a reduced local flexibility in region D1, which is marked by a left shifted width distribution profile (Fig. 5). On the other hand, the higher flexibility observed in region D1 for MAP when bound to predisposing protein is linked to the higher configurational entropy of MAP (Table 1). The calculation of entropy upon peptide binding is beyond the scope of this work. Nevertheless, to get some insight into the entropic aspects of binding we used the peptide configuration entropy inside the binding groove as a surrogate of the binding entropy. Our results suggest the existence of an entropic threshold[52] (20-30 kcal/mol), that differentiates peptide presentation between predisposing and protective proteins, and an enthalpy increase to compensate the entropy augment.[54] However, an higher value of entropy of MAP in the binding groove of predisposing proteins was not destructive for peptide binding, indeed we observed in these cases an higher value of binding free energy (~ 3 kcal/mol) with respect to the protective proteins, that is consistent with a previous peptide-MHC

binding experiment.[55] Concerning free HLA simulations, we observe in the four proteins an higher flexibility in the amino-terminal region (D1, Fig. 5) of the peptide-binding groove, with respect to the peptide bound simulations. Indeed, the amino-terminal region has been noted to be more flexible and involved in significant structural modification in the absence of peptide.[56] The pocket 1 is localized in region D1, relevant for successful peptide molecular recognition,[57] where the predisposing complexes exhibit a broader binding groove width compared to free proteins; while the opposite seems to hold for the protective ones. Therefore, steric effect of residual types, coordinated motion and flexibility analysis confirm that the predisposing proteins are able to bind slightly better to the MAP peptide than the protective ones, and localize the main source of this effect in D1 region. This behavior is exactly the same that we observed in our previous simulation with EBNA1 peptide,[22] thus confirming that at least in part multiple sclerosis susceptibility is linked to an higher ability of the predisposing proteins to recognize pathogenic invaders.

In general, hydrogen bond network in peptide-HLA class II complexes is considered an important component in adaptive immune system, including HLA transport process (where binding with CLIP peptide is required to avoid protein unfolding) that precedes the antigen presentation step.[58] Three binding site residues, Y60, H81, and N82, are found to be involved in H-bond interaction networks with the MAP peptide residues in all the four proteins (Table 2). The role of H-bond interactions involving residues 81, 82 has also been demonstrated by previous experimental studies.[59] MAP residue V306, which was predicted to have the highest antigenic index value[19] displayed persistent H-bond interactions with binding groove residues R13, Q70 for the predisposing *15:01 and the protective *15:02 proteins, and with the basic residue R71 for the predisposing *03:01 and protective *16:01 proteins (Table 2). On a careful

inspection, residue Ala at polymorphic position 71 in the predisposing *15:01 and the protective *15:02 protein (Fig. 1) can explain the absence of H-bond interaction in these two proteins. A notable difference between predisposing and protective complexes is the presence of a persistent H-bond interactions between binding site residue Q64 and MAP residue D309 only in the case of the two predisposing proteins (Table 2).

Buried surface area has been suggested to be an important parameter that controls the intrinsic flexibility in proteins[50], and in particular the reciprocal conformational plasticity resulting in an optimal binding between the T-cell receptor and the peptide-HLA complex.[60] In our case BSA analysis revealed a higher values of BSA for the predisposing proteins (Fig. 6). Buried surface area analysis can depend strongly on amino acid sequence. In the four proteins investigated here (*03:01, *15:01, *16:01, *15:02), we found sequence of *03:01 protein significantly differing from the other three protein sequences (See Figure 1). Thus this protein is potentially subject to bias from sequence dependence on the BSA. However, the other three proteins (predisposing *15:01, protective *16:01 and *15:02) share a very high sequence identity (Figure 1), thus the noted difference in the BSA must have different reasons, and our analysis pointed to these potential distinct contributions. In a previous MD simulation studies,[60, 61] the author found a direct relation between BSA, H-bond interaction and hydration of peptide, and our simulations confirm a similar relationship. The number of H-bond interactions in the three proteins *15:01, *16:01, *15:02 were identical (6-7), thus pinpointing a possible link of BSA to peptide hydration. Indeed, lower values of BSA found in the two protective proteins can be related to a higher number of water molecules bound to the peptide (Table 2), with respect to the *15:01 predisposing protein. This fact is consistent with slightly preferential peptide-HLA binding estimate observed for the predisposing proteins (Table 1) and with binding groove distinctive

flexibility in region D1. On the other hand, water molecules close to peptide-HLA interaction interface were suggested to play a key role in interactions by contributing to a more favorable binding of peptides to HLA class I proteins.[62, 63] These authors also showed the different roles performed by water molecules inside HLA class I binding groove. They demonstrated the presence of water molecules with long residence time, that are the only ones involved in bridging interaction between the peptide backbone and HLA groove residues.[63] The same effect is also observed in our simulations, with slow water molecules (Table 4, Fig.7) bridging peptide-HLA H-bond interactions. We confirmed that the slow water molecules, for all the four proteins we analyzed, contributed to a more favorable peptide binding. The total number of water molecules bound to peptide was higher for predisposing complexes as expected from our previous observations.

