

Molecular modeling of class I and II alleles of the major histocompatibility complex in *Salmo salar*

Constanza Cárdenas · Axel Bidon-Chanal ·
Pablo Conejeros · Gloria Arenas · Sergio Marshall ·
F. Javier Luque

Received: 13 October 2009 / Accepted: 22 September 2010 / Published online: 10 October 2010
© Springer Science+Business Media B.V. 2010

Abstract Knowledge of the 3D structure of the binding groove of major histocompatibility (MHC) molecules, which play a central role in the immune response, is crucial to shed light into the details of peptide recognition and polymorphism. This work reports molecular modeling studies aimed at providing 3D models for two class I and two class II MHC alleles from *Salmo salar* (*Sasa*), as the lack of experimental structures of fish MHC molecules represents a serious limitation to understand the specific preferences for peptide binding. The reliability of the

structural models built up using bioinformatic tools was explored by means of molecular dynamics simulations of their complexes with representative peptides, and the energetics of the MHC-peptide interaction was determined by combining molecular mechanics interaction energies and implicit continuum solvation calculations. The structural models revealed the occurrence of notable differences in the nature of residues at specific positions in the binding groove not only between human and *Sasa* MHC proteins, but also between different *Sasa* alleles. Those differences lead to distinct trends in the structural features that mediate the binding of peptides to both class I and II MHC molecules, which are qualitatively reflected in the relative binding affinities. Overall, the structural models presented here are a valuable starting point to explore the interactions between MHC receptors and pathogen-specific interactions and to design vaccines against viral pathogens.

Electronic supplementary material The online version of this article (doi:10.1007/s10822-010-9387-8) contains supplementary material, which is available to authorized users.

C. Cárdenas (✉)
Laboratorio de Genética e Inmunología Molecular, Facultad de Ciencias, Pontificia Universidad Católica de Valparaíso, Chile,
Av. Brasil 2950, Valparaíso, Chile
e-mail: constanza.cardenas@ucv.cl

P. Conejeros · G. Arenas · S. Marshall (✉)
Laboratorio de Genética e Inmunología Molecular, Instituto de Biología, Facultad de Ciencias, Pontificia Universidad Católica de Valparaíso, Chile, Av. Brasil 2950, Valparaíso, Chile
e-mail: smarshal@ucv.cl

C. Cárdenas · P. Conejeros · G. Arenas · S. Marshall
NBC-Núcleo de Biotecnología Curauma, Campus Curauma
PUCV, Valparaíso, Chile

S. Marshall
Centro Regional Estudios Alimentación Saludable (CREAS),
PUCV, Valparaíso, Chile

A. Bidon-Chanal · F. J. Luque (✉)
Departament de Físicoquímica and Institut de Biomedicina (IBUB), Facultat de Farmàcia, Universitat de Barcelona, Av. Diagonal 643, 08028 Barcelona, Spain
e-mail: fjluque@ub.edu

Keywords Major histocompatibility complex · Molecular modeling · Protein–peptide interaction · Peptide recognition · Molecular dynamics simulation

Introduction

The major histocompatibility complex (MHC) plays an important role in alerting the immune system, as the heterodimeric proteins encoded by MHC genes are involved in the presentation of protein antigens [1]. It has been noticed that about 10% of all possible peptide sequences of appropriate length may bind with high affinity to the same MHC protein [2]. Besides the pivotal role exerted in the immune system, the unusual combination of low specificity and high affinity properties of peptide binding explains the intense research effort devoted to understand the molecular basis of

antigen recognition by MHC molecules, leading to a variety of tools that comprise databases of MHC alleles and antigenic peptide variants [3–6], algorithms for the prediction of peptide binding [7–12] and sequence processing [9, 13, 14].

Understanding of peptide recognition by MHC proteins at the atomic level has been facilitated by X-ray structures of peptide-MHC complexes, especially in humans and mice. This information has shed light into the molecular basis of the immune surveillance and the origins of certain immunological diseases, thus spurring the design of MHC-based reagents that might be used as immunospecific therapeutics and diagnostics. However, similar progress in other higher organisms is seriously hampered by the lack of 3D structural data. This is the case of MHC molecules in fish, particularly in salmonids [15, 16], which have been the subject of increasing interest due to the association between specific alleles with resistance to certain pathogens [17–22] and the severe economic losses caused by infectious diseases in fish farms [23].

In spite of the similar physiological role played by fish and mammalian MHC molecules [24, 25], there are relevant differences in their genetic organization [25–27], as the components of fish MHC do not strictly form a complex because they are located in different linkage groups [26]. Class I MHC molecules are formed by a structurally highly conserved β 2-microglobulin protein and a polymorphic α chain involved in antigen binding [25, 28]. Though there are more than 10 salmonid class I genes, only one (known as UBA) has been indisputably identified to possess classical MHC functionality [29, 30]. Class II MHC molecules are composed of α and β chains (named DAA and DAB, respectively), which participate in peptide binding. In humans the α chain is much more conserved than the β chain, while in salmonids polymorphism affects similarly the two subunits [31]. The functional form of MHC proteins requires the presence of a bound peptide, which can be endogenous or exogenous. Class I MHC molecules bind mostly endogenous peptides acquired directly in the endoplasmic reticulum [32, 33]. Class II MHC proteins are found in mammals associated with an invariant chain, which assists in proper folding, trafficking and protection of the antigen-binding groove [34, 35]. Cleavage of this protein at the endocytic compartment leaves a peptide named CLIP in the binding groove, which is replaced by peptides of putative antigens for their presentation.

The evolutionary homology of mammalian and fish MHC molecules supports the assumption that the global fold should be preserved [24, 31]. Nevertheless, a detailed knowledge of the nature and distribution of residues that delineate the peptide-binding groove is critical to ascertain the molecular factors responsible for the binding and selectivity preferences of antigenic peptides. To the best of our knowledge, only one in silico modeling study has been

conducted for sea bass class II MHC, but limited to the β chain [36]. Regarding class I MHC molecules, Grimholt et al. [25] have examined the conservation of the binding groove in terms of residues and electrostatic features of the unbound α chain in Atlantic salmon, but without paying attention to the interaction with peptides in the binding groove. Finally, other studies have dealt with interactions outside the binding groove in sea bream and grass carp [37, 38].

This study provides the first structural models of *Salmo salar* (*Sasa*) class I (UBA) and class II (DAA-DAB) MHC proteins, taking into account the structural features of both α and β chains and their interaction with peptides. With this aim, *Sasa* class I and II MHC proteins were modeled taking advantage of the available structural data of MHC proteins in mice and human. Particularly, models of two UBA and two DAA-DAB receptors bound to several peptides are examined. The structural integrity of the MHC-peptide complexes and the differential trends involved in the binding of antigenic peptides are explored by means of molecular dynamics simulations, and the energetics of the MHC-peptide interaction is examined by combining molecular mechanics interaction energies with implicit continuum solvation computations. Overall, the modeled structures should be valuable for the design of novel drugs or peptide-based vaccines.

Methods

System setup

The amino acid sequences of the alleles reported for *Sasa* class I (UBA) and II (DAA and DAB) MHC molecules were obtained from the IPD database [4]. For each class two alleles known to be associated with disease resistance [17, 18] were chosen for modeling their 3-D structure: UBA*0201 and UBA*0301 for class I, and DAA*0201-DAB*0201 and DAA*0501-DAB*0301 for class II.

The choice of the peptide bound to the peptide-binding groove was more delicate due not only to the enormous diversity of peptide sequences that bind MHC proteins, but also to the lack of quantitative data regarding the binding affinities of peptides to salmonid MHC molecules. Therefore, the strategy adopted here was to define *consensus* peptides based on the analysis of the available binding motifs experimentally identified for class I and II MHC proteins, further subjected to a size restriction of 9–11 residues according to the usual length of peptides bound to human and murine MHCs [11, 39]. Certainly this strategy calls for caution on the quantitative interpretation of the binding of peptides to *Sasa* MHC proteins. Nevertheless, it must be stressed that our main interest was to calibrate the reliability of the models by examining the overall structural

integrity of *Sasa* MHC complexes and their ability to discriminate selectively between different peptides.

Two consensus sequences were chosen for binding to class I MHC proteins. The first one relied on the preferred sequence motifs for interaction of peptides with *Sasa* UBA*0301 detected in binding assays by Zhao et al. [40], who identified two sets of 107 7-mer and 74 12-mer peptides recognized by this allele. The second consensus sequence was based on the binding preferences of peptides to murine H-2Kb complexes (chosen according to PDB structures 1FZO, 1WBZ, 1G7P, 1G7Q, 1KPV, 1LK2, and 1N59), as Zhao et al. [40] suggested that both *Sasa* UBA*0301 and murine H-2Kb might have a very similar peptide recognition motif at the C-terminus.

In contrast with the consensus peptides chosen for class I MHC proteins, which are known to bind specifically the *Sasa* UBA*0301 allele, it is worth noting that a proper definition of the consensus peptides for class II MHC is hampered by the lack of a precise knowledge of the CLIP peptide in the invariant chain of fishes, as noted in the limited sequence similarity observed for this region in alignment studies [41]. Accordingly, two consensus sequences were chosen based on two alternative sequence alignments. The first was based on the alignment reported by Dixon and coworkers [41] for human Ia antigen-associated invariant chain (p33 isoform), the invariant chains of other vertebrates (mouse and chicken) and versions of invariant chains in fish (Uniprot codes for the aligned fish proteins are Q8JFN4, Q8JFN5 and Q8JFN6 for *O. Mykiss*, Q9PUT2 and Q9PVD9 for *D. Rerio*). Since the region of the chicken invariant chain aligned with the CLIP region (residues 82–107) of the human invariant chain is not extremely well conserved and particularly there is an insertion of 3 residues compared to the human and mouse counterparts [42], an alternative alignment was performed by excluding the chicken sequence in order to define an alternative consensus peptide.

Homology modeling

The 3D model structures of *Sasa* UBA*0201 and UBA*0301 (named UBA2 and UBA3 hereafter, respectively) were built up using as template the X-ray crystallographic structure of a natural mutant of the murine allele H-2Kb (PDB entry 1FZO [43], solved at 1.8 Å resolution), which is bound to the 9-mer viral nucleocapsid peptide F₁APGNYPAL₉. For class II MHC proteins the structure of *Sasa* alleles DAA*0201-DAB*0201 and DAA*0501-DAB*0301 (denoted ahead as DA2B2 and DA5B3, respectively) were modeled based on the X-ray structure of human HLA-DRβ1*0101 (PDB entry 1DLH [44]; 2.8 Å resolution), which is bound to the influenza virus peptide K₁YVKQNTLKL₁₁.

Alignment of *Sasa* MHC proteins and the reference templates was performed using ClustalW [45], leading to

sequence similarities in the ranges 47–54% and 53–63% for α and β chains, respectively, considering both identities and positive changes between residues (see Fig. S1 in Supporting Information). The starting 3D models of *Sasa* complexes were generated using Swiss-Model in the ExPASy server [46]. Comparison of the modeled structures and the template proteins suggests that gaps/insertions of residues detected in the alignment of sequences should have no relevant effect on the structural features of the peptide binding site.

