HAL
open science

Analysis of Key Parameters for Molecular Dynamics of pMHC Molecules

Ulrich Omasits, Bernhard Knapp, Martin Neumann, Othmar Steinhauser, Hannes Stockinger, Rene Kobler, Wolfgang Schreiner

To cite this version:

Ulrich Omasits, Bernhard Knapp, Martin Neumann, Othmar Steinhauser, Hannes Stockinger, et al.. Analysis of Key Parameters for Molecular Dynamics of pMHC Molecules. Molecular Simulation, 2008, 34 (08), pp.781-793. 10.1080/08927020802256298 . hal-00515047

HAL Id: hal-00515047

https://hal.science/hal-00515047v1

Submitted on 4 Sep 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

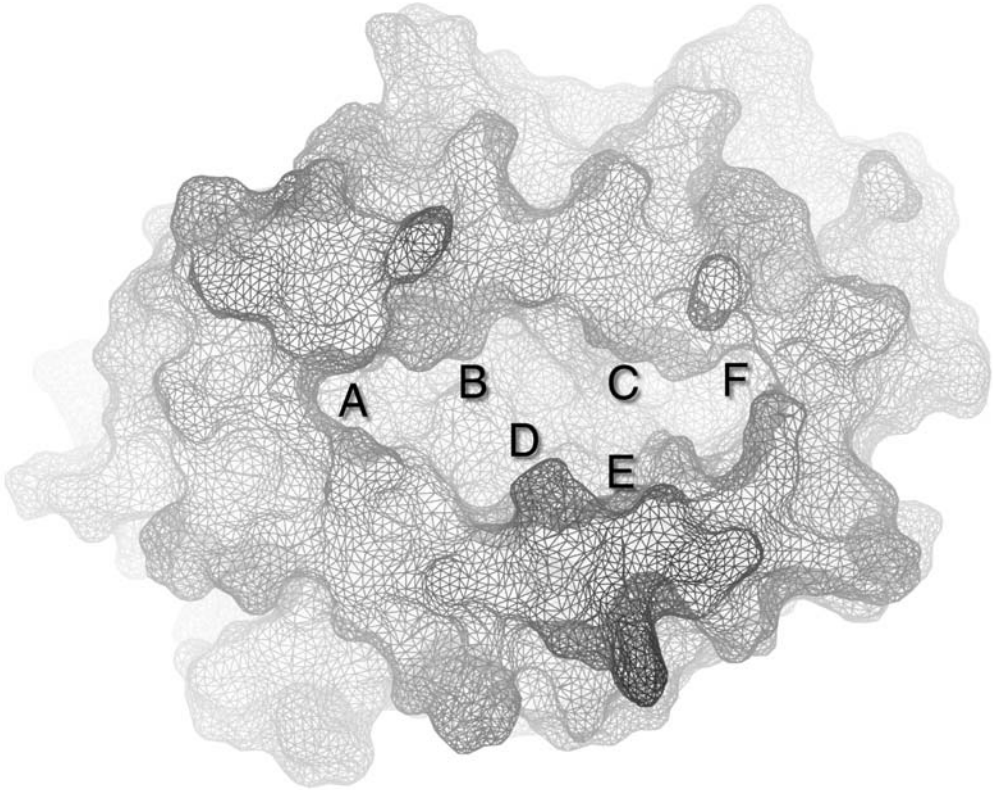
L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Molecular Simulation
Journal of
Experimental Nanoscience
Taylor & Francis
Taylor & Francis Group

Analysis of Key Parameters for Molecular Dynamics of pMHC Molecules

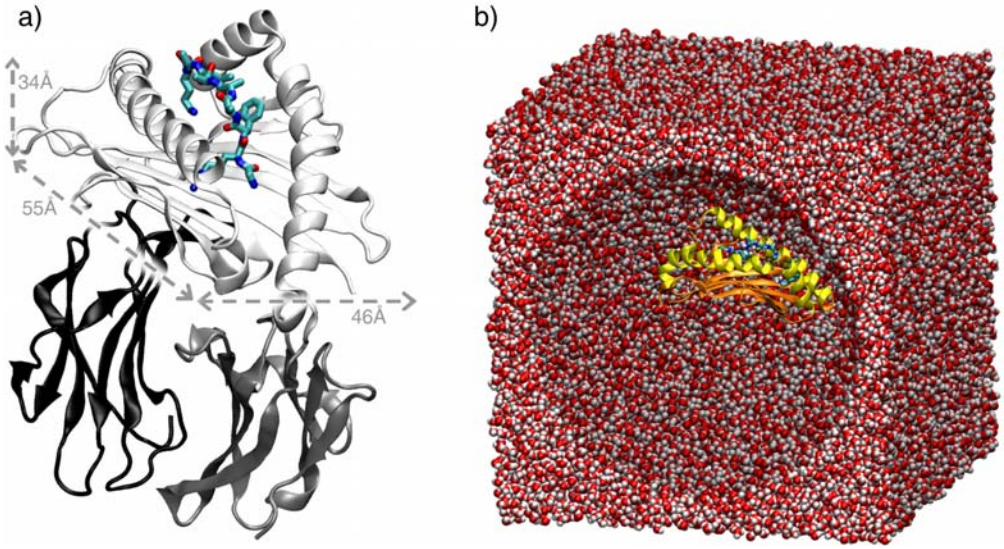
Journal:	*Molecular Simulation/Journal of Experimental Nanoscience*
Manuscript ID:	GMOS-2008-0089.R1
Journal:	Molecular Simulation
Date Submitted by the Author:	06-Jun-2008
Complete List of Authors:	Omasits, Ulrich; Medical University of Vienna, Core Unit for Medical Statistics and Informatics Knapp, Bernhard; Medical University of Vienna, Core Unit for Medical Statistics and Informatics Neumann, Martin; University of Vienna, Institute of Experimental Physics Steinhauser, Othmar; University of Vienna, Department of Computational Biological Chemistry Stockinger, Hannes; Medical University of Vienna, Department of Molecular Immunology Kobler, Rene; Johannes Kepler University Linz, Institute of Graphics and Parallel Processing Schreiner, Wolfgang; Medical University of Vienna, Core Unit for Medical Statistics and Informatics
Keywords:	molecular dynamics, peptide dynamics, conformational substates, major histocompatibility complex, T cell activation
Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.	
sm1.avi	

SCHOLARONE™
Manuscripts



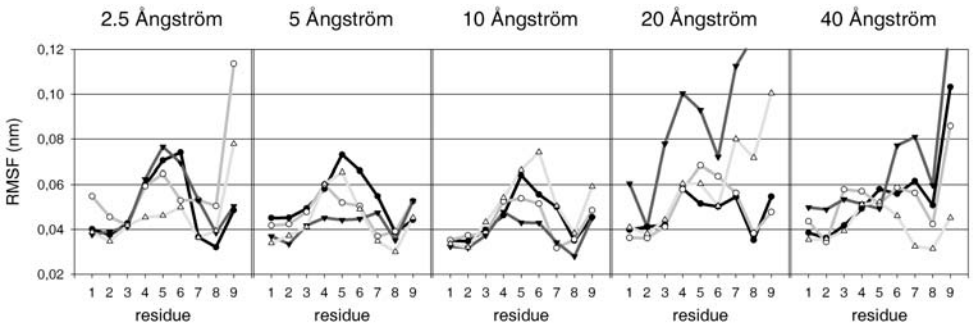
HLA-B*2705 molecular surface and its binding pockets.
82x66mm (600 x 600 DPI)

ew Only



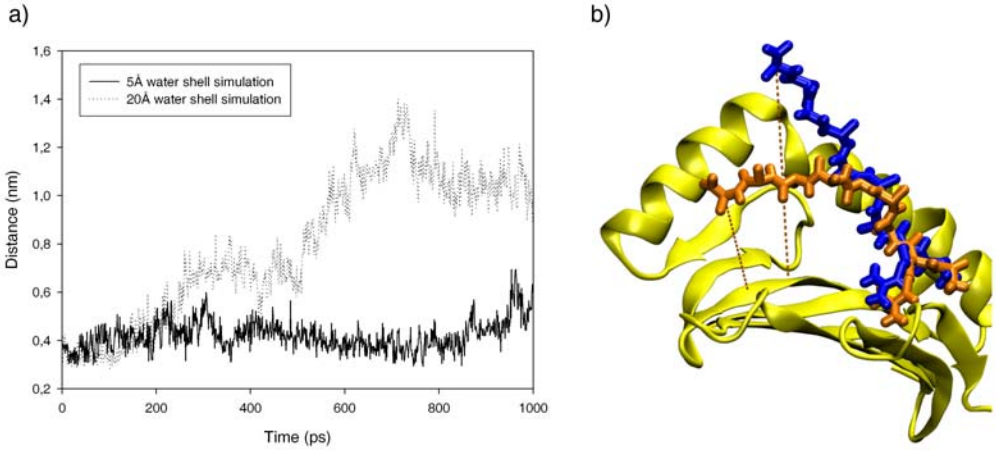
Crystal structure and simulated system.
177x97mm (300 x 300 DPI)

Review Only



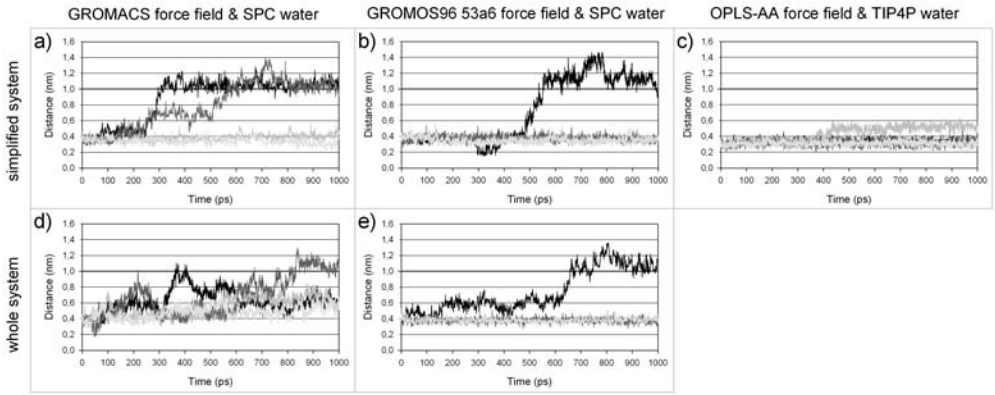
RMSF of peptide residues' backbone atoms.
177x57mm (600 x 600 DPI)

Peer Review Only



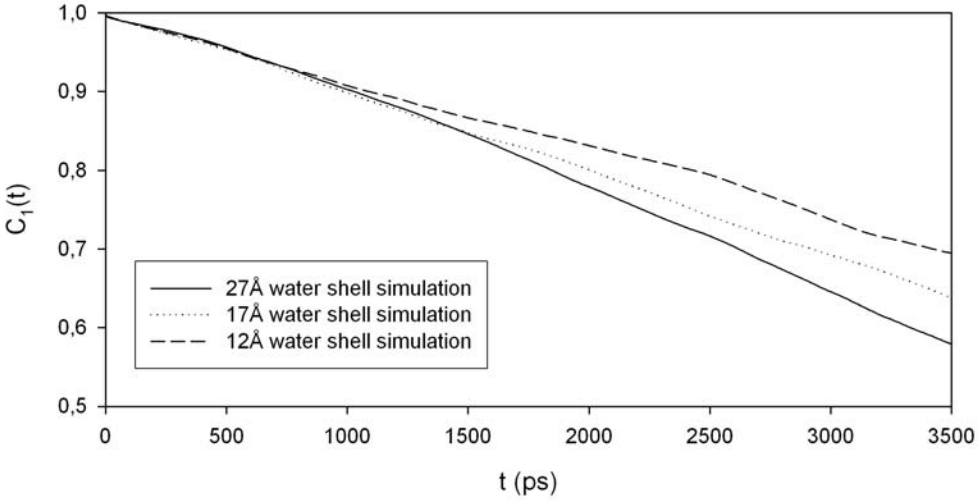
Demonstration of the minimum distance between epitope C-terminus and MHC β-sheet.
177x80mm (600 x 600 DPI)

er Review Only



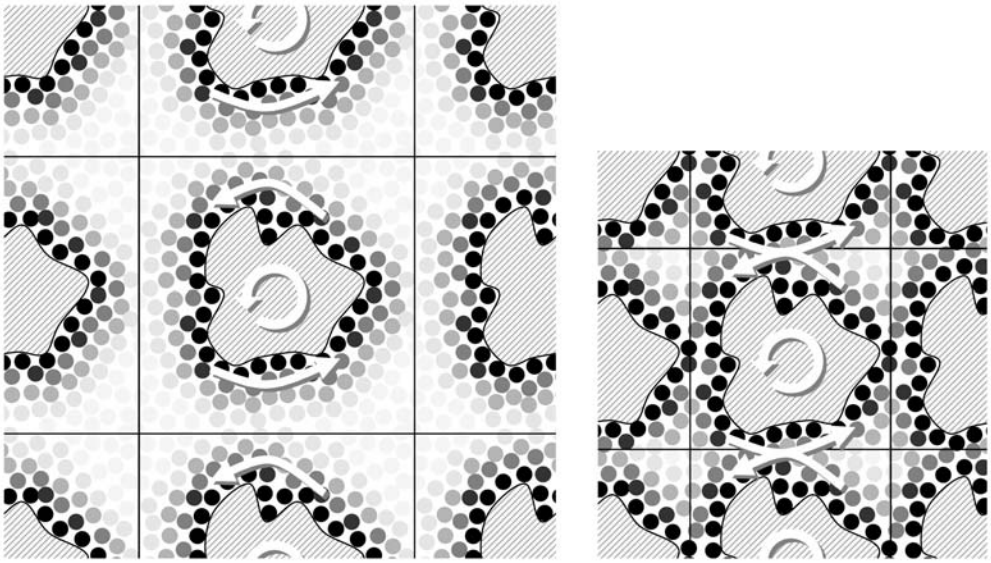
Minimum distance between peptide C-terminus and MHC β-sheet during different simulations of the "single site binding" peptide.
177x75mm (600 x 600 DPI)

er Review Only



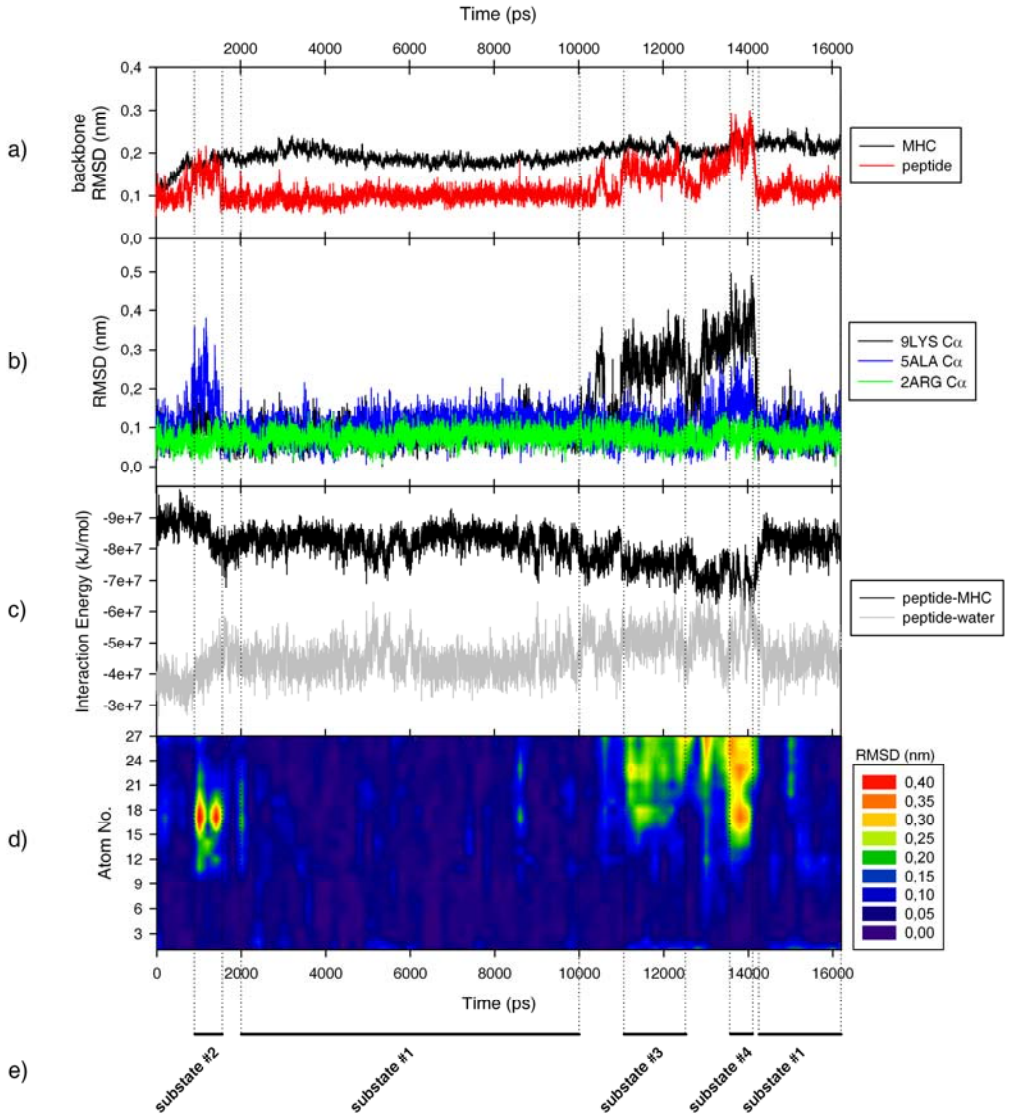
Orientational relaxation (RACF) of the pMHC complex for different sizes of the simulation box.
82x41mm (600 x 600 DPI)

Review Only

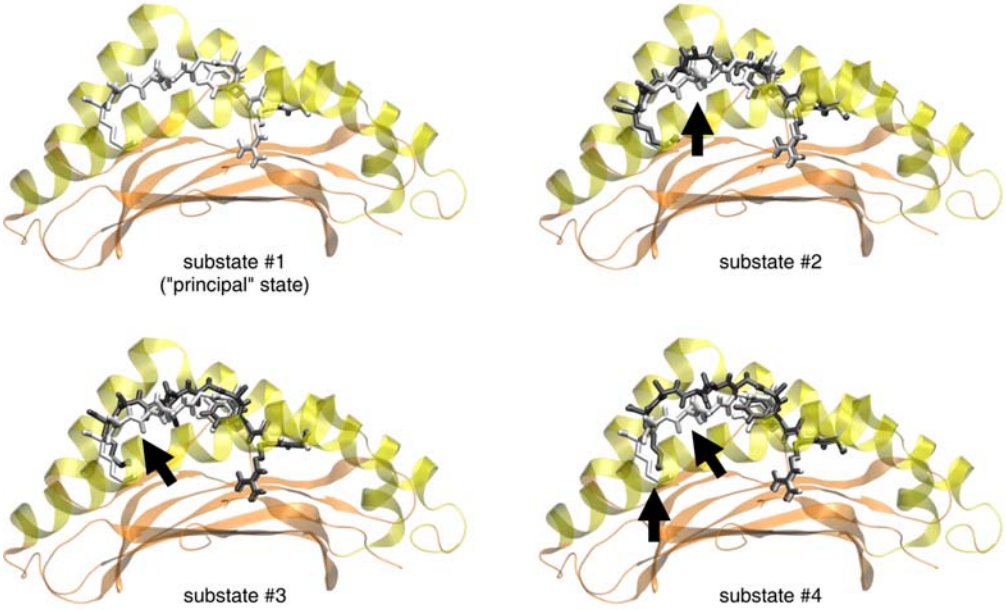


Model for system size influencing rotational diffusion.
156x89mm (600 x 600 DPI)

Review Only

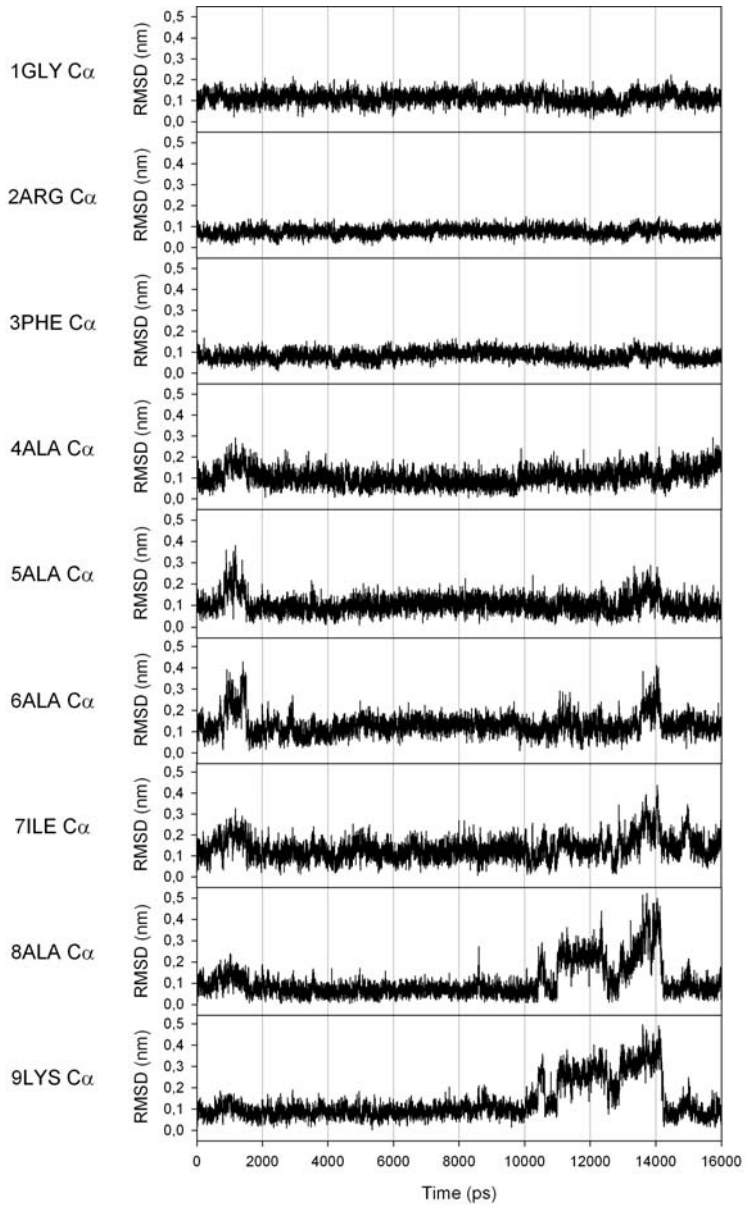


Analysis of the long time simulation (16ns) of the "double site binding" peptide complexed to HLA-B*2705 (1jge) within a 27Å water box.
177x198mm (600 x 600 DPI)



Average structures of peptide's conformational substates.
177x109mm (300 x 300 DPI)

For Review Only



Individual RMSD curves for all the peptide's Cα atoms.
176x286mm (300 x 300 DPI)

Analysis of Key Parameters for Molecular Dynamics of pMHC Molecules

Ulrich Omasits[a], Bernhard Knapp[a], Martin Neumann[b], Othmar Steinhauser[c], Hannes Stockinger[d], Rene Kobler[e], and Wolfgang Schreiner[a]*

The work described in this paper is supported by the Austrian Grid Project, funded by the Austrian BMBWK (Federal Ministry for Education, Science and Culture) under contract GZ 4003/2-VI/4c/2004.

[a] Core Unit for Medical Statistics and Informatics, Medical University of Vienna, Austria

[b] Institute of Experimental Physics, University of Vienna, Austria

[c] Department of Computational Biological Chemistry, University of Vienna, Austria

[d] Department of Molecular Immunology, Centre of Biomolecular Medicine and Pharmacology, Medical University of Vienna, Austria

[e] Institute of Graphics and Parallel Processing, Johannes Kepler University Linz, Austria

[*] Corresponding author: fax: +43 1 40400-6677 – phone: +43 1 40400-6673– email: bernhard.knapp@meduniwien.ac.at – address: Spitalgasse 23, 1090 Vienna, Austria

For Peer Review Only

Abstract

Molecular dynamics (MD) studies of human major histocompatibility complex (MHC) HLA-B*2705 complexing two different peptides were performed. During simulation one peptide partially detached from the MHC while the other peptide switched back and forth between several different configurations. These different configurations relate to conformational substates and can be assigned to different levels of chemical activity or even the molecular mechanisms of immunological signalling. To ensure reliable immunological conclusions from MD simulations we prepare the methodological tools by carefully evaluating initial conditions, system simplification, solvation shell thickness, water model / force field combination, and simulation length. We also derive a guideline for appropriate model selection. This kind of quality assessment is seen a mandatory prerequisite for coming studies linking pMHC dynamics to T cell activation.