Immunological assays results for MAP2694 were reported for a 9-mer long peptide,[21] and we selected the same residues for our investigation. We are aware that this does not allow an easy quantitative comparison with our previous studies of 14-mer self and non-self peptides.[22, 23] Nevertheless our study confirms some qualitative similarities with the results published and the susceptibility mechanisms we proposed in the case of myelin and EBNA1 derived peptides. [22, 23]

5.   Conclusions

Overall, a plausible interpretation emerging from all these modeling evidences is that: (i) MAP peptide displays a slight preferential binding to the predisposing protein, which thus shows an higher presentation efficiency and therefore a higher probability of a T-cell activation; (ii) the higher configurational entropy (Table 1) accompanied by an higher conformational sampling of the binding groove in the predisposing proteins can be a bonus for subsequent peptide-HLA

plastic adaptation to T-cell receptor;[64] (iii) the residues and the dynamical characteristics of the binding groove region bound to N-terminal peptide fragment, particularly pocket 1, jointly with pocket 7, play an important role in conferring disease susceptibility. These points converge towards a molecular picture suggesting an higher predisposing peptide-HLA recognition efficiency by the T-cell receptor, with a potentially higher T-cell triggering efficiency, eventually leading to an autoimmune reaction against self proteins.[24, 65]

Altogether, the structural and dynamical information gathered from our molecular analysis, allowed us to characterize the predisposing or the protective nature of the proteins associated to multiple sclerosis, in complex with MAP peptide. We believe the integration of our results on physicochemical properties of peptide-HLA interaction with the clinical and experimental data, their extension to longer antigenic MAP peptides and the quantitative comparison to our previous results, will lead to a better comprehension of the finest disease susceptibility mechanisms, both general and protein serotype family specific, and will facilitate the screening of therapeutical peptides.[66]

Supplementary Material

Supporting information can be found in the file "Supplementary_material_Kumar_et.al.pdf ",
which has been attached separately.

REFERENCES

1. J. R. Oksenberg and S. E. Baranzini, *Nat Rev Neurol*, 2010, **6**, 429-437.
2. K. Kakalacheva and J. D. Lunemann, *FEBS letters*, 2011, **585**, 3724-3729.
3. G. Disanto, J. M. Morahan and S. V. Ramagopalan, *CNS Neurol Disord Drug Targets*, 2012, **11**, 545-555.
4. J. H. Wang, D. Pappas, P. L. De Jager, D. Pelletier, P. I. de Bakker, L. Kappos, C. H. Polman, Australian, C. New Zealand Multiple Sclerosis Genetics, L. B. Chibnik, D. A. Hafler, P. M. Matthews, S. L. Hauser, S. E. Baranzini and J. R. Oksenberg, *Genome Med*, 2011, **3**, 3.
5. S. E. Baranzini, J. Wang, R. A. Gibson, N. Galwey, Y. Naegelin, F. Barkhof, E. W. Radue, R. L. Lindberg, B. M. Uitdehaag, M. R. Johnson, A. Angelakopoulou, L. Hall, J. C. Richardson, R. K. Prinjha, A. Gass, J. J. Geurts, J. Kragt, M. Sombekke, H. Vrenken, P. Qualley, R. R. Lincoln, R. Gomez, S. J. Caillier, M. F. George, H. Mousavi, R. Guerrero, D. T. Okuda, B. A. Cree, A. J. Green, E. Waubant, D. S. Goodin, D. Pelletier, P. M. Matthews, S. L. Hauser, L. Kappos, C. H. Polman and J. R. Oksenberg, *Hum Mol Genet*, 2009, **18**, 767-778.
6. J. R. Oksenberg, S. E. Baranzini, S. Sawcer and S. L. Hauser, *Nat Rev Genet*, 2008, **9**, 516-526.
7. D. Ontaneda, M. Hyland and J. A. Cohen, *Annu Rev Med*, 2012, **63**, 389-404.
8. L. F. Barcellos, S. Sawcer, P. P. Ramsay, S. E. Baranzini, G. Thomson, F. Briggs, B. C. Cree, A. B. Begovich, P. Villoslada, X. Montalban, A. Uccelli, G. Savettieri, R. R. Lincoln, C. DeLoa, J. L. Haines, M. A. Pericak-Vance, A. Compston, S. L. Hauser and J. R. Oksenberg, *Hum Mol Genet*, 2006, **15**, 2813-2824.
9. J. A. Traherne, *Int J Immunogenet*, 2008, **35**, 179-192.
10. E. Y. Jones, L. Fugger, J. L. Strominger and C. Siebold, *Nat Rev Immunol*, 2006, **6**, 271-282.
11. K. H. Piepenbrink, S. J. Blevins, D. R. Scott and B. M. Baker, *Nature communications*, 2013, **4**, 1948.
12. M. G. Marrosu, R. Murru, M. R. Murru, G. Costa, P. Zavattari, M. Whalen, E. Cocco, C. Mancosu, L. Schirru, E. Solla, E. Fadda, C. Melis, I. Porru, M. Rolesu and F. Cucca, *Hum Mol Genet*, 2001, **10**, 2907-2916.
13. E. Cocco, C. Sardu, E. Pieroni, M. Valentini, R. Murru, G. Costa, S. Tranquilli, J. Frau, G. Coghe, N. Carboni, M. Floris, P. Contu and M. G. Marrosu, *PloS one*, 2012, **7**, e33972.
14. R. Lampis, L. Morelli, S. De Virgiliis, M. Congia and F. Cucca, *Tissue Antigens*, 2000, **56**, 515-521.
15. E. J. Manning and M. T. Collins, *Rev Sci Tech*, 2001, **20**, 133-150.
16. C. Thomas Dow, *Medical Hypotheses*, 2008, **71**, 858-861.
17. L. A. Sechi, M. Gazouli, J. Ikonomopoulos, J. C. Lukas, A. M. Scanu, N. Ahmed, G. Fadda and S. Zanetti, *J Clin Microbiol*, 2005, **43**, 5275-5277.
18. S. Masala, D. Paccagnini, D. Cossu, V. Brezar, A. Pacifico, N. Ahmed, R. Mallone and L. A. Sechi, *PloS one*, 2011, **6**, e26931.
19. D. Cossu, E. Cocco, D. Paccagnini, S. Masala, N. Ahmed, J. Frau, M. G. Marrosu and L. A. Sechi, *PloS one*, 2011, **6**, e18482.