Positioning of the consensus peptides along the peptide-binding groove was guided by the skeleton of the bound peptides found in X-ray structures of the template MHC molecules, whereas the orientation of the side chains was made taking into account the rotameric preferences of the residues [47]. In all cases, nevertheless, visual inspection of the complexes allowed further refinement of the side chains, which were manually adjusted to avoid unfavorable clashes in the modeled structures. Finally, the structure of the complexes was further relaxed by means of molecular dynamics simulations (see below). The adoption of this strategy was motivated by the similar positions found for the C_α carbon atoms of peptides having the same total length upon binding to different mutants of murine class I MHC H-2Kb, and upon binding of different viral peptides to human class I MHC molecule HLA-A2, specially at the peptide termini ([48]; see also Fig. S2 in Supporting Information). Thus, alignment of the 9-mer peptides bound to the peptide-binding groove of both murine and human MHC proteins reveals a nearly perfect matching of the C_α carbons in the N- and C-terminal sides, whilst the main structural differences in the peptide backbone are limited to the central region. Similar findings are found in the comparison of peptides bound to different human class II MHC molecules (see Fig. S3 in Supporting Information).

Noteworthy, this strategy was also assumed in previous studies aimed at discriminating between binder and non-binder peptides, as modeling of the side chains was found sufficient to explore the substitution of peptides to a given MHC molecule, while very minor adjustments in the peptide main chain were made through energy minimizations [49, 50]. However, keeping in mind the mutations between *Sasa* alleles and the template structures (see below), energy minimization does not seem well suited to ensure a proper relaxation of the peptide-MHC complexes, which was then conducted by molecular dynamics simulations, as done in previous studies [51].

The complete list of *Sasa* MHC complexes simulated here is given in Table 1, which also includes the murine (1FZO) and human (1DLH) complexes chosen as references for comparative purposes in molecular simulations.

To further calibrate the reliability of the modeled MHC complexes, additional computations were performed for human HLA-DRβ1*0101 and HLA-DRβ1*0301 alleles

Table 1 MHC molecules used to build up structural models of *Sasa* class I and II MHC complexes together with murine and human MHC proteins used as reference systems, the sequences of the peptides

| MHC molecule | Peptide name | Peptide sequence | Code |
|---|--------------------------------|--|---------|
| <i>Class I</i> | | | |
| Murine H-2Kb (1FZO) ^a | Sendai virus nucleoprotein | F ₁ APGNYPAL ₉ (PDB entry 1FZO) | 1FZO |
| <i>Sasa</i> UBA*0201 | H-2Kb-like peptide | F ₁ SGYNATAL ₉ | UBA2_1 |
| | UBA*0301 peptide | M ₁ PSQLYTLM ₉ | UBA2_2 |
| <i>Sasa</i> UBA*0301 | H-2Kb like peptide | F ₁ SGYNATAL ₉ | UBA3_1 |
| | UBA*0301 peptide | M ₁ PSQLYTLM ₉ | UBA3_2 |
| <i>Class II</i> | | | |
| Human HLA-DRβ1*0101 (1DLH) ^a | Hemagglutinin: HA | K ₁ YVKQNTLKLA ₁₁ (PDB entry 1DLH) | 1DLH |
| <i>Sasa</i> DAA*0201-DAB*0201 | Invariant chain-like peptide 1 | P ₁ MKMHMPLNM ₁₁ | DA2B2_1 |
| | Invariant chain-like peptide 2 | H ₁ LPMLNMARLI ₁₁ | DA2B2_2 |
| <i>Sasa</i> DAA*0501-DAB*0301 | Invariant chain-like peptide 1 | P ₁ MKMHMPLNM ₁₁ | DA5B3_1 |
| | Invariant chain-like peptide 2 | H ₁ LPMLNMARLI ₁₁ | DA5B3_2 |

Table 2 Human class II MHC complexes bound to *ideal* and *poor*-binding peptides used as reference systems

| MHC molecule | Peptide type | Peptide sequence |
|---------------|---------------|---|
| HLA-DRβ1*0101 | Control/ideal | K ₁ YVKQNTLKLA ₁₁ (PDB entry 1DLH) |
| | Poor | K ₁ ATMDRFGEKA ₁₁ |
| HLA-DRβ1*0301 | Control | K ₁ MRMATPLLMQ ₁₁ (PDB entry 1A6A) |
| | Ideal | K ₁ LKSDGKIKYQ ₁₁ |
| | Poor | K ₁ ADGKFEQLTQ ₁₁ |

bound to peptides that exhibit *ideal* and *poor* binding features (Table 2). These peptides were chosen by exploiting the known preferences of amino acids for the anchoring positions of MHC peptides as reported in the SYFPEITHI database [52] (<http://www.syfpeithi.de>). Ideal peptides should have sequences expected to fulfill the preferences at four main anchoring positions (see Table S1 in Supporting Information). This is accomplished by peptide K₁**YVKQNTL**KLA₁₁ for HLA-DRβ1*0101 (anchoring positions are marked in bold in the peptide sequence). This peptide is found in the X-ray crystallographic structure 1DLH, which was one of the reference complexes mentioned above (see Table 1), and corresponds to the fragment 306–318 of hemagglutinin HA peptide (PKYVKQNTLKLAT). For HLA-DRβ1*0301 an *ideal* binding peptide is fulfilled by K₁**LKSDGKIKY**Q₁₁, which reflects the sequence NLFLKSDGRIKYTL found in several peptides reported as examples of ligands binding to HLA-DRβ1*0301 in the SYFPEITHI database (see Table S2 in Supporting Information). Furthermore, as an additional control for the complexes with HLA-DRβ1*0301, we have also simulated the complex with K₁MRMATPLLMQ₁₁,

considered in MHC complexes, and the code used to designate each MHC-peptide complex in this study

which is the peptide found in the X-ray crystallographic structure 1A6A. Several peptides containing the whole or partial sequence of the control peptide are collected in the SYFPEITHI database as ligands able to bind HLA-DRβ1*0301 (see Table S2).

Finding sequences in the database suitable as *poor* binding peptides is less feasible. Accordingly, we imposed that the first and last residues in the 11-mer peptides were unaltered compared to the control ones (i.e., K at position 1, A at position 11 for HLA-DRβ1*0101; K at position 1, Q at position 11 for HLA-DRβ1*0301). Then, one of the anchoring sites was imposed to satisfy the anchoring preferences for binding. The precise anchoring site and the nature of the required amino acid were chosen randomly. The rest of the anchoring positions were randomly filled with amino acids other than the preferred ones. Finally, a similar procedure was adopted for the vacant positions in the 11-mer peptides subject to visual inspection of the complexes and eventual manual adjustment of the side chains in order to minimize the impact of steric clashes. This procedure led to the assignment of K₁ATMDRFGEKA₁₁ and K₁ADGKFEQLTQ₁₁ as poor binding peptides for the interaction with HLA-DRβ1*0101 and HLA-DRβ1*0301, respectively.

Molecular dynamics simulations

Refinement of the 3D models was performed by means of atomistic molecular dynamics (MD) simulations using the parm99 force field of the Amber9 suite of programs [53]. Preparation of the systems involved the addition of missing atoms and embedding of the system in a solvent-filled truncated octahedral box spanning 12 Å from the protein using the Leap module of the AMBER package.

Assignment of the most adequate ionization state for ionizable residues in the acidic physiological conditions of endosomes (pH 5.5–6.0) was performed using PROPKA algorithm [54] (see Tables S3 and S4 in Supporting Information) and further checked upon visual inspection of the local environment and solvent accessibility in the modeled structures. In particular, all the titratable (Asp, Glu, Lys) residues at the peptide-binding groove were assigned the standard ionization state, and histidine side chains were protonated (see Supporting Information). Nevertheless, this assignment might be dubious in two cases. First, for human 1DLH residue E9 in α chain was predicted to have a pK_a of 5.7, which suggests the potential involvement of both neutral and charged states. Second, whereas a low pK_a (4.9) was estimated for α H5 in the modeled structure of DA5B3, the pK_a predicted for the analogous histidine in DA2B2 was 9.5. Inspection of the modeled structures suggested that the protonated state of α H5 seemed more reasonable, as the imidazole ring of that residue is at short distance (<2.8 Å) from an aspartate residue (α D20), suggesting the formation of a strong salt bridge, whereas the difference in pK_a appeared to stem from the different orientation of a spatially close arginine residue (β R9). Considering the mildly acidic conditions of the endosome, the larger uncertainty of PROPKA algorithm in predicting the pK_a values of histidines, and the dependence of the estimated pK_a on the orientation of side chains in the modeled structures, which neglect the flexibility of the surrounding environment, those residues were considered in their ionized form in molecular simulations. This choice seemed to be supported by the structural integrity observed along the trajectories (see below). Finally, chlorine or sodium ions were added to obtain an electrostatically neutral system.

A multi-stage protocol was adopted for minimization and equilibration of the system. Minimization of hydrogen atoms, water molecules and counterions, and the complete system was performed in separate steps that combine 2000 steepest-descent and 8000 conjugate-gradient steps. Thermalization was performed in four steps where the temperature of the system was increased from 150 to 298 K, each involving a 50 ps MD at constant volume, followed by a MD simulation at constant temperature (298 K) and pressure (1 atm) covering up to 3 ns. The productive part of the simulation ranged from 10 to 15 ns depending on the particular behavior of each complex. Nevertheless, to further check the structural stability of the simulated complexes one of the trajectories run for DA5B3 was enlarged up to 30 ns. All simulations were carried out under periodic boundary conditions. The coordinates of all atoms in the system were saved every 1 ps for further analysis. Calculations were performed at the Barcelona Supercomputing Centre.

Binding affinities

Molecular mechanics coupled to Poisson-Boltzmann-Solvent Accessible (MM/PBSA) calculations were used to estimate the affinity of the different peptides for binding to class I and class II MHC proteins. Computations were performed for a subset of 400 snapshots taken regularly during the last 5 ns of the trajectories. In all cases an energy minimization of the peptide-MHC complex was carried out previously in order to reduce steric clashes that might arise from thermal fluctuations during the simulations. The binding affinities ($\Delta G_{\text{binding}}$) were determined using Eq. 1.

$$\Delta G_{\text{binding}} = \Delta G_{\text{MM}} + \Delta G_{\text{ele}}^{\text{sol}} + \Delta G_{\text{non-polar}}^{\text{sol}} - T\Delta S_{\text{conf}} \quad (1)$$

where the terms in the right-hand side of the equation stand for the internal conformational energy (ΔG_{MM}), the electrostatic ($\Delta G_{\text{ele}}^{\text{sol}}$) and non-electrostatic ($\Delta G_{\text{non-polar}}^{\text{sol}}$) components of the solvation free energy, and the entropic contribution ($-T\Delta S_{\text{conf}}$).

The internal conformational energy (ΔG_{MM}) was determined using the standard formalism and parameters implemented in Amber. The electrostatic contribution ($\Delta G_{\text{ele}}^{\text{sol}}$) was determined using a dielectric constant of 78.4 for the aqueous environment. The choice of the internal dielectric constant is still a subject of debate, as it must account for factors such as the electronic, vibrational and orientational relaxation of the environment, including both protein residues and solvent molecules [55]. Therefore, a diverse range of values have been proposed depending on factors such as the structural fluctuation of the snapshots used in the analysis (i.e., a “static” single structure vs. multiple “dynamic” snapshots), or the solvent accessibility. Moreover, the accuracy of continuum models also depends on parameters such as the solute charges and the atomic radii that define the dielectric boundary. In this context, we have computed $\Delta G_{\text{ele}}^{\text{sol}}$ using the optimized radii developed by Swanson et al. [56] for PB calculations with the atomic charges implemented in AMBER force field. Those radii were adjusted to reproduce charging free energies determined with free energy calculations using explicit solvent simulations. Following the protocol of the parameterization, PB computations were performed using a dielectric constant (ϵ_{int}) of 1 for the interior of the protein and zero bulk ionic strength. The electrostatic component of the solvation free energy was calculated from the energetic difference in the solvated ($\epsilon = 78.4$) and gas phase systems ($\epsilon = 1$). The electrostatic potentials were calculated using a grid spacing of 0.25 Å, and the interior of the solutes was defined as the volume inaccessible to a solvent probe sphere of radius 1.4 Å. For the sake of comparison, additional calculations were also performed using both Parse and standard AMBER radii.