Keywords: HLA; major histocompatibility complex; peptide dynamics; conformational substates; T cell activation

Abbreviations: β_2m: β_2-microglobulin – HLA: human leukocyte antigen – MD: molecular dynamics – MHC: major histocompatibility complex – PBC: periodic boundary conditions – PME: Particle Mesh Ewald – pMHC: peptide loaded MHC – RACF: rotational autocorrelation function – RMSD: root mean square deviation – RMSF: root mean square fluctuation – SASA: solvent accessible surface area – TCR: T cell receptor

For Peer Review Only

1 Introduction

Class I major histocompatibility complexes (MHC) bind antigenic peptides within the α1-α2 domain in a long narrow cleft located between two α-helices on the top of an antiparallel β-sheet. Binding is mainly mediated through sequence independent contacts of the peptide's charged N- and C-terminus to conserved "pockets" (Figure 1) in the MHC's binding groove. The contacts are maintained by networks of hydrogen bonds that are very similar in all class I pMHC complexes and make the largest contribution to the binding energy of the peptide. In addition, polymorphic pockets, so-called primary anchor regions, interact with specific sequence dependent amino acids (anchor residues) of the peptide and thus are responsible for the allele specificity of the recognition process. The central part of the peptide bulges out of the binding groove and serves as primary recognition site for the T cell receptor (TCR) (see Figure 2a).

Deleted: Figure 1

Deleted: Figure 2

Molecular dynamics (MD) is the technique of numerically solving the equations of motion of an assembly of particles (atoms). The total force acting on each particle depends on the positions and properties of all other particles in the system and, in addition, may include externally imposed forces. MD programs calculate these forces and solve Newton's equation of motion for each particle. For details refer to reviews on MD of biomolecules [1;2].

Various MD studies on pMHC have been reported in the literature [3-16]. All these MD studies use some force field, some water model, some initial condition, and some size of water box, in most cases without a particular justification or comment on their choices. The aim of this study is to vary these parameters and analyze their effects and possible pitfalls regarding a pMHC complex simulation. To our knowledge no such detailed investigation has been reported in the literature. We are going to answer the following questions:

For Peer Review Only

- Should parallel simulations be performed using different initial conditions?
- What size of solvation box should be used?
- Which force field and water model should be chosen?
- What is the minimum simulation time necessary?
- Are system simplifications admissible?

Our aim is to evaluate those key parameters giving a reasonable chance to cope with the rather large pMHC molecule in a meaningful way. All in all, the present work lays the basis for MD as a method of investigation in pMHC interactions.

Besides the technical evaluation we also present a minute mobility analysis of a peptide adopting different configurational substates within the MHC-cleft.

1.1 Environment representation in MD

In order to achieve a manageable representation of the natural pMHC complex a trade-off between computational cost and simulation accuracy has to be made, especially concerning the amount of solvation water included in the simulation, the boundary conditions, and protein simplifications. Since proteins evolved to function in aqueous environments, water molecules binding directly to the protein and forming the first layer of the hydration shell may be considered as an integral part of protein structure. It is also clear that solute-solvent interactions extend beyond the first hydration layer. Although an extended hydration shell may be represented by simplified models, the most reliable results on structure and dynamics are achieved by inclusion of a sufficient number of additional explicit solvent molecules. To avoid an abrupt border with a vacuum, periodic boundary conditions (PBC) repeat the simulation box periodically in all three space dimensions, creating a continuous system. PBC allows a very useful method for calculating long-range electrostatics: the Particle Mesh Ewald (PME) summation [17]. Currently PBC/PME represents the standard in MD simulations of

For Peer Review Only

biomolecules since it is the most accurate method. On the other hand, the imposition of periodicity may lead to artefacts since all atoms now may interact with their periodic images. These artefacts can be minimized using larger solvation shells and an appropriate water model / force field combination, i.e. a combination to stay on the "safe side" for a particular simulation experiment (see Discussion section).

1.1.1 Size of solvation shell

A minimal system is claimed to need a 5-15Å thick solvation shell [18]. In other words, the system needs to be large enough to allow for the formation of a 5-15Å "buffer zone" between the solute and the boundaries of the simulation box. We show that for pMHC simulations a solvation layer thickness of more than 10Å, ideally 20Å, is necessary (Figure 2b). It is known that the aqueous environment facilitates the protein sampling the conformational space [19]. Water, due to its high relative dielectric permittivity, is very effective in screening long-range electrostatic interactions and dipoles. Thus the thickness of the solvation shell essentially determines the degree of electrostatic shielding, which is seen crucial in the presence of significant dipoles induced by the α-helices of the MHC [20;21]. PBC/PME imposes an artificial periodicity on the grossly inhomogeneous solute-solvent system. This periodicity limits the configurational flexibility of a protein [22] by overstabilization of the starting configuration which is normally taken to be that of a periodic crystal. Obviously, the larger the simulation box, the better the water molecules in the solvation shell (and the protein itself) will be screened from interactions with the protein's periodic images. Thus crystalline effects are minimized, sampling is less hindered [23], and a more realistic dynamics of the system can be observed.

Deleted: Figure 2

1.1.2 Structural simplifications

It is quite common to simplify a simulation system in order to minimize the number of atoms. In the case of pMHC complexes frequently only the MHC's peptide binding domain (α1 and α2 domain) is simulated since this is the site of interesting peptide-MHC interactions. In vivo

For Peer Review Only

such a truncated pMHC complex would not be stable at all and unfolding would occur. MHC simplification has been found not to alter the results of MD simulations on the timescale feasible [24]. However, there is disagreement about this simplification and there are authors [25] claiming that the α3 domain and the β_2m subunit are essential for peptide MHC interactions even within one nanosecond simulation time. According to our results, simplification plays a minor role. The α3 domain and the β_2m subunit are certainly essential for a simulation of a system in its full complexity. However, we are interested in processes taking place within the binding groove at a much shorter timescale than MHC unfolding would occur.

For Peer Review Only

2 Materials and Methods

2.1 Hard- and Software

The project was carried out on a SGI Altix 350 system within the framework of the Austrian Grid [26] and on a computational cluster of the Technical University of Vienna. For a typical 2ns simulation run a total of 800 CPU hours are required.

We use GROMACS [27;28] because of its detailed documentation of computational methodology and very fast procedure to treat water molecules. Its high scalar performance outweighed the fact that (on our system) GROMACS scales efficiently only up to 16 nodes.

2.2 Setup of coordinates

For simulations we selected the MHC class I B*2705 protein. Almost all binding epitopes for this MHC show an arginine at position 2, defined as a dominant anchor residue. Positions 1, 3, and 9 provide the remaining main anchoring positions with a preference for hydrophobic amino acids at P3, and hydrophobic or positively charged amino acids at P9. The dominant anchor residue P2 binds deep into the MHC pocket B, while P9 binds into MHC pocket F (see Figure 1). HLA-B27 subtypes are strongly associated with the occurrence of several diseases, in particular inflammatory diseases of the joints called spondyloarthropathies [29] and thus lends itself as a most interesting candidate for computational screening.

Deleted: Figure 1

Starting coordinates were taken from crystal structures deposited in the Protein Data Bank [30]. We chose two structures representing the same MHC complexing two different, but substantially similar peptides. The first structure was obtained by Madden et al. from an epitope mix. As a “minimal binding” peptide the consensus backbone was modelled with the epitope ARAAAAAAA. The resulting structure was deposited as entry 1hsa [31]. The second structure containing the epitope GRFAAAIAK was obtained by Hülsmeyer et al. and deposited as entry 1jge [32]. We chose those two complexes due to their chemically similar

For Peer Review Only

epitopes. Both consist of mainly hydrophobic residues (A,F,I,G) and both have the R anchor at position 2, typical for HLA-B*2705 epitopes. The crucial difference is the secondary anchor (K) in 1jge at position 9 which is missing in 1hsa. Therefore we will refer to the epitope of 1jge as "double site binding" peptide and to the epitope of 1hsa as "single site binding" peptide.

For most simulations only the antigen binding α1 and α2 domains were considered. The C-terminal end of the α2 domain was protonated to render it uncharged while all the other termini and side chains were assigned typical charge states at pH 7. Water molecules observed in the crystal structure were explicitly taken into account in the simulations. Polar hydrogens were added and the complexes were centred in cubic boxes of different sizes, to allow for water shells ranging from 2.5Å to more than 40Å thickness between the solute and the boundaries of the simulation box. The created space around the solute was initially filled with an equilibrium configuration of bulk water. The water models used were SPC and TIP4P, respectively, both recommended for usage in biomolecular systems [33]. To electrically neutralize the system, an appropriate number of Na^+ counterions were added.

2.3 Simulation methods and parameters

In order to investigate the influence of the force field and water model on our results we selected three different parametrizations from the suite of force fields available with GROMACS. The GROMACS force field (a modified GROMOS87 force field) [34] and the GROMOS96 53a6 force field [35] were used for SPC solvated pMHC complexes while the OPLS-AA/L all-atom force field (version 2001) [36] was used for TIP4P solvated pMHC complexes, as recommended in GROMACS documentation [34]. After the initial coordinate setup, as described above, the system's potential energy was first minimized by a steepest descent procedure, in order to allow the water molecules to adjust to the presence of the pMHC complex. The coordinates so obtained served as a starting structure for the subsequent

For Peer Review Only

MD simulations. Periodic boundaries were employed and the system was then heated linearly from 0K to 300K during the first 40ps and left at 300K for the last 10ps using a Berendsen thermostat with a temperature coupling constant of 0.1 ps [37]. For this warm-up phase only, position restraints (force constant of 1000 kJ mol^{-1} nm^{-2}) were applied to each atom of the pMHC complex. Subsequent production runs without any position restraints used Berendsen pressure coupling (reference pressure: 1 bar, pressure coupling constant: 0.5 ps) [37], the Particle Mesh Ewald (PME) [17] method for long-range electrostatic interactions, and an integration step of 2fs. All bond lengths were constrained to their equilibrium values using the SHAKE algorithm [38]. Simulations started with random velocities distributed to the atoms (i.e. initial conditions) and total simulation times between 1ns and 16ns were used. For subsequent analysis, configurations were saved in trajectory files every 2ps for long runs (16ns total) and every 0.5ps for short runs (1ns total). An overview of all performed simulation runs can be found in Table 1.

Deleted: Table 1

2.4 *Methods of analysis*

All simulation runs were evaluated via the following analysis modules provided by the GROMACS package:

2.4.1 SASA

The solvent accessible surface area (SASA) was computed using the Double Cube Lattice Method [39]. The radius of the solvent probe was set to 1.4 Ångström.

2.4.2 RMSD

The root mean square deviation (RMSD) gives the average deviation of a structure relative to a reference structure (e.g. crystal structure). In order to subtract global motions each structure is least squares fitted to the reference structure before RMSD calculation.

2.4.3 RMSF

The root mean square fluctuations (RMSF) represent the standard deviation of atomic positions during (a part of) a trajectory after first fitting to a reference structure. The RMSF of

For Peer Review Only

the crystal structure is calculated out of the crystallographic B-factors [40]. Contact residues have systematically low RMSF values while surface side chains protruding from the protein show large RMSF values.

2.4.4 Rotational autocorrelation function (RACF)

Given a unit vector $\vec{n}(t)$ characterizing a key direction of a molecule, the RACF is computed via

$$C_1(t) = \langle P_1(\vec{n}(0) \cdot \vec{n}(t)) \rangle \qquad \text{(Equation 1)}$$

with t being the time lag and $P_1(x) = x$ the first order Legendre polynomial. Values of 1 indicate "perfect memory" of orientation, a value of 0 indicates loss of orientational correlation.

For Peer Review Only

3 Results

Table 1 summarizes the simulations performed in order to analyze the key parameters of MD such as size of water box, initial conditions, system simplification, simulation time, and water model / force field combination. In particular, the following results were obtained:

Deleted: Table 1

3.1 "Single site binding" peptide

In order to point out the possible impacts of simplifications and limitations, we chose the rather labile system of the "single site binding" peptide (ARAAAAAAA) complexed to B*2705 MHC (PDB-id: 1hsa). This complex features an artificially modelled epitope with only one anchor residue (arginine at position 2).

3.1.1 Solvation shell thickness & initial conditions

We performed several MD simulations of the "single site binding" peptide loaded MHC with different initial conditions (randomly distributed starting velocities) and with solvation shell thicknesses ranging from 2.5Å to 40Å and computed the RMSF, see Figure 3. The RMSF values for the peptide's C-terminal alanine (residue 9), show the most pronounced variability with box size: Using the 5Å and the 10Å water shell box, the RMSF values are low. Usage of the smallest box (2.5Å shell) results in larger fluctuations. This may be due to direct interactions – via the periodic boundary conditions – between the epitope and the "bottom" (i.e. the part diametrically opposed to the binding cleft) of the simplified MHC. On the other hand, the RMSF graphs for the systems with a 20Å or a 40Å box show extremely high fluctuations of the peptide's C-terminal, alanine (residue 9). This indicates an interesting phenomenon: Simulating this rather artificial system revealed a detachment of the epitope's C-terminal end (a video of this simulation can be found in supplementary material S1). This finding strongly indicates that this artificial "single site binding" peptide is in fact not able to bind to the MHC. This is not surprising since there is no anchor residue at the C-terminal end.

Deleted: Figure 3

For Peer Review Only

The salt bridges between the epitope's C-terminus and the MHC's F-pocket are just not stable enough to retain the hydrophobic epitope completely inside the hydrophilic binding groove. The gradual detachment of the epitope can be conveniently visualized by plotting the time dependence of the minimum distance between all the atoms of the peptide's C-terminal alanine and all the atoms of the MHC's β-sheet, which forms the bottom of the binding cleft (Figure 4).

Deleted: Figure 4

Such a labile system is very sensible to different initial conditions, in particular if one uses too small water shell sizes. For each water shell thickness we performed four parallel simulations which were absolute identical, except for the randomly distributed initial velocities. These parallel simulations evolved very differently, as shown by the RMSF graph (Figure 3). For the system with a 20Å box, the peptide's C-terminus detached in two out of four runs. In the 40Å box, the peptide detached in even three out of four runs. Thus, performing only one simulation could produce misleading results.

Deleted: Figure 3

Within the first nanosecond of simulation time detachment was observed with the 40Å and the 20Å water shell while smaller boxes (10Å and 5Å shell) forced the epitope to stay inside the binding groove. However, extending the simulation time for another 5ns the protein also detached within a 10Å water shell (data not shown). A larger water shell somehow speeds up the relaxation process and – together with long simulation times seems to be necessary for plausible results of pMHC molecular dynamics simulations.

3.1.2 Force fields, water models & system simplification

We simulated the "single site binding" peptide loaded MHC in a 20Å hydration shell with different force fields and different water models. Surprisingly, for some simulations, we found a stable binding behaviour. Figure 5 shows minimum distances between the peptide's C-terminal alanine and the MHC β-sheet as a measure for peptide detachment (see also Figure 4). Different force fields and recommended compatible water models were used: The GROMACS force field (which is based on a GROMOS force field) and the most recent

Deleted: Figure 5

Deleted: Figure 4

For Peer Review Only

GROMOS force field (GROMOS96 53a6) both combined with SPC water model, and the OPLS-AA/L force field combined with TIP4P water model, respectively. The OPLS-AA force field together with TIP4P water (shown in Figure 5c) yields such low distance values that one could conclude that this artificial peptide actually is a binding epitope for the HLA-B*2705 MHC. This is in severe contrast to the simulations using the GROMACS force field together with SPC water described in previous sections, where the C-terminal end of the peptide detached in two out of four simulations (Figure 5a). Using the GROMOS96 53a6 force field, detachment of the peptide was observed too (Figure 5b).

Deleted: Figure 5

Deleted: Figure 5

Deleted: Figure 5

Regarding system simplifications we found almost no difference between the whole and the simplified pMHC complex, where the α3 and the β_2m domains have been left out. Neither the GROMACS force field (see Figure 5a and d) nor the GROMOS96 53a6 force field (see Figure 5b and e) show any effect of simplification.

Deleted: Figure 5

Deleted: Figure 5

Taking all in all we may conclude: Size of the solvation shell and the choice of water model / force field combination, in particular the water model, shows an important impact on simulation results. pMHC simplification, however, is shown to have less effect.

3.2 "Double site binding" peptide

The "double site binding" peptide (GRFAAAIAK) complexed to B*2705 MHC (PDB-id: 1jge) shows rather stable binding, and thus longer simulation runs are needed to study the influences of key parameters.

3.2.1 Solvation

We performed MD simulations for at least 12ns of the "double site binding" peptide loaded MHC, again systematically varying the solvation layer thickness from 12Å to 27Å. In addition to the properties described above, we also computed the rotational autocorrelation function (RACF) of the MHC. This was done to check if reorientation of the pMHC complex is hampered by an incomplete solvation shell, possibly in combination with the periodic

For Peer Review Only

boundary conditions. Orientation of the MHC molecule was defined by a vector perpendicular to the plane of three selected atoms belonging to the β-sheet forming the bottom of the groove (Cα atoms of V25, N97, and G100). The loss of orientational memory is characterized by the decay of the RACF (Equation 1, and Figure 6). Figure 6

Figure 6reveals that under periodic boundary conditions the loss of memory depends on the size of the water box. The results are only approximate since only one molecule was monitored for a rather short time, which may cause the non-exponential decays. However, a qualitative trend is clearly observable: While all three box sizes allow for rotation, rotational relaxation of the pMHC complex is less hindered in the 27Å system than in the 17Å system which in turn is less hindered than in the 12Å system.

The source of the observed rotational movement is rotational diffusion. The surrounding water hitting the pMHC complex randomly will cause to some extent a rotational movement of the complex. Figure 7 illustrates the way a smaller simulation box hinders the rotational movement of the pMHC complex: If the protein starts to rotate, the surrounding water network will follow the moving protein surface. In a large box there is less hindrance because the outer layers of the water network are almost unaffected by the rotation of the protein. Contrary, in the small box the rotation is more hindered because the outer water layers are bound to both, the protein and its periodic image. Thus, relatively tightly bound water molecules are forced to move in opposite directions, which effectively hinders rotation. Physically, tightly bound water means higher viscosity, resulting in longer relaxation times and thus less rotation.

3.2.2 Peptide's configurational substates

In our simulations of the MHC bound "double site binding" peptide within a 27Å water shell we observed that the bound peptide does not retain its original shape throughout the entire simulation run (16ns). It rather seems to switch back and forth between several different "substates" [19], which are temporarily stable (Figure 8a). Conformational substates of

Deleted: Figure 6

Deleted: Figure 6

Deleted: ¶ Figure 6

Deleted: Figure 7

Deleted: Figure 8

For Peer Review Only

proteins and polypeptides are also known from experiments [41]. Figure 8b shows the RMSD of Cα atoms for three specific residues. Some residues show low RMSD values throughout the simulation. Other residues show phases of high RMSD values at certain time intervals. One can conclude that several residues make small contributions to the peptide's total RMSD (Figure 8a), while others contribute significantly more. Hence the average RMSD, as usually displayed, hides the marked differences in mobility of atoms within the peptide. These specific movements and substates are made transparent using the following procedure:

We define the peptide's "principal" conformation as substate #1. The "principal" structure is defined by averaging the individual atomic positions between t=2ns and t=10ns and subsequently performing an "in vacuo" energy minimization. Then, positional deviations can conveniently be measured for each peptide's backbone atom relative to substate #1 (rather than relative to the crystal structure). Individual RMSD values can be shown grouped to a 3D plot over time (Figure 8d). The "hot spots" (tops and plateaus) indicate large deviations from the "principal" conformation and can be used to define further substates (see Figure 8e, Figure 9, and Table 2).

The observed substates can be characterized by several parameters (Table 2). Substate #2 (Figure 9, upper right corner) occurred very early and may be a reaction of the crystal structure to the surrounding water bath. Indeed, in this substate the centre part of the peptide (notably 6ALA) gets kinked up, causing the 6ALA residue not to face the lateral helix but protrude right up, away from the MHC. This movement generates a small hole beneath 6ALA, which allows for water molecules to enter the peptide-MHC interface. Actually, through the period of substate #2, the peptide-water interaction energy rises continually, while the peptide-MHC interaction energy decreases concomitantly (Figure 8c). In substate #1 ("principal" state – Figure 9, upper left corner) 6ALA is facing the lateral helix again and the hole is closed. During this state the peptide seems very stable and shows little fluctuations

Deleted: Figure 8
Deleted: Figure 8
Deleted: Figure 8
Deleted: Figure 8
Deleted: Figure 9
Deleted: Table 2
Deleted: Table 2
Deleted: Figure 9
Deleted: Figure 8
Deleted: Figure 9

For Peer Review Only

(RMSF). It remains in this substate for more than 8ns and comes back again to substate #1 after substate #3 and #4. Substate #3 (Figure 9, lower left corner) shows the C-terminal half of the peptide pointing away from the MHC. The peptide-water interaction energy increases again, causing a concomitant decrease in peptide-MHC interaction energy. The C-terminal residue 9LYS loses its interaction partner (116ASP in the β-sheet) and finds a new one (77ASP in the lateral α-helix). In substate #4 (Figure 9, lower right corner) the peptide lifts even more. The residue of 9LYS has to be extended almost totally to stay in contact with the F pocket (77ASP). More than 42% of the peptide's surface is exposed to water during this substate – a remarkable increase compared to 32.7% during substate #1. After this short period the peptide lowers again, 9LYS contacts 116ASP as before, and the peptide adopts the substate #1 for the rest of simulation time.

Deleted: Figure 9

Deleted: Figure 9

For Peer Review Only

4 Discussion

Conclusions from molecular dynamics simulations may reveal important mechanisms but can be delicate. A pMHC system is too large for a "full simulation", given the computational power available today. However, cautious choice of parameters still allows drawing conclusions from simulations. In this study we analyzed following key parameters for molecular dynamics simulations of pMHC molecules:

1) Initial conditions

   Minor differences in initial conditions can have a major effect on the result, in particular for labile systems as shown in the parallel simulations of the MHC bound "single site binding" peptide. Therefore, parallel simulations of identical structures starting from different initial conditions can help to identify labile systems. This seems crucial to avoid premature conclusions – in our case regarding weak binding or non binding epitopes.

2) Size of solvation shell

   Using periodic boundary conditions, the simulated system is embedded as the unit cell of a "supercrystal" within copies of itself. However, this crystalline environment may induce artefacts when the protein is comparable in size to the simulation box. Usually a bigger water shell minimizes these artefacts due to shielding effects of water: a polar solvent generates an electrostatic reaction field which screens the charges and dipoles [20]. We demonstrated the effects of a bigger water shell by several simulations of the MHC bound "single site binding" peptide. Using a small water shell (5Å or 10Å), the effect of the "supercrystal" was seen strong enough to artificially stabilize the whole structure: no detachment of the epitope was observed. Using a sufficiently large water shell (20Å), detachment of the epitope can be observed.

For Peer Review Only

Even if the system is large enough to solvate the whole protein, rotability might still be hindered. Analyzing simulations of the MHC bound "double site binding" peptide, we found that small water shells obviously allow for a direct interaction of the periodic images of the complexes through the water network, thereby reducing rotability. A sufficiently large water shell is mandatory to minimize water network interactions and enable realistic rotability.

Concluding, an enlargement of the water shell raises the computational workload but yields more realistic dynamics of the whole system.

3) Water model / force field combination

Using different models for force field and water representation induces drastic differences. Since SPC water is distinctly polar ($\varepsilon_r=65$, M.Neumann, unpublished), it is highly able to substitute the binding partners of the epitope and facilitates the "single site binding" epitope's detachment. With a less polar water model, like TIP4P ($\varepsilon_r=53\pm2$, [42]), solvation of the "single site binding" epitope is not able to disrupt its binding to the MHC: the epitope stays in the binding groove, displaying very low RMSD values. Rather "inert" water causes a pseudo-stabilization of the pMHC complex. Additionally, the capability of shielding electrostatic forces (see point 2) depends on the relative dielectric permittivity ε_r [20].

Usage of different force fields and water models provides different insights into the system's dynamics. Since there is no "one and only" model, variation is essential for reliable results. However, for a particular simulation experiment we can propose a guideline to choose a model that allows "staying on the safe side":

Simulations using a GROMOS-type force field with SPC water will tend to predict a "border case" epitope as non-binding. In contrast, the TIP4P water coming with the OPLS-AA force field shows a certain tendency that "border case" epitopes may be

For Peer Review Only

categorized as binding. Thus, when testing for binding (e.g. vaccine design) take the model that proved "looser" binding, and when testing for non-binding (e.g. allergenicity) take the model that proved "tighter" binding.

4) Simulation time

Long simulation runs are needed to observe all (or at least the most) relevant thermally accessible configurations. Sometimes a few hundred picoseconds simulation time will suffice. In other systems, representative sampling will take several seconds, which may be out of reach for molecular dynamics studies for a long time to come. Since we discovered the epitope within the MHC binding groove adopting different substates on the timescale of nanoseconds, pMHC simulations have to be at least several tens or even hundreds of nanoseconds long. However, in many cases it remains hard to decide whether a particular simulation has run long enough.

5) System simplification

The legitimacy of MHC simplification in MD simulations has already been covered by various authors [43;44] and there is no consensus about its effect on peptide dynamics on a timescale of nanoseconds. Our simulations show almost no difference between the simplified pMHC complex, where the α3 and the β_2m are left out of simulation, and the whole pMHC complex, neither when using the GROMACS force field (see Figure 5a and d) nor when using the GROMOS96 53a6 force field (see Figure 5b and e).