20.   D. Cossu, S. Masala and L. A. Sechi, *Future Microbiol*, 2013, **8**, 223-232.
21.   D. Cossu, S. Masala, J. Frau, G. Mameli, M. G. Marrosu, E. Cocco and L. A. Sechi, *Mol Immunol*, 2014, **57**, 138-140.
22.   A. Kumar, E. Cocco, L. Atzori, M. G. Marrosu and E. Pieroni, *PloS one*, 2013, **8**, e59711.
23.   A. Kumar, P. Melis, V. Genna, E. Cocco, M. G. Marrosu and E. Pieroni, *Mol Biosyst*, 2014, **10**, 2043-2054.
24.   E. M. Chastain and S. D. Miller, *Immunol Rev*, 2012, **245**, 227-238.
25.   E. Cocco, R. Murru, G. Costa, A. Kumar, E. Pieroni, C. Melis, L. Barberini, C. Sardu, L. Lorefice, G. Fenu, J. Frau, G. Coghe, N. Carboni and M. G. Marrosu, *PloS one*, 2013, **8**, e59790.
26.   A. Sali and T. L. Blundell, *Journal of molecular biology*, 1993, **234**, 779-815.
27.   R. A. Laskowski, M. W. MacArthur, D. S. Moss and J. M. Thornton, *Journal of Applied Crystallography*, 1993, **26**, 283-291.
28.   K. Arnold, L. Bordoli, J. Kopp and T. Schwede, *Bioinformatics*, 2006, **22**, 195-201.
29.   A. A. Canutescu, A. A. Shelenkov and R. L. Dunbrack, Jr., *Protein science : a publication of the Protein Society*, 2003, **12**, 2001-2014.
30.   G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell and A. J. Olson, *Journal of computational chemistry*, 2009, **30**, 2785-2791.
31.   W. Humphrey, A. Dalke and K. Schulten, *J Mol Graph*, 1996, **14**, 33-38, 27-38.
32.   W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *The Journal of chemical physics*, 1983, **79**, 926.
33.   A. D. MacKerell, Jr., N. Banavali and N. Foloppe, *Biopolymers*, 2000, **56**, 257-265.
34.   H. Li, A. D. Robertson and J. H. Jensen, *Proteins*, 2005, **61**, 704-721.
35.   R. Yaneva, S. Springer and M. Zacharias, *Biopolymers*, 2009, **91**, 14-27.
36.   S. E. Feller, Y. Zhang, R. W. Pastor and B. R. Brooks, *The Journal of chemical physics*, 1995, **103**, 4613.
37.   J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale and K. Schulten, *Journal of computational chemistry*, 2005, **26**, 1781-1802.
38.   A. Altis, P. H. Nguyen, R. Hegger and G. Stock, *The Journal of chemical physics*, 2007, **126**, 244111.
39.   N. M. Glykos, *Journal of computational chemistry*, 2006, **27**, 1765-1768.
40.   T. Aleksiev, R. Potestio, F. Pontiggia, S. Cozzini and C. Micheletti, *Bioinformatics*, 2009, **25**, 2743-2744.
41.   R. Potestio, F. Pontiggia and C. Micheletti, *Biophysical journal*, 2009, **96**, 4993-5002.
42.   C. Micheletti, P. Carloni and A. Maritan, *Proteins*, 2004, **55**, 635-645.
43.   B. Lee and F. M. Richards, *Journal of molecular biology*, 1971, **55**, 379-400.
44.   J. Schlitter, *Chemical Physics Letters*, 1993, **215**, 617-621.
45.   M. Naim, S. Bhat, K. N. Rankin, S. Dennis, S. F. Chowdhury, I. Siddiqi, P. Drabik, T. Sulea, C. I. Bayly, A. Jakalian and E. O. Purisima, *Journal of chemical information and modeling*, 2007, **47**, 122-133.
46.   Q. Cui, T. Sulea, J. D. Schrag, C. Munger, M. N. Hung, M. Naim, M. Cygler and E. O. Purisima, *Journal of molecular biology*, 2008, **379**, 787-802.
47.   F. Sterpone, M. Ceccarelli and M. Marchi, *Journal of molecular biology*, 2001, **311**, 409-419.