The non-polar contribution ($\Delta G_{\text{non-polar}}^{\text{sol}}$) was calculated using a linear dependence with the solvent-accessible surface [57]. In particular, we have used the expression $\Delta G_{\text{non-polar}}^{\text{sol}} = \gamma \text{SAS} + b$, with SAS being the solvent-accessible surface. We have used two different sets of parameters: $\gamma = 0.0072 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$, and $b = 0.0 \text{ kcal mol}^{-1}$, and $\gamma = 0.0054 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ and $b = 0.86 \text{ kcal mol}^{-1}$ [57].

The entropic term ($-T\Delta S_{\text{conf}}$) describes the change in configurational entropy due to the loss of backbone and side chain torsional freedom. These terms have been computed using the computational scheme adopted by Honig and coworkers [58], which separates backbone and side-chain components. For the backbone, an entropic penalty of 2 kcal mol^{-1} per residue is considered. For the side chain, Honig et al. adopted the empirical scale defined by Pickett and Stenberg [59], which gives the configurational entropies of the amino acids obtained from the distribution of rotamers in non-homologous protein crystal structures. Here, nevertheless, we have used the mean values reported by Doig and Sternberg [60]. The difference is expected to be minimal, as the average contribution per frozen residue is -1.09 and $-0.95 \text{ kcal mol}^{-1}$. Following Froloff et al. [58], we have assumed that a solvent-exposed side chain (relative accessibility greater than 60%) rotates freely, whereas a buried side chain (relative accessibility lower than 60%) is restrained to one rotamer.

Finally, it is worth noting that this computational scheme should overestimate the binding affinity, as Eq. 1 omits the contributions arising from the conformational strain and the loss of translational and rotational degrees of freedom upon complexation, which are not explicitly considered. Those terms can make sizable contributions to the binding affinity [61, 62], though their magnitude is difficult to be estimated accurately. However, it seems reasonable to assume that those terms are likely to be similar for the binding of structurally related proteins to peptides with the same number of residues. Therefore, the analysis of MM/PBSA computations will be performed separately for the binding of 9-mer or 11-mer peptides.

Results and discussion

Consensus peptides

Two consensus sequences were chosen to calibrate the reliability of the structural models built up for *Sasa* class I and II MHC proteins. For class I MHC they rely on the preferred binding motifs determined specifically from the analysis of 7-mer and 12-mer peptides able to interact with *Sasa* UBA*0301, but also of 9-mer peptides interacting with murine H-2Kb complexes, as previous studies [40]

have pointed out that they share a similar peptide recognition pattern at the C-terminus region. The logo representation (Fig. 1) [63, 64] clearly illustrates the enormous structural diversity of the peptidic ligands that can bind MHC proteins, as there is a large degeneracy in the preferences for residues along the sequence of the peptides, as well as for the preferential recognition of amino acids at a given site along the sequence.

Although it was difficult to identify a unique consensus sequence, some amino acids seem to have a slight preference for specific positions, such as the presence of proline at position 2 in the 7-mer peptides, which was suggested as a putative anchoring site [40]. The preference for apolar residues at the C-terminus is found in the analysis of the 12-mer peptides, as noted in an enrichment of leucine, methionine and valine. This trend has been related to a similar binding motif at the C-terminus in murine class I H-2Kb [40]. Choice of the consensus peptides was then made by combining the propensities suggested by the sequence alignments in conjunction with a careful inspection of the most feasible arrangement of the side chains in the peptide-binding groove of the putative structural models of *Sasa* UBA*0201 and UBA*0301 proteins, yielding the sequences $M_1\text{PSQLYTLM}_9$ and $F_1\text{SGYNTAL}_9$ as tentative peptide candidates.

The aligned sequence of peptides corresponding to the invariant chain-associated class II MHC molecules were chosen to define two consensus sequences for class II MHC [41]. They arise from two alternative alignments of fish and

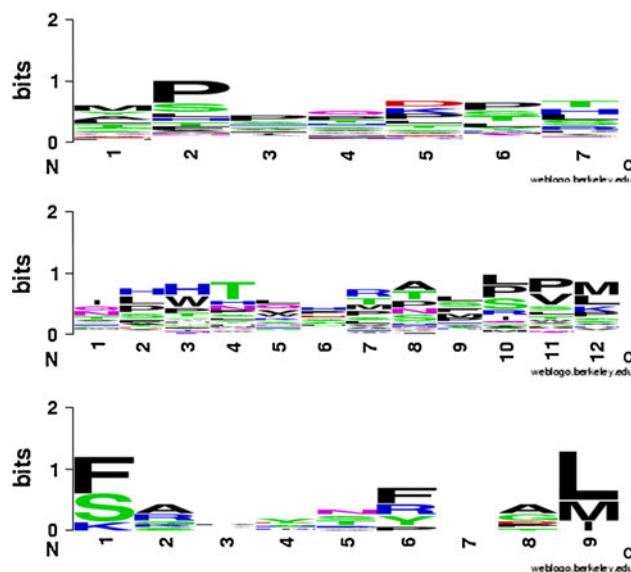


Fig. 1 Logo of the aligned (top) 7-mer and (middle) 12-mer residue sequences experimentally known to bind UBA*0301, and (bottom) 9-mer residue sequences identified in H-2Kb complexes. The propensity of a given residue is related to the size of the corresponding character [63, 64]

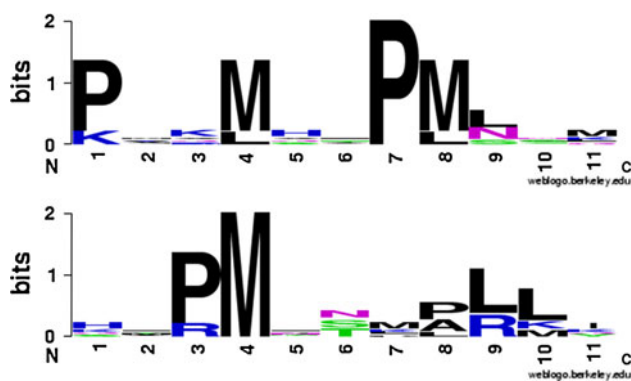


Fig. 2 Logo of the consensus peptides derived from two alternative alignments of fish invariant chain-related sequences with the human, mouse and chicken invariant chains. The peptides were derived upon exclusion (*top*) or inclusion (*bottom*) of the chicken sequence (see text and Table S5 in supporting information for more details). The propensity of a given residue is related to the size of the corresponding character [63, 64]

human invariant chains (see above and Table S5 in Supporting Information for more details), which in turn reflect the uncertainty in the identification of the CLIP region within the invariant chain in fishes. Propensities for specific residues at certain positions can be identified from the comparison of human and fish peptide sequences, such as apolar residues at positions 4, 7 and 8 in the 11-mer peptides (Fig. 2). However, a larger variability was observed at other sites, such as position 1, which contains proline and lysine in human peptides, but histidine in fish peptide sequences, and 3, where the lysine residue found in human sequences is replaced by proline (and at less extent by arginine) in fish peptides. For our purposes here, this information was used in conjunction with the manual docking of the peptidic side chains in the binding groove leading to the 11-mer sequences P₁MKMHMPMLNM₁₁ and H₁LPMLNMARLI₁₁ as consensus peptides.

Global structural analysis

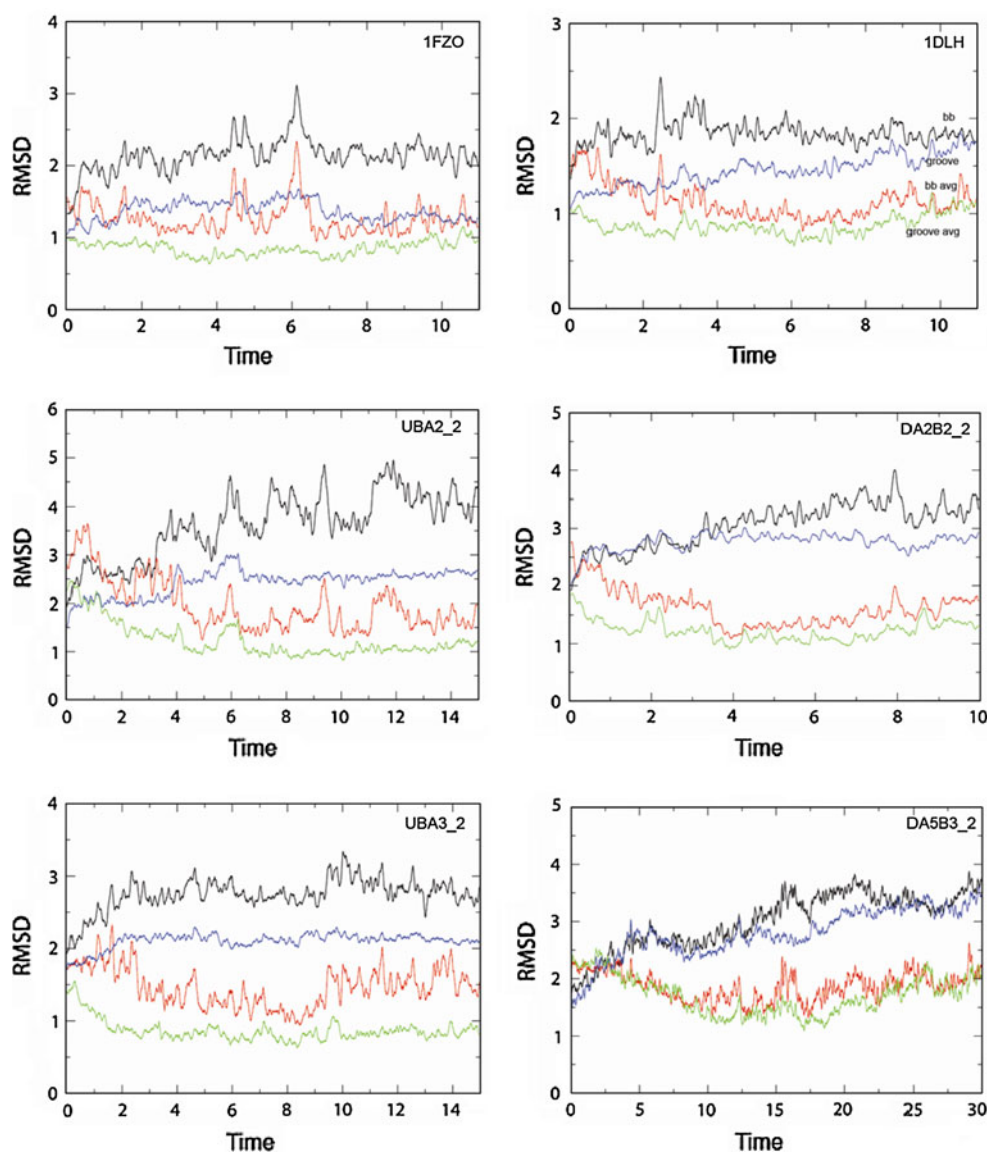
The structural integrity of the 3D structural model built up for each *Sasa* complex was examined by means of MD simulations. In all cases the trajectories were stable, as noted by the time evolution of both the potential energy (data not shown) and the positional root-mean square deviation (RMSD) of the protein skeleton (Fig. 3; see also Fig. S4 in Supporting Information). The structural integrity of the reference complexes 1FZO and 1DLH was noted by RMSD values of 2.2 and 1.8 Å, respectively, for the backbone of the whole protein, which decreases to around 1.5 Å for the residues that delineate the peptide-binding groove in the two cases, thus giving support to the computational protocol used for MD simulations.