Deleted: Figure 5

Deleted: Figure 5

The computational discovery of the epitope's substates raises some very interesting questions regarding a possible immunological meaning: The "outbreaking" characteristic of substate #4 (C-terminal half slightly protruding from the cleft) could represent the very trigger for a TCR screening the pMHC complex. This hypothesis fits well into the hitherto known framework: Although T cell activation is to some extent correlated to the half-life of pMHC-TCR interaction, half-life alone is insufficient to accurately predict activation potencies. In

For Peer Review Only

addition, conformational changes within the TCR-peptide-MHC interface may essentially contribute to T cell activation [45]. Pöhlmann et al. [46] and Starikov et al. [47] suggested that even a more flexible peptide may have a noticeable effect on T cell activation. This is precisely the point where the presented simulations are capable to provide additional insight and evidence: In substate #4 the peptide is not only more flexible (RMSF of 0.045 nm instead of 0.037 nm in substate #1) but there is also an actual movement of the total C-terminal half of the peptide out of the cleft, towards a possible TCR. This conformational change in the peptide could be transmitted to a bound TCR, possibly contributing to T cell activation [48].

Some authors have described peptide flexibilities, but the current pMHC-analysis is – to the best of our knowledge – the first to describe substates of the MHC bound peptide. Pöhlmann et al. [49] and Starikov et al. [50] performed MD simulations of the same system (PDB-id: 1jge). However, they observed no specifiable conformational substates for the bound peptide. Pöhlmann et al. used the OPLS-AA/L force field and TIP4P water model and Starikov et al. used a 7 Å water shell, both may suffer from a limited configurational sampling.

Further simulation studies can be envisaged to find out whether the existence of "protruding substates" actually correlates to the activation potency of TCRpMHC complexes, as obtained from experimental measurements. The present work evaluates the computational tools and prepares their future application in the prediction of "protruding substates" for a given pMHC complex. In addition, single peptide residues could be exchanged and the impact on "protruding substates" computationally evaluated for vaccine design.

For Peer Review Only

5 Acknowledgments

We are indebted to Prof. Dr. Anneliese Schimpl for continuous support and advice in preparing the manuscript.

The computational work for the current investigation has been supported by Prof. Dr. Rudolf Karch, Core Unit for Medical Statistics and Informatics, Medical University of Vienna, Austria.

For Peer Review Only

6 References

[1] M. Karplus and G.A. Petsko, *Molecular dynamics simulations in biology*, Nature 347 (1990), pp. 631-639.

[2] T. Hansson, C. Oostenbrink, and W. van Gunsteren, *Molecular dynamics simulations*, Curr. Opin. Struct. Biol. 12 (2002), pp. 190-196.

[3] D. Rognan, L. Scapozza, G. Folkers, and A. Daser, *Molecular dynamics simulation of MHC-peptide complexes as a tool for predicting potential T cell epitopes*, Biochemistry 33 (1994), pp. 11476-11485.

[4] D. Rognan, S. Krebs, O. Kuonen, J.R. Lamas, J.A. Lopez de Castro, and G. Folkers, *Fine specificity of antigen binding to two class I major histocompatibility proteins (B*2705 and B*2703) differing in a single amino acid residue*, J. Comput. Aided Mol. Des 11 (1997), pp. 463-478.

[5] H. Toh, N. Kamikawaji, T. Tana, T. Sasazuki, and S. Kuhara, *Molecular dynamics simulations of HLA-DR4 (DRB1*0405) complexed with analogue peptide: conformational changes in the putative T-cell receptor binding regions*, Protein Eng 11 (1998), pp. 1027-1032.

[6] T. Pöhlmann, R.A. Bockmann, H. Grubmuller, B. Uchanska-Ziegler, A. Ziegler, and U. Alexiev, *Differential peptide dynamics is linked to major histocompatibility complex polymorphism*, J. Biol. Chem. 279 (2004), pp. 28197-28201.

[7] E.B. Starikov, L. Nilsson, and M. Hulsmeyer, *A single residue exchange between two HLA-B27 alleles triggers increased peptide flexibility*, Eur. Biophys. J. 33 (2004), pp. 651-655.

[8] U. Omasits, M. Neumann, O. Steinhauser, R. Valenta, R. Kobler, and W. Schreiner, *Molecular Dynamics Simulations in Immunology*, in *Proceedings of the 2nd Austrian Grid Symposium*, J. Volkert, T. Fahringer, D. Kranzlmüller, and W. Schreiner, eds., Oesterreichische Computer Gesellschaft, Vienna, 2007, pp. 81-90.

[9] W.S. Meng, H. von Grafenstein, and I.S. Haworth, *A model of water structure inside the HLA-A2 peptide binding groove*, Int. Immunol. 9 (1997), pp. 1339-1346.

[10] P.M. Petrone and A.E. Garcia, *MHC-peptide binding is assisted by bound water molecules*, J. Mol. Biol. 338 (2004), pp. 419-435.

[11] M. Zacharias and S. Springer, *Conformational flexibility of the MHC class I alpha1-alpha2 domain in peptide bound and free states: a molecular dynamics simulation study*, Biophys. J. 87 (2004), pp. 2203-2214.

[12] O. Michielin and M. Karplus, *Binding free energy differences in a TCR-peptide-MHC complex induced by a peptide mutation: a simulation analysis*, J. Mol. Biol. 324 (2002), pp. 547-569.

[13] V. Zoete and O. Michielin, *Comparison between computational alanine scanning and per-residue binding free energy decomposition for protein-protein association using*

For Peer Review Only

MM-GBSA: Application to the TCR-p-MHC complex, Proteins 67 (2007), pp. 1026-1047.

[14] O. Michielin and M. Karplus, *Binding free energy differences in a TCR-peptide-MHC complex induced by a peptide mutation: a simulation analysis*, J. Mol. Biol. 324 (2002), pp. 547-569.

[15] S. Wan, P.V. Coveney, and D.R. Flower, *Molecular basis of peptide recognition by the TCR: affinity differences calculated using large scale computing*, J. Immunol. 175 (2005), pp. 1715-1723.

[16] V. Zoete and O. Michielin, *Comparison between computational alanine scanning and per-residue binding free energy decomposition for protein-protein association using MM-GBSA: Application to the TCR-p-MHC complex*, Proteins 67 (2007), pp. 1026-1047.

[17] T. Darden, D. York, and L. Pedersen, *Particle mesh Ewald: An N.log(N) method for Ewald sums in large systems*, J. Chem. Phys. 98 (1993), pp. 10089-10092.

[18] T.E. Cheatham and B.R. Brooks, *Recent advances in molecular dynamics simulation towards the realistic representation of biomolecules in solution*, Theor. Chem. Acc. 99 (1998), pp. 279-288.

[19] B.K. Andrews, T. Romo, J.B. Clarage, B.M. Pettitt, and J. Phillips, *Characterizing global substates of myoglobin*, Structure 6 (1998), pp. 587-594.

[20] D. Sengupta, R.N. Behera, J.C. Smith, and G.M. Ullmann, *The alpha helix dipole: Screened out?*, Structure 13 (2005), pp. 849-855.

[21] A. Wada, *The alpha-helix as an electric macro-dipole*, Advances in Biophysics (1976), pp. 1-63.

[22] W. Weber, P.H. Hünenberger, and J.A. McCammon, *Molecular dynamics simulations of a polyalanine octapeptide under Ewald boundary conditions: Influence of artificial periodicity on peptide conformation*, J. Phys. Chem. B 104 (2000), pp. 3668-3675.

[23] G.M. Torrie and J.P. Valleau, *Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling*, J. Comput. Phys. 23 (1977), pp. 187-199.

[24] D. Rognan, N. Zimmermann, G. Jung, and G. Folkers, *Molecular dynamics study of a complex between the human histocompatibility antigen HLA-A2 and the IMP58-66 nonapeptide from influenza virus matrix protein*, Eur. J. Biochem. 208 (1992), pp. 101-113.

[25] S. Wan, P. Coveney, and D.R. Flower, *Large-scale molecular dynamics simulations of HLA-A*0201 complexed with a tumor-specific antigenic peptide: can the alpha3 and beta2m domains be neglected?*, J. Comput. Chem. 25 (2004), pp. 1803-1813.

[26] The Austrian Grid Initiative, http://www.austriangrid.at, cited 2008

[27] H.J.C. Berendsen, D. van der Spoel, and R. van Drunen, *GROMACS: A message-passing parallel molecular dynamics implementation*, Comput. Phys. Commun. 91 (1995), pp. 43-56.

For Peer Review Only

[28] E. Lindahl, B. Hess, and D. van der Spoel, *GROMACS 3.0: a package for molecular simulation and trajectory analysis*, J. Mol. Model. 7 (2001), pp. 306-317.

[29] G. Kingsley and J. Sieper, *Current perspectives in reactive arthritis*, Immunol. Today 14 (1993), pp. 387-391.

[30] H.M. Bergman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, and P.E. Bourne, *The Protein Data Bank*, Nucleic Acids. Res. 28 (2000), pp. 235-242.

[31] D.R. Madden, J.C. Gorga, J.L. Strominger, and D.C. Wiley, *The three-dimensional structure of HLA-B27 at 2.1 A resolution suggests a general mechanism for tight peptide binding to MHC*, Cell 70 (1992), pp. 1035-1048.

[32] M. Hulsmeyer, R.C. Hillig, A. Volz, M. Ruhl, W. Schroder, W. Saenger, A. Ziegler, and B. Uchanska-Ziegler, *HLA-B27 subtypes differentially associated with disease exhibit subtle structural alterations*, J. Biol. Chem. 277 (2002), pp. 47844-47853.

[33] J. Zielkiewicz, *Structural properties of water: comparison of the SPC, SPCE, TIP4P, and TIP5P models of water*, J. Chem. Phys. 123 (2005), pp. 1-6.

[34] D. van der Spoel, E. Lindahl, B. Hess, A.R. van Buuren, E. Apol, P.J. Meulenhoff, D.P. Tieleman, A.L.T.M. Sijbers, K.A. Feenstra, R. van Drunen, and H.J.C. Berendsen, *Gromacs User Manual version 3.2*, 2004.

[35] C. Oostenbrink, A. Villa, A.E. Mark, and W.F. Van Gunsteren, *A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6*, J. Comput. Chem. 25 (2004), pp. 1656-1676.

[36] G.A. Kaminski, R.A. Friesner, J. Tirado-Rives, and W.L. Jorgensen, *Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides*, J. Phys. Chem. B 105 (2001), pp. 6474-6487.

[37] H.J.C. Berendsen, J.P.M. Postma, W.F. Van Gunsteren, A. DiNola, and J.R. Haak, *Molecular dynamics with coupling to an external bath*, J. Chem. Phys. 81 (1984), pp. 3684-3690.

[38] J.P. Ryckaert, G. Ciccotti, and H.J.C. Berendsen, *Numerical integration of the cartesian equations of motion of a system with constraints; molecular dynamics of n-alkanes.*, J. Comput. Phys. 23 (1977), pp. 327-341.

[39] F. Eisenhaber, P. Lijnzaad, P. Argos, C. Sander, and M. Scharf, *The Double Cube Lattice Method: Efficient Approaches to Numerical Integration of Surface Area and Volume and to Dot Surface Contouring of Molecular Assemblies*, J. Comput. Chem. 16 (1995), pp. 273-284.

[40] M. Karplus and G.A. Petsko, *Molecular dynamics simulations in biology*, Nature 347 (1990), pp. 631-639.

[41] H. Frauenfelder, F. Parak, and R.D. Young, *Conformational substates in proteins*, Annual review of biophysics and biophysical chemistry 17 (1988), pp. 451-479.

For Peer Review Only

[42] M. Neumann, *Dielectric relaxation in water. Computer simulations with the TIP4P potential*, J. Chem. Phys. 85 (1986), pp. 1567-1580.

[43] D. Rognan, N. Zimmermann, G. Jung, and G. Folkers, *Molecular dynamics study of a complex between the human histocompatibility antigen HLA-A2 and the IMP58-66 nonapeptide from influenza virus matrix protein*, Eur. J. Biochem. 208 (1992), pp. 101-113.

[44] S. Wan, P. Coveney, and D.R. Flower, *Large-scale molecular dynamics simulations of HLA-A*0201 complexed with a tumor-specific antigenic peptide: can the alpha3 and beta2m domains be neglected?*, J. Comput. Chem. 25 (2004), pp. 1803-1813.

[45] M. Krogsgaard, N. Prado, E.J. Adams, X.L. He, D.C. Chow, D.B. Wilson, K.C. Garcia, and M.M. Davis, *Evidence that structural rearrangements and/or flexibility during TCR binding can contribute to T cell activation*, Mol. Cell 12 (2003), pp. 1367-1378.

[46] T. Pöhlmann, R.A. Bockmann, H. Grubmuller, B. Uchanska-Ziegler, A. Ziegler, and U. Alexiev, *Differential peptide dynamics is linked to major histocompatibility complex polymorphism*, J. Biol. Chem. 279 (2004), pp. 28197-28201.

[47] E.B. Starikov, L. Nilsson, and M. Hulsmeyer, *A single residue exchange between two HLA-B27 alleles triggers increased peptide flexibility*, Eur. Biophys. J. 33 (2004), pp. 651-655.

[48] J.B. Reiser, C. Gregoire, C. Darnault, T. Mosser, A. Guimezanes, A.M. Schmitt-Verhulst, J.C. Fontecilla-Camps, G. Mazza, B. Malissen, and D. Housset, *A T cell receptor CDR3beta loop undergoes conformational changes of unprecedented magnitude upon binding to a peptide/MHC class I complex*, Immunity. 16 (2002), pp. 345-354.

[49] T. Pöhlmann, R.A. Bockmann, H. Grubmuller, B. Uchanska-Ziegler, A. Ziegler, and U. Alexiev, *Differential peptide dynamics is linked to major histocompatibility complex polymorphism*, J. Biol. Chem. 279 (2004), pp. 28197-28201.

[50] E.B. Starikov, L. Nilsson, and M. Hulsmeyer, *A single residue exchange between two HLA-B27 alleles triggers increased peptide flexibility*, Eur. Biophys. J. 33 (2004), pp. 651-655.

[51] M.A. Saper, P.J. Bjorkman, and D.C. Wiley, *Refined structure of the human histocompatibility antigen HLA-A2 at 2.6 A resolution*, J. Mol. Biol. 219 (1991), pp. 277-319.

[52] W. Humphrey, A. Dalke, and K. Schulten, *VMD: visual molecular dynamics*, J. Mol. Graph. 14 (1996), pp. 33-38.

[53] M. Hulsmeyer, R.C. Hillig, A. Volz, M. Ruhl, W. Schroder, W. Saenger, A. Ziegler, and B. Uchanska-Ziegler, *HLA-B27 subtypes differentially associated with disease exhibit subtle structural alterations*, J. Biol. Chem. 277 (2002), pp. 47844-47853.

For Peer Review Only

7 Tables

Table 1. Overview of molecular dynamics simulations performed. Peptide main anchor residues are printed bold.

peptide	MHC	water model / force field combination	size of water box	parallel simulations	simulation time
A**R**AAAAAAA ("single site binding peptide")	B*2705 whole	SPC / GROMACS	20Å	4x	1ns
	B*2705 whole	SPC / GROMOS 53a6	20Å	4x	1ns
	B*2705 simplified	SPC / GROMACS	2.5Å	4x	1ns
	B*2705 simplified	SPC / GROMACS	5Å	4x	1ns
	B*2705 simplified	SPC / GROMACS	10Å	4x	1ns
	B*2705 simplified	SPC / GROMACS	10Å	1x	10ns
	B*2705 simplified	SPC / GROMACS	20Å	4x	1ns
	B*2705 simplified	SPC / GROMACS	40Å	4x	1ns
	B*2705 simplified	SPC / GROMOS 53a6	20Å	4x	1ns
	B*2705 simplified	TIP4P / OPLS-AA	20Å	4x	1ns
G**R**FAAAIA**K** ("double site binding peptide")	B*2705 simplified	SPC / GROMACS	12Å	1x	12ns
	B*2705 simplified	SPC / GROMACS	17Å	1x	12ns
	B*2705 simplified	SPC / GROMACS	27Å	1x	16ns

For Peer Review Only

Table 2. Characteristics of the "double site binding" peptide in crystal structure and in different substates occurring during the 16ns simulation. SASA is taken relative to SASA of the peptide in water.

configuration	time (ps)	mainly affected residues	SASA (relative)	interaction energy (kJ/mol)		RMSF (nm)
				peptide-water	peptide-MHC	
crystal	0		29.7%	$-3.8\ 10^7$	$-8.8\ 10^7$	0.037
substate #1	2,000 - 10,000		32.7%	$-4.4\ 10^7$	$-8.3\ 10^7$	0.037
substate #2	890 - 1,550	5A, 6A, 7I	36.8%	$-4.3\ 10^7$	$-8.5\ 10^7$	0.045
substate #3	11,050 - 12,500	8A, 9K	39.9%	$-5.1\ 10^7$	$-7.5\ 10^7$	0.038
substate #4	13,550 - 14,100	5A, 6A, 7I, 8A, 9K	42.3%	$-5.2\ 10^7$	$-7.0\ 10^7$	0.045

For Peer Review Only

8 Figure captions

Figure 1. HLA-B*2705 molecular surface and its binding pockets.

The pockets definition is according to [51]. Pocket A (M5, Y7, Y59, E63, Y159, E163, W167, Y171), pocket B (H9, T24, E45, L66, C67, Y99), pocket C (H9, K70, T73, D74, R97), pocket D (Y99, H114, L156, Y159, L160), pocket E (H114, W133, W147, V152, L156), pocket F (D77, T80, L81, Y84, D116, Y123, T143, K146, W147). The peptide's primary anchors P2 (conserved, Arg) and P9 form salt bridges and hydrogen bonds to the complementary negatively charged MHC pockets B and F, respectively. P3 interacts with the upper part of the hydrophobic pocket D. This one and all subsequent graphical representations of molecules were produced using the VMD package [52].

Figure 2. Crystal structure and simulated system.

*a) Crystal structure of a pMHC complex (m9 peptide complexed to HLA-B*2705 [53]). The peptide binding domain (α1 and α2 domain, white), the α3 domain (grey) and the associated β-2-microglobulin (black) are represented as "cartoons". The bound peptide is shown as "licorice". The dimensions of the MHC's peptide binding domain are given in Ångström.*

b) A simplified pMHC complex placed in the centre of a simulation box allowing for a 20 Ångström solvation shell resulting in altogether 99,101 atoms (water molecules in the front of the cube are carved out in order to make the pMHC complex visible). This system could be used as input for a MD simulation "as is".

Figure 3. RMSF of peptide residues' backbone atoms.

For Peer Review Only

We performed four independent 1ns runs (same starting structure, different starting velocities) for each simulation box size. The RMSF values shown are averages for the time interval from 600ps to 1000ps, which represents the equilibrated part of a stable run. Results for the individual runs are indicated by different line types. Low RMSF values indicate low positional fluctuations of the backbone atoms during the simulation, which means that these residues are likely to be tightly bound by the MHC. A residue displaying high RMSF values may be loosely bound or not bound at all.

Figure 4. Demonstration of the minimum distance between epitope C-terminus and MHC β-sheet.

a) Minimum distances of the epitope's C-terminal alanine to the MHC's β-sheet during 1ns simulations. In the simulation of the complex with a 20Å surrounding water shell (dashed line) the distance first rises gradually from 0.37nm to 0.7nm. Then the C-terminus totally detaches and moves as far as 1.4nm away. Although there are large fluctuations for this distance, in the simulation employing a 5Å water shell (continuous line) no detachment occurs during the 1ns run.

b) The dashed line marks the minimum distance of all the atoms of the epitope's C-terminal alanine to all the atoms of the MHC's β-sheet. The orange (light grey) peptide shows the distance in the crystal structure. The blue (dark grey) peptide depicts the increased distance after C-terminal detachment. For an animated display of the simulation please refer to (supplementary material S1).

Figure 5. Minimum distance between peptide C-terminus and MHC β-sheet during different simulations of the "single site binding" peptide.

For Peer Review Only

Three different force fields and two different water models were used for simulations of the simplified pMHC complex, and two different force fields with the same water model were used for the whole pMHC complex. Each combination was simulated in four independent runs for 1ns within a 20Å hydration shell (black, dark grey, grey, and light grey line).

a) Simulating the simplified pMHC complex enclosed in a 20Å hydration shell results for two out of the four simulations (black and dark grey) in the previously described partial detachment of the peptide (see Figure 4) when the GROMACS force field and the SPC water model are used. The rising distance of the peptide's C-terminal alanine residue to 1nm and above indicates total detachment from the MHC binding cleft.

Deleted: *Figure 4*

b) Application of the GROMOS96 53a6 force field on the simplified system leads to peptide detachment in one out of four simulations (black).

c) When the OPLS-AA force field and the TIP4P water model are used on the simplified system, for all four simulations the peptide calmly resides in the binding groove and the minimum distance measure retains its initial low value for the entire length of the 1ns simulation.

d) Simulating the whole pMHC complex using the GROMACS force field and the SPC water model results in an onset of peptide detachment in two out of four simulations (black and dark grey).

e) Using the GROMOS96 53a6 force field on the whole pMHC complex leads to peptide detachment in one out of four simulations (black).

Figure 6. Orientational relaxation (RACF) of the pMHC complex for different sizes of the simulation box.

All the pMHC systems lose their orientational "memory" over time, which is shown by a decay of the RACF $C_1(t)$. The larger the surrounding water shell the faster the loss of memory. All three systems have been simulated for at least 12ns.

For Peer Review Only

Figure 7. Model for system size influencing rotational diffusion.

Both illustrations show a simulation box containing a protein solved in water using periodic boundaries but with different simulation box sizes. The darker the water molecules, the more structured is the water network. A small box hinders rotation by forcing a relatively tightly bound water layer to move in opposite directions.

Figure 8. Analysis of the long time simulation (16ns) of the "double site binding" peptide complexed to HLA-B*2705 (1jge) within a 27Å water box.

a) RMSD of the peptide's backbone (red line) and the MHC's backbone (black line) relative to the initial crystal structure. The peptide spends most of the time in a state (characterized by low RMSD values) that is very similar to the crystal structure, but seems to flip to a different conformational substates around simulation time t≈1ns and between t≈10ns and t≈14ns.

b) RMSD of specific peptide backbone atoms. The Cα atom of the primary anchor arginine (green line) remains close to its initial position throughout the simulation (low RSMD values). The Cα atom of the central alanine (position 5, blue line) temporarily moves farther off its position in the crystal structure after t≈1ns (higher RMSD values) while the Cα atom of the secondary anchor lysine (black line) undergoes an even larger dislocation (large RMSD values) between t≈10ns and t≈14ns. Therefore the peptide's residues differently contribute to the substates. Individual RMSD curves for all the peptide's Cα atoms can be found in (supplementary material S2).

c) Interaction energies of the peptide with the MHC (black line) and with water (grey line). Energies are the sum of Coulomb and Lennard Jones potential.

d) RMSD of each epitope's backbone atom relative to substate #1. This plot shows the substates and backbone atoms involved.

e) Conformational substates defined using above graphs.

For Peer Review Only

Figure 9. Average structures of peptide's conformational substates.

While the white peptide representations depict the substate #1, the grey representations depict the average structures of corresponding substates. The MHC structure shown is an average structure over the whole simulation. Arrows indicate major structural deviations to substate #1

For Peer Review Only

9 Word count

33 pages

8237 words

45002 characters

For Peer Review Only

Analysis of Key Parameters for Molecular Dynamics of pMHC Molecules

Ulrich Omasits[a], Bernhard Knapp[a], Martin Neumann[b], Othmar Steinhauser[c], Hannes Stockinger[d], Rene Kobler[e], and Wolfgang Schreiner[a]*

For Peer Review Only

The work described in this paper is supported by the Austrian Grid Project, funded by the Austrian BMBWK (Federal Ministry for Education, Science and Culture) under contract GZ 4003/2-VI/4c/2004.

[a] Core Unit for Medical Statistics and Informatics, Medical University of Vienna, Austria

[b] Institute of Experimental Physics, University of Vienna, Austria

[c] Department of Computational Biological Chemistry, University of Vienna, Austria

[d] Department of Molecular Immunology, Centre of Biomolecular Medicine and Pharmacology, Medical University of Vienna, Austria

[e] Institute of Graphics and Parallel Processing, Johannes Kepler University Linz, Austria

[*] Corresponding author: fax: +43 1 40400-6677 – phone: +43 1 40400-6673– email: bernhard.knapp@meduniwien.ac.at – address: Spitalgasse 23, 1090 Vienna, Austria

Abstract

Molecular dynamics (MD) studies of human major histocompatibility complex (MHC) HLA-B*2705 complexing two different peptides were performed. During simulation one peptide partially detached from the MHC while the other peptide switched back and forth between several different configurations. These different configurations relate to conformational substates and can be assigned to different levels of chemical activity or even the molecular mechanisms of immunological signalling. To ensure reliable immunological conclusions from MD simulations we prepare the methodological tools by carefully evaluating initial conditions, system simplification, solvation shell thickness, water model / force field combination, and simulation length. We also derive a guideline for appropriate model selection. This kind of quality assessment is seen a mandatory prerequisite for coming studies linking pMHC dynamics to T cell activation.