48. A. Kumar, E. Hajjar, P. Ruggerone and M. Ceccarelli, *The journal of physical chemistry. B*, 2010, **114**, 9608-9616.
49. P. I. Koukos and N. M. Glykos, *Journal of chemical information and modeling*, 2014, **54**, 209-217.
50. J. A. Marsh, *Journal of molecular biology*, 2013, **425**, 3250-3263.
51. D. Narzi, C. M. Becker, M. T. Fiorillo, B. Uchanska-Ziegler, A. Ziegler and R. A. Böckmann, *Journal of molecular biology*, 2012, **415**, 429-442.
52. T. Pohlmann, R. A. Bockmann, H. Grubmuller, B. Uchanska-Ziegler, A. Ziegler and U. Alexiev, *Journal of Biological Chemistry*, 2004, **279**, 28197-28201.
53. C. J. Holland, P. J. Rizkallah, S. Vollers, J. M. Calvo-Calle, F. Madura, A. Fuller, A. K. Sewell, L. J. Stern, A. Godkin and D. K. Cole, *Sci Rep*, 2012, **2**, 629.
54. A. Ferrante and J. Gorski, *Journal of molecular biology*, 2012, **417**, 454-467.
55. E. C. de Haan, M. H. Wauben, M. C. Grosfeld-Stulemeyer, J. A. Kruijtzer, R. M. Liskamp and E. E. Moret, *Bioorganic & medicinal chemistry*, 2002, **10**, 1939-1945.
56. A. Ferrante, *Immunol Res*, 2013, **56**, 85-95.
57. T. S. Jardetzky, J. C. Gorga, R. Busch, J. Rothbard, J. L. Strominger and D. C. Wiley, *EMBO J*, 1990, **9**, 1797-1803.
58. B. J. McFarland, C. Beeson and A. J. Sant, *J Immunol*, 1999, **163**, 3567-3571.
59. A. V. Chervonsky, L. Gordon and A. J. Sant, *Int Immunol*, 1994, **6**, 973-982.
60. A. Stavrakoudis, *FEBS letters*, 2011, **585**, 485-491.
61. A. Stavrakoudis, *Cell biochemistry and biophysics*, 2011, **60**, 283-295.
62. W. S. Meng, H. von Grafenstein and I. S. Haworth, *Int Immunol*, 1997, **9**, 1339-1346.
63. P. M. Petrone and A. E. Garcia, *Journal of molecular biology*, 2004, **338**, 419-435.
64. J. J. Adams, S. Narayanan, B. Liu, M. E. Birnbaum, A. C. Kruse, N. A. Bowerman, W. Chen, A. M. Levin, J. M. Connolly, C. Zhu, D. M. Kranz and K. C. Garcia, *Immunity*, 2011, **35**, 681-693.
65. E. M. Chastain, D. S. Duncan, J. M. Rodgers and S. D. Miller, *Biochimica et biophysica acta*, 2011, **1812**, 265-274.
66. F. K. Insaidoo, O. Y. Borbulevych, M. Hossain, S. M. Santhanagopolan, T. K. Baxter and B. M. Baker, *The Journal of biological chemistry*, 2011, **286**, 40163-40173.

Table 1. Configurational Entropy and Binding Free Energy Calculation. MAP peptide entropy (in kcal/mol) and binding free energy (in kcal/mol) for four MAP-protein complexes.

Protein (DRB1)	MAP Entropy TΔS (kcal/mol)	Binding free energy ΔG (kcal/mol)
*03:01	202	-11.7 ± 0.6
*15:01	198	-12.9 ± 0.5
*16:01	168	-10.1 ± 0.7
*15:02	162	-9.8 ± 0.6

Table 2. MAP-DRB1 H-bond network analysis. Only persistent H-bond interactions are reported together with pocket localization. The conserved interaction among the proteins is indicated with ALL; empty boxes represent absence of H-bond interaction.

DRB1/Map	D301	T303	N305	V306	K307	D309
T12					*03:01	
R13 (P4, P6)			*16:01	*15:01,*15:02		
Y26			*03:01			
D28 (P7)			*16:01		*03:01	
Y60						ALL
W61						*15:02
Q64 (P7)						*03:01,*15:01
Q70 (P4)				*15:01*15:02		
K71/R71 (P7)				*03:01/*16:01		
T77/N77		ALL				
Y78			*03:01			
H81	ALL					
N82		ALL				

Table 3. Dynamics of water bound to MAP peptide. We report the average number of water molecules (#) bound to MAP residues in bulk water simulation and in four HLA-peptide complexes. The three time regimes are denoted as fast (F), medium (M) and slow (S).

MAP sections	Bulk water #	Number (#) and percentage (%) of water molecules															
		*03:01				*15:01				*16:01				*15:02			
		#	F	M	S	#	F	M	S	#	F	M	S	#	F	M	S
N-ter	8.9	6.7	100	0	0	6.9	100	0	0	7.9	100	0	0	7.2	100	0	0
Central	10.9	7.9	78	22	0	6.9	83	0	17	8.9	56	26	18	10	88	0	12
C-ter	9.1	10.6	78	18	4	8.8	78	0	22	10.5	82	4	14	10.3	100	0	0
Total #	28.9	25.6				22.6				27.3				27.5			

Table 4. Slow water molecules bridge peptide-HLA interactions. We report slow water molecules that are involved in bridging interactions between the peptide residues and DRB1 binding groove residues. The sections of MAP peptide that participate are denoted as central (c) and C-terminal (cter).