Not surprisingly, larger RMSDs are found for the backbone atoms of the modeled *Sasa* proteins when compared to the energy-minimized homology models. The RMSD profiles determined for the protein backbone were stable after the first few nanoseconds and do not exhibit drastic fluctuations that might arise from artifactual changes in the modeled MHC complexes. In fact, the structural deviation found with regard to the average structure was similar for both the reference (1FZO and 1DLH) complexes and for the different *Sasa* alleles considering either the overall protein skeleton (RMSD values ranging from 1.0 to 1.9 Å) or the peptide-binding groove (RMSD values comprised between 0.8 and 1.4 Å). To further check the structural stability of the *Sasa* complexes, the simulation ran for DA5B3 bound to peptide H₁LPMLNMARLI₁₁ was enlarged up to 30 ns, but no drastic changes in the RMSD profiles were observed along the trajectory (see Fig. 3).

Superposition of the average structures for class I and II MHC proteins showed that the largest structural differences were associated with fluctuations in the loops, while the integrity of the domains formed by α -helices and β -strands is highly preserved (Fig. 4). The overall integrity of the models was also supported by several indexes of the structural quality of the sampled structures available in Swiss-model server (<http://swissmodel.expasy.org>) determined from a set of 25 snapshots taken regularly along the last 4 ns of the trajectories (Table 3). The content of secondary structural elements was determined from the atomic coordinates of the collected snapshots with DSSP [65] and found to be very similar in all cases, ranging from 0.75 to 0.79. In addition, the stability of the complexes was examined from the pseudo-energies determined with D-Fire [66] and Q-mean [67]. D-Fire is an all-atom statistical potential that estimates non-bonded interactions, and Q-mean is a scoring function whose pseudo-energy relies on different local and global structural descriptors and a solvation potential. Even though a direct comparison between the proteins from distinct organisms was limited by the differences in protein sequence, the scores obtained from both methods supported the quality of the structural models built up for the complexes of *Sasa* alleles.

Overall, the preceding results suggest that the global structural features of the *Sasa* MHC molecules were satisfactorily preserved in the models. Nevertheless, this does not suffice to warrant a proper description of the fine details associated with the differential interactions involved in the binding of antigenic peptides. Accordingly, further assessment on the quality of the structural models must come from the analysis of the nature and structural integrity of the interactions at the peptide-binding groove along MD simulations.

Fig. 3 RMSD values (Å) for the trajectories sampled for selected class I and class II MHC complexes. *Top:* 1FZO (left) and 1DLH (right). *Middle:* UBA2_2 (left) and DA2B2_2 (right). *Bottom:* UBA3_2 (left) and DA5B3_2 (right). The RMSD profiles are determined for the backbone of the whole protein versus the energy-minimized starting structures (the X-ray for 1FZO and 1DLH, or the homology model for *Sasa* alleles; *black*) or the averaged structure of the simulation (*red*), as well as for the residues that define the peptide-binding groove (in *blue* and *green* for the starting or averaged structure, respectively)



Peptide binding to class I MHC proteins

The binding groove in class I MHC complexes consists of five pockets, which are denoted A-F [24, 68]. The N- and C-terminal sides of the bound peptides fit pockets A and F, respectively. Relevant residues for interaction with the peptide were conserved in MHC proteins: α Y7(7), α Y59(57), α Y171(169) in pocket A, and α W147(144) in pocket F (the numbering is made using 1FZO as reference, whereas the sequence numbering in *Sasa* proteins is given in parenthesis). Other conserved residues were α F22(21) in pocket B, α F74(72) in pocket C, α G100(97), α C101(98), α Y159(157) in pocket D, and α K146(143) in pocket F.

In spite of the preceding similarities, a relevant number of mutations were also found in the peptide-binding grooves of murine and *Sasa* class I MHC proteins. Few involve conserved mutations, such as α M52 \rightarrow I(49) in

UBA2, α R170 \rightarrow K(168) in both UBA2 and UBA3, and α N70 \rightarrow T(68) in UBA2 and S(68) in UBA3. However, most of the mutations imply drastic changes in the steric and electrostatic features of the groove (see Fig. 5). Thus, the shape and size in the middle of the groove were notably altered by mutations α V9 \rightarrow Y(9) and α S99 \rightarrow Y(96). Similarly, the pair of residues α V97 and α Y116, which were spatially close, were mutated to K(94) and D(113), respectively, while residues α Q114 and α L156 were mutated to D(111) and R(154). The reverse change is found in residues α E152 and α R155, which were changed to Q(150) and Y(153), respectively. Other relevant changes were the replacement of α D77 by N(75) and α Y84 by R(82). Finally, it is worth noting the existence of differences in specific residues between alleles UBA2 and UBA3. Among those differences, some involve similar changes, such as the mutations α R62 \rightarrow T(60; UBA2)

Fig. 4 Superposition of the average structures obtained from class I (left) and II (right) MHC complexes. Color code: 1FZO and 1DLH (blue); UBA2_1 and DA2B2_1 (orange); UBA2_2 and DA2B2_2 (red); UBA3_1 and DA5B3_1 (yellow); UBA3_2 and DA5B3_2 (green). Figures were made with PyMOL

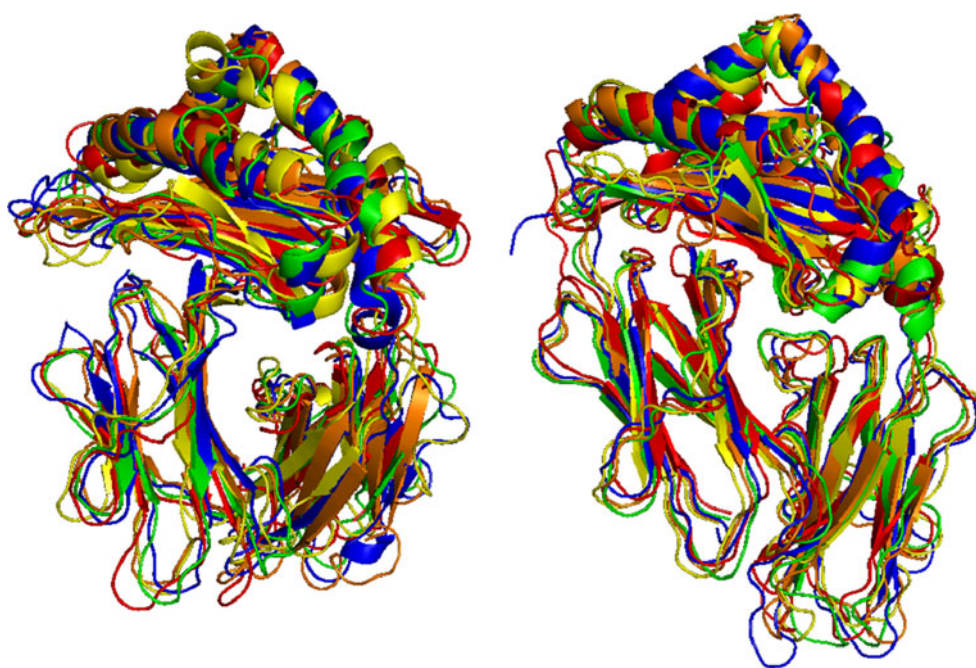


Table 3 Indexes of the model quality as calculated in Swiss-Model workspace

| Complex | D-fire | Q-mean | |
|-------------------|-----------------|---------------|----------------------------------|
| | | Score | Secondary structure ^b |
| 1FZO ^a | -512.2 (-559.8) | -80.3 (-80.8) | 0.79 (0.85) |
| UBA2_1 | -451.0 | -64.1 | 0.75 |
| UBA2_2 | -402.1 | -65.3 | 0.76 |
| UBA3_1 | -423.5 | -63.0 | 0.77 |
| UBA3_2 | -438.1 | -65.3 | 0.74 |
| 1DLH ^a | -497.1 (-535.3) | -66.2 (-68.0) | 0.75 (0.83) |
| DA2B2_1 | -451.0 | -59.2 | 0.75 |
| DA2B2_2 | -454.3 | -58.4 | 0.75 |
| DA5B3_1 | -465.7 | -65.7 | 0.79 |
| DA5B3_2 | -455.0 | -63.3 | 0.75 |

^a Values derived for the X-ray structure are given in parenthesis

^b Agreement between predicted and observed secondary structure

or S(UBA3), α E63 \rightarrow N(61; UBA2) or Q(UBA3), and α K66 \rightarrow I(64; UBA2) or V(UBA3), while other imply more drastic changes, such as the mutation α Y45 \rightarrow A(42; UBA2) or E(UBA3), and the replacement of α E55 by V(53) in UBA2, whereas a glutamic residue was retained in UBA3.

The mutations found at specific positions of the peptide-binding groove should have a drastic effect on the recognition of antigenic peptides not only between murine and *Sasa* class I MHC proteins, but even between the two *Sasa* alleles. Therefore, it is worth to examine the stability of the

interactions that mediate the binding of peptides for the complexes of UBA2 and UBA3 with the consensus peptides F₁SGYNATAL₉ and M₁PSQLYTLM₉, and compare them with those determined for the complex between murine H-2 Kb and the peptide (F₁SGYNATAL₉) found in the X-ray crystallographic structure 1FZO, which was used as reference system.

1FZO. A large fraction of the interactions that mediate the binding of F₁APGNYPAL₉ to H-2 Kb are preserved along the simulation (see Fig. S5 in Supporting Information). In pocket A the amino group of the terminal F₁ residue is linked to the carboxylate group of α E63 (average value of 2.8 Å in the MD simulation; 3.4 Å in the X-ray structure), which in turn is hydrogen-bonded to the NH group of A₂ (MD: 2.9 Å; X-ray: 2.9 Å), and forms transient interactions with the hydroxyl groups of α Y7 and α Y171. Moreover, the carbonyl oxygen of F₁ is hydrogen-bonded to α Y159 (MD: 3.4 Å; X-ray: 2.7 Å). Compared to the X-ray structure, the enlargement of this latter interaction likely stems from the weakening of the cation- π interaction between α K66 and F₁ due to the large exposure to the aqueous solvent. In pocket F L₉ is stabilized by the salt bridge between the terminal carboxylate group and the amino group of α K146 (MD: 2.9 Å; X-ray: 2.8 Å), and by the hydrogen bond between the backbone NH and the carboxylate group of α D77 (MD: 3.3 Å; X-ray: 2.8 Å). Likewise, the indole NH unit of α W147 interacts with the (C=O) group of A₈ (MD: 2.8 Å; X-ray: 2.8 Å). A number of interactions are also retained in the central part of the peptide. For instance, the carbonyl oxygen of P₃ is hydrogen-bonded to α N70 (MD: 3.3 Å; X-ray: 2.9 Å),

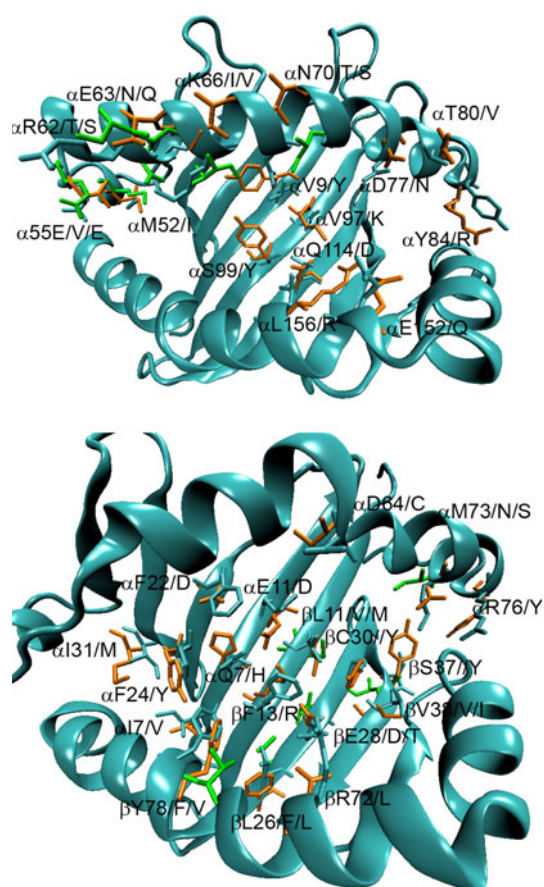


Fig. 5 Representation of selected mutations in the binding groove of *Sasa* class I (*top*) and class II (*bottom*) MHC proteins in comparison with murine H-2Kb and human HLA-DR β 1*0101 proteins. Labels are referred to the murine or human protein followed by the corresponding residue in *Sasa* UBA2 and UBA3 proteins or DA2B2 and DA5B3 (a single label is used in those cases where the same residue is present in both). For clarity only the backbone skeleton of the reference proteins (1FZO and 1DLH) is shown in cyan. Mutated residues between *Sasa* proteins are shown in orange and green, respectively. Figures were made with VMD

while the ring of P₃ fills a pocket defined by α Y7 and α Y159. In addition, Y₆ protrudes into the groove filling the space delineated by residues α Y7, α V9, α V97 and α S99, while its backbone NH group interacts with α N70 (MD: 3.3 Å; X-ray: 2.9 Å), and the (C=O) group of P₇ forms transient interactions with α Y116 (MD: 3.5 Å; X-ray: 2.7 Å). Overall, the results indicate that the features that mediate binding of peptide F₁APGNYPAL₉ are generally well retained along the simulation.