Keywords: HLA; major histocompatibility complex; peptide dynamics; conformational substates; T cell activation

Abbreviations: β_2m: β_2-microglobulin – HLA: human leukocyte antigen – MD: molecular dynamics – MHC: major histocompatibility complex – PBC: periodic boundary conditions – PME: Particle Mesh Ewald – pMHC: peptide loaded MHC – RACF: rotational autocorrelation function – RMSD: root mean square deviation – RMSF: root mean square fluctuation – SASA: solvent accessible surface area – TCR: T cell receptor

For Peer Review Only

1 Introduction

Class I major histocompatibility complexes (MHC) bind antigenic peptides within the α1-α2 domain in a long narrow cleft located between two α-helices on the top of an antiparallel β-sheet. Binding is mainly mediated through sequence independent contacts of the peptide's charged N- and C-terminus to conserved "pockets" (Figure 1) in the MHC's binding groove. The contacts are maintained by networks of hydrogen bonds that are very similar in all class I pMHC complexes and make the largest contribution to the binding energy of the peptide. In addition, polymorphic pockets, so-called primary anchor regions, interact with specific sequence dependent amino acids (anchor residues) of the peptide and thus are responsible for the allele specificity of the recognition process. The central part of the peptide bulges out of the binding groove and serves as primary recognition site for the T cell receptor (TCR) (see Figure 2a).

Molecular dynamics (MD) is the technique of numerically solving the equations of motion of an assembly of particles (atoms). The total force acting on each particle depends on the positions and properties of all other particles in the system and, in addition, may include externally imposed forces. MD programs calculate these forces and solve Newton's equation of motion for each particle. For details refer to reviews on MD of biomolecules [1;2].

Various MD studies on pMHC have been reported in the literature [3-16]. All these MD studies use some force field, some water model, some initial condition, and some size of water box, in most cases without a particular justification or comment on their choices. The aim of this study is to vary these parameters and analyze their effects and possible pitfalls regarding a pMHC complex simulation. To our knowledge no such detailed investigation has been reported in the literature. We are going to answer the following questions:

For Peer Review Only

- Should parallel simulations be performed using different initial conditions?
- What size of solvation box should be used?
- Which force field and water model should be chosen?
- What is the minimum simulation time necessary?
- Are system simplifications admissible?

Our aim is to evaluate those key parameters giving a reasonable chance to cope with the rather large pMHC molecule in a meaningful way. All in all, the present work lays the basis for MD as a method of investigation in pMHC interactions.

Besides the technical evaluation we also present a minute mobility analysis of a peptide adopting different configurational substates within the MHC-cleft.

1.1 Environment representation in MD

In order to achieve a manageable representation of the natural pMHC complex a trade-off between computational cost and simulation accuracy has to be made, especially concerning the amount of solvation water included in the simulation, the boundary conditions, and protein simplifications. Since proteins evolved to function in aqueous environments, water molecules binding directly to the protein and forming the first layer of the hydration shell may be considered as an integral part of protein structure. It is also clear that solute-solvent interactions extend beyond the first hydration layer. Although an extended hydration shell may be represented by simplified models, the most reliable results on structure and dynamics are achieved by inclusion of a sufficient number of additional explicit solvent molecules. To avoid an abrupt border with a vacuum, periodic boundary conditions (PBC) repeat the simulation box periodically in all three space dimensions, creating a continuous system. PBC allows a very useful method for calculating long-range electrostatics: the Particle Mesh Ewald (PME) summation [17]. Currently PBC/PME represents the standard in MD simulations of

http://mc.manuscriptcentral.com/tandf/jenmol

For Peer Review Only

biomolecules since it is the most accurate method. On the other hand, the imposition of periodicity may lead to artefacts since all atoms now may interact with their periodic images. These artefacts can be minimized using larger solvation shells and an appropriate water model / force field combination, i.e. a combination to stay on the "safe side" for a particular simulation experiment (see Discussion section).

1.1.1 Size of solvation shell

A minimal system is claimed to need a 5-15Å thick solvation shell [18]. In other words, the system needs to be large enough to allow for the formation of a 5-15Å "buffer zone" between the solute and the boundaries of the simulation box. We show that for pMHC simulations a solvation layer thickness of more than 10Å, ideally 20Å, is necessary (Figure 2b). It is known that the aqueous environment facilitates the protein sampling the conformational space [19]. Water, due to its high relative dielectric permittivity, is very effective in screening long-range electrostatic interactions and dipoles. Thus the thickness of the solvation shell essentially determines the degree of electrostatic shielding, which is seen crucial in the presence of significant dipoles induced by the α-helices of the MHC [20;21]. PBC/PME imposes an artificial periodicity on the grossly inhomogeneous solute-solvent system. This periodicity limits the configurational flexibility of a protein [22] by overstabilization of the starting configuration which is normally taken to be that of a periodic crystal. Obviously, the larger the simulation box, the better the water molecules in the solvation shell (and the protein itself) will be screened from interactions with the protein's periodic images. Thus crystalline effects are minimized, sampling is less hindered [23], and a more realistic dynamics of the system can be observed.

1.1.2 Structural simplifications

It is quite common to simplify a simulation system in order to minimize the number of atoms. In the case of pMHC complexes frequently only the MHC's peptide binding domain (α1 and α2 domain) is simulated since this is the site of interesting peptide-MHC interactions. In vivo

For Peer Review Only

such a truncated pMHC complex would not be stable at all and unfolding would occur. MHC simplification has been found not to alter the results of MD simulations on the timescale feasible [24]. However, there is disagreement about this simplification and there are authors [25] claiming that the α3 domain and the β_2m subunit are essential for peptide MHC interactions even within one nanosecond simulation time. According to our results, simplification plays a minor role. The α3 domain and the β_2m subunit are certainly essential for a simulation of a system in its full complexity. However, we are interested in processes taking place within the binding groove at a much shorter timescale than MHC unfolding would occur.

For Peer Review Only

2 Materials and Methods

2.1 Hard- and Software

The project was carried out on a SGI Altix 350 system within the framework of the Austrian Grid [26] and on a computational cluster of the Technical University of Vienna. For a typical 2ns simulation run a total of 800 CPU hours are required.

We use GROMACS [27;28] because of its detailed documentation of computational methodology and very fast procedure to treat water molecules. Its high scalar performance outweighed the fact that (on our system) GROMACS scales efficiently only up to 16 nodes.

2.2 Setup of coordinates

For simulations we selected the MHC class I B*2705 protein. Almost all binding epitopes for this MHC show an arginine at position 2, defined as a dominant anchor residue. Positions 1, 3, and 9 provide the remaining main anchoring positions with a preference for hydrophobic amino acids at P3, and hydrophobic or positively charged amino acids at P9. The dominant anchor residue P2 binds deep into the MHC pocket B, while P9 binds into MHC pocket F (see Figure 1). HLA-B27 subtypes are strongly associated with the occurrence of several diseases, in particular inflammatory diseases of the joints called spondyloarthropathies [29] and thus lends itself as a most interesting candidate for computational screening.

Starting coordinates were taken from crystal structures deposited in the Protein Data Bank [30]. We chose two structures representing the same MHC complexing two different, but substantially similar peptides. The first structure was obtained by Madden et al. from an epitope mix. As a “minimal binding” peptide the consensus backbone was modelled with the epitope ARAAAAAAA. The resulting structure was deposited as entry 1hsa [31]. The second structure containing the epitope GRFAAAIAK was obtained by Hülsmeyer et al. and deposited as entry 1jge [32]. We chose those two complexes due to their chemically similar

http://mc.manuscriptcentral.com/tandf/jenmol

For Peer Review Only

epitopes. Both consist of mainly hydrophobic residues (A,F,I,G) and both have the R anchor at position 2, typical for HLA-B*2705 epitopes. The crucial difference is the secondary anchor (K) in 1jge at position 9 which is missing in 1hsa. Therefore we will refer to the epitope of 1jge as “double site binding” peptide and to the epitope of 1hsa as “single site binding” peptide.

For most simulations only the antigen binding α1 and α2 domains were considered. The C-terminal end of the α2 domain was protonated to render it uncharged while all the other termini and side chains were assigned typical charge states at pH 7. Water molecules observed in the crystal structure were explicitly taken into account in the simulations. Polar hydrogens were added and the complexes were centred in cubic boxes of different sizes, to allow for water shells ranging from 2.5Å to more than 40Å thickness between the solute and the boundaries of the simulation box. The created space around the solute was initially filled with an equilibrium configuration of bulk water. The water models used were SPC and TIP4P, respectively, both recommended for usage in biomolecular systems [33]. To electrically neutralize the system, an appropriate number of Na^+ counterions were added.

2.3 Simulation methods and parameters

In order to investigate the influence of the force field and water model on our results we selected three different parametrizations from the suite of force fields available with GROMACS. The GROMACS force field (a modified GROMOS87 force field) [34] and the GROMOS96 53a6 force field [35] were used for SPC solvated pMHC complexes while the OPLS-AA/L all-atom force field (version 2001) [36] was used for TIP4P solvated pMHC complexes, as recommended in GROMACS documentation [34]. After the initial coordinate setup, as described above, the system’s potential energy was first minimized by a steepest descent procedure, in order to allow the water molecules to adjust to the presence of the pMHC complex. The coordinates so obtained served as a starting structure for the subsequent

For Peer Review Only

MD simulations. Periodic boundaries were employed and the system was then heated linearly from 0K to 300K during the first 40ps and left at 300K for the last 10ps using a Berendsen thermostat with a temperature coupling constant of 0.1 ps [37]. For this warm-up phase only, position restraints (force constant of 1000 kJ mol^{-1} nm^{-2}) were applied to each atom of the pMHC complex. Subsequent production runs without any position restraints used Berendsen pressure coupling (reference pressure: 1 bar, pressure coupling constant: 0.5 ps) [37], the Particle Mesh Ewald (PME) [17] method for long-range electrostatic interactions, and an integration step of 2fs. All bond lengths were constrained to their equilibrium values using the SHAKE algorithm [38]. Simulations started with random velocities distributed to the atoms (i.e. initial conditions) and total simulation times between 1ns and 16ns were used. For subsequent analysis, configurations were saved in trajectory files every 2ps for long runs (16ns total) and every 0.5ps for short runs (1ns total). An overview of all performed simulation runs can be found in Table 1.

2.4 Methods of analysis

All simulation runs were evaluated via the following analysis modules provided by the GROMACS package:

2.4.1 SASA

The solvent accessible surface area (SASA) was computed using the Double Cube Lattice Method [39]. The radius of the solvent probe was set to 1.4 Ångström.

2.4.2 RMSD

The root mean square deviation (RMSD) gives the average deviation of a structure relative to a reference structure (e.g. crystal structure). In order to subtract global motions each structure is least squares fitted to the reference structure before RMSD calculation.

2.4.3 RMSF

The root mean square fluctuations (RMSF) represent the standard deviation of atomic positions during (a part of) a trajectory after first fitting to a reference structure. The RMSF of

For Peer Review Only

the crystal structure is calculated out of the crystallographic B-factors [40]. Contact residues have systematically low RMSF values while surface side chains protruding from the protein show large RMSF values.

2.4.4 Rotational autocorrelation function (RACF)

Given a unit vector $\vec{n}(t)$ characterizing a key direction of a molecule, the RACF is computed via

$$C_1(t) = \langle P_1(\vec{n}(0) \cdot \vec{n}(t)) \rangle \qquad \text{(Equation 1)}$$

with t being the time lag and $P_1(x) = x$ the first order Legendre polynomial. Values of 1 indicate "perfect memory" of orientation, a value of 0 indicates loss of orientational correlation.

For Peer Review Only

3 Results

Table 1 summarizes the simulations performed in order to analyze the key parameters of MD such as size of water box, initial conditions, system simplification, simulation time, and water model / force field combination. In particular, the following results were obtained:

3.1 "Single site binding" peptide

In order to point out the possible impacts of simplifications and limitations, we chose the rather labile system of the "single site binding" peptide (ARAAAAAAA) complexed to B*2705 MHC (PDB-id: 1hsa). This complex features an artificially modelled epitope with only one anchor residue (arginine at position 2).

3.1.1 Solvation shell thickness & initial conditions

We performed several MD simulations of the "single site binding" peptide loaded MHC with different initial conditions (randomly distributed starting velocities) and with solvation shell thicknesses ranging from 2.5Å to 40Å and computed the RMSF, see Figure 3. The RMSF values for the peptide's C-terminal alanine (residue 9), show the most pronounced variability with box size: Using the 5Å and the 10Å water shell box, the RMSF values are low. Usage of the smallest box (2.5Å shell) results in larger fluctuations. This may be due to direct interactions – via the periodic boundary conditions – between the epitope and the "bottom" (i.e. the part diametrically opposed to the binding cleft) of the simplified MHC. On the other hand, the RMSF graphs for the systems with a 20Å or a 40Å box show extremely high fluctuations of the peptide's C-terminal, alanine (residue 9). This indicates an interesting phenomenon: Simulating this rather artificial system revealed a detachment of the epitope's C-terminal end (a video of this simulation can be found in supplementary material S1). This finding strongly indicates that this artificial "single site binding" peptide is in fact not able to bind to the MHC. This is not surprising since there is no anchor residue at the C-terminal end.

For Peer Review Only

The salt bridges between the epitope's C-terminus and the MHC's F-pocket are just not stable enough to retain the hydrophobic epitope completely inside the hydrophilic binding groove. The gradual detachment of the epitope can be conveniently visualized by plotting the time dependence of the minimum distance between all the atoms of the peptide's C-terminal alanine and all the atoms of the MHC's β-sheet, which forms the bottom of the binding cleft (Figure 4).

Such a labile system is very sensible to different initial conditions, in particular if one uses too small water shell sizes. For each water shell thickness we performed four parallel simulations which were absolute identical, except for the randomly distributed initial velocities. These parallel simulations evolved very differently, as shown by the RMSF graph (Figure 3). For the system with a 20Å box, the peptide's C-terminus detached in two out of four runs. In the 40Å box, the peptide detached in even three out of four runs. Thus, performing only one simulation could produce misleading results.

Within the first nanosecond of simulation time detachment was observed with the 40Å and the 20Å water shell while smaller boxes (10Å and 5Å shell) forced the epitope to stay inside the binding groove. However, extending the simulation time for another 5ns the protein also detached within a 10Å water shell (data not shown). A larger water shell somehow speeds up the relaxation process and – together with long simulation times seems to be necessary for plausible results of pMHC molecular dynamics simulations.

3.1.2 Force fields, water models & system simplification

We simulated the "single site binding" peptide loaded MHC in a 20Å hydration shell with different force fields and different water models. Surprisingly, for some simulations, we found a stable binding behaviour. Figure 5 shows minimum distances between the peptide's C-terminal alanine and the MHC β-sheet as a measure for peptide detachment (see also Figure 4). Different force fields and recommended compatible water models were used: The GROMACS force field (which is based on a GROMOS force field) and the most recent

For Peer Review Only

GROMOS force field (GROMOS96 53a6) both combined with SPC water model, and the OPLS-AA/L force field combined with TIP4P water model, respectively. The OPLS-AA force field together with TIP4P water (shown in Figure 5c) yields such low distance values that one could conclude that this artificial peptide actually is a binding epitope for the HLA-B*2705 MHC. This is in severe contrast to the simulations using the GROMACS force field together with SPC water described in previous sections, where the C-terminal end of the peptide detached in two out of four simulations (Figure 5a). Using the GROMOS96 53a6 force field, detachment of the peptide was observed too (Figure 5b).

Regarding system simplifications we found almost no difference between the whole and the simplified pMHC complex, where the α3 and the β_2m domains have been left out. Neither the GROMACS force field (see Figure 5a and d) nor the GROMOS96 53a6 force field (see Figure 5b and e) show any effect of simplification.

Taking all in all we may conclude: Size of the solvation shell and the choice of water model / force field combination, in particular the water model, shows an important impact on simulation results. pMHC simplification, however, is shown to have less effect.

3.2 "Double site binding" peptide

The "double site binding" peptide (GRFAAAIAK) complexed to B*2705 MHC (PDB-id: 1jge) shows rather stable binding, and thus longer simulation runs are needed to study the influences of key parameters.

3.2.1 Solvation

We performed MD simulations for at least 12ns of the "double site binding" peptide loaded MHC, again systematically varying the solvation layer thickness from 12Å to 27Å. In addition to the properties described above, we also computed the rotational autocorrelation function (RACF) of the MHC. This was done to check if reorientation of the pMHC complex is hampered by an incomplete solvation shell, possibly in combination with the periodic

For Peer Review Only

boundary conditions. Orientation of the MHC molecule was defined by a vector perpendicular to the plane of three selected atoms belonging to the β-sheet forming the bottom of the groove (Cα atoms of V25, N97, and G100). The loss of orientational memory is characterized by the decay of the RACF (Equation 1, and Figure 6). Figure 6

Figure 6reveals that under periodic boundary conditions the loss of memory depends on the size of the water box. The results are only approximate since only one molecule was monitored for a rather short time, which may cause the non-exponential decays. However, a qualitative trend is clearly observable: While all three box sizes allow for rotation, rotational relaxation of the pMHC complex is less hindered in the 27Å system than in the 17Å system which in turn is less hindered than in the 12Å system.

The source of the observed rotational movement is rotational diffusion. The surrounding water hitting the pMHC complex randomly will cause to some extent a rotational movement of the complex. Figure 7 illustrates the way a smaller simulation box hinders the rotational movement of the pMHC complex: If the protein starts to rotate, the surrounding water network will follow the moving protein surface. In a large box there is less hindrance because the outer layers of the water network are almost unaffected by the rotation of the protein. Contrary, in the small box the rotation is more hindered because the outer water layers are bound to both, the protein and its periodic image. Thus, relatively tightly bound water molecules are forced to move in opposite directions, which effectively hinders rotation. Physically, tightly bound water means higher viscosity, resulting in longer relaxation times and thus less rotation.

3.2.2 Peptide's configurational substates

In our simulations of the MHC bound "double site binding" peptide within a 27Å water shell we observed that the bound peptide does not retain its original shape throughout the entire simulation run (16ns). It rather seems to switch back and forth between several different "substates" [19], which are temporarily stable (Figure 8a). Conformational substates of

http://mc.manuscriptcentral.com/tandf/jenmol

For Peer Review Only

proteins and polypeptides are also known from experiments [41]. Figure 8b shows the RMSD of Cα atoms for three specific residues. Some residues show low RMSD values throughout the simulation. Other residues show phases of high RMSD values at certain time intervals. One can conclude that several residues make small contributions to the peptide's total RMSD (Figure 8a), while others contribute significantly more. Hence the average RMSD, as usually displayed, hides the marked differences in mobility of atoms within the peptide. These specific movements and substates are made transparent using the following procedure:

We define the peptide's "principal" conformation as substate #1. The "principal" structure is defined by averaging the individual atomic positions between t=2ns and t=10ns and subsequently performing an "in vacuo" energy minimization. Then, positional deviations can conveniently be measured for each peptide's backbone atom relative to substate #1 (rather than relative to the crystal structure). Individual RMSD values can be shown grouped to a 3D plot over time (Figure 8d). The "hot spots" (tops and plateaus) indicate large deviations from the "principal" conformation and can be used to define further substates (see Figure 8e, Figure 9, and Table 2).

The observed substates can be characterized by several parameters (Table 2). Substate #2 (Figure 9, upper right corner) occurred very early and may be a reaction of the crystal structure to the surrounding water bath. Indeed, in this substate the centre part of the peptide (notably 6ALA) gets kinked up, causing the 6ALA residue not to face the lateral helix but protrude right up, away from the MHC. This movement generates a small hole beneath 6ALA, which allows for water molecules to enter the peptide-MHC interface. Actually, through the period of substate #2, the peptide-water interaction energy rises continually, while the peptide-MHC interaction energy decreases concomitantly (Figure 8c). In substate #1 ("principal" state – Figure 9, upper left corner) 6ALA is facing the lateral helix again and the hole is closed. During this state the peptide seems very stable and shows little fluctuations

For Peer Review Only

(RMSF). It remains in this substate for more than 8ns and comes back again to substate #1 after substate #3 and #4. Substate #3 (Figure 9, lower left corner) shows the C-terminal half of the peptide pointing away from the MHC. The peptide-water interaction energy increases again, causing a concomitant decrease in peptide-MHC interaction energy. The C-terminal residue 9LYS loses its interaction partner (116ASP in the β-sheet) and finds a new one (77ASP in the lateral α-helix). In substate #4 (Figure 9, lower right corner) the peptide lifts even more. The residue of 9LYS has to be extended almost totally to stay in contact with the F pocket (77ASP). More than 42% of the peptide's surface is exposed to water during this substate – a remarkable increase compared to 32.7% during substate #1. After this short period the peptide lowers again, 9LYS contacts 116ASP as before, and the peptide adopts the substate #1 for the rest of simulation time.

For Peer Review Only

4 Discussion

Conclusions from molecular dynamics simulations may reveal important mechanisms but can be delicate. A pMHC system is too large for a "full simulation", given the computational power available today. However, cautious choice of parameters still allows drawing conclusions from simulations. In this study we analyzed following key parameters for molecular dynamics simulations of pMHC molecules:

1) Initial conditions

   Minor differences in initial conditions can have a major effect on the result, in particular for labile systems as shown in the parallel simulations of the MHC bound "single site binding" peptide. Therefore, parallel simulations of identical structures starting from different initial conditions can help to identify labile systems. This seems crucial to avoid premature conclusions – in our case regarding weak binding or non binding epitopes.

2) Size of solvation shell

   Using periodic boundary conditions, the simulated system is embedded as the unit cell of a "supercrystal" within copies of itself. However, this crystalline environment may induce artefacts when the protein is comparable in size to the simulation box. Usually a bigger water shell minimizes these artefacts due to shielding effects of water: a polar solvent generates an electrostatic reaction field which screens the charges and dipoles [20]. We demonstrated the effects of a bigger water shell by several simulations of the MHC bound "single site binding" peptide. Using a small water shell (5Å or 10Å), the effect of the "supercrystal" was seen strong enough to artificially stabilize the whole structure: no detachment of the epitope was observed. Using a sufficiently large water shell (20Å), detachment of the epitope can be observed.

For Peer Review Only

Even if the system is large enough to solvate the whole protein, rotability might still be hindered. Analyzing simulations of the MHC bound "double site binding" peptide, we found that small water shells obviously allow for a direct interaction of the periodic images of the complexes through the water network, thereby reducing rotability. A sufficiently large water shell is mandatory to minimize water network interactions and enable realistic rotability.

Concluding, an enlargement of the water shell raises the computational workload but yields more realistic dynamics of the whole system.

3) Water model / force field combination

Using different models for force field and water representation induces drastic differences. Since SPC water is distinctly polar ($\varepsilon_r=65$, M.Neumann, unpublished), it is highly able to substitute the binding partners of the epitope and facilitates the "single site binding" epitope's detachment. With a less polar water model, like TIP4P ($\varepsilon_r=53\pm2$, [42]), solvation of the "single site binding" epitope is not able to disrupt its binding to the MHC: the epitope stays in the binding groove, displaying very low RMSD values. Rather "inert" water causes a pseudo-stabilization of the pMHC complex. Additionally, the capability of shielding electrostatic forces (see point 2) depends on the relative dielectric permittivity ε_r [20].

Usage of different force fields and water models provides different insights into the system's dynamics. Since there is no "one and only" model, variation is essential for reliable results. However, for a particular simulation experiment we can propose a guideline to choose a model that allows "staying on the safe side":

Simulations using a GROMOS-type force field with SPC water will tend to predict a "border case" epitope as non-binding. In contrast, the TIP4P water coming with the OPLS-AA force field shows a certain tendency that "border case" epitopes may be

For Peer Review Only

categorized as binding. Thus, when testing for binding (e.g. vaccine design) take the model that proved "looser" binding, and when testing for non-binding (e.g. allergenicity) take the model that proved "tighter" binding.

4) Simulation time

Long simulation runs are needed to observe all (or at least the most) relevant thermally accessible configurations. Sometimes a few hundred picoseconds simulation time will suffice. In other systems, representative sampling will take several seconds, which may be out of reach for molecular dynamics studies for a long time to come. Since we discovered the epitope within the MHC binding groove adopting different substates on the timescale of nanoseconds, pMHC simulations have to be at least several tens or even hundreds of nanoseconds long. However, in many cases it remains hard to decide whether a particular simulation has run long enough.

5) System simplification

The legitimacy of MHC simplification in MD simulations has already been covered by various authors [43;44] and there is no consensus about its effect on peptide dynamics on a timescale of nanoseconds. Our simulations show almost no difference between the simplified pMHC complex, where the α3 and the β_2m are left out of simulation, and the whole pMHC complex, neither when using the GROMACS force field (see Figure 5a and d) nor when using the GROMOS96 53a6 force field (see Figure 5b and e).