Proteins	Water molecule	HLA-DRB1 residue	Map peptide residue
DRB1*03:01	w2960	S13	K307 cter
DRB1*15:01	w8193	Q70, D28	N305 c
	w6441	W61, Y30	G308 cter
	w4393	R13	K307 cter
DRB1*16:01	w8526	Y78	N305 cter
	w4456	R13	N305 c, K307 cter
	w10298	W9	K307 cter
DRB1*15:02	w2197	N82	R304 c

Figure legends

Fig. 1. Sequence analysis of DRB1 proteins. Sequence alignment of the binding cleft residue (1:90) for all the proteins under investigation with respect to DRB1*16:01. Residues color is based on polarity. Identical residues with respect to DRB1*16:01 are indicated by dots, and gaps are indicated by dash.

Fig. 2. Antigenic MAP peptide structure (DPTRNVKGD). The peptide backbone is shown using ball-stick with C-alpha atoms colored in blue.

Fig. 3. Dihedral angle principal component analysis of the HLA binding groove. The two dimensional density maps are shown here and corresponds to the projection of dihedral angles (Φ, Ψ) fluctuations (from MD simulation trajectory) of HLA binding groove residues on the plane defined by first two principal components (PC1, PC2). Conformations with similar fluctuations are grouped into different clusters (circled in white), with high density shown in blue. The range of values for the vertical axis (PC1) and horizontal axis (PC2) in the four density maps are the same and lies between [-4.0 to +4.0] kcal/mol. (A) *03:01 protein, (B) *15:01 protein, (C) *16:01 protein and (D) *15:02 protein.

Fig. 4. Collective Motion analysis. The binding groove is decomposed into 5 domains based on the collective motion of the residues. Each domain is represented with a different color. (A) *03:01 protein, (B) *15:01 protein, (C) *16:01 protein, and (D) *15:02 protein.

Fig. 5. Binding groove width fluctuation. Probability distribution of center of mass distance of heavy atoms between the flanking residues of chain A and chain B in HLA-MAP complexes. The distribution are shown for each groove domain: D1 (α 50–51 and β 85-86), D2 (α 53–55 and β 78–83), D3 (α 50–51 and β 85–86) and D4 (α 50–51 and β 85–86). In black we show the profile for the free protein in the four regions, while in red for the predisposing protein in complex with MAP and in green for the protective protein in complex with MAP.

Fig. 6. Buried surface area analysis. BSA calculations on the HLA binding groove residues during 100 ns MD simulations, for the four HLA-MAP complexes are shown here. (A) *03:01 protein in (B) *15:01 protein in (C) *16:01 protein and (D) *15:02 protein. BSA values smoothened (every 10th step) in the two predisposing proteins are shown in red and in green for the two protective proteins.

Fig. 7. Role of slow water molecules in the binding groove. Slow water molecules involved in
bridging interactions between peptide backbone and HLA residues are shown here. The
HLA binding groove is shown using cartoon representation (white), the peptide backbone
using ball-stick representation, in blue the C-alpha atoms, and the water molecules in Van
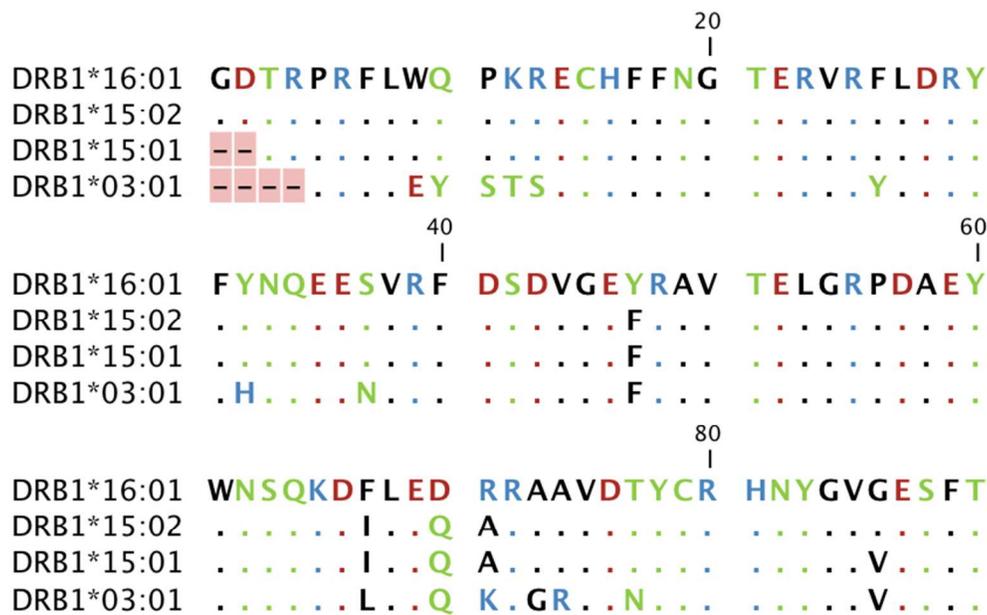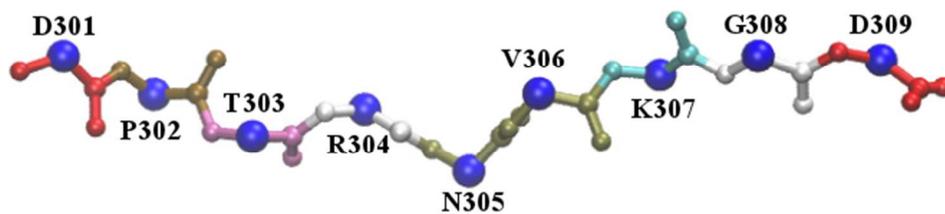der Walls representation.