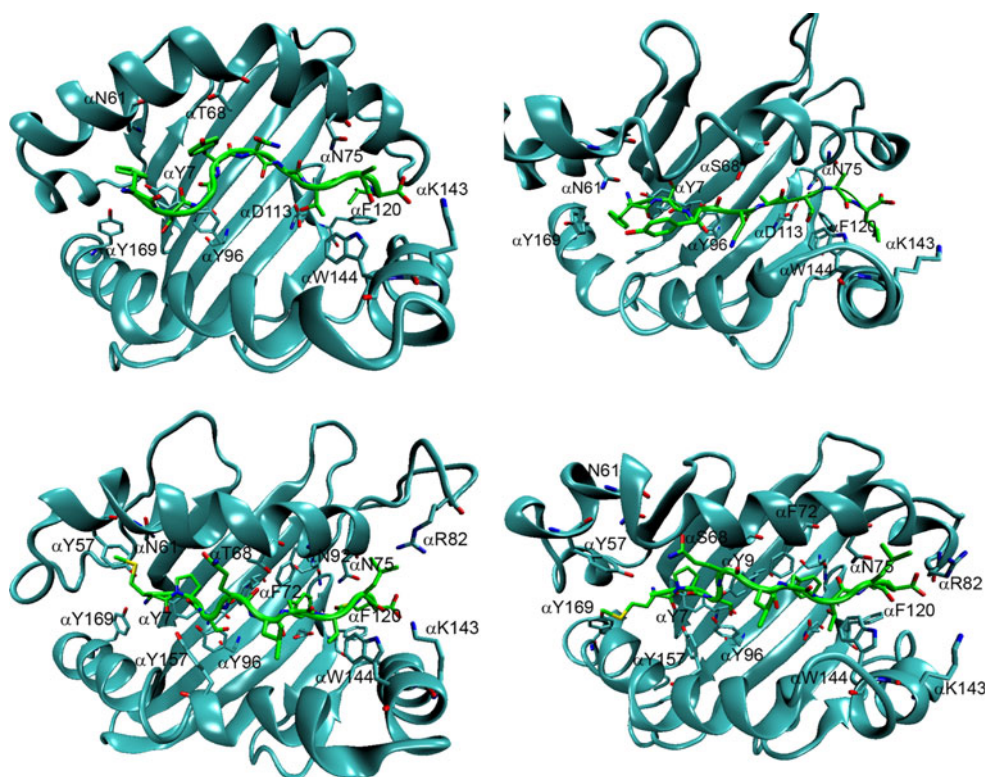
UBA2_1 and UBA3_1. Binding of F₁SGYNATAL₉ to *Sasa* UBA*0201 (UBA2_1) exhibits differences in pocket A and similarities in pocket F compared to F₁APGNYPAL₉ in 1FZO (see Fig. 6). In pocket F the terminal carboxylate group of L₉ interacts with α K143 (3.4 Å), and the aliphatic side chain is close to the benzene ring of α F120. Moreover, the (C=O) group of A₈ retains the hydrogen-

bond with α W144 (3.3 Å), and α D113, which replaces Y116 in FZO, contributes to the anchoring of the peptide through hydrogen-bonds with the backbone NH (2.8 Å) of T₇ and with its hydroxyl moiety (2.8 Å). In contrast, the salt bridge formed between F₁ and α E63 in 1FZO (see above) is lost due to the mutation of this residue to N. In turn, the terminal amino group is exposed to the solvent, whereas the benzene ring protrudes into the pocket, which alters the arrangement of the peptide backbone along the groove compared to the 1FZO complex.

In contrast to the preceding results, a more dense interaction pattern is observed for the binding of F₁SGYNATAL₉ to UBA*0301 (UBA3_1; see Fig. 6). In pocket A the terminal amino group mimicks the interactions found for 1FZO, as it forms a salt bridge with α E42 (2.8 Å); the carboxylate moieties of α E63 in 1FZO and α E42 in UBA3 were spatially close, but this latter residue replaces α Y45 in H-2 Kb) and is also hydrogen-bonded to α Y7 (2.9 Å) and α Q61 (2.7 Å). Whereas the benzene ring of F₁ in 1FZO is exposed to the solvent, this orientation is impeded here due to the steric clash with Y₄, and the benzene ring points towards the edge of the groove lying partially stacked with α Y169. Additional anchoring points are the hydrogen-bonds between the (C=O) group of F₁ and α Y157 (2.8 Å), and between the hydroxyl group of S₂ and α E42 (2.6 Å). In the central part of the peptide binding is assisted by the hydrogen bond between the NH unit of G₃ and α Y96 (2.9 Å), the electrostatic stabilization between α R154 and the carbonyl oxygen of G₃ (3.0 Å) and N₅ (2.9 Å), and the hydrogen-bond between T₇ and α N75 (3.4 Å). Finally, in pocket F the terminal carboxylate group transiently contacts α K143 (4.1 Å), with the side chain of L₉ lying onto α F120, and the (C=O) group of A₈ hydrogen-bonded to α W144 (3.1 Å).

UBA2_2 and UBA3_2. Binding of M₁PSQLYTLML₉ to UBA*0201 (UBA2_2) reflects qualitatively the trends noted above for UBA2_1 (Fig. 6). Thus, there is no salt bridge interaction in pocket A between the terminal amino group in M₁ and α E42. Instead, the amino group forms hydrogen-bonds with the hydroxyl groups of α Y7 (3.1 Å) and α Y57 (3.4 Å), while the thiomethyl group lies onto the aromatic ring of α Y57. In pocket F the terminal carboxylate group of M₆ forms a salt bridge with α R82 (3.7 Å) and α K143 (3.9 Å), and the side chain interacts with α F120. A series of hydrogen bonds is observed along the peptide, such as those formed between the (C=O) groups of P₂, S₃ and L₈ with α Y157 (3.3 Å), α Y9 (3.0 Å) and α W144 (2.9 Å), respectively, or the backbone NH of S₃ and Q₄ with α Y96 (3.1 Å) and α Y153 (3.4 and 3.6 Å). Other interactions involve the hydrogen-bonded bridge between Y₆ with α N92 (3.1 Å) and α D113 (2.7 Å), the stacking between the aromatic rings of Y₆ and α F72, and the hydrophobic contact between the side chains of L₅ and α Y153.

Fig. 6 Selection of representative interactions found in the peptide-binding groove for *Sasa* UBA2 and UBA3 bound to consensus peptides. *Top* UBA2 (*left*) and UBA3 (*right*) complexed to F₁APGNYPAL₉. *Bottom* UBA2 (*left*) and UBA3 (*right*) complexed to M₁PSQLYTLM₉. Figures were made with VMD



Most of the preceding interactions are also found in the binding of M₁PSQLYTLM₉ to UBA3 (UBA3_2). However, as noted before for the binding of F₁SGYNATAL₉, the most remarkable difference is found in pocket A, where the protonated amino group of M₁ forms a salt bridge with α E42 (3.1 Å) and hydrogen bonds with α Y7 (3.0 Å) and α Y57 (3.1 Å). Moreover, this structural rearrangement is accompanied with a conformational change of α W165, which can thus retain the cation- π interaction with the terminal amino group.

In summary, the preceding discussion suggests that the two consensus peptides interact preferentially with the allele UBA*0301, and that this preference arise from a better accommodation of the peptide to the features of pocket A. Therefore, this finding agrees with the fact that the peptides were chosen based on the preferred motifs determined for peptides experimentally known to bind *Sasa* UBA*0301 [40].

Peptide binding to class II MHC proteins

The interaction of the peptide with class II MHC proteins involves residues located in five pockets named 1, 4, 6, 7 and 9 [43]. The peptide-binding region is delineated by residues in both α and β chains, and some residues relevant for peptide binding in human leukocyte antigens have been proposed [69–72]. In pocket 1, α E55 and β H81 in HLA proteins (numbering relative to PDB entry 1DLH; [43] see Fig. S1),

which contribute to the anchoring of the peptide, correspond to residues α E57 and β H77 in *Sasa* proteins. Similarly, residue α N69, which participates in hydrogen bonding with the peptide backbone in pockets 6 and 7, was conserved in *Sasa* proteins (residue α N69), while residues β W61 and β L67 in pocket 6 were also preserved (residues β W58 and β L64 in *Sasa* proteins). However, there were also numerous differences between human and *Sasa* alleles (Fig. 5), including not only conserved changes (for instance, α I7 \rightarrow α V(3), α F24 \rightarrow α Y(22), α E11 \rightarrow α D(7), and β V38 \rightarrow β V(35; DA2B2) or β I(35; DA5B3)), but also changes that imply substantial changes in both the shape and size and the polarity of the groove, such as β F13 \rightarrow β R(9), β R71 \rightarrow β E(67), α Q9 \rightarrow α H(5) and β S37 \rightarrow β Y(34). Finally, it is worth noting the existence of mutations leading to differences between the *Sasa* alleles, such as β W(9) \rightarrow β E(5; DA2B2) or β Y(DA5B3), β E(28) \rightarrow β D(25; DA2B2) or β T(DA5B3), and α M(73; 1DLH) \rightarrow α N(73; DA2B2) or α S(DA5B3).

In order to explore the impact of those mutations on the binding properties, MD simulations were run for the X-ray crystallographic structure 1DLH (bound to the peptide K₁YVKQNTLKL_{A11}) as control system, and the modeled complexes of *Sasa* alleles DA2B2 and DA5B3 bound to peptides P₁MKMHMPMLN_{M11} and H₁LPMLN_{MARLI11}.