The computational discovery of the epitope's substates raises some very interesting questions regarding a possible immunological meaning: The "outbreaking" characteristic of substate #4 (C-terminal half slightly protruding from the cleft) could represent the very trigger for a TCR screening the pMHC complex. This hypothesis fits well into the hitherto known framework: Although T cell activation is to some extent correlated to the half-life of pMHC-TCR interaction, half-life alone is insufficient to accurately predict activation potencies. In

For Peer Review Only

addition, conformational changes within the TCR-peptide-MHC interface may essentially contribute to T cell activation [45]. Pöhlmann et al. [46] and Starikov et al. [47] suggested that even a more flexible peptide may have a noticeable effect on T cell activation. This is precisely the point where the presented simulations are capable to provide additional insight and evidence: In substate #4 the peptide is not only more flexible (RMSF of 0.045 nm instead of 0.037 nm in substate #1) but there is also an actual movement of the total C-terminal half of the peptide out of the cleft, towards a possible TCR. This conformational change in the peptide could be transmitted to a bound TCR, possibly contributing to T cell activation [48].

Some authors have described peptide flexibilities, but the current pMHC-analysis is – to the best of our knowledge – the first to describe substates of the MHC bound peptide. Pöhlmann et al. [49] and Starikov et al. [50] performed MD simulations of the same system (PDB-id: 1jge). However, they observed no specifiable conformational substates for the bound peptide. Pöhlmann et al. used the OPLS-AA/L force field and TIP4P water model and Starikov et al. used a 7 Å water shell, both may suffer from a limited configurational sampling.

Further simulation studies can be envisaged to find out whether the existence of “protruding substates” actually correlates to the activation potency of TCRpMHC complexes, as obtained from experimental measurements. The present work evaluates the computational tools and prepares their future application in the prediction of “protruding substates” for a given pMHC complex. In addition, single peptide residues could be exchanged and the impact on “protruding substates” computationally evaluated for vaccine design.

For Peer Review Only

5 Acknowledgments

We are indebted to Prof. Dr. Anneliese Schimpl for continuous support and advice in preparing the manuscript.

The computational work for the current investigation has been supported by Prof. Dr. Rudolf Karch, Core Unit for Medical Statistics and Informatics, Medical University of Vienna, Austria.

For Peer Review Only

6 References

[1] M. Karplus and G.A. Petsko, *Molecular dynamics simulations in biology*, Nature 347 (1990), pp. 631-639.

[2] T. Hansson, C. Oostenbrink, and W. van Gunsteren, *Molecular dynamics simulations*, Curr. Opin. Struct. Biol. 12 (2002), pp. 190-196.

[3] D. Rognan, L. Scapozza, G. Folkers, and A. Daser, *Molecular dynamics simulation of MHC-peptide complexes as a tool for predicting potential T cell epitopes*, Biochemistry 33 (1994), pp. 11476-11485.

[4] D. Rognan, S. Krebs, O. Kuonen, J.R. Lamas, J.A. Lopez de Castro, and G. Folkers, *Fine specificity of antigen binding to two class I major histocompatibility proteins (B*2705 and B*2703) differing in a single amino acid residue*, J. Comput. Aided Mol. Des 11 (1997), pp. 463-478.

[5] H. Toh, N. Kamikawaji, T. Tana, T. Sasazuki, and S. Kuhara, *Molecular dynamics simulations of HLA-DR4 (DRB1*0405) complexed with analogue peptide: conformational changes in the putative T-cell receptor binding regions*, Protein Eng 11 (1998), pp. 1027-1032.

[6] T. Pöhlmann, R.A. Bockmann, H. Grubmuller, B. Uchanska-Ziegler, A. Ziegler, and U. Alexiev, *Differential peptide dynamics is linked to major histocompatibility complex polymorphism*, J. Biol. Chem. 279 (2004), pp. 28197-28201.

[7] E.B. Starikov, L. Nilsson, and M. Hulsmeyer, *A single residue exchange between two HLA-B27 alleles triggers increased peptide flexibility*, Eur. Biophys. J. 33 (2004), pp. 651-655.

[8] U. Omasits, M. Neumann, O. Steinhauser, R. Valenta, R. Kobler, and W. Schreiner, *Molecular Dynamics Simulations in Immunology*, in *Proceedings of the 2nd Austrian Grid Symposium*, J. Volkert, T. Fahringer, D. Kranzlmüller, and W. Schreiner, eds., Oesterreichische Computer Gesellschaft, Vienna, 2007, pp. 81-90.

[9] W.S. Meng, H. von Grafenstein, and I.S. Haworth, *A model of water structure inside the HLA-A2 peptide binding groove*, Int. Immunol. 9 (1997), pp. 1339-1346.

[10] P.M. Petrone and A.E. Garcia, *MHC-peptide binding is assisted by bound water molecules*, J. Mol. Biol. 338 (2004), pp. 419-435.

[11] M. Zacharias and S. Springer, *Conformational flexibility of the MHC class I alpha1-alpha2 domain in peptide bound and free states: a molecular dynamics simulation study*, Biophys. J. 87 (2004), pp. 2203-2214.

[12] O. Michielin and M. Karplus, *Binding free energy differences in a TCR-peptide-MHC complex induced by a peptide mutation: a simulation analysis*, J. Mol. Biol. 324 (2002), pp. 547-569.

[13] V. Zoete and O. Michielin, *Comparison between computational alanine scanning and per-residue binding free energy decomposition for protein-protein association using*

For Peer Review Only

MM-GBSA: Application to the TCR-p-MHC complex, Proteins 67 (2007), pp. 1026-1047.

[14] O. Michielin and M. Karplus, *Binding free energy differences in a TCR-peptide-MHC complex induced by a peptide mutation: a simulation analysis*, J. Mol. Biol. 324 (2002), pp. 547-569.

[15] S. Wan, P.V. Coveney, and D.R. Flower, *Molecular basis of peptide recognition by the TCR: affinity differences calculated using large scale computing*, J. Immunol. 175 (2005), pp. 1715-1723.

[16] V. Zoete and O. Michielin, *Comparison between computational alanine scanning and per-residue binding free energy decomposition for protein-protein association using MM-GBSA: Application to the TCR-p-MHC complex*, Proteins 67 (2007), pp. 1026-1047.

[17] T. Darden, D. York, and L. Pedersen, *Particle mesh Ewald: An N.log(N) method for Ewald sums in large systems*, J. Chem. Phys. 98 (1993), pp. 10089-10092.

[18] T.E. Cheatham and B.R. Brooks, *Recent advances in molecular dynamics simulation towards the realistic representation of biomolecules in solution*, Theor. Chem. Acc. 99 (1998), pp. 279-288.

[19] B.K. Andrews, T. Romo, J.B. Clarage, B.M. Pettitt, and J. Phillips, *Characterizing global substates of myoglobin*, Structure 6 (1998), pp. 587-594.

[20] D. Sengupta, R.N. Behera, J.C. Smith, and G.M. Ullmann, *The alpha helix dipole: Screened out?*, Structure 13 (2005), pp. 849-855.

[21] A. Wada, *The alpha-helix as an electric macro-dipole*, Advances in Biophysics (1976), pp. 1-63.

[22] W. Weber, P.H. Hünenberger, and J.A. McCammon, *Molecular dynamics simulations of a polyalanine octapeptide under Ewald boundary conditions: Influence of artificial periodicity on peptide conformation*, J. Phys. Chem. B 104 (2000), pp. 3668-3675.

[23] G.M. Torrie and J.P. Valleau, *Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling*, J. Comput. Phys. 23 (1977), pp. 187-199.

[24] D. Rognan, N. Zimmermann, G. Jung, and G. Folkers, *Molecular dynamics study of a complex between the human histocompatibility antigen HLA-A2 and the IMP58-66 nonapeptide from influenza virus matrix protein*, Eur. J. Biochem. 208 (1992), pp. 101-113.

[25] S. Wan, P. Coveney, and D.R. Flower, *Large-scale molecular dynamics simulations of HLA-A*0201 complexed with a tumor-specific antigenic peptide: can the alpha3 and beta2m domains be neglected?*, J. Comput. Chem. 25 (2004), pp. 1803-1813.

[26] The Austrian Grid Initiative, http://www.austriangrid.at, cited 2008

[27] H.J.C. Berendsen, D. van der Spoel, and R. van Drunen, *GROMACS: A message-passing parallel molecular dynamics implementation*, Comput. Phys. Commun. 91 (1995), pp. 43-56.

For Peer Review Only

[28] E. Lindahl, B. Hess, and D. van der Spoel, *GROMACS 3.0: a package for molecular simulation and trajectory analysis*, J. Mol. Model. 7 (2001), pp. 306-317.

[29] G. Kingsley and J. Sieper, *Current perspectives in reactive arthritis*, Immunol. Today 14 (1993), pp. 387-391.

[30] H.M. Bergman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, and P.E. Bourne, *The Protein Data Bank*, Nucleic Acids. Res. 28 (2000), pp. 235-242.

[31] D.R. Madden, J.C. Gorga, J.L. Strominger, and D.C. Wiley, *The three-dimensional structure of HLA-B27 at 2.1 A resolution suggests a general mechanism for tight peptide binding to MHC*, Cell 70 (1992), pp. 1035-1048.

[32] M. Hulsmeyer, R.C. Hillig, A. Volz, M. Ruhl, W. Schroder, W. Saenger, A. Ziegler, and B. Uchanska-Ziegler, *HLA-B27 subtypes differentially associated with disease exhibit subtle structural alterations*, J. Biol. Chem. 277 (2002), pp. 47844-47853.

[33] J. Zielkiewicz, *Structural properties of water: comparison of the SPC, SPCE, TIP4P, and TIP5P models of water*, J. Chem. Phys. 123 (2005), pp. 1-6.

[34] D. van der Spoel, E. Lindahl, B. Hess, A.R. van Buuren, E. Apol, P.J. Meulenhoff, D.P. Tieleman, A.L.T.M. Sijbers, K.A. Feenstra, R. van Drunen, and H.J.C. Berendsen, *Gromacs User Manual version 3.2*, 2004.

[35] C. Oostenbrink, A. Villa, A.E. Mark, and W.F. Van Gunsteren, *A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6*, J. Comput. Chem. 25 (2004), pp. 1656-1676.

[36] G.A. Kaminski, R.A. Friesner, J. Tirado-Rives, and W.L. Jorgensen, *Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides*, J. Phys. Chem. B 105 (2001), pp. 6474-6487.

[37] H.J.C. Berendsen, J.P.M. Postma, W.F. Van Gunsteren, A. DiNola, and J.R. Haak, *Molecular dynamics with coupling to an external bath*, J. Chem. Phys. 81 (1984), pp. 3684-3690.

[38] J.P. Ryckaert, G. Ciccotti, and H.J.C. Berendsen, *Numerical integration of the cartesian equations of motion of a system with constraints; molecular dynamics of n-alkanes.*, J. Comput. Phys. 23 (1977), pp. 327-341.

[39] F. Eisenhaber, P. Lijnzaad, P. Argos, C. Sander, and M. Scharf, *The Double Cube Lattice Method: Efficient Approaches to Numerical Integration of Surface Area and Volume and to Dot Surface Contouring of Molecular Assemblies*, J. Comput. Chem. 16 (1995), pp. 273-284.

[40] M. Karplus and G.A. Petsko, *Molecular dynamics simulations in biology*, Nature 347 (1990), pp. 631-639.

[41] H. Frauenfelder, F. Parak, and R.D. Young, *Conformational substates in proteins*, Annual review of biophysics and biophysical chemistry 17 (1988), pp. 451-479.

For Peer Review Only

[42] M. Neumann, *Dielectric relaxation in water. Computer simulations with the TIP4P potential*, J. Chem. Phys. 85 (1986), pp. 1567-1580.

[43] D. Rognan, N. Zimmermann, G. Jung, and G. Folkers, *Molecular dynamics study of a complex between the human histocompatibility antigen HLA-A2 and the IMP58-66 nonapeptide from influenza virus matrix protein*, Eur. J. Biochem. 208 (1992), pp. 101-113.

[44] S. Wan, P. Coveney, and D.R. Flower, *Large-scale molecular dynamics simulations of HLA-A*0201 complexed with a tumor-specific antigenic peptide: can the alpha3 and beta2m domains be neglected?*, J. Comput. Chem. 25 (2004), pp. 1803-1813.

[45] M. Krogsgaard, N. Prado, E.J. Adams, X.L. He, D.C. Chow, D.B. Wilson, K.C. Garcia, and M.M. Davis, *Evidence that structural rearrangements and/or flexibility during TCR binding can contribute to T cell activation*, Mol. Cell 12 (2003), pp. 1367-1378.

[46] T. Pöhlmann, R.A. Bockmann, H. Grubmuller, B. Uchanska-Ziegler, A. Ziegler, and U. Alexiev, *Differential peptide dynamics is linked to major histocompatibility complex polymorphism*, J. Biol. Chem. 279 (2004), pp. 28197-28201.

[47] E.B. Starikov, L. Nilsson, and M. Hulsmeyer, *A single residue exchange between two HLA-B27 alleles triggers increased peptide flexibility*, Eur. Biophys. J. 33 (2004), pp. 651-655.

[48] J.B. Reiser, C. Gregoire, C. Darnault, T. Mosser, A. Guimezanes, A.M. Schmitt-Verhulst, J.C. Fontecilla-Camps, G. Mazza, B. Malissen, and D. Housset, *A T cell receptor CDR3beta loop undergoes conformational changes of unprecedented magnitude upon binding to a peptide/MHC class I complex*, Immunity. 16 (2002), pp. 345-354.

[49] T. Pöhlmann, R.A. Bockmann, H. Grubmuller, B. Uchanska-Ziegler, A. Ziegler, and U. Alexiev, *Differential peptide dynamics is linked to major histocompatibility complex polymorphism*, J. Biol. Chem. 279 (2004), pp. 28197-28201.

[50] E.B. Starikov, L. Nilsson, and M. Hulsmeyer, *A single residue exchange between two HLA-B27 alleles triggers increased peptide flexibility*, Eur. Biophys. J. 33 (2004), pp. 651-655.

[51] M.A. Saper, P.J. Bjorkman, and D.C. Wiley, *Refined structure of the human histocompatibility antigen HLA-A2 at 2.6 A resolution*, J. Mol. Biol. 219 (1991), pp. 277-319.

[52] W. Humphrey, A. Dalke, and K. Schulten, *VMD: visual molecular dynamics*, J. Mol. Graph. 14 (1996), pp. 33-38.

[53] M. Hulsmeyer, R.C. Hillig, A. Volz, M. Ruhl, W. Schroder, W. Saenger, A. Ziegler, and B. Uchanska-Ziegler, *HLA-B27 subtypes differentially associated with disease exhibit subtle structural alterations*, J. Biol. Chem. 277 (2002), pp. 47844-47853.

For Peer Review Only

7 Tables

Table 1. Overview of molecular dynamics simulations performed. Peptide main anchor residues are printed bold.

peptide	MHC	water model / force field combination	size of water box	parallel simulations	simulation time
A**R**AAAAAAA (“single site binding peptide”)	B*2705 whole	SPC / GROMACS	20Å	4x	1ns
	B*2705 whole	SPC / GROMOS 53a6	20Å	4x	1ns
	B*2705 simplified	SPC / GROMACS	2.5Å	4x	1ns
	B*2705 simplified	SPC / GROMACS	5Å	4x	1ns
	B*2705 simplified	SPC / GROMACS	10Å	4x	1ns
	B*2705 simplified	SPC / GROMACS	10Å	1x	10ns
	B*2705 simplified	SPC / GROMACS	20Å	4x	1ns
	B*2705 simplified	SPC / GROMACS	40Å	4x	1ns
	B*2705 simplified	SPC / GROMOS 53a6	20Å	4x	1ns
	B*2705 simplified	TIP4P / OPLS-AA	20Å	4x	1ns
G**R**FAAAIA**K** (“double site binding peptide”)	B*2705 simplified	SPC / GROMACS	12Å	1x	12ns
	B*2705 simplified	SPC / GROMACS	17Å	1x	12ns
	B*2705 simplified	SPC / GROMACS	27Å	1x	16ns

For Peer Review Only

Table 2. Characteristics of the "double site binding" peptide in crystal structure and in different substates occurring during the 16ns simulation. SASA is taken relative to SASA of the peptide in water.

configuration	time (ps)	mainly affected residues	SASA (relative)	interaction energy (kJ/mol)		RMSF (nm)
				peptide-water	peptide-MHC	
crystal	0		29.7%	$-3.8\ 10^{7}$	$-8.8\ 10^{7}$	0.037
substate #1	2,000 - 10,000		32.7%	$-4.4\ 10^{7}$	$-8.3\ 10^{7}$	0.037
substate #2	890 - 1,550	5A, 6A, 7I	36.8%	$-4.3\ 10^{7}$	$-8.5\ 10^{7}$	0.045
substate #3	11,050 - 12,500	8A, 9K	39.9%	$-5.1\ 10^{7}$	$-7.5\ 10^{7}$	0.038
substate #4	13,550 - 14,100	5A, 6A, 7I, 8A, 9K	42.3%	$-5.2\ 10^{7}$	$-7.0\ 10^{7}$	0.045

For Peer Review Only

8 Figure captions

Figure 1. HLA-B*2705 molecular surface and its binding pockets.

The pockets definition is according to [51]. Pocket A (M5, Y7, Y59, E63, Y159, E163, W167, Y171), pocket B (H9, T24, E45, L66, C67, Y99), pocket C (H9, K70, T73, D74, R97), pocket D (Y99, H114, L156, Y159, L160), pocket E (H114, W133, W147, V152, L156), pocket F (D77, T80, L81, Y84, D116, Y123, T143, K146, W147). The peptide's primary anchors P2 (conserved, Arg) and P9 form salt bridges and hydrogen bonds to the complementary negatively charged MHC pockets B and F, respectively. P3 interacts with the upper part of the hydrophobic pocket D. This one and all subsequent graphical representations of molecules were produced using the VMD package [52].

Figure 2. Crystal structure and simulated system.

*a) Crystal structure of a pMHC complex (m9 peptide complexed to HLA-B*2705 [53]). The peptide binding domain (α1 and α2 domain, white), the α3 domain (grey) and the associated β-2-microglobulin (black) are represented as "cartoons". The bound peptide is shown as "licorice". The dimensions of the MHC's peptide binding domain are given in Ångström.*

b) A simplified pMHC complex placed in the centre of a simulation box allowing for a 20 Ångström solvation shell resulting in altogether 99,101 atoms (water molecules in the front of the cube are carved out in order to make the pMHC complex visible). This system could be used as input for a MD simulation "as is".

Figure 3. RMSF of peptide residues' backbone atoms.

For Peer Review Only

We performed four independent 1ns runs (same starting structure, different starting velocities) for each simulation box size. The RMSF values shown are averages for the time interval from 600ps to 1000ps, which represents the equilibrated part of a stable run. Results for the individual runs are indicated by different line types. Low RMSF values indicate low positional fluctuations of the backbone atoms during the simulation, which means that these residues are likely to be tightly bound by the MHC. A residue displaying high RMSF values may be loosely bound or not bound at all.

Figure 4. Demonstration of the minimum distance between epitope C-terminus and MHC β-sheet.

a) Minimum distances of the epitope's C-terminal alanine to the MHC's β-sheet during 1ns simulations. In the simulation of the complex with a 20Å surrounding water shell (dashed line) the distance first rises gradually from 0.37nm to 0.7nm. Then the C-terminus totally detaches and moves as far as 1.4nm away. Although there are large fluctuations for this distance, in the simulation employing a 5Å water shell (continuous line) no detachment occurs during the 1ns run.

b) The dashed line marks the minimum distance of all the atoms of the epitope's C-terminal alanine to all the atoms of the MHC's β-sheet. The orange (light grey) peptide shows the distance in the crystal structure. The blue (dark grey) peptide depicts the increased distance after C-terminal detachment. For an animated display of the simulation please refer to (supplementary material S1).

Figure 5. Minimum distance between peptide C-terminus and MHC β-sheet during different simulations of the "single site binding" peptide.

For Peer Review Only

Three different force fields and two different water models were used for simulations of the simplified pMHC complex, and two different force fields with the same water model were used for the whole pMHC complex. Each combination was simulated in four independent runs for 1ns within a 20Å hydration shell (black, dark grey, grey, and light grey line).

a) Simulating the simplified pMHC complex enclosed in a 20Å hydration shell results for two out of the four simulations (black and dark grey) in the previously described partial detachment of the peptide (see Figure 4) when the GROMACS force field and the SPC water model are used. The rising distance of the peptide's C-terminal alanine residue to 1nm and above indicates total detachment from the MHC binding cleft.

b) Application of the GROMOS96 53a6 force field on the simplified system leads to peptide detachment in one out of four simulations (black).

c) When the OPLS-AA force field and the TIP4P water model are used on the simplified system, for all four simulations the peptide calmly resides in the binding groove and the minimum distance measure retains its initial low value for the entire length of the 1ns simulation.

d) Simulating the whole pMHC complex using the GROMACS force field and the SPC water model results in an onset of peptide detachment in two out of four simulations (black and dark grey).

e) Using the GROMOS96 53a6 force field on the whole pMHC complex leads to peptide detachment in one out of four simulations (black).

For Peer Review Only

Figure 6. Orientational relaxation (RACF) of the pMHC complex for different sizes of the simulation box.

All the pMHC systems lose their orientational "memory" over time, which is shown by a decay of the RACF $C_1(t)$. The larger the surrounding water shell the faster the loss of memory. All three systems have been simulated for at least 12ns.

Figure 7. Model for system size influencing rotational diffusion.

Both illustrations show a simulation box containing a protein solved in water using periodic boundaries but with different simulation box sizes. The darker the water molecules, the more structured is the water network. A small box hinders rotation by forcing a relatively tightly bound water layer to move in opposite directions.

Figure 8. Analysis of the long time simulation (16ns) of the "double site binding" peptide complexed to HLA-B*2705 (1jge) within a 27Å water box.

a) RMSD of the peptide's backbone (red line) and the MHC's backbone (black line) relative to the initial crystal structure. The peptide spends most of the time in a state (characterized by low RMSD values) that is very similar to the crystal structure, but seems to flip to a different conformational substates around simulation time t≈1ns and between t≈10ns and t≈14ns.

b) RMSD of specific peptide backbone atoms. The Cα atom of the primary anchor arginine (green line) remains close to its initial position throughout the simulation (low RSMD values). The Cα atom of the central alanine (position 5, blue line) temporarily moves farther off its position in the crystal structure after t≈1ns (higher RMSD values) while the Cα atom of the secondary anchor lysine (black line) undergoes an even larger dislocation (large RMSD values) between t≈10ns and t≈14ns. Therefore the peptide's residues differently contribute to the substates. Individual RMSD curves for all the peptide's Cα atoms can be found in (supplementary material S2).

c) Interaction energies of the peptide with the MHC (black line) and with water (grey line). Energies are the sum of Coulomb and Lennard Jones potential.

d) RMSD of each epitope's backbone atom relative to substate #1. This plot shows the substates and backbone atoms involved.

e) Conformational substates defined using above graphs.

For Peer Review Only

Figure 9. Average structures of peptide's conformational substates.

While the white peptide representations depict the substate #1, the grey representations depict the average structures of corresponding substates. The MHC structure shown is an average structure over the whole simulation. Arrows indicate major structural deviations to substate #1

For Peer Review Only

9 Word count

33 pages

8237 words

45002 characters

For Peer Review Only

Analysis of Key Parameters for Molecular Dynamics of pMHC Molecules

Ulrich Omasits[a], Bernhard Knapp[a], Martin Neumann[b], Othmar Steinhauser[c], Hannes Stockinger[d], Rene Kobler[e], and Wolfgang Schreiner[a]*

For Peer Review Only

The work described in this paper is supported by the Austrian Grid Project, funded by the Austrian BMBWK (Federal Ministry for Education, Science and Culture) under contract GZ 4003/2-VI/4c/2004.

[a] Core Unit for Medical Statistics and Informatics, Medical University of Vienna, Austria

[b] Institute of Experimental Physics, University of Vienna, Austria

[c] Department of Computational Biological Chemistry, University of Vienna, Austria

[d] Department of Molecular Immunology, Centre of Biomolecular Medicine and Pharmacology, Medical University of Vienna, Austria

[e] Institute of Graphics and Parallel Processing, Johannes Kepler University Linz, Austria

[*] Corresponding author: fax: +43 1 40400-6677 – phone: +43 1 40400-6673– email: bernhard.knapp@meduniwien.ac.at – address: Spitalgasse 23, 1090 Vienna, Austria

Abstract

Molecular dynamics (MD) studies of human major histocompatibility complex (MHC) HLA-B*2705 complexing two different peptides were performed. During simulation one peptide partially detached from the MHC while the other peptide switched back and forth between several different configurations. These different configurations relate to conformational substates and can be assigned to different levels of chemical activity or even the molecular mechanisms of immunological signalling. To ensure reliable immunological conclusions from MD simulations we prepare the methodological tools by carefully evaluating initial conditions, system simplification, solvation shell thickness, water model / force field combination, and simulation length. We also derive a guideline for appropriate model selection. This kind of quality assessment is seen a mandatory prerequisite for coming studies linking pMHC dynamics to T cell activation.