Fig.1 Sequence analysis of DRB1 alleles. Sequence alignment of the binding cleft residues (1:90) for all the alleles under investigation with respect to DRB1*16:01. Residues color is based on polarity. Identical residues with respect to DRB1*16:01 are indicated by dots, and gaps are indicated by dash.
72x45mm (300 x 300 DPI)

Fig. 2 Antigenic MAP peptide structure (DPTRNVKGD). The peptide backbone is shown using ball-stick with
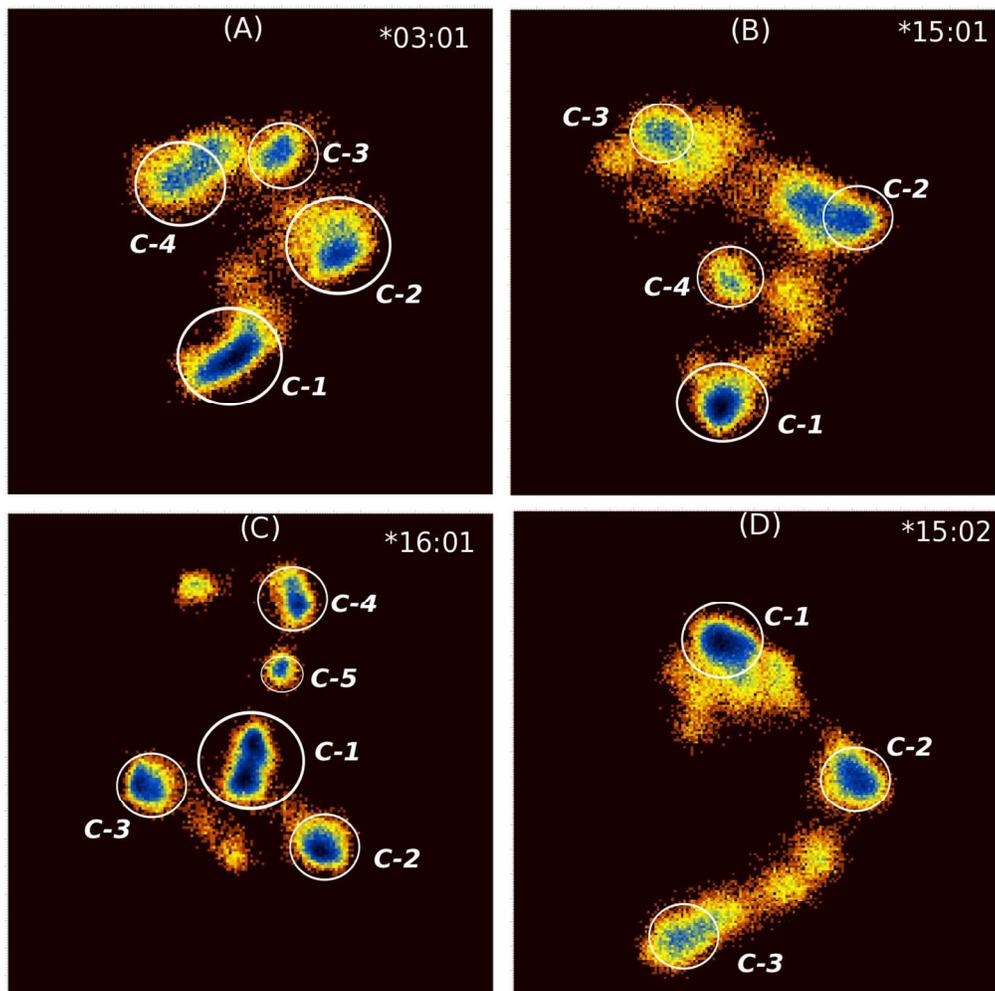C-alpha atoms colored in blue.
244x63mm (72 x 72 DPI)