1DLH. There is a close resemblance between the arrangement of the peptide K₁YVKQNTLKL_{A11} in the snapshots taken along the MD simulation and the X-ray

crystallographic structure (PDB entry 1DLH; see Fig. S5 in Supporting Information), which reflects the fact that most of the interactions found between the peptide and the binding groove are preserved along the trajectory. Thus, the position of K₁ is maintained by a hydrogen bond between the carbonyl oxygen and the imidazole ring of β H81 (MD: 2.8 Å; X-ray: 2.7 Å), and from transient interactions between the side chain amino group and α E55 (MD: 4.8 Å; X-ray: 4.5 Å). A variety of hydrogen bonds mediate binding of the peptide backbone to the groove: NH in Y₂ with (C=)O in α S53 (MD: 2.9 Å; X-ray: 2.7 Å); NH in V₃ with the amido (C=)O in β N82 (MD: 2.9 Å; X-ray: 2.7 Å); (C=)O in V₃ with the amido NH in β N82 (MD: 2.8 Å; X-ray: 2.7 Å); NH in Q₅ with the amido (C=)O in α Q9 (MD: 3.0 Å; X-ray: 2.7 Å); (C=)O in N₇ with β R71 (MD: 2.9 Å; X-ray: 2.8 Å); (C=)O in L₈ with the amido NH in α N69 (MD: 3.2 Å; X-ray: 2.7 Å); (C=)O in K₉ with the indole NH in β W61; and NH in L₁₀ with the amido (C=)O in α N69 (MD: 3.1 Å; X-ray: 2.8 Å). Other stabilizing contacts found in both the X-ray structure and along the trajectory are the hydrophobic contacts of Y₂, which fills the pocket formed by α W43, α F32, α F24 and β V85, as well as L₈ and L₁₀, which interact with β W61, β W9 and α M36. Finally, the terminal residue A₁₁ is fixed through a hydrogen bond between the backbone NH group and the carboxylate moiety of β D57 (MD: 2.9 Å; X-ray: 3.0 Å) and a salt bridge with α R74 (MD: 3.2 Å; note that in the X-ray structure this interaction is mimicked by a contact between the backbone (C=)O and α R74: 3.4 Å). Overall, these results indicate that the main features that mediate binding of the peptide K₁YVKQNTLKLA₁₁ to human HLA-DR β 1*0101 in the X-ray structure are properly retained along the trajectory.

DA2B2_1 and DA5B3_1. The binding of P₁MKMHMPMLNM₁₁ to DAA*0201-DAB*0201 (DA2B2_1) and DAA*0501-DAB*0301 (DA5B3_1) exhibits notable differences compared to the interaction of K₁YVKQNTLKLA₁₁ in 1DLH.

For DA5B3_1 the position of P₁ is only fixed by a hydrogen bond between the carbonyl oxygen and β H77 (3.0 Å). Moreover, unlike K₁ in 1DLH, the protonated proline ring does not interact with the binding groove. Similarly, the terminal carboxylate group in M₁₁ is exposed to the aqueous solvent, which can be explained by the mutation of α R74 in 1DLH to Y. In addition, a reduced number of hydrogen bonds with the peptide backbone is found, such as NH in M₂ with (C=)O in α T51 (3.0 Å); NH in K₃ with β N78 (3.0 Å); (C=)O in K₃ with β N78 (2.9 Å); (C=)O in L₉ with β W58; and NH in N₁₀ with α N69 (2.8 Å). Other interactions are altered due to specific mutations, such as the lost of the interaction between (C=)O in position 6 and β R71 in 1DLH due to the mutation of this latter residue to β E67, or the formation of a hydrogen-bond

between NH in M₈ and the hydroxyl group in β Y27 (3.5 Å), which replaces β C30 in 1DLH.

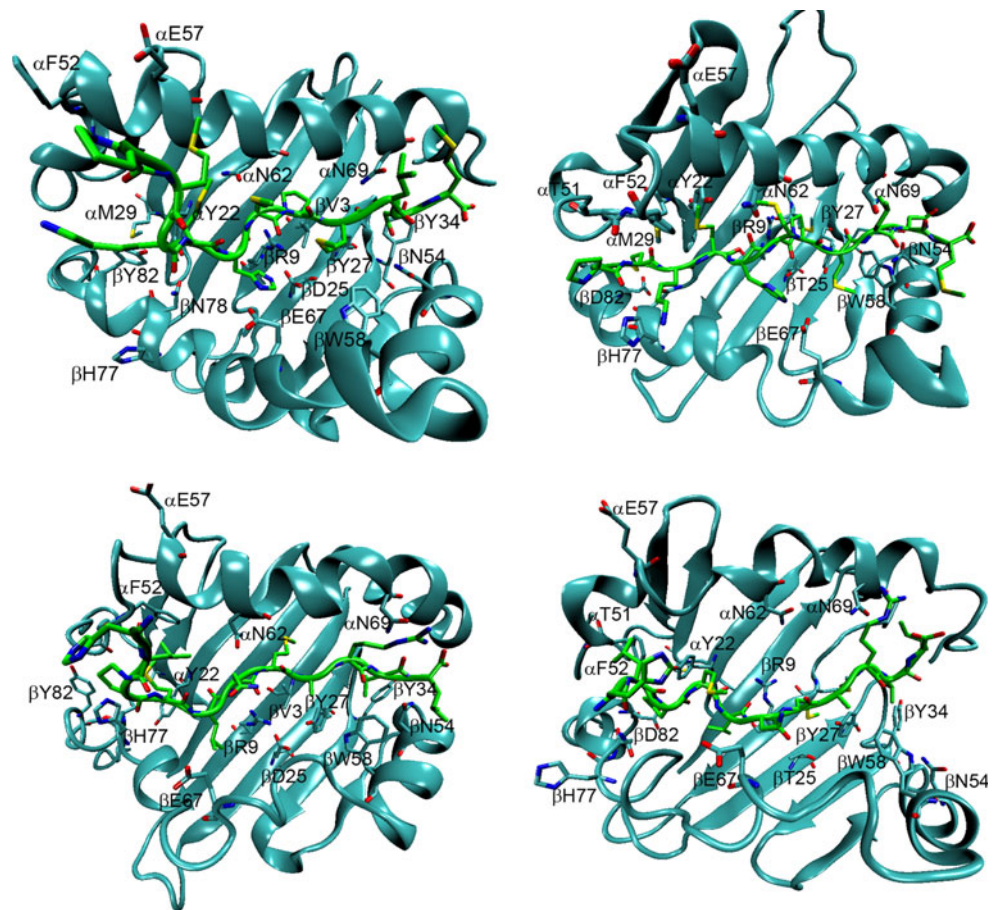
The structural integrity of the N-terminal part of the peptide P₁MKMHMPMLNM₁₁ in DA2B2_1 is less satisfactory, as there seems to be a weaker binding to the groove. This structural perturbation can be attributed to β Y82, which protrudes into the binding groove in DA2B2, and replaces β G86 in 1DLH. Accordingly, whereas a tyrosine residue is found in front of β G86 in the X-ray structure of 1DLH, the side chain of M₂ cannot be accommodated due to steric clash with β Y82 in DA2B2 (this residue is replaced by D in DA5B3, whose side chain is pointing towards H₁ and thus does not impede the binding of the peptide). The rest of the peptide backbone is quite stable, which likely reflects the role played by electrostatic interactions in the middle region of the groove, as the protonated imidazole ring in H₅ bridges β E67 (3.9 Å) and β D25 (2.8 Å), which in turn is linked to β R9 (3.3 Å). In addition, the guanidinium moiety of β R9 interacts with the (C=)O groups of H₅ (2.9 Å) and M₆ (3.2 Å).

DA2B2_2 and DA5B3_2. The arrangement of the backbone of H₁LPMLNMARLI₁₁ in the binding groove of DAA*0501-DAB*0301 (DA5B3_2) is rather well preserved along the 30 ns trajectory. Compared to the simulation run for 1DLH, the most flexible region is the N-terminal part, which reflects the absence of the stabilizing interaction between the backbone carbonyl group and β H77. However, a number of hydrogen bonds are found along the peptide: (C=)O in M₄ and α H5 (2.7 Å), (C=)O in L₅ and β R9 (3.1 Å), NH in A₈ and β Y27 (3.0 Å), (C=)O in A₈ and α N69 (3.2 Å), (C=)O in R₉ and β W58 (3.1 Å), NH in L₁₀ and α N69 (3.3 Å), and NH in I₁₁ and β N54 (3.2 Å). Moreover, additional interactions are found between the side chain of N₆ and β E67, and between the terminal carboxylate group in I₁₁ and both α Y77 and β N54, even though these latter interactions are lost in the last part of the trajectory. Finally, binding is also assisted by hydrophobic contacts of L₂, which fills the pocket formed by α Y22, α M29 and α F52, and of M₇, which interacts with α V65 and β M7.

As noted above for the interaction of P₁MKMHMPMLNM₁₁ in DA2B2, binding of the N-terminal region of H₁LPMLNMARLI₁₁ to DA2B2 is affected by the protrusion of β Y82 into the groove (Fig. 7), which hinders the proper positioning of the side chain of L₂. Likewise, the rest of the peptide remains more ordered due to different interactions, involving several hydrogen bonds with the peptide backbone ((C=)O in N₆ with β R9 (2.8 Å); NH in M₇ with α N62 (2.9 Å) and (C=)O in L₁₀ with α N69 (2.9 Å)) and the hydrophobic contact between L₁₀ and β Y34.

In summary, even though a series of interactions between the peptides and the residues that define the binding groove in *Sasa* models are preserved along the

Fig. 7 Selection of representative interactions found in the peptide-binding groove for *Sasa* DA2B2 and DA5B3 bound to consensus peptides. *Top* DA2B2 (*left*) and DA5B3 (*right*) complexed to P₁MKMHMPMLNM₁₁. *Bottom* DA2B2 (*left*) and DA5B3 (*right*) complexed to H₁LPMLNMARLL₁₁. Figures were made with VMD



trajectories, the structural integrity of the complexes seems to be slightly less satisfactory, specially regarding the binding of the N-terminal region of the peptides to the allele DAA*0201-DAB*0201, which reflects the steric hindrance due to the presence of $\beta Y82$.

Energetic analysis

In order to explore the correspondence between the structural features that mediate peptide binding with the relative binding affinities, MM-PBSA computations were performed for a series of snapshots collected along the last part of the trajectories (see “Methods”). This technique has been largely used to re-rank the ligand poses in docking studies [73–75]. Regarding peptide-MHC complexes, Honig and coworkers used MM/PBSA computations to examine the binding affinities of 8 MHC class I complexes, including murine MHC class I protein H-2K and human MHC class I HLA-A2 proteins, and concluded that the results are successful in reproducing a fairly small range of binding affinities [58]. More recently, Schafroth and Floudas have compared the reliability of MM-PBSA, MMSED and RRIGS methods to predict peptide binding

to MHC pockets, concluding that the former is the most accurate approach [76].

The lack of precise data for the binding of peptides to fish MHC molecules limits the capability to perform a quantitative analysis of the calculated binding affinities. Moreover, since the binding to *Sasa* MHC molecules has been performed using *consensus* peptides, it is convenient to restrict such analysis to a qualitative level. In this context, we have performed MM/PBSA computations to estimate upper and lower limits to the binding of peptides to human class II HLA-DR $\beta 1$ *0101 and HLA-DR $\beta 1$ *0301 alleles. This strategy seems more adequate for our purposes as there is notable discrepancy in the affinities reported in the literature, which would then affect the suitability of a quantitative analysis. For instance, a search for the binding affinities to human HLA-DR $\beta 1$ *0101 of the fragment 306–318 of hemagglutinin HA peptide (PKYVKQNTLKLAT) compiled in the immune epitope database [77] (<http://www.immuneepitope.org/>) rendered IC₅₀ values ranging from 2 to 600 nM.