Keywords: HLA; major histocompatibility complex; peptide dynamics; conformational substates; T cell activation

Abbreviations: β_2m: β_2-microglobulin – HLA: human leukocyte antigen – MD: molecular dynamics – MHC: major histocompatibility complex – PBC: periodic boundary conditions – PME: Particle Mesh Ewald – pMHC: peptide loaded MHC – RACF: rotational autocorrelation function – RMSD: root mean square deviation – RMSF: root mean square fluctuation – SASA: solvent accessible surface area – TCR: T cell receptor

For Peer Review Only

1 Introduction

Class I major histocompatibility complexes (MHC) bind antigenic peptides within the α1-α2 domain in a long narrow cleft located between two α-helices on the top of an antiparallel β-sheet. Binding is mainly mediated through sequence independent contacts of the peptide's charged N- and C-terminus to conserved "pockets" (Figure 1) in the MHC's binding groove. The contacts are maintained by networks of hydrogen bonds that are very similar in all class I pMHC complexes and make the largest contribution to the binding energy of the peptide. In addition, polymorphic pockets, so-called primary anchor regions, interact with specific sequence dependent amino acids (anchor residues) of the peptide and thus are responsible for the allele specificity of the recognition process. The central part of the peptide bulges out of the binding groove and serves as primary recognition site for the T cell receptor (TCR) (see Figure 2a).

Molecular dynamics (MD) is the technique of numerically solving the equations of motion of an assembly of particles (atoms). The total force acting on each particle depends on the positions and properties of all other particles in the system and, in addition, may include externally imposed forces. MD programs calculate these forces and solve Newton's equation of motion for each particle. For details refer to reviews on MD of biomolecules [1;2].

Various MD studies on pMHC have been reported in the literature [3-14]. All these MD studies use some force field, some water model, some initial condition, and some size of water box, in most cases without a particular justification or comment on their choices. The aim of this study is to vary these parameters and analyze their effects and possible pitfalls regarding a pMHC complex simulation. To our knowledge no such detailed investigation has been reported in the literature. We are going to answer the following questions:

For Peer Review Only

- Should parallel simulations be performed using different initial conditions?
- What size of solvation box should be used?
- Which force field and water model should be chosen?
- What is the minimum simulation time necessary?
- Are system simplifications admissible?

Our aim is to evaluate those key parameters giving a reasonable chance to cope with the rather large pMHC molecule in a meaningful way. **Studies to reliably evaluate the performance of MD simulations have been carried out in smaller systems, based on sufficient sampling over configurational space. As opposed to this, the present study deals with a very large system which is interesting for medical reasons and the task can only be to find out to which extent reliability can be obtained.** All in all, the present work lays the basis for MD as a method of investigation in pMHC interactions.

Besides the technical evaluation we also present a minute mobility analysis of a peptide adopting different configurational substates within the MHC-cleft.

1.1 Environment representation in MD

In order to achieve a manageable representation of the natural pMHC complex a trade-off between computational cost and simulation accuracy has to be made, especially concerning the amount of solvation water included in the simulation, the boundary conditions, and protein simplifications. Since proteins evolved to function in aqueous environments, water molecules binding directly to the protein and forming the first layer of the hydration shell may be considered as an integral part of protein structure. It is also clear that solute-solvent interactions extend beyond the first hydration layer. Although an extended hydration shell may be represented by simplified models, the most reliable results on structure and dynamics are achieved by inclusion of a sufficient number of additional explicit solvent molecules. To

http://mc.manuscriptcentral.com/tandf/jenmol

For Peer Review Only

avoid an abrupt border with a vacuum, periodic boundary conditions (PBC) repeat the simulation box periodically in all three space dimensions, creating a continuous system. PBC allows a very useful method for calculating long-range electrostatics: the Particle Mesh Ewald (PME) summation [15]. Currently PBC/PME represents the standard in MD simulations of biomolecules since it is the most accurate method. On the other hand, the imposition of periodicity may lead to artefacts since all atoms now may interact with their periodic images. These artefacts can be minimized using larger solvation shells and an appropriate water model / force field combination, i.e. a combination to stay on the "safe side" for a particular simulation experiment (see Discussion section).

1.1.1 Size of solvation shell

A minimal system is claimed to need a 5-15Å thick solvation shell [16]. In other words, the system needs to be large enough to allow for the formation of a 5-15Å "buffer zone" between the solute and the boundaries of the simulation box. We show that for pMHC simulations a solvation layer thickness of more than 10Å, ideally 20Å, is necessary (Figure 2b). It is known that the aqueous environment facilitates the protein sampling the conformational space [17]. Water, due to its high relative dielectric permittivity, is very effective in screening long-range electrostatic interactions and dipoles. Thus the thickness of the solvation shell essentially determines the degree of electrostatic shielding, which is seen crucial in the presence of significant dipoles induced by the α-helices of the MHC [18;19]. PBC/PME imposes an artificial periodicity on the grossly inhomogeneous solute-solvent system. This periodicity limits the configurational flexibility of a protein [20] by overstabilization of the starting configuration which is normally taken to be that of a periodic crystal. Obviously, the larger the simulation box, the better the water molecules in the solvation shell (and the protein itself) will be screened from interactions with the protein's periodic images. Thus crystalline effects are minimized, sampling is less hindered [21], and a more realistic dynamics of the system can be observed.

For Peer Review Only

1.1.2 Structural simplifications

It is quite common to simplify a simulation system in order to minimize the number of atoms. In the case of pMHC complexes frequently only the MHC's peptide binding domain (α1 and α2 domain) is simulated since this is the site of interesting peptide-MHC interactions. In vivo such a truncated pMHC complex would not be stable at all and unfolding would occur. MHC simplification has been found not to alter the results of MD simulations on the timescale feasible [22]. However, there is disagreement about this simplification and there are authors [23] claiming that the α3 domain and the β_2m subunit are essential for peptide MHC interactions even within one nanosecond simulation time. According to our results, simplification plays a minor role. The α3 domain and the β_2m subunit are certainly essential for a simulation of a system in its full complexity. However, we are interested in processes taking place within the binding groove at a much shorter timescale than MHC unfolding would occur.

For Peer Review Only

2 Materials and Methods

2.1 Hard- and Software

The project was carried out on a SGI Altix 350 system within the framework of the Austrian Grid [24] and on a computational cluster of the Technical University of Vienna. For a typical 2ns simulation run a total of 800 CPU hours are required.

We use GROMACS [25;26] because of its detailed documentation of computational methodology and very fast procedure to treat water molecules. Its high scalar performance outweighed the fact that (on our system) GROMACS scales efficiently only up to 16 nodes.

2.2 Setup of coordinates

For simulations we selected the MHC class I B*2705 protein. Almost all binding epitopes for this MHC show an arginine at position 2, defined as a dominant anchor residue. Positions 1, 3, and 9 provide the remaining main anchoring positions with a preference for hydrophobic amino acids at P3, and hydrophobic or positively charged amino acids at P9. The dominant anchor residue P2 binds deep into the MHC pocket B, while P9 binds into MHC pocket F (see Figure 1). HLA-B27 subtypes are strongly associated with the occurrence of several diseases, in particular inflammatory diseases of the joints called spondyloarthropathies [27] and thus lends itself as a most interesting candidate for computational screening.

Starting coordinates were taken from crystal structures deposited in the Protein Data Bank [28]. We chose two structures representing the same MHC complexing two different, but substantially similar peptides. The first structure was obtained by Madden et al. from an epitope mix. As a “minimal binding” peptide the consensus backbone was modelled with the epitope ARAAAAAAA. The resulting structure was deposited as entry 1hsa [29]. The second structure containing the epitope GRFAAAIAK was obtained by Hülsmeyer et al. and deposited as entry 1jge [30]. We chose those two complexes due to their chemically similar

http://mc.manuscriptcentral.com/tandf/jenmol

For Peer Review Only

epitopes. Both consist of mainly hydrophobic residues (A,F,I,G) and both have the R anchor at position 2, typical for HLA-B*2705 epitopes. The crucial difference is the secondary anchor (K) in 1jge at position 9 which is missing in 1hsa. Therefore we will refer to the epitope of 1jge as "double site binding" peptide and to the epitope of 1hsa as "single site binding" peptide.

For most simulations only the antigen binding α1 and α2 domains were considered. The C-terminal end of the α2 domain was protonated to render it uncharged while all the other termini and side chains were assigned typical charge states at pH 7. Water molecules observed in the crystal structure were explicitly taken into account in the simulations. Polar hydrogens were added and the complexes were centred in cubic boxes of different sizes, to allow for water shells ranging from 2.5Å to more than 40Å thickness between the solute and the boundaries of the simulation box. The created space around the solute was initially filled with an equilibrium configuration of bulk water. The water models used were SPC and TIP4P, respectively, both recommended for usage in biomolecular systems [31]. To electrically neutralize the system, an appropriate number of Na^+ counterions were added.

2.3 Simulation methods and parameters

In order to investigate the influence of the force field and water model on our results we selected three different parametrizations from the suite of force fields available with GROMACS. The GROMACS force field (a modified GROMOS87 force field) [32] and the GROMOS96 53a6 force field [33] were used for SPC solvated pMHC complexes while the OPLS-AA/L all-atom force field (version 2001) [34] was used for TIP4P solvated pMHC complexes, as recommended in GROMACS documentation [32]. After the initial coordinate setup, as described above, the system's potential energy was first minimized by a steepest descent procedure, in order to allow the water molecules to adjust to the presence of the pMHC complex. The coordinates so obtained served as a starting structure for the subsequent

For Peer Review Only

MD simulations. Periodic boundaries were employed and the system was then heated linearly from 0K to 300K during the first 40ps and left at 300K for the last 10ps using a Berendsen thermostat with a temperature coupling constant of 0.1 ps [35]. For this warm-up phase only, position restraints (force constant of 1000 kJ mol^{-1} nm^{-2}) were applied to each atom of the pMHC complex. Subsequent production runs without any position restraints used Berendsen pressure coupling (reference pressure: 1 bar, pressure coupling constant: 0.5 ps) [35], the Particle Mesh Ewald (PME) [15] method for long-range electrostatic interactions, and an integration step of 2fs. All bond lengths were constrained to their equilibrium values using the SHAKE algorithm [36]. Simulations started with random velocities distributed to the atoms (i.e. initial conditions) and total simulation times between 1ns and 16ns were used. For subsequent analysis, configurations were saved in trajectory files every 2ps for long runs (16ns total) and every 0.5ps for short runs (1ns total). An overview of all performed simulation runs can be found in Table 1.

2.4 Methods of analysis

All simulation runs were evaluated via the following analysis modules provided by the GROMACS package:

2.4.1 SASA

The solvent accessible surface area (SASA) was computed using the Double Cube Lattice Method [37]. The radius of the solvent probe was set to 1.4 Ångström.

2.4.2 RMSD

The root mean square deviation (RMSD) gives the average deviation of a structure relative to a reference structure (e.g. crystal structure). In order to subtract global motions each structure is least squares fitted to the reference structure before RMSD calculation.

2.4.3 RMSF

The root mean square fluctuations (RMSF) represent the standard deviation of atomic positions during (a part of) a trajectory. **To eliminate overall protein translation and**

http://mc.manuscriptcentral.com/tandf/jenmol

For Peer Review Only

rotation each structure is least squares fitted to the initial structure before RMSF calculation. The RMSF of the crystal structure is calculated out of the crystallographic B-factors [1]. Contact residues have systematically low RMSF values while surface side chains protruding from the protein show large RMSF values.

2.4.4 Rotational autocorrelation function (RACF)

Given a unit vector $\vec{n}(t)$ characterizing a key direction of a molecule, the RACF is computed via

$$C_1(t) = \langle P_1(\vec{n}(0) \cdot \vec{n}(t)) \rangle \qquad \text{(Equation 1)}$$

with t being the time lag and $P_1(x) = x$ the first order Legendre polynomial. Values of 1 indicate "perfect memory" of orientation, a value of 0 indicates loss of orientational correlation.

For Peer Review Only

3 Results

Table 1 summarizes the simulations performed in order to analyze the key parameters of MD such as size of water box, initial conditions, system simplification, simulation time, and water model / force field combination. In particular, the following results were obtained:

3.1 "Single site binding" peptide

In order to point out the possible impacts of simplifications and limitations, we chose the rather labile system of the "single site binding" peptide (ARAAAAAAA) complexed to B*2705 MHC (PDB-id: 1hsa). This complex features an artificially modelled epitope with only one anchor residue (arginine at position 2).

3.1.1 Solvation shell thickness & initial conditions

We performed several MD simulations of the "single site binding" peptide loaded MHC with different initial conditions (randomly distributed starting velocities) and with solvation shell thicknesses ranging from 2.5Å to 40Å and computed the RMSF, see Figure 3. The RMSF values for the peptide's C-terminal alanine (residue 9), show the most pronounced variability with box size: Using the 5Å and the 10Å water shell box, the RMSF values are low. Usage of the smallest box (2.5Å shell) results in larger fluctuations. This may be due to direct interactions – via the periodic boundary conditions – between the epitope and the "bottom" (i.e. the part diametrically opposed to the binding cleft) of the simplified MHC. On the other hand, the RMSF graphs for the systems with a 20Å or a 40Å box show extremely high fluctuations of the peptide's C-terminal, alanine (residue 9). This indicates an interesting phenomenon: Simulating this rather artificial system revealed a detachment of the epitope's C-terminal end (a video of this simulation can be found in supplementary material S1). This finding strongly indicates that this artificial "single site binding" peptide is in fact not able to bind to the MHC. This is not surprising since there is no anchor residue at the C-terminal end.

For Peer Review Only

The salt bridges between the epitope's C-terminus and the MHC's F-pocket are just not stable enough to retain the hydrophobic epitope completely inside the hydrophilic binding groove. The gradual detachment of the epitope can be conveniently visualized by plotting the time dependence of the minimum distance between all the atoms of the peptide's C-terminal alanine and all the atoms of the MHC's β-sheet, which forms the bottom of the binding cleft (Figure 4).

Such a labile system is very sensible to different initial conditions, in particular if one uses too small water shell sizes. For each water shell thickness we performed four parallel simulations which were absolute identical, except for the randomly distributed initial velocities. These parallel simulations evolved very differently, as shown by the RMSF graph (Figure 3). For the system with a 20Å box, the peptide's C-terminus detached in two out of four runs. In the 40Å box, the peptide detached in even three out of four runs. Thus, performing only one simulation could produce misleading results.

Within the first nanosecond of simulation time detachment was observed with the 40Å and the 20Å water shell while smaller boxes (10Å and 5Å shell) forced the epitope to stay inside the binding groove. However, extending the simulation time for another 5ns the protein also detached within a 10Å water shell (data not shown). A larger water shell somehow speeds up the relaxation process and – together with long simulation times – seems to be necessary for plausible results of pMHC molecular dynamics simulations.

3.1.2 Force fields, water models & system simplification

We simulated the "single site binding" peptide loaded MHC in a 20Å hydration shell with different force fields and different water models. Surprisingly, for some simulations, we found a stable binding behaviour. Figure 5 shows minimum distances between the peptide's C-terminal alanine and the MHC β-sheet as a measure for peptide detachment (see also Figure 4). Different force fields and recommended compatible water models were used: The GROMACS force field (which is based on a GROMOS force field) and the most recent

For Peer Review Only

GROMOS force field (GROMOS96 53a6) both combined with SPC water model, and the OPLS-AA/L force field combined with TIP4P water model, respectively. The OPLS-AA force field together with TIP4P water (shown in Figure 5c) yields such low distance values that one could conclude that this artificial peptide actually is a binding epitope for the HLA-B*2705 MHC. This is in severe contrast to the simulations using the GROMACS force field together with SPC water described in previous sections, where the C-terminal end of the peptide detached in two out of four simulations (Figure 5a). Using the GROMOS96 53a6 force field, detachment of the peptide was observed too (Figure 5b).

Regarding system simplifications we found almost no difference between the whole and the simplified pMHC complex, where the α3 and the β_2m domains have been left out. Neither the GROMACS force field (see Figure 5a and d) nor the GROMOS96 53a6 force field (see Figure 5b and e) show any effect of simplification.

Taking all in all we may conclude: Size of the solvation shell and the choice of water model / force field combination, in particular the water model, shows an important impact on simulation results. pMHC simplification, however, is shown to have less effect.

3.2 "Double site binding" peptide

The "double site binding" peptide (GRFAAAIAK) complexed to B*2705 MHC (PDB-id: 1jge) shows rather stable binding, and thus longer simulation runs are needed to study the influences of key parameters.

3.2.1 Solvation

We performed MD simulations for at least 12ns of the "double site binding" peptide loaded MHC, again systematically varying the solvation layer thickness from 12Å to 27Å. In addition to the properties described above, we also computed the rotational autocorrelation function (RACF) of the MHC. This was done to check if reorientation of the pMHC complex is hampered by an incomplete solvation shell, possibly in combination with the periodic

For Peer Review Only

boundary conditions. Orientation of the MHC molecule was defined by a vector perpendicular to the plane of three selected atoms belonging to the β-sheet forming the bottom of the groove (Cα atoms of V25, N97, and G100). The loss of orientational memory is characterized by the decay of the RACF (Equation 1, and Figure 6). Figure 6 reveals that under periodic boundary conditions the loss of memory depends on the size of the water box. The results are only approximate since only one molecule was monitored for a rather short time, which may cause the non-exponential decays. However, a qualitative trend is clearly observable: While all three box sizes allow for rotation, rotational relaxation of the pMHC complex is less hindered in the 27Å system than in the 17Å system which in turn is less hindered than in the 12Å system.

The source of the observed rotational movement is rotational diffusion. The surrounding water hitting the pMHC complex randomly will cause to some extent a rotational movement of the complex. Figure 7 illustrates the way a smaller simulation box hinders the rotational movement of the pMHC complex: If the protein starts to rotate, the surrounding water network will follow the moving protein surface. In a large box there is less hindrance because the outer layers of the water network are almost unaffected by the rotation of the protein. Contrary, in the small box the rotation is more hindered because the outer water layers are bound to both, the protein and its periodic image. Thus, relatively tightly bound water molecules are forced to move in opposite directions, which effectively hinders rotation. Physically, tightly bound water means higher viscosity, resulting in longer relaxation times and thus less rotation.

3.2.2 Peptide's configurational substates

In our simulations of the MHC bound "double site binding" peptide within a 27Å water shell we observed that the bound peptide does not retain its original shape throughout the entire simulation run (16ns). It rather seems to switch back and forth between several different "substates" [17], which are temporarily stable (Figure 8a). Conformational substates of proteins and polypeptides are also known from experiments [38]. Figure 8b shows the RMSD

http://mc.manuscriptcentral.com/tandf/jenmol

For Peer Review Only

of Cα atoms for three specific residues. Some residues show low RMSD values throughout the simulation. Other residues show phases of high RMSD values at certain time intervals. One can conclude that several residues make small contributions to the peptide's total RMSD (Figure 8a), while others contribute significantly more. Hence the average RMSD, as usually displayed, hides the marked differences in mobility of atoms within the peptide. These specific movements and substates are made transparent using the following procedure:

We define the peptide's "principal" conformation as substate #1. The "principal" structure is defined by averaging the individual atomic positions between t=2ns and t=10ns and subsequently performing an "in vacuo" energy minimization. Then, positional deviations can conveniently be measured for each peptide's backbone atom relative to substate #1 (rather than relative to the crystal structure). Individual RMSD values can be shown grouped to a 3D plot over time (Figure 8d). The "hot spots" (tops and plateaus) indicate large deviations from the "principal" conformation and can be used to define further substates (see Figure 8e, Figure 9, and Table 2).

The observed substates can be characterized by several parameters (Table 2). Substate #2 (Figure 9, upper right corner) occurred very early and may be a reaction of the crystal structure to the surrounding water bath. Indeed, in this substate the centre part of the peptide (notably 6ALA) gets kinked up, causing the 6ALA residue not to face the lateral helix but protrude right up, away from the MHC. This movement generates a small hole beneath 6ALA, which allows for water molecules to enter the peptide-MHC interface. Actually, through the period of substate #2, the peptide-water interaction energy rises continually, while the peptide-MHC interaction energy decreases concomitantly (Figure 8c). In substate #1 ("principal" state – Figure 9, upper left corner) 6ALA is facing the lateral helix again and the hole is closed. During this state the peptide seems very stable and shows little fluctuations (RMSF). It remains in this substate for more than 8ns and comes back again to substate #1

http://mc.manuscriptcentral.com/tandf/jenmol

For Peer Review Only

after substate #3 and #4. Substate #3 (Figure 9, lower left corner) shows the C-terminal half of the peptide pointing away from the MHC. The peptide-water interaction energy increases again, causing a concomitant decrease in peptide-MHC interaction energy. The C-terminal residue 9LYS loses its interaction partner (116ASP in the β-sheet) and finds a new one (77ASP in the lateral α-helix). In substate #4 (Figure 9, lower right corner) the peptide lifts even more. The residue of 9LYS has to be extended almost totally to stay in contact with the F pocket (77ASP). More than 42% of the peptide's surface is exposed to water during this substate – a remarkable increase compared to 32.7% during substate #1. After this short period the peptide lowers again, 9LYS contacts 116ASP as before, and the peptide adopts the substate #1 for the rest of simulation time.

For Peer Review Only

4 Discussion

Conclusions from molecular dynamics simulations may reveal important mechanisms but can be delicate. A pMHC system is too large for a "full simulation", given the computational power available today. However, cautious choice of parameters still allows drawing conclusions from simulations. In this study we analyzed following key parameters for molecular dynamics simulations of pMHC molecules:

1) Initial conditions

   Minor differences in initial conditions can have a major effect on the result, in particular for labile systems as shown in the parallel simulations of the MHC bound "single site binding" peptide. Therefore, parallel simulations of identical structures starting from different initial conditions can help to identify labile systems. This seems crucial to avoid premature conclusions – in our case regarding weak binding or non binding epitopes.

2) Size of solvation shell

   Using periodic boundary conditions, the simulated system is embedded as the unit cell of a "supercrystal" within copies of itself. However, this crystalline environment may induce artefacts when the protein is comparable in size to the simulation box. Usually a bigger water shell minimizes these artefacts due to shielding effects of water: a polar solvent generates an electrostatic reaction field which screens the charges and dipoles [18]. We demonstrated the effects of a bigger water shell by several simulations of the MHC bound "single site binding" peptide. Using a small water shell (5Å or 10Å), the effect of the "supercrystal" was seen strong enough to artificially stabilize the whole structure: no detachment of the epitope was observed. Using a sufficiently large water shell (20Å), detachment of the epitope can be observed.

For Peer Review Only

Even if the system is large enough to solvate the whole protein, rotability might still be hindered. Analyzing simulations of the MHC bound "double site binding" peptide, we found that small water shells obviously allow for a direct interaction of the periodic images of the complexes through the water network, thereby reducing rotability. A sufficiently large water shell is mandatory to minimize water network interactions and enable realistic rotability.

Concluding, an enlargement of the water shell raises the computational workload but yields more realistic dynamics of the whole system.

3) Water model / force field combination

Using different models for force field and water representation induces drastic differences. Since SPC water is distinctly polar ($\varepsilon_r=65$, M.Neumann, unpublished), it is highly able to substitute the binding partners of the epitope and facilitates the "single site binding" epitope's detachment. With a less polar water model, like TIP4P ($\varepsilon_r=53\pm2$, [39]), solvation of the "single site binding" epitope is not able to disrupt its binding to the MHC: the epitope stays in the binding groove, displaying very low RMSD values. Rather "inert" water causes a pseudo-stabilization of the pMHC complex. Additionally, the capability of shielding electrostatic forces (see point 2) depends on the relative dielectric permittivity ε_r [18].

Usage of different force fields and water models provides different insights into the system's dynamics. Since there is no "one and only" model, variation is essential for reliable results. However, for a particular simulation experiment we can propose a guideline to choose a model that allows "staying on the safe side":

Simulations using a GROMOS-type force field with SPC water will tend to predict a "border case" epitope as non-binding. In contrast, the TIP4P water coming with the OPLS-AA force field shows a certain tendency that "border case" epitopes may be

For Peer Review Only

categorized as binding. Thus, when testing for binding (e.g. vaccine design) take the model that proved "looser" binding, and when testing for non-binding (e.g. allergenicity) take the model that proved "tighter" binding.

4) Simulation time

Long simulation runs are needed to observe all (or at least the most) relevant thermally accessible configurations. Sometimes a few hundred picoseconds simulation time will suffice. In other systems, representative sampling will take several seconds, which may be out of reach for molecular dynamics studies for a long time to come. Since we discovered the epitope within the MHC binding groove adopting different substates on the timescale of nanoseconds, pMHC simulations have to be at least several tens or even hundreds of nanoseconds long. However, in many cases it remains hard to decide whether a particular simulation has run long enough.

5) System simplification

The legitimacy of MHC simplification in MD simulations has already been covered by various authors [22;23] and there is no consensus about its effect on peptide dynamics on a timescale of nanoseconds. Our simulations show almost no difference between the simplified pMHC complex, where the α3 and the β_2m are left out of simulation, and the whole pMHC complex, neither when using the GROMACS force field (see Figure 5a and d) nor when using the GROMOS96 53a6 force field (see Figure 5b and e).