Fig. 3 Dihedral principal component analysis of the HLA binding groove. The two dimensional density maps are shown here and corresponds to the projection of dihedral angles (Φ,Ψ) fluctuations (from MD simulation trajectory) of HLA binding groove residues on the plane defined by first two principal components (PC1, PC2). Conformations with similar fluctuations are grouped into different clusters (circled in white), with high density shown in blue. The range of values for the vertical axis (PC1) and horizontal axis (PC2) in the four density maps are the same and lies between [-4.0 to +4.0] kcal/mol. (A) *03:01 allele, (B) *15:01 allele, (C) *16:01 allele and (D) *15:02 allele.
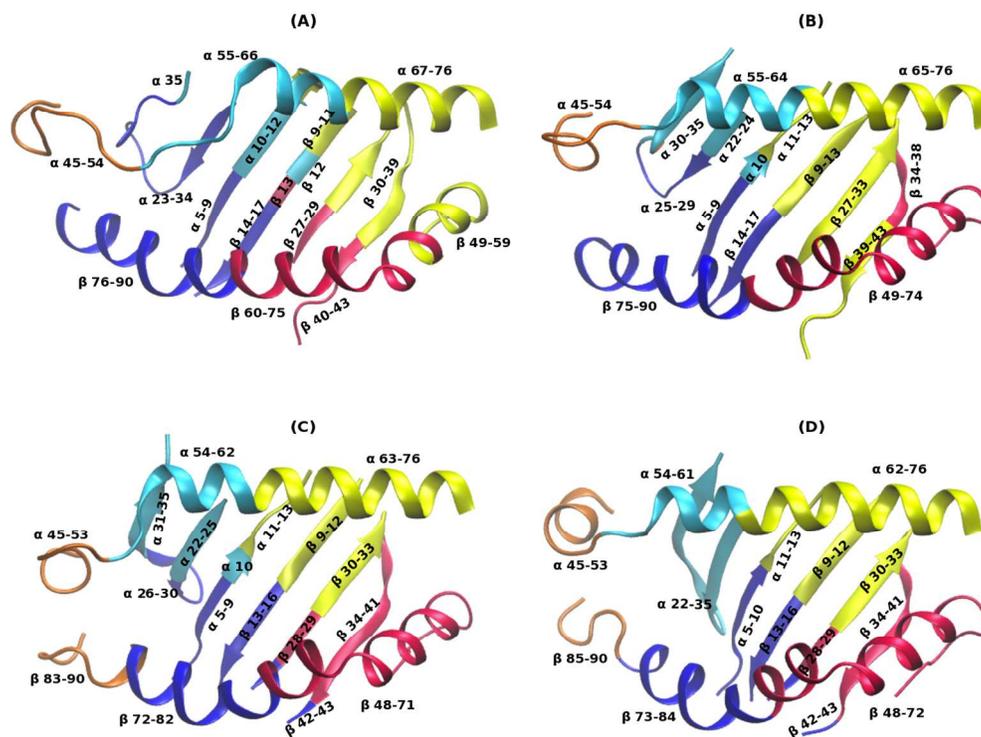
523x523mm (72 x 72 DPI)

Fig. 4 Collective Motion analysis. The binding groove is decomposed into 5 domains based on the collective motion of the residues. Each domain is represented with a different color. (A) *03:01 allele, (B) *15:01 allele, (C) *16:01 allele, and (D) *15:02 allele.
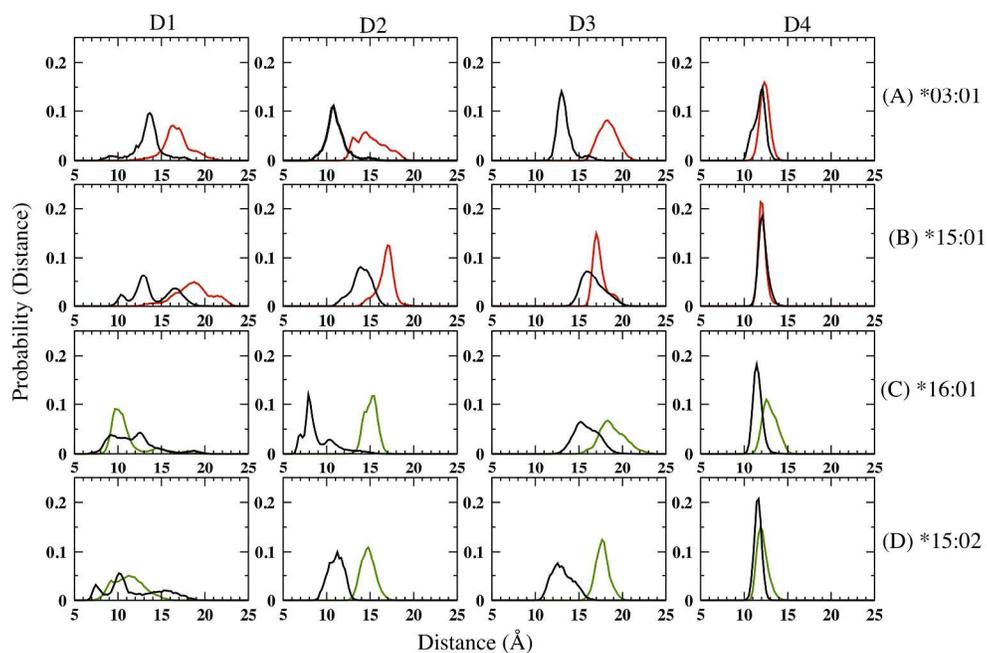541x400mm (72 x 72 DPI)