The trajectories sampled during MD simulations for the complexes with human class II MHC molecules were stable, as noted by the time evolution of the potential energy

(data not shown) and the positional RMSD of the protein skeleton, which ranged from 1.7 to 2.4 Å (see Fig. S6 in Supporting Information). The results shown in Table 4 indicate that the binding affinities determined for the ideal binding peptides are significantly more negative than the expected range of experimental values. As noted in “Methods”, this deviation can be attributed to the fact that the positive contributions arising from the conformational strain energy and the entropic cost of fixing translational and rotational degrees of freedom upon complexation have been omitted. However, the overestimated binding affinities can also be attributed to the difficulty in keeping a precise balance between the different components of the binding affinity in MM/PBSA. For instance, even though the atomic radii reported by Swanson et al. [56] were optimized by fitting the charging free energies for peptides in solution, to the best of our knowledge there is no systematic analysis about the suitability of atomic tensions to be used in conjunction with the solvent-accessible surface for calculating the non-polar contribution to the solvation free energy.

To further check the reliability of the MM/PBSA values, additional computations were performed using PARSE radii in conjunction with an internal dielectric constant of 2, which was already considered in the derivation of the PARSE parameters [57], and the standard radii implemented in AMBER. The results in Table 5 reveal a sizable dependence between the calculated binding affinity and the atomic radii, especially when the standard AMBER parameters are used as they lead to too negative binding affinities. Nevertheless, it is also worth noting that the

Table 4 Calculated binding free energies determined for the binding of peptides to human class II MHC molecules

| Complex | ΔG_{MM}^a | ΔG_{ele}^{sol} | ΔG_{n-ele}^{sol} ^b | $-T\Delta S_{conf}$ | $\Delta G_{binding}^b$ |
|---------------|----------------------------|------------------------|---------------------------------------|---------------------|------------------------|
| <i>IDLH</i> | | | | | |
| Control/ideal | −606.3 (−495.2; −111.1) | 561.6 | −16.1 −11.2 | 34.0 | −26.8 −21.9 |
| Poor | −368.2 (−118.2; −82.2) | 298.7 | −11.4 −7.7 | 33.9 | 2.6 6.3 |
| <i>IA6A</i> | | | | | |
| Control | −357.2 (−236.0; −121.3) | 306.9 | −13.7 −9.4 | 38.4 | −25.6 −21.3 |
| Ideal | −556.4 (−441.8; −114.7) | 485.6 | −14.4 −9.9 | 39.2 | −46.0 −41.5 |
| Poor | −320.4 (−247.2; −73.2) | 302.8 | −11.0 −7.4 | 34.4 | 5.8 9.4 |

^a The electrostatic (left) and van der Waals (right) components are given in parenthesis

^b Values determined using two sets of parameters for the linear dependence with the solvent-accessible surface (see “Methods”)

Table 5 Calculated binding free energies determined for the binding of peptides to human class II MHC molecules using different atomic radii and internal dielectric constants^a

| Complex | Optimized AMBER ^b | PARSE ^c | Standard AMBER ^d |
|---------------|------------------------------|--------------------|-----------------------------|
| <i>IDLH</i> | | | |
| Control/ideal | −26.8 −21.9 | −20.1 −15.3 | −68.3 −63.4 |
| Poor | 2.6 6.3 | 17.2 20.9 | −23.9 −20.2 |
| <i>IA6A</i> | | | |
| Control | −25.6 −21.3 | 3.7 7.9 | −47.7 −43.4 |
| Ideal | −46.0 −41.5 | −23.9 −19.4 | −118.3 −113.8 |
| Poor | 5.8 9.4 | 9.5 13.1 | −23.7 −20.1 |

^a Values determined using two sets of parameters for the linear dependence with the solvent-accessible surface (see “Methods”)

^b $\epsilon_{int} = 1$; Ref. [56]

^c $\epsilon_{int} = 2$; Ref. [57]

^d $\epsilon_{int} = 2$

results reflect adequately the distinction in the relative binding affinities estimated for *ideal* and *poor* binding peptides. Thus, in agreement with the findings by Swanson et al. [56], although changes in either the solute dielectric or the boundary between solute and solvent can have a remarkable effect on the calculated solvation free energy, they are less relevant for the comparison of relative solvation free energies between structurally related peptides.

Table 6 reports the calculated components of the binding affinity determined according to Eq. 1 for the *Sasa* MHC-peptide-MHC complexes and the reference systems. With regard to class I MHC complexes, the most relevant finding is the larger affinity predicted for the binding of peptides F₁SGYNATAL₉ and M₁PSQLYTLM₉ to the *Sasa* allele UBA*0301, which can be ascribed to the internal conformational energy (ΔG_{MM}) of the MHC-peptide complex. Such a preference is mainly due to the larger stabilization arising from electrostatic interactions, which is partially compensated by the desolvation cost. Moreover, the van der Waals component also favors the binding of the peptides to UBA*0301, which suggests a larger steric complementarity with the binding groove. These findings are not completely unexpected, as the two consensus peptides were chosen according to the preferred motifs determined experimentally from binding assays to *Sasa* UBA*0301 (see above; [40]). Thus, whereas the binding affinity predicted for the interaction of F₁SGYNATAL₉ and M₁PSQLYTLM₉ to the *Sasa* allele UBA*0201 compares with the value obtained for the binding of F₁APGNYPAL₉ to murine H-2Kb,

Table 6 Calculated binding free energies determined for the binding of peptides to *Sasa* class I and II MHC molecules

| Complex | ΔG_{MM}^a | ΔG_{ele}^{sol} | ΔG_{n-ele}^{sol} ^b | $-T\Delta S_{conf}$ | $\Delta G_{binding}^b$ |
|-------------------|----------------------------|------------------------|---------------------------------------|---------------------|------------------------|
| IFZO ^a | −253.1 (−167.5; −85.7) | 224.6 | −10.9 −7.3 | 28.4 | −11.1 −7.5 |
| UBA2_1 | −312.6 (−230.8; −81.9) | 277.0 | −10.4 −6.9 | 28.0 | −18.0 −14.5 |
| UBA2_2 | −293.7 (−197.2; −96.6) | 259.2 | −12.3 −8.4 | 30.5 | −16.3 −12.4 |
| UBA3_1 | −453.0 (−350.7; −102.3) | 379.7 | −11.8 −8.0 | 26.9 | −58.3 −54.5 |
| UBA3_2 | −486.9 (−379.1; −107.8) | 408.9 | −12.4 −8.4 | 32.8 | −57.6 −53.6 |
| 1DLH ^a | −606.3 (−495.2; −111.1) | 561.6 | −16.1 −11.2 | 34.0 | −26.8 −21.9 |
| DA2B2_1 | −392.0 (−324.1; −68.0) | 354.9 | −8.8 −5.7 | 31.5 | −14.4 −11.4 |
| DA2B2_2 | −368.2 (−274.2; −94.0) | 334.5 | −11.7 −7.9 | 32.7 | −12.6 −8.8 |
| DA5B3_1 | −331.1 (−226.6; −104.5) | 289.3 | −16.4 −11.4 | 35.5 | −22.7 −17.7 |
| DA5B3_2 | −350.5 (−255.5; −94.9) | 320.1 | −16.6 −11.6 | 33.0 | −14.0 −8.9 |

^a The electrostatic (left) and van der Waals (right) components are given in parenthesis

^b Values determined using two sets of parameters for the linear dependence with the solvent-accessible surface (see “Methods”)

binding to *Sasa* UBA*0301 is significantly more favorable. The preferential binding determined from MM/PBSA computations for the two consensus peptides to UBA*0301 gives confidence to the overall quality of the 3D models built up for the *Sasa* alleles, as the structural and energetic analysis allow us to discriminate the binding of the peptides to *Sasa* UBA*0201 and UBA*0301.

Regarding class II MHC molecules, binding of the consensus peptides to the *Sasa* alleles is less favored than the interaction of peptide K₁YVKQNTLKLA₁₁ to human HLA-DRβ1*0101 (1DLH), but are much higher than the value obtained for the *poor* binding peptide (see Table 4). The exception to this general trend is the interaction of peptide P₁MKMHMPMLNM₁₁ to DAA*0501-DAB*0301 (DA5B3_1), since the calculated binding affinity compares with that determined for 1DLH. According to the structural analysis of peptide-MHC complexes derived from molecular dynamics simulations, the lower binding affinity predicted for the consensus peptides can be attributed to two main structural features. First, whereas favorable contacts are observed between residues K₁ and A₁₁ in the peptide K₁YVKQNTLKLA₁₁ bound to HLA-DRβ1*0101, the electrostatic interactions of the capping groups in the consensus peptides are weaker, which can be explained by the mutation of αR74 in 1DLH to Y in the two *Sasa* alleles. Second, the replacement of βG82 in 1DLH by Y, which protrudes into the groove, impedes the correct positioning of the N-terminus, thus contributing to the difference in the

van der Waals contribution to the binding of peptides to DA2B2 and DA5B3.

Overall, the structural integrity and energetic stability found for the binding of consensus peptides F₁SGYNA-TAL₉ and M₁PSQLYTLM₉ to class I *Sasa* UBA*0301 reflects the correspondence that underlies those peptides and the preferred binding motifs determined for this allele. At this point, it is worth noting that those peptides were derived based on the sequences of 107 7-mer and 74 12-mer peptides known to interact with this allele in binding assays reported by Zhao et al. [40]. Present results, therefore, give confidence to the computational strategy adopted here to derive the structural model for UBA*0301. Regarding class II alleles, the binding predicted for P₁MKMHMPMLNM₁₁ and H₁LPMLNMARLI₁₁ to *Sasa* class II alleles is intermediate between the range of values determined for *ideal* and *poor*-binding peptides to human HLA-DRβ1*0101 and HLA-DRβ1*0301. This trend likely reflects the fact that in this case the consensus peptides were derived by using a smaller number of invariant chains in the sequence alignment, and to the larger uncertainty associated with the precise definition of the CLIP region in the MHC class II-associated invariant chain in fishes. Nevertheless, the binding affinity predicted for P₁MKMHMPMLNM₁₁ to *Sasa* DAA*0501-DAB*0301 (DA5B3_1) compares with the binding of K₁YVKQNTLKLA₁₁ to human HLA-DRβ1*0101 (1DLH), which also lends support to the structural model built up for the *Sasa* allele.

Final remarks

Understanding of the molecular basis of peptide recognition by MHC molecules is necessary to enhance our ability to design novel drugs and peptide-based vaccines that might be used as immunospecific therapeutics. In the case of fishes, this information could be extremely valuable to develop vaccines against infectious diseases, which will have a relevant economical impact for fish farms. The lack of atomistic 3D structural data at the microscopic level represents a serious limitation, which can nevertheless be attenuated by the use of bioinformatic tools for the development of structural models of peptide-MHC complexes. In this context, combination of homology techniques and molecular dynamics simulations has allowed us to build up the first complete 3D structural models for two MHC class I (UBA*0201, UBA*0301) and class II (DAA*0201-DAB*0201, DAA*0501-DAB*0301) alleles from *Salmo salar* (the energy-minimized structures of the starting peptide-bound homology models are given as Supporting Information; additional data available upon request to the authors).

The model relies on the hypothesis that these molecules are well preserved through evolution, which thus justifies the exploration of structure–function properties by resorting to suitable reference MHC proteins. The global structural fold of the *Sasa* MHC proteins is similar to that found in the reference proteins (murine H-2Kb and human HLA-DR β 1*0101), which supports the structural integrity of the models. However, the results also reveal the existence of a sizable number of differences in the nature and distribution of the residues that delineate the peptide-binding groove between the template structures and the *Sasa* alleles. While those differences should be reflected in distinct trends regarding the preferred motifs for peptide recognition between the template structures and the *Sasa* alleles, they make it necessary to allow for structural relaxation of the contacts between the bound peptide and the MHC residues in the binding groove, which is performed here by extensive molecular dynamics simulations.