Beside the above mentioned key parameters (initial conditions, size of solvation shell, water model / force field combination, simulation time, system simplification) the binding free energy resulting from such a simulation would also be an interesting target to evaluate. The feasibility of its computation has been demonstrated in several studies [12-14;23;33;40;41] and challenges involved thereby can not be ignored: In order to reliably compute the binding free energy of a system as large as the pMHC complex,

For Peer Review Only

special techniques [40;41] have to be applied since the system's whole configurational space may be too large to be fully sampled. Thus, binding free energy computation will definitely be the scope of further publications.

The computational discovery of the epitope's substates raises some very interesting questions regarding a possible immunological meaning: The "outbreaking" characteristic of substate #4 (C-terminal half slightly protruding from the cleft) could represent the very trigger for a TCR screening the pMHC complex. This hypothesis fits well into the hitherto known framework: Although T cell activation is to some extent correlated to the half-life of pMHC-TCR interaction, half-life alone is insufficient to accurately predict activation potencies. In addition, conformational changes within the TCR-peptide-MHC interface may essentially contribute to T cell activation [42]. Pöhlmann et al. [6] and Starikov et al. [7] suggested that even a more flexible peptide may have a noticeable effect on T cell activation. This is precisely the point where the presented simulations are capable to provide additional insight and evidence: In substate #4 the peptide is not only more flexible (RMSF of 0.045 nm instead of 0.037 nm in substate #1) but there is also an actual movement of the total C-terminal half of the peptide out of the cleft, towards a possible TCR. This conformational change in the peptide could be transmitted to a bound TCR, possibly contributing to T cell activation [43].

Some authors have described peptide flexibilities, but the current pMHC-analysis is – to the best of our knowledge – the first to describe substates of the MHC bound peptide. Pöhlmann et al. [6] and Starikov et al. [7] performed MD simulations of the same system (PDB-id: 1jge). However, they observed no specifiable conformational substates for the bound peptide. Pöhlmann et al. used the OPLS-AA/L force field and TIP4P water model and Starikov et al. used a 7 Å water shell, both may suffer from a limited configurational sampling.

Further simulation studies can be envisaged to find out whether the existence of "protruding substates" actually correlates to the activation potency of TCRpMHC complexes, as obtained

http://mc.manuscriptcentral.com/tandf/jenmol

For Peer Review Only

from experimental measurements. The present work evaluates the computational tools and prepares their future application in the prediction of "protruding substates" for a given pMHC complex. In addition, single peptide residues could be exchanged and the impact on "protruding substates" computationally evaluated for vaccine design.

For Peer Review Only

5 Acknowledgments

We are indebted to Prof. Dr. Anneliese Schimpl for continuous support and advice in preparing the manuscript.

The computational work for the current investigation has been supported by Prof. Dr. Rudolf Karch, Core Unit for Medical Statistics and Informatics, Medical University of Vienna, Austria.

For Peer Review Only

6 References

[1] M. Karplus and G.A. Petsko, *Molecular dynamics simulations in biology*, Nature 347 (1990), pp. 631-639.

[2] T. Hansson, C. Oostenbrink, and W. van Gunsteren, *Molecular dynamics simulations*, Curr. Opin. Struct. Biol. 12 (2002), pp. 190-196.

[3] D. Rognan, L. Scapozza, G. Folkers, and A. Daser, *Molecular dynamics simulation of MHC-peptide complexes as a tool for predicting potential T cell epitopes*, Biochemistry 33 (1994), pp. 11476-11485.

[4] D. Rognan, S. Krebs, O. Kuonen, J.R. Lamas, J.A. Lopez de Castro, and G. Folkers, *Fine specificity of antigen binding to two class I major histocompatibility proteins (B*2705 and B*2703) differing in a single amino acid residue*, J. Comput. Aided Mol. Des 11 (1997), pp. 463-478.

[5] H. Toh, N. Kamikawaji, T. Tana, T. Sasazuki, and S. Kuhara, *Molecular dynamics simulations of HLA-DR4 (DRB1*0405) complexed with analogue peptide: conformational changes in the putative T-cell receptor binding regions*, Protein Eng 11 (1998), pp. 1027-1032.

[6] T. Pöhlmann, R.A. Bockmann, H. Grubmuller, B. Uchanska-Ziegler, A. Ziegler, and U. Alexiev, *Differential peptide dynamics is linked to major histocompatibility complex polymorphism*, J. Biol. Chem. 279 (2004), pp. 28197-28201.

[7] E.B. Starikov, L. Nilsson, and M. Hulsmeyer, *A single residue exchange between two HLA-B27 alleles triggers increased peptide flexibility*, Eur. Biophys. J. 33 (2004), pp. 651-655.

[8] U. Omasits, M. Neumann, O. Steinhauser, R. Valenta, R. Kobler, and W. Schreiner, *Molecular Dynamics Simulations in Immunology*, in *Proceedings of the 2nd Austrian Grid Symposium*, J. Volkert, T. Fahringer, D. Kranzlmüller, and W. Schreiner, eds., Oesterreichische Computer Gesellschaft, Vienna, 2007, pp. 81-90.

[9] W.S. Meng, H. von Grafenstein, and I.S. Haworth, *A model of water structure inside the HLA-A2 peptide binding groove*, Int. Immunol. 9 (1997), pp. 1339-1346.

[10] P.M. Petrone and A.E. Garcia, *MHC-peptide binding is assisted by bound water molecules*, J. Mol. Biol. 338 (2004), pp. 419-435.

[11] M. Zacharias and S. Springer, *Conformational flexibility of the MHC class I alpha1-alpha2 domain in peptide bound and free states: a molecular dynamics simulation study*, Biophys. J. 87 (2004), pp. 2203-2214.

[12] O. Michielin and M. Karplus, *Binding free energy differences in a TCR-peptide-MHC complex induced by a peptide mutation: a simulation analysis*, J. Mol. Biol. 324 (2002), pp. 547-569.

[13] V. Zoete and O. Michielin, *Comparison between computational alanine scanning and per-residue binding free energy decomposition for protein-protein association using*

For Peer Review Only

MM-GBSA: Application to the TCR-p-MHC complex, Proteins 67 (2007), pp. 1026-1047.

[14] S. Wan, P.V. Coveney, and D.R. Flower, *Molecular basis of peptide recognition by the TCR: affinity differences calculated using large scale computing*, J. Immunol. 175 (2005), pp. 1715-1723.

[15] T. Darden, D. York, and L. Pedersen, *Particle mesh Ewald: An N.log(N) method for Ewald sums in large systems*, J. Chem. Phys. 98 (1993), pp. 10089-10092.

[16] T.E. Cheatham and B.R. Brooks, *Recent advances in molecular dynamics simulation towards the realistic representation of biomolecules in solution*, Theor. Chem. Acc. 99 (1998), pp. 279-288.

[17] B.K. Andrews, T. Romo, J.B. Clarage, B.M. Pettitt, and J. Phillips, *Characterizing global substates of myoglobin*, Structure 6 (1998), pp. 587-594.

[18] D. Sengupta, R.N. Behera, J.C. Smith, and G.M. Ullmann, *The alpha helix dipole: Screened out?*, Structure 13 (2005), pp. 849-855.

[19] A. Wada, *The alpha-helix as an electric macro-dipole*, Advances in Biophysics (1976), pp. 1-63.

[20] W. Weber, P.H. Hünenberger, and J.A. McCammon, *Molecular dynamics simulations of a polyalanine octapeptide under Ewald boundary conditions: Influence of artificial periodicity on peptide conformation*, J. Phys. Chem. B 104 (2000), pp. 3668-3675.

[21] G.M. Torrie and J.P. Valleau, *Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling*, J. Comput. Phys. 23 (1977), pp. 187-199.

[22] D. Rognan, N. Zimmermann, G. Jung, and G. Folkers, *Molecular dynamics study of a complex between the human histocompatibility antigen HLA-A2 and the IMP58-66 nonapeptide from influenza virus matrix protein*, Eur. J. Biochem. 208 (1992), pp. 101-113.

[23] S. Wan, P. Coveney, and D.R. Flower, *Large-scale molecular dynamics simulations of HLA-A*0201 complexed with a tumor-specific antigenic peptide: can the alpha3 and beta2m domains be neglected?*, J. Comput. Chem. 25 (2004), pp. 1803-1813.

[24] The Austrian Grid Initiative, http://www.austriangrid.at, cited 2008

[25] H.J.C. Berendsen, D. van der Spoel, and R. van Drunen, *GROMACS: A message-passing parallel molecular dynamics implementation*, Comput. Phys. Commun. 91 (1995), pp. 43-56.

[26] E. Lindahl, B. Hess, and D. van der Spoel, *GROMACS 3.0: a package for molecular simulation and trajectory analysis*, J. Mol. Model. 7 (2001), pp. 306-317.

[27] G. Kingsley and J. Sieper, *Current perspectives in reactive arthritis*, Immunol. Today 14 (1993), pp. 387-391.

For Peer Review Only

[28] H.M. Bergman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, and P.E. Bourne, *The Protein Data Bank*, Nucleic Acids. Res. 28 (2000), pp. 235-242.

[29] D.R. Madden, J.C. Gorga, J.L. Strominger, and D.C. Wiley, *The three-dimensional structure of HLA-B27 at 2.1 A resolution suggests a general mechanism for tight peptide binding to MHC*, Cell 70 (1992), pp. 1035-1048.

[30] M. Hulsmeyer, R.C. Hillig, A. Volz, M. Ruhl, W. Schroder, W. Saenger, A. Ziegler, and B. Uchanska-Ziegler, *HLA-B27 subtypes differentially associated with disease exhibit subtle structural alterations*, J. Biol. Chem. 277 (2002), pp. 47844-47853.

[31] J. Zielkiewicz, *Structural properties of water: comparison of the SPC, SPCE, TIP4P, and TIP5P models of water*, J. Chem. Phys. 123 (2005), pp. 1-6.

[32] D. van der Spoel, E. Lindahl, B. Hess, A.R. van Buuren, E. Apol, P.J. Meulenhoff, D.P. Tieleman, A.L.T.M. Sijbers, K.A. Feenstra, R. van Drunen, and H.J.C. Berendsen, *Gromacs User Manual version 3.2*, 2004.

[33] C. Oostenbrink, A. Villa, A.E. Mark, and W.F. Van Gunsteren, *A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6*, J. Comput. Chem. 25 (2004), pp. 1656-1676.

[34] G.A. Kaminski, R.A. Friesner, J. Tirado-Rives, and W.L. Jorgensen, *Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides*, J. Phys. Chem. B 105 (2001), pp. 6474-6487.

[35] H.J.C. Berendsen, J.P.M. Postma, W.F. Van Gunsteren, A. DiNola, and J.R. Haak, *Molecular dynamics with coupling to an external bath*, J. Chem. Phys. 81 (1984), pp. 3684-3690.

[36] J.P. Ryckaert, G. Ciccotti, and H.J.C. Berendsen, *Numerical integration of the cartesian equations of motion of a system with constraints; molecular dynamics of n-alkanes.*, J. Comput. Phys. 23 (1977), pp. 327-341.

[37] F. Eisenhaber, P. Lijnzaad, P. Argos, C. Sander, and M. Scharf, *The Double Cube Lattice Method: Efficient Approaches to Numerical Integration of Surface Area and Volume and to Dot Surface Contouring of Molecular Assemblies*, J. Comput. Chem. 16 (1995), pp. 273-284.

[38] H. Frauenfelder, F. Parak, and R.D. Young, *Conformational substates in proteins*, Annual review of biophysics and biophysical chemistry 17 (1988), pp. 451-479.

[39] M. Neumann, *Dielectric relaxation in water. Computer simulations with the TIP4P potential*, J. Chem. Phys. 85 (1986), pp. 1567-1580.

[40] M.A. Cuendet, *The Jarzynski identity derived from general Hamiltonian or non-Hamiltonian dynamics reproducing NVT or NPT ensembles*, J. Chem. Phys. 125 (2006).

[41] S. Wan, P.V. Coveney, and D.R. Flower, *Peptide recognition by the T cell receptor: comparison of binding free energies from thermodynamic integration, Poisson-*

For Peer Review Only

Boltzmann and linear interaction energy approximations, Philos. Transact. A Math. Phys. Eng Sci. 363 (2005), pp. 2037-2053.

[42] M. Krogsgaard, N. Prado, E.J. Adams, X.L. He, D.C. Chow, D.B. Wilson, K.C. Garcia, and M.M. Davis, *Evidence that structural rearrangements and/or flexibility during TCR binding can contribute to T cell activation*, Mol. Cell 12 (2003), pp. 1367-1378.

[43] J.B. Reiser, C. Gregoire, C. Darnault, T. Mosser, A. Guimezanes, A.M. Schmitt-Verhulst, J.C. Fontecilla-Camps, G. Mazza, B. Malissen, and D. Housset, *A T cell receptor CDR3beta loop undergoes conformational changes of unprecedented magnitude upon binding to a peptide/MHC class I complex*, Immunity. 16 (2002), pp. 345-354.

[44] M.A. Saper, P.J. Bjorkman, and D.C. Wiley, *Refined structure of the human histocompatibility antigen HLA-A2 at 2.6 A resolution*, J. Mol. Biol. 219 (1991), pp. 277-319.

[45] W. Humphrey, A. Dalke, and K. Schulten, *VMD: visual molecular dynamics*, J. Mol. Graph. 14 (1996), pp. 33-38.

For Peer Review Only

7 Tables

Table 1. Overview of molecular dynamics simulations performed. Peptide main anchor residues are printed bold.

peptide	MHC	water model / force field combination	size of water box	parallel simulations	simulation time
ARAAAAAAA ("single site binding peptide")	B*2705 whole	SPC / GROMACS	20Å	4x	1ns
	B*2705 whole	SPC / GROMOS 53a6	20Å	4x	1ns
	B*2705 simplified	SPC / GROMACS	2.5Å	4x	1ns
	B*2705 simplified	SPC / GROMACS	5Å	4x	1ns
	B*2705 simplified	SPC / GROMACS	10Å	4x	1ns
	B*2705 simplified	SPC / GROMACS	10Å	1x	10ns
	B*2705 simplified	SPC / GROMACS	20Å	4x	1ns
	B*2705 simplified	SPC / GROMACS	40Å	4x	1ns
	B*2705 simplified	SPC / GROMOS 53a6	20Å	4x	1ns
	B*2705 simplified	TIP4P / OPLS-AA	20Å	4x	1ns
GRFAAAIAK ("double site binding peptide")	B*2705 simplified	SPC / GROMACS	12Å	1x	12ns
	B*2705 simplified	SPC / GROMACS	17Å	1x	12ns
	B*2705 simplified	SPC / GROMACS	27Å	1x	16ns

For Peer Review Only

Table 2. Characteristics of the "double site binding" peptide in crystal structure and in different substates occurring during the 16ns simulation. SASA is taken relative to SASA of the peptide in water.

configuration	time (ps)	mainly affected residues	SASA (relative)	interaction energy (kJ/mol)		RMSF (nm)
				peptide-water	peptide-MHC	
crystal	0		29.7%	-3.8 10^7	-8.8 10^7	0.037
substate #1	2,000 - 10,000		32.7%	-4.4 10^7	-8.3 10^7	0.037
substate #2	890 - 1,550	5A, 6A, 7I	36.8%	-4.3 10^7	-8.5 10^7	0.045
substate #3	11,050 - 12,500	8A, 9K	39.9%	-5.1 10^7	-7.5 10^7	0.038
substate #4	13,550 - 14,100	5A, 6A, 7I, 8A, 9K	42.3%	-5.2 10^7	-7.0 10^7	0.045

For Peer Review Only

8 Figure captions

Figure 1. HLA-B*2705 molecular surface and its binding pockets.

The pockets definition is according to [44]. Pocket A (M5, Y7, Y59, E63, Y159, E163, W167, Y171), pocket B (H9, T24, E45, L66, C67, Y99), pocket C (H9, K70, T73, D74, R97), pocket D (Y99, H114, L156, Y159, L160), pocket E (H114, W133, W147, V152, L156), pocket F (D77, T80, L81, Y84, D116, Y123, T143, K146, W147). The peptide's primary anchors P2 (conserved, Arg) and P9 form salt bridges and hydrogen bonds to the complementary negatively charged MHC pockets B and F, respectively. P3 interacts with the upper part of the hydrophobic pocket D. This one and all subsequent graphical representations of molecules were produced using the VMD package [45].

Figure 2. Crystal structure and simulated system.

*a) Crystal structure of a pMHC complex (m9 peptide complexed to HLA-B*2705 [30]). The peptide binding domain (α1 and α2 domain, white), the α3 domain (grey) and the associated β-2-microglobulin (black) are represented as "cartoons". The bound peptide is shown as "licorice". The dimensions of the MHC's peptide binding domain are given in Ångström.*

b) A simplified pMHC complex placed in the centre of a simulation box allowing for a 20 Ångström solvation shell resulting in altogether 99,101 atoms (water molecules in the front of the cube are carved out in order to make the pMHC complex visible). This system could be used as input for a MD simulation "as is".

Figure 3. RMSF of peptide residues' backbone atoms.

For Peer Review Only

We performed four independent 1ns runs (same starting structure, different starting velocities) for each simulation box size. The RMSF values shown are averages for the time interval from 600ps to 1000ps, which represents the equilibrated part of a stable run. Results for the individual runs are indicated by different line types. Low RMSF values indicate low positional fluctuations of the backbone atoms during the simulation, which means that these residues are likely to be tightly bound by the MHC. A residue displaying high RMSF values may be loosely bound or not bound at all.

Figure 4. Demonstration of the minimum distance between epitope C-terminus and MHC β-sheet.

a) Minimum distances of the epitope's C-terminal alanine to the MHC's β-sheet during 1ns simulations. In the simulation of the complex with a 20Å surrounding water shell (dashed line) the distance first rises gradually from 0.37nm to 0.7nm. Then the C-terminus totally detaches and moves as far as 1.4nm away. Although there are large fluctuations for this distance, in the simulation employing a 5Å water shell (continuous line) no detachment occurs during the 1ns run.

b) The dashed line marks the minimum distance of all the atoms of the epitope's C-terminal alanine to all the atoms of the MHC's β-sheet. The orange (light grey) peptide shows the distance in the crystal structure. The blue (dark grey) peptide depicts the increased distance after C-terminal detachment. For an animated display of the simulation please refer to (supplementary material S1).

Figure 5. Minimum distance between peptide C-terminus and MHC β-sheet during different simulations of the "single site binding" peptide.

For Peer Review Only

Three different force fields and two different water models were used for simulations of the simplified pMHC complex, and two different force fields with the same water model were used for the whole pMHC complex. Each combination was simulated in four independent runs for 1ns within a 20Å hydration shell (black, dark grey, grey, and light grey line).

a) Simulating the simplified pMHC complex enclosed in a 20Å hydration shell results for two out of the four simulations (black and dark grey) in the previously described partial detachment of the peptide (see Figure 4) when the GROMACS force field and the SPC water model are used. The rising distance of the peptide's C-terminal alanine residue to 1nm and above indicates total detachment from the MHC binding cleft.

b) Application of the GROMOS96 53a6 force field on the simplified system leads to peptide detachment in one out of four simulations (black).

c) When the OPLS-AA force field and the TIP4P water model are used on the simplified system, for all four simulations the peptide calmly resides in the binding groove and the minimum distance measure retains its initial low value for the entire length of the 1ns simulation.

d) Simulating the whole pMHC complex using the GROMACS force field and the SPC water model results in an onset of peptide detachment in two out of four simulations (black and dark grey).

e) Using the GROMOS96 53a6 force field on the whole pMHC complex leads to peptide detachment in one out of four simulations (black).

Figure 6. Orientational relaxation (RACF) of the pMHC complex for different sizes of the simulation box.

All the pMHC systems lose their orientational "memory" over time, which is shown by a decay of the RACF $C_1(t)$. The larger the surrounding water shell the faster the loss of memory. All three systems have been simulated for at least 12ns.

For Peer Review Only

Figure 7. Model for system size influencing rotational diffusion.

Both illustrations show a simulation box containing a protein solved in water using periodic boundaries but with different simulation box sizes. The darker the water molecules, the more structured is the water network. A small box hinders rotation by forcing a relatively tightly bound water layer to move in opposite directions.

Figure 8. Analysis of the long time simulation (16ns) of the "double site binding" peptide complexed to HLA-B*2705 (1jge) within a 27Å water box.

a) RMSD of the peptide's backbone (red line) and the MHC's backbone (black line) relative to the initial crystal structure. The peptide spends most of the time in a state (characterized by low RMSD values) that is very similar to the crystal structure, but seems to flip to a different conformational substates around simulation time t≈1ns and between t≈10ns and t≈14ns.

b) RMSD of specific peptide backbone atoms. The Cα atom of the primary anchor arginine (green line) remains close to its initial position throughout the simulation (low RSMD values). The Cα atom of the central alanine (position 5, blue line) temporarily moves farther off its position in the crystal structure after t≈1ns (higher RMSD values) while the Cα atom of the secondary anchor lysine (black line) undergoes an even larger dislocation (large RMSD values) between t≈10ns and t≈14ns. Therefore the peptide's residues differently contribute to the substates. Individual RMSD curves for all the peptide's Cα atoms can be found in (supplementary material S2).

c) Interaction energies of the peptide with the MHC (black line) and with water (grey line). Energies are the sum of Coulomb and Lennard Jones potential.

d) RMSD of each epitope's backbone atom relative to substate #1. This plot shows the substates and backbone atoms involved.

e) Conformational substates defined using above graphs.

http://mc.manuscriptcentral.com/tandf/jenmol

For Peer Review Only

Figure 9. Average structures of peptide's conformational substates.

While the white peptide representations depict the substate #1, the grey representations depict the average structures of corresponding substates. The MHC structure shown is an average structure over the whole simulation. Arrows indicate major structural deviations to substate #1

For Peer Review Only

9 Word count

34 pages

8180 words

44563 characters

For Peer Review Only

Analysis of Key Parameters for Molecular Dynamics of pMHC Molecules

Ulrich Omasits[a], Bernhard Knapp[a], Martin Neumann[b], Othmar Steinhauser[c], Hannes Stockinger[d], Rene Kobler[e], and Wolfgang Schreiner[a]*

For Peer Review Only

The work described in this paper is supported by the Austrian Grid Project, funded by the Austrian BMBWK (Federal Ministry for Education, Science and Culture) under contract GZ 4003/2-VI/4c/2004.

[a] Core Unit for Medical Statistics and Informatics, Medical University of Vienna, Austria

[b] Institute of Experimental Physics, University of Vienna, Austria

[c] Department of Computational Biological Chemistry, University of Vienna, Austria

[d] Department of Molecular Immunology, Centre of Biomolecular Medicine and Pharmacology, Medical University of Vienna, Austria

[e] Institute of Graphics and Parallel Processing, Johannes Kepler University Linz, Austria

[*] Corresponding author: fax: +43 1 40400-6677 – phone: +43 1 40400-6673– email: bernhard.knapp@meduniwien.ac.at – address: Spitalgasse 23, 1090 Vienna, Austria

Abstract

Molecular dynamics (MD) studies of human major histocompatibility complex (MHC) HLA-B*2705 complexing two different peptides were performed. During simulation one peptide partially detached from the MHC while the other peptide switched back and forth between several different configurations. These different configurations relate to conformational substates and can be assigned to different levels of chemical activity or even the molecular mechanisms of immunological signalling. To ensure reliable immunological conclusions from MD simulations we prepare the methodological tools by carefully evaluating initial conditions, system simplification, solvation shell thickness, water model / force field combination, and simulation length. We also derive a guideline for appropriate model selection. This kind of quality assessment is seen a mandatory prerequisite for coming studies linking pMHC dynamics to T cell activation.

Keywords: HLA; major histocompatibility complex; peptide dynamics; conformational substates; T cell activation

Abbreviations: β_2m: β_2-microglobulin – HLA: human leukocyte antigen – MD: molecular dynamics – MHC: major histocompatibility complex – PBC: periodic boundary conditions – PME: Particle Mesh Ewald – pMHC: peptide loaded MHC – RACF: rotational autocorrelation function – RMSD: root mean square deviation – RMSF: root mean square fluctuation – SASA: solvent accessible surface area – TCR: T cell receptor

For Peer Review Only

1 Introduction

Class I major histocompatibility complexes (MHC) bind antigenic peptides within the α1-α2 domain in a long narrow cleft located between two α-helices on the top of an antiparallel β-sheet. Binding is mainly mediated through sequence independent contacts of the peptide's charged N- and C-terminus to conserved "pockets" (Figure 1) in the MHC's binding groove. The contacts are maintained by networks of hydrogen bonds that are very similar in all class I pMHC complexes and make the largest contribution to the binding energy of the peptide. In addition, polymorphic pockets, so-called primary anchor regions, interact with specific sequence dependent amino acids (anchor residues) of the peptide and thus are responsible for the allele specificity of the recognition process. The central part of the peptide bulges out of the binding groove and serves as primary recognition site for the T cell receptor (TCR) (see Figure 2a).

Molecular dynamics (MD) is the technique of numerically solving the equations of motion of an assembly of particles (atoms). The total force acting on each particle depends on the positions and properties of all other particles in the system and, in addition, may include externally imposed forces. MD programs calculate these forces and solve Newton's equation of motion for each particle. For details refer to reviews on MD of biomolecules [1;2].