Fig. 5 Binding groove width fluctuation. Probability distribution of center of mass distance of heavy atoms between the flanking residues of chain A and chain B in HLA-MAP complexes. The distribution are shown for each groove domain: D1 (α 50–51 and β 85–86), D2 (α 53–55 and β 78–83), D3 (α 50–51 and β 85–86) and D4 (α 50–51 and β 85–86). In black we show the profile for the free allele in the four regions, while in red for the predisposing alleles in complex with MAP and in green for the protective alleles in complex with MAP.
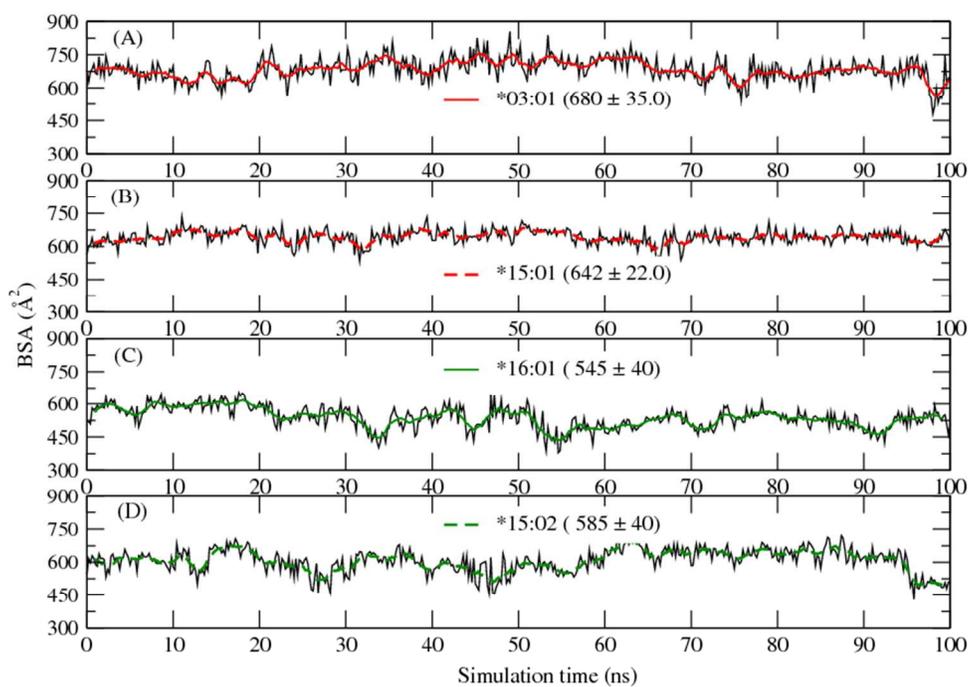
259x172mm (300 x 300 DPI)

Fig. 6 Buried surface area analysis. BSA calculations on the HLA binding groove residues during 100 ns MD simulations, for the four HLA-MAP complexes are shown here. (A) *03:01 allele in (B) *15:01 allele in (C) *16:01 allele and (D) *15:02 allele. BSA values smoothened (every 10th step) in the two predisposing alleles are shown in red and in green for the two protective alleles.
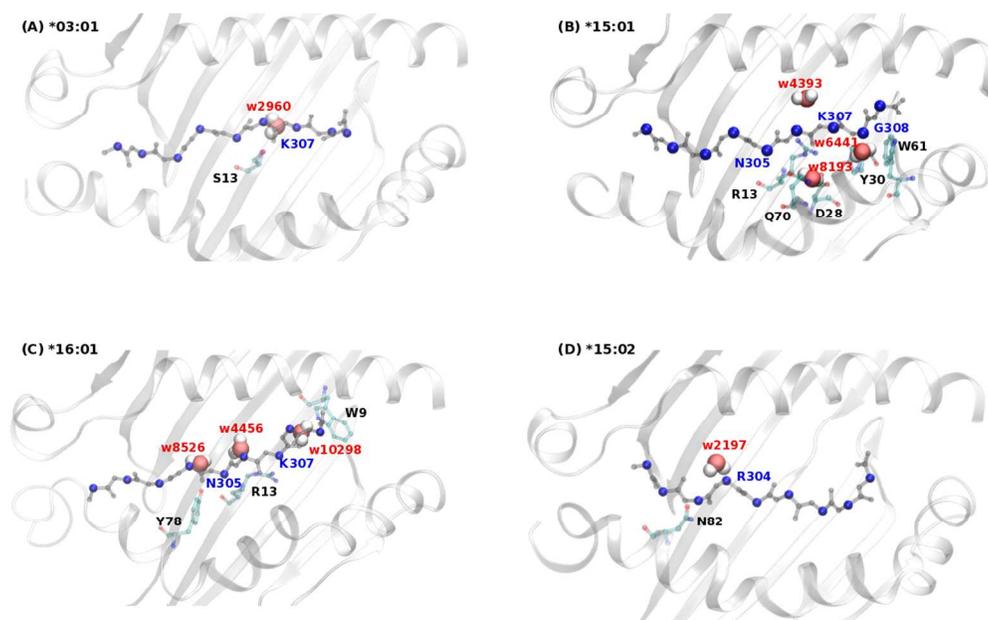322x225mm (72 x 72 DPI)

Fig. 7 Role of slow water molecules in the binding groove. Slow water molecules involved in bridging interactions between peptide backbone and HLA residues are shown here. The HLA binding groove is shown using cartoon representation (white), the peptide backbone using ball-stick representation, in blue the C-alpha atoms, and the water molecules in Van der Walls representation.

465x285mm (72 x 72 DPI)