The suitability of this computational strategy is supported by the structural and energetic stability predicted for the binding of peptides F₁SGYNATAL₉ and M₁PSQLYTLM₉ to the class I *Sasa* UBA*0301, which were chosen as representative peptides based on the sequences of a large number of 7-mer and 12-mer peptides experimentally identified as binders to the UBA*0301 allele. Our studies also reveal a larger structural instability for the recognition of peptides P₁MKMHMPMLNM₁₁ and H₁LPMLNMLL₁₁ to *Sasa* class II alleles, which can be associated with specific mutations compared to the human HLA-DR β 1*0101 structure, such as the replacement of α R74 by Y in both DAA*0201-DAB*0201 and DAA*

0501-DAB*0301, and of β G86 by Y in DAA*0201-DAB*0201 and by D in DAA*0501-DAB*0301. Such instability is not completely unexpected, as it can be realized from the limited number of invariant chains considered in the sequence alignment and to the lack of a detailed knowledge of the CLIP region in the invariant chain in fishes.

The structural models presented here represent a suitable starting point for the study of pathogen-specific interactions in *Salmo salar* MHC molecules, which could provide useful guidelines to gain insight into the interactions between fish MHC receptors and antigenic peptides. Finally, the results also support the suitability of the computational strategy presented here to explore the binding preferences of antigenic peptides by MHC molecules in other higher organisms.

Acknowledgments We kindly acknowledge one of the reviewers for stimulating discussions. Financial support from the Spanish Ministerio de Ciencia e Innovación (SAF2008-05595), Generalitat de Catalunya (2009-SGR00298), and Chilean grant 07CN13PBT165 INNOVA-CORFO, and computational facilities from the Barcelona Supercomputer Center are also acknowledged. CC was partially supported by the Bicentennial Program PSD-08 (CONICYT/Chile).

References

- Vyas JM, Van der Veen AG, Ploegh HL (2008) Nat Rev Immunol 8:607–618
- Ruppert J, Sidney J, Celis E, Kubo RT, Grey HM, Sette A (1993) Cell 74:929–937
- NCBI (<http://www.ncbi.nlm.nih.gov/gv/mhc/>), Inc., 2009
- Ellis SA, Bontrop RE, Antczak DF, Ballingall K, Davies CJ, Kaufman J, Kennedy LJ, Robinson J, Smith DM, Stear MJ, Stet RJ, Waller MJ, Walter L, Marsh SG (2006) Immunogenetics 57:953–958
- Robinson J, Waller MJ, Fail SC, McWilliam H, Lopez R, Parham P, Marsh SG (2009) Nucleic Acids Res 37:D1013–D1017
- Peters B, Bui H-H, Frankild S, Nielsen M, Lundegaard C, Kostem E, Basch D, Lamberth K, Harndahl M, Fleri W, Wilson SS, Sidney J, Lund O, Buus S, Sette A (2006) PLOS Comput Biol 2:574–584
- Hoof I, Peters B, Sidney J, Pedersen LE, Sette A, Lund O, Buus S, Nielsen M (2009) Immunogenetics 61:1–13
- Rapin N, Hoof I, Lund O, Nielsen M (2008) Immunogenetics 60:759–765
- Reche PA, Glutting JP, Reinherz EL (2002) Hum Immunol 63:701–709
- Tsurui H, Takahashi T (2007) J Pharmacol Sci 105:299–316
- Nielsen M, Lundegaard C, Lund O (2007) BMC Bioinformatics 8:238
- Singh H, Raghava GP (2001) Bioinformatics 17:1236–1237
- Lundegaard C, Lamberth K, Harndahl M, Buus S, Lund O, Nielsen M (2008) Nucleic Acids Res 36:W509–W512
- Tenzer S, Peters B, Bulik S, Schoor O, Lemmel C, Schatz MM, Kloetzel PM, Rammensee HG, Schild H, Holzhtter HG (2005) Cell Mol Life Sci 62:1025–1037
- Dixon B, van Erp SHM, Rodrigues PNS, Egberts E, Stet RJ (1995) Dev Comp Immunol 19:109–133
- Harstad H, Lukacs MF, Bakke HG, Grimholt U (2008) BMC Genomics 9:193

17. Kjolglum S, Larsen S, Bakke HG, Grimholt U (2006) Fish Shellfish Immunol 21:431–441
18. Jorgensen SM, Syvertsen BL, Lukacs M, Grimholt U, Gjoen T (2006) Fish Shellfish Immunol 21:548–560
19. Rakus KL, Wiegertjes GF, Adamek M, Siwicki AK, Lepa A, Irnazarow I (2009) Fish Shellfish Immunol 26:737–743
20. Kjolglum S, Larsen S, Bakke HG, Grimholt U (2008) Scand J Immunol 67:160–168
21. Wynne JW, Cook MT, Nowak BF, Elliott NG (2007) Fish Shellfish Immunol 22:707–717
22. Xu JY, Chen SL, Ding H (2009) J Fish Dis 32:637–640
23. Fishpathogens database (2009) European Community Reference Laboratory for Fish Diseases (<http://www.crl-fish.eu/>)
24. Stet RJ, Dixon B, Van Erp SHM, Van Lierop MJC, Rodrigues PNS, Egberts E (1996) Fish Shellfish Immunol 6:305–318
25. Grimholt U, Drablos F, Jorgensen SM, Hoyheim B, Stet RJ (2002) Immunogenetics 54:570–581
26. Sato A, Figueroa F, Murray BW, Malaga-Trillo E, Zaleska-Rutczynska Z, Sultmann H, Toyosawa S, Wedekind C, Steck N, Klein J (2000) Immunogenetics 51:108–116
27. Phillips RB, Zimmerman A, Noakes MA, Palti Y, Morasch MR, Eiben L, Ristow SS, Thorgaard GH, Hansen JD (2003) Immunogenetics 55:561–569
28. Zhao H, Stet RJ, Skjodt K, Savelkoul HF (2008) Fish Shellfish Immunol 24:459–466
29. Stet RJ, Kruiswijk CP, Dixon B (2003) Crit Rev Immunol 23:441–471
30. Miller KM, Li S, Ming TJ, Kaukinen KH, Schulze AD (2006) Immunogenetics 58:571–589
31. Ruiz I, Fernández AB, de Blas I (2003) Aquatic 18:33–38
32. Elliott T, Williams A (2005) Immunol Rev 207:89–99
33. Leng Q, Bentwich Z (2002) Scand J Immunol 56:224–232
34. Bryant P, Ploegh H (2004) Curr Opin Immunol 16:96–102
35. Rocha N, Neeffjes J (2008) EMBO J 27:1–5
36. Buonocore F, Randelli E, Casani D, Costantini S, Facchiano A, Scapigliati G, Stet RJ (2007) Fish Shellfish Immunol 23:853–866
37. Costantini S, Buonocore F, Facchiano AM (2008) Fish Shellfish Immunol 25:782–790
38. Hao HF, Li XS, Gao FS, Wu WX, Xia C (2007) Protein Expr Purif 51:120–125
39. Stevens J, Wiesmuller KH, Walden P, Joly E (1998) Eur J Immunol 28:1272–1279
40. Zhao H, Hermsen T, Stet RJ, Skjodt K, Savelkoul HF (2008) Mol Immunol 45:1658–1664
41. Fujiki K, Smith CM, Liu L, Sundick RS, Dixon B (2003) Dev Comp Immunol 27:377–391
42. Bremnes B, Rode M, Gedde-Dahl M, Nordeng TW, Jacobsen J, Ness SA, Bakke O (2000) Exp Cell Res 259:360–369
43. Rudolph MG, Speir JA, Brunmark A, Mattsson N, Jackson MR, Peterson PA, Teyton L, Wilson IA (2001) Immunity 14:231–242
44. Stern LJ, Brown JH, Jardetzky TS, Gorga JC, Urban RG, Strominger JL, Wiley DC (1994) Nature 368:215–221
45. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Bioinformatics 23:2947–2948
46. Arnold K, Bordoli L, Kopp J, Schwede T (2006) Bioinformatics 22:195–201
47. Dunbrack RL Jr, Cohen FE (1997) Protein Sci 6:1661–1681
48. Madden DR, Garboczi DN, Wiley DC (1993) Cell 75:693–708
49. Knapp B, Omasits U, Schreiner W (2008) Protein Sci 17:977–982
50. Knapp B, Omasits U, Frantal S, Schreiner W (2009) J Comput Aided Mol Des 23:301–307
51. Rognan D, Stryhn A, Fugger L, Lyngbaek S, Engberg J, Andersen PS, Buus S (2000) J Comput Aided Mol Des 14:53–69
52. Rammensee HG, Bachmann J, Emmerich NN, Bachor OA, Stevanovic S (1999) Immunogenetics 50:213–219
53. Case DA, Darden TA, Cheatham I TE, Simmerling CL, Wang J, Duke RE, Luo R, Merz KM, Pearlman DA, Crowley M, Walker RC, Zhang W, Wang B, Hayik S, Roitberg A, Seabra G, Wong KF, Paesani F, Wu X, Brozell S, Tsui V, Gohlke H, Yang L, Tan C, Mongan J, Hornak V, Cui G, Beroza P, Mathews DH, Schafmeister C, Ross WS, Kollman PA (2006) AMBER 9. University of California, San Francisco
54. Li H, Robertson AD, Jensen JH (2005) Proteins 61:704–721
55. Schutz CN, Warshel A (2001) Proteins 44:400–417
56. Swanson JMJ, Adcock SA, McCammon JA (2005) J Chem Theory Comput 1:484–493
57. Sitkoff D, Sharp KA, Honig B (1994) J Phys Chem 98:1978–1988
58. Froloff N, Windemuth A, Honig B (1997) Protein Sci 6:1293–1301
59. Pickett SD, Stenberg MJE (1993) J Mol Biol 231:825–839
60. Doig AJ, Stenberg MJE (1995) Prot Sci 4:2247–2251
61. Chang CE, Chen W, Gilson MK (2007) Proc Natl Acad Sci USA 104:1534–1539
62. Chen W, Chang CE, Gilson MK (2004) Biophys J 87:3035–3049
63. Schneider TD, Stephens RM (1990) Nucleic Acids Res 18:6097–6100
64. Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) Genome Res 14:1188–1190
65. Kabsch W, Sander C (1983) Biopolymers 22:2577–2637
66. Zhou H, Zhou Y (2002) Protein Sci 11:2714–2726
67. Benkert P, Tosatto SC, Schomburg D (2008) Proteins 71:261–277
68. Madden DR (1995) Annu Rev Immunol 13:587–622
69. Cardenas C, Villaveces JL, Bohorquez H, Llanos E, Suarez C, Obregon M, Patarroyo ME (2004) Biochem Biophys Res Commun 323:1265–1277
70. Cardenas C, Villaveces JL, Suarez C, Obregon M, Ortiz M, Patarroyo ME (2005) J Struct Biol 149:38–52
71. Cardenas C, Ortiz M, Balbin A, Villaveces JL, Patarroyo ME (2005) Biochem Biophys Res Commun 330:1162–1167
72. Cardenas C, Obregon M, Balbin A, Villaveces JL, Patarroyo ME (2007) J Mol Graph Model 25:605–615
73. Wang J, Morin P, Wang W, Kollman PA (2001) J Am Chem Soc 123:5221–5230
74. Lyne PD, Lamb ML, Saeh JC (2006) J Med Chem 49:4805–4808
75. Stoica I, Sadiq SK, Coveney PV (2008) J Am Chem Soc 130:2639–2648
76. Schafroth HD, Floudas CA (2004) Proteins 54:534–556
77. Vita R, Zarebski L, Greenbaum JA, Emami H, Hoof I, Salimi N, Damle R, Sette A, Peters B (2010) Nucleic Acids Res 38:D854–D862