Various MD studies on pMHC have been reported in the literature [3-14]. All these MD studies use some force field, some water model, some initial condition, and some size of water box, in most cases without a particular justification or comment on their choices. The aim of this study is to vary these parameters and analyze their effects and possible pitfalls regarding a pMHC complex simulation. To our knowledge no such detailed investigation has been reported in the literature. We are going to answer the following questions:

For Peer Review Only

- Should parallel simulations be performed using different initial conditions?
- What size of solvation box should be used?
- Which force field and water model should be chosen?
- What is the minimum simulation time necessary?
- Are system simplifications admissible?

Our aim is to evaluate those key parameters giving a reasonable chance to cope with the rather large pMHC molecule in a meaningful way. **Studies to reliably evaluate the performance of MD simulations have been carried out in smaller systems, based on sufficient sampling over configurational space. As opposed to this, the present study deals with a very large system which is interesting for medical reasons and the task can only be to find out to which extent reliability can be obtained.** All in all, the present work lays the basis for MD as a method of investigation in pMHC interactions.

Besides the technical evaluation we also present a minute mobility analysis of a peptide adopting different configurational substates within the MHC-cleft.

1.1 Environment representation in MD

In order to achieve a manageable representation of the natural pMHC complex a trade-off between computational cost and simulation accuracy has to be made, especially concerning the amount of solvation water included in the simulation, the boundary conditions, and protein simplifications. Since proteins evolved to function in aqueous environments, water molecules binding directly to the protein and forming the first layer of the hydration shell may be considered as an integral part of protein structure. It is also clear that solute-solvent interactions extend beyond the first hydration layer. Although an extended hydration shell may be represented by simplified models, the most reliable results on structure and dynamics are achieved by inclusion of a sufficient number of additional explicit solvent molecules. To

http://mc.manuscriptcentral.com/tandf/jenmol

For Peer Review Only

avoid an abrupt border with a vacuum, periodic boundary conditions (PBC) repeat the simulation box periodically in all three space dimensions, creating a continuous system. PBC allows a very useful method for calculating long-range electrostatics: the Particle Mesh Ewald (PME) summation [15]. Currently PBC/PME represents the standard in MD simulations of biomolecules since it is the most accurate method. On the other hand, the imposition of periodicity may lead to artefacts since all atoms now may interact with their periodic images. These artefacts can be minimized using larger solvation shells and an appropriate water model / force field combination, i.e. a combination to stay on the "safe side" for a particular simulation experiment (see Discussion section).

1.1.1 Size of solvation shell

A minimal system is claimed to need a 5-15Å thick solvation shell [16]. In other words, the system needs to be large enough to allow for the formation of a 5-15Å "buffer zone" between the solute and the boundaries of the simulation box. We show that for pMHC simulations a solvation layer thickness of more than 10Å, ideally 20Å, is necessary (Figure 2b). It is known that the aqueous environment facilitates the protein sampling the conformational space [17]. Water, due to its high relative dielectric permittivity, is very effective in screening long-range electrostatic interactions and dipoles. Thus the thickness of the solvation shell essentially determines the degree of electrostatic shielding, which is seen crucial in the presence of significant dipoles induced by the α-helices of the MHC [18;19]. PBC/PME imposes an artificial periodicity on the grossly inhomogeneous solute-solvent system. This periodicity limits the configurational flexibility of a protein [20] by overstabilization of the starting configuration which is normally taken to be that of a periodic crystal. Obviously, the larger the simulation box, the better the water molecules in the solvation shell (and the protein itself) will be screened from interactions with the protein's periodic images. Thus crystalline effects are minimized, sampling is less hindered [21], and a more realistic dynamics of the system can be observed.

For Peer Review Only

1.1.2 Structural simplifications

It is quite common to simplify a simulation system in order to minimize the number of atoms. In the case of pMHC complexes frequently only the MHC's peptide binding domain (α1 and α2 domain) is simulated since this is the site of interesting peptide-MHC interactions. In vivo such a truncated pMHC complex would not be stable at all and unfolding would occur. MHC simplification has been found not to alter the results of MD simulations on the timescale feasible [22]. However, there is disagreement about this simplification and there are authors [23] claiming that the α3 domain and the β_2m subunit are essential for peptide MHC interactions even within one nanosecond simulation time. According to our results, simplification plays a minor role. The α3 domain and the β_2m subunit are certainly essential for a simulation of a system in its full complexity. However, we are interested in processes taking place within the binding groove at a much shorter timescale than MHC unfolding would occur.

For Peer Review Only

2 Materials and Methods

2.1 Hard- and Software

The project was carried out on a SGI Altix 350 system within the framework of the Austrian Grid [24] and on a computational cluster of the Technical University of Vienna. For a typical 2ns simulation run a total of 800 CPU hours are required.

We use GROMACS [25;26] because of its detailed documentation of computational methodology and very fast procedure to treat water molecules. Its high scalar performance outweighed the fact that (on our system) GROMACS scales efficiently only up to 16 nodes.

2.2 Setup of coordinates

For simulations we selected the MHC class I B*2705 protein. Almost all binding epitopes for this MHC show an arginine at position 2, defined as a dominant anchor residue. Positions 1, 3, and 9 provide the remaining main anchoring positions with a preference for hydrophobic amino acids at P3, and hydrophobic or positively charged amino acids at P9. The dominant anchor residue P2 binds deep into the MHC pocket B, while P9 binds into MHC pocket F (see Figure 1). HLA-B27 subtypes are strongly associated with the occurrence of several diseases, in particular inflammatory diseases of the joints called spondyloarthropathies [27] and thus lends itself as a most interesting candidate for computational screening.

Starting coordinates were taken from crystal structures deposited in the Protein Data Bank [28]. We chose two structures representing the same MHC complexing two different, but substantially similar peptides. The first structure was obtained by Madden et al. from an epitope mix. As a “minimal binding” peptide the consensus backbone was modelled with the epitope ARAAAAAAA. The resulting structure was deposited as entry 1hsa [29]. The second structure containing the epitope GRFAAAIAK was obtained by Hülsmeyer et al. and deposited as entry 1jge [30]. We chose those two complexes due to their chemically similar

http://mc.manuscriptcentral.com/tandf/jenmol

For Peer Review Only

epitopes. Both consist of mainly hydrophobic residues (A,F,I,G) and both have the R anchor at position 2, typical for HLA-B*2705 epitopes. The crucial difference is the secondary anchor (K) in 1jge at position 9 which is missing in 1hsa. Therefore we will refer to the epitope of 1jge as “double site binding” peptide and to the epitope of 1hsa as “single site binding” peptide.

For most simulations only the antigen binding α1 and α2 domains were considered. The C-terminal end of the α2 domain was protonated to render it uncharged while all the other termini and side chains were assigned typical charge states at pH 7. Water molecules observed in the crystal structure were explicitly taken into account in the simulations. Polar hydrogens were added and the complexes were centred in cubic boxes of different sizes, to allow for water shells ranging from 2.5Å to more than 40Å thickness between the solute and the boundaries of the simulation box. The created space around the solute was initially filled with an equilibrium configuration of bulk water. The water models used were SPC and TIP4P, respectively, both recommended for usage in biomolecular systems [31]. To electrically neutralize the system, an appropriate number of Na^+ counterions were added.

2.3 Simulation methods and parameters

In order to investigate the influence of the force field and water model on our results we selected three different parametrizations from the suite of force fields available with GROMACS. The GROMACS force field (a modified GROMOS87 force field) [32] and the GROMOS96 53a6 force field [33] were used for SPC solvated pMHC complexes while the OPLS-AA/L all-atom force field (version 2001) [34] was used for TIP4P solvated pMHC complexes, as recommended in GROMACS documentation [32]. After the initial coordinate setup, as described above, the system’s potential energy was first minimized by a steepest descent procedure, in order to allow the water molecules to adjust to the presence of the pMHC complex. The coordinates so obtained served as a starting structure for the subsequent

For Peer Review Only

MD simulations. Periodic boundaries were employed and the system was then heated linearly from 0K to 300K during the first 40ps and left at 300K for the last 10ps using a Berendsen thermostat with a temperature coupling constant of 0.1 ps [35]. For this warm-up phase only, position restraints (force constant of 1000 kJ mol^{-1} nm^{-2}) were applied to each atom of the pMHC complex. Subsequent production runs without any position restraints used Berendsen pressure coupling (reference pressure: 1 bar, pressure coupling constant: 0.5 ps) [35], the Particle Mesh Ewald (PME) [15] method for long-range electrostatic interactions, and an integration step of 2fs. All bond lengths were constrained to their equilibrium values using the SHAKE algorithm [36]. Simulations started with random velocities distributed to the atoms (i.e. initial conditions) and total simulation times between 1ns and 16ns were used. For subsequent analysis, configurations were saved in trajectory files every 2ps for long runs (16ns total) and every 0.5ps for short runs (1ns total). An overview of all performed simulation runs can be found in Table 1.

2.4 Methods of analysis

All simulation runs were evaluated via the following analysis modules provided by the GROMACS package:

2.4.1 SASA

The solvent accessible surface area (SASA) was computed using the Double Cube Lattice Method [37]. The radius of the solvent probe was set to 1.4 Ångström.

2.4.2 RMSD

The root mean square deviation (RMSD) gives the average deviation of a structure relative to a reference structure (e.g. crystal structure). In order to subtract global motions each structure is least squares fitted to the reference structure before RMSD calculation.

2.4.3 RMSF

The root mean square fluctuations (RMSF) represent the standard deviation of atomic positions during (a part of) a trajectory. **To eliminate overall protein translation and**

http://mc.manuscriptcentral.com/tandf/jenmol

For Peer Review Only

rotation each structure is least squares fitted to the initial structure before RMSF calculation. The RMSF of the crystal structure is calculated out of the crystallographic B-factors [1]. Contact residues have systematically low RMSF values while surface side chains protruding from the protein show large RMSF values.

2.4.4 Rotational autocorrelation function (RACF)

Given a unit vector $\vec{n}(t)$ characterizing a key direction of a molecule, the RACF is computed via

$$C_1(t) = \langle P_1(\vec{n}(0) \cdot \vec{n}(t)) \rangle \qquad \text{(Equation 1)}$$

with t being the time lag and $P_1(x) = x$ the first order Legendre polynomial. Values of 1 indicate "perfect memory" of orientation, a value of 0 indicates loss of orientational correlation.

For Peer Review Only

3 Results

Table 1 summarizes the simulations performed in order to analyze the key parameters of MD such as size of water box, initial conditions, system simplification, simulation time, and water model / force field combination. In particular, the following results were obtained:

3.1 "Single site binding" peptide

In order to point out the possible impacts of simplifications and limitations, we chose the rather labile system of the "single site binding" peptide (ARAAAAAAA) complexed to B*2705 MHC (PDB-id: 1hsa). This complex features an artificially modelled epitope with only one anchor residue (arginine at position 2).

3.1.1 Solvation shell thickness & initial conditions

We performed several MD simulations of the "single site binding" peptide loaded MHC with different initial conditions (randomly distributed starting velocities) and with solvation shell thicknesses ranging from 2.5Å to 40Å and computed the RMSF, see Figure 3. The RMSF values for the peptide's C-terminal alanine (residue 9), show the most pronounced variability with box size: Using the 5Å and the 10Å water shell box, the RMSF values are low. Usage of the smallest box (2.5Å shell) results in larger fluctuations. This may be due to direct interactions – via the periodic boundary conditions – between the epitope and the "bottom" (i.e. the part diametrically opposed to the binding cleft) of the simplified MHC. On the other hand, the RMSF graphs for the systems with a 20Å or a 40Å box show extremely high fluctuations of the peptide's C-terminal, alanine (residue 9). This indicates an interesting phenomenon: Simulating this rather artificial system revealed a detachment of the epitope's C-terminal end (a video of this simulation can be found in supplementary material S1). This finding strongly indicates that this artificial "single site binding" peptide is in fact not able to bind to the MHC. This is not surprising since there is no anchor residue at the C-terminal end.

For Peer Review Only

The salt bridges between the epitope's C-terminus and the MHC's F-pocket are just not stable enough to retain the hydrophobic epitope completely inside the hydrophilic binding groove. The gradual detachment of the epitope can be conveniently visualized by plotting the time dependence of the minimum distance between all the atoms of the peptide's C-terminal alanine and all the atoms of the MHC's β-sheet, which forms the bottom of the binding cleft (Figure 4).

Such a labile system is very sensible to different initial conditions, in particular if one uses too small water shell sizes. For each water shell thickness we performed four parallel simulations which were absolute identical, except for the randomly distributed initial velocities. These parallel simulations evolved very differently, as shown by the RMSF graph (Figure 3). For the system with a 20Å box, the peptide's C-terminus detached in two out of four runs. In the 40Å box, the peptide detached in even three out of four runs. Thus, performing only one simulation could produce misleading results.

Within the first nanosecond of simulation time detachment was observed with the 40Å and the 20Å water shell while smaller boxes (10Å and 5Å shell) forced the epitope to stay inside the binding groove. However, extending the simulation time for another 5ns the protein also detached within a 10Å water shell (data not shown). A larger water shell somehow speeds up the relaxation process and – together with long simulation times – seems to be necessary for plausible results of pMHC molecular dynamics simulations.

3.1.2 Force fields, water models & system simplification

We simulated the "single site binding" peptide loaded MHC in a 20Å hydration shell with different force fields and different water models. Surprisingly, for some simulations, we found a stable binding behaviour. Figure 5 shows minimum distances between the peptide's C-terminal alanine and the MHC β-sheet as a measure for peptide detachment (see also Figure 4). Different force fields and recommended compatible water models were used: The GROMACS force field (which is based on a GROMOS force field) and the most recent

For Peer Review Only

GROMOS force field (GROMOS96 53a6) both combined with SPC water model, and the OPLS-AA/L force field combined with TIP4P water model, respectively. The OPLS-AA force field together with TIP4P water (shown in Figure 5c) yields such low distance values that one could conclude that this artificial peptide actually is a binding epitope for the HLA-B*2705 MHC. This is in severe contrast to the simulations using the GROMACS force field together with SPC water described in previous sections, where the C-terminal end of the peptide detached in two out of four simulations (Figure 5a). Using the GROMOS96 53a6 force field, detachment of the peptide was observed too (Figure 5b).

Regarding system simplifications we found almost no difference between the whole and the simplified pMHC complex, where the α3 and the β_2m domains have been left out. Neither the GROMACS force field (see Figure 5a and d) nor the GROMOS96 53a6 force field (see Figure 5b and e) show any effect of simplification.

Taking all in all we may conclude: Size of the solvation shell and the choice of water model / force field combination, in particular the water model, shows an important impact on simulation results. pMHC simplification, however, is shown to have less effect.

3.2 "Double site binding" peptide

The "double site binding" peptide (GRFAAAIAK) complexed to B*2705 MHC (PDB-id: 1jge) shows rather stable binding, and thus longer simulation runs are needed to study the influences of key parameters.

3.2.1 Solvation

We performed MD simulations for at least 12ns of the "double site binding" peptide loaded MHC, again systematically varying the solvation layer thickness from 12Å to 27Å. In addition to the properties described above, we also computed the rotational autocorrelation function (RACF) of the MHC. This was done to check if reorientation of the pMHC complex is hampered by an incomplete solvation shell, possibly in combination with the periodic

For Peer Review Only

boundary conditions. Orientation of the MHC molecule was defined by a vector perpendicular to the plane of three selected atoms belonging to the β-sheet forming the bottom of the groove (Cα atoms of V25, N97, and G100). The loss of orientational memory is characterized by the decay of the RACF (Equation 1, and Figure 6). Figure 6 reveals that under periodic boundary conditions the loss of memory depends on the size of the water box. The results are only approximate since only one molecule was monitored for a rather short time, which may cause the non-exponential decays. However, a qualitative trend is clearly observable: While all three box sizes allow for rotation, rotational relaxation of the pMHC complex is less hindered in the 27Å system than in the 17Å system which in turn is less hindered than in the 12Å system.

The source of the observed rotational movement is rotational diffusion. The surrounding water hitting the pMHC complex randomly will cause to some extent a rotational movement of the complex. Figure 7 illustrates the way a smaller simulation box hinders the rotational movement of the pMHC complex: If the protein starts to rotate, the surrounding water network will follow the moving protein surface. In a large box there is less hindrance because the outer layers of the water network are almost unaffected by the rotation of the protein. Contrary, in the small box the rotation is more hindered because the outer water layers are bound to both, the protein and its periodic image. Thus, relatively tightly bound water molecules are forced to move in opposite directions, which effectively hinders rotation. Physically, tightly bound water means higher viscosity, resulting in longer relaxation times and thus less rotation.

3.2.2 Peptide's configurational substates

In our simulations of the MHC bound "double site binding" peptide within a 27Å water shell we observed that the bound peptide does not retain its original shape throughout the entire simulation run (16ns). It rather seems to switch back and forth between several different "substates" [17], which are temporarily stable (Figure 8a). Conformational substates of proteins and polypeptides are also known from experiments [38]. Figure 8b shows the RMSD

http://mc.manuscriptcentral.com/tandf/jenmol

For Peer Review Only

of Cα atoms for three specific residues. Some residues show low RMSD values throughout the simulation. Other residues show phases of high RMSD values at certain time intervals. One can conclude that several residues make small contributions to the peptide's total RMSD (Figure 8a), while others contribute significantly more. Hence the average RMSD, as usually displayed, hides the marked differences in mobility of atoms within the peptide. These specific movements and substates are made transparent using the following procedure:

We define the peptide's "principal" conformation as substate #1. The "principal" structure is defined by averaging the individual atomic positions between t=2ns and t=10ns and subsequently performing an "in vacuo" energy minimization. Then, positional deviations can conveniently be measured for each peptide's backbone atom relative to substate #1 (rather than relative to the crystal structure). Individual RMSD values can be shown grouped to a 3D plot over time (Figure 8d). The "hot spots" (tops and plateaus) indicate large deviations from the "principal" conformation and can be used to define further substates (see Figure 8e, Figure 9, and Table 2).

The observed substates can be characterized by several parameters (Table 2). Substate #2 (Figure 9, upper right corner) occurred very early and may be a reaction of the crystal structure to the surrounding water bath. Indeed, in this substate the centre part of the peptide (notably 6ALA) gets kinked up, causing the 6ALA residue not to face the lateral helix but protrude right up, away from the MHC. This movement generates a small hole beneath 6ALA, which allows for water molecules to enter the peptide-MHC interface. Actually, through the period of substate #2, the peptide-water interaction energy rises continually, while the peptide-MHC interaction energy decreases concomitantly (Figure 8c). In substate #1 ("principal" state – Figure 9, upper left corner) 6ALA is facing the lateral helix again and the hole is closed. During this state the peptide seems very stable and shows little fluctuations (RMSF). It remains in this substate for more than 8ns and comes back again to substate #1

For Peer Review Only

after substate #3 and #4. Substate #3 (Figure 9, lower left corner) shows the C-terminal half of the peptide pointing away from the MHC. The peptide-water interaction energy increases again, causing a concomitant decrease in peptide-MHC interaction energy. The C-terminal residue 9LYS loses its interaction partner (116ASP in the β-sheet) and finds a new one (77ASP in the lateral α-helix). In substate #4 (Figure 9, lower right corner) the peptide lifts even more. The residue of 9LYS has to be extended almost totally to stay in contact with the F pocket (77ASP). More than 42% of the peptide's surface is exposed to water during this substate – a remarkable increase compared to 32.7% during substate #1. After this short period the peptide lowers again, 9LYS contacts 116ASP as before, and the peptide adopts the substate #1 for the rest of simulation time.

For Peer Review Only

4 Discussion

Conclusions from molecular dynamics simulations may reveal important mechanisms but can be delicate. A pMHC system is too large for a "full simulation", given the computational power available today. However, cautious choice of parameters still allows drawing conclusions from simulations. In this study we analyzed following key parameters for molecular dynamics simulations of pMHC molecules:

1) Initial conditions

   Minor differences in initial conditions can have a major effect on the result, in particular for labile systems as shown in the parallel simulations of the MHC bound "single site binding" peptide. Therefore, parallel simulations of identical structures starting from different initial conditions can help to identify labile systems. This seems crucial to avoid premature conclusions – in our case regarding weak binding or non binding epitopes.

2) Size of solvation shell

   Using periodic boundary conditions, the simulated system is embedded as the unit cell of a "supercrystal" within copies of itself. However, this crystalline environment may induce artefacts when the protein is comparable in size to the simulation box. Usually a bigger water shell minimizes these artefacts due to shielding effects of water: a polar solvent generates an electrostatic reaction field which screens the charges and dipoles [18]. We demonstrated the effects of a bigger water shell by several simulations of the MHC bound "single site binding" peptide. Using a small water shell (5Å or 10Å), the effect of the "supercrystal" was seen strong enough to artificially stabilize the whole structure: no detachment of the epitope was observed. Using a sufficiently large water shell (20Å), detachment of the epitope can be observed.

For Peer Review Only

Even if the system is large enough to solvate the whole protein, rotability might still be hindered. Analyzing simulations of the MHC bound "double site binding" peptide, we found that small water shells obviously allow for a direct interaction of the periodic images of the complexes through the water network, thereby reducing rotability. A sufficiently large water shell is mandatory to minimize water network interactions and enable realistic rotability.

Concluding, an enlargement of the water shell raises the computational workload but yields more realistic dynamics of the whole system.

3) Water model / force field combination

Using different models for force field and water representation induces drastic differences. Since SPC water is distinctly polar ($\varepsilon_r=65$, M.Neumann, unpublished), it is highly able to substitute the binding partners of the epitope and facilitates the "single site binding" epitope's detachment. With a less polar water model, like TIP4P ($\varepsilon_r=53\pm2$, [39]), solvation of the "single site binding" epitope is not able to disrupt its binding to the MHC: the epitope stays in the binding groove, displaying very low RMSD values. Rather "inert" water causes a pseudo-stabilization of the pMHC complex. Additionally, the capability of shielding electrostatic forces (see point 2) depends on the relative dielectric permittivity ε_r [18].

Usage of different force fields and water models provides different insights into the system's dynamics. Since there is no "one and only" model, variation is essential for reliable results. However, for a particular simulation experiment we can propose a guideline to choose a model that allows "staying on the safe side":

Simulations using a GROMOS-type force field with SPC water will tend to predict a "border case" epitope as non-binding. In contrast, the TIP4P water coming with the OPLS-AA force field shows a certain tendency that "border case" epitopes may be

For Peer Review Only

categorized as binding. Thus, when testing for binding (e.g. vaccine design) take the model that proved "looser" binding, and when testing for non-binding (e.g. allergenicity) take the model that proved "tighter" binding.

4) Simulation time

Long simulation runs are needed to observe all (or at least the most) relevant thermally accessible configurations. Sometimes a few hundred picoseconds simulation time will suffice. In other systems, representative sampling will take several seconds, which may be out of reach for molecular dynamics studies for a long time to come. Since we discovered the epitope within the MHC binding groove adopting different substates on the timescale of nanoseconds, pMHC simulations have to be at least several tens or even hundreds of nanoseconds long. However, in many cases it remains hard to decide whether a particular simulation has run long enough.

5) System simplification

The legitimacy of MHC simplification in MD simulations has already been covered by various authors [22;23] and there is no consensus about its effect on peptide dynamics on a timescale of nanoseconds. Our simulations show almost no difference between the simplified pMHC complex, where the α3 and the β_2m are left out of simulation, and the whole pMHC complex, neither when using the GROMACS force field (see Figure 5a and d) nor when using the GROMOS96 53a6 force field (see Figure 5b and e).

Beside the above mentioned key parameters (initial conditions, size of solvation shell, water model / force field combination, simulation time, system simplification) the binding free energy resulting from such a simulation would also be an interesting target to evaluate. The feasibility of its computation has been demonstrated in several studies [12-14;23;33;40;41] and challenges involved thereby can not be ignored: In order to reliably compute the binding free energy of a system as large as the pMHC complex,

http://mc.manuscriptcentral.com/tandf/jenmol

For Peer Review Only

special techniques [40;41] have to be applied since the system's whole configurational space may be too large to be fully sampled. Thus, binding free energy computation will definitely be the scope of further publications.

The computational discovery of the epitope's substates raises some very interesting questions regarding a possible immunological meaning: The "outbreaking" characteristic of substate #4 (C-terminal half slightly protruding from the cleft) could represent the very trigger for a TCR screening the pMHC complex. This hypothesis fits well into the hitherto known framework: Although T cell activation is to some extent correlated to the half-life of pMHC-TCR interaction, half-life alone is insufficient to accurately predict activation potencies. In addition, conformational changes within the TCR-peptide-MHC interface may essentially contribute to T cell activation [42]. Pöhlmann et al. [6] and Starikov et al. [7] suggested that even a more flexible peptide may have a noticeable effect on T cell activation. This is precisely the point where the presented simulations are capable to provide additional insight and evidence: In substate #4 the peptide is not only more flexible (RMSF of 0.045 nm instead of 0.037 nm in substate #1) but there is also an actual movement of the total C-terminal half of the peptide out of the cleft, towards a possible TCR. This conformational change in the peptide could be transmitted to a bound TCR, possibly contributing to T cell activation [43].

Some authors have described peptide flexibilities, but the current pMHC-analysis is – to the best of our knowledge – the first to describe substates of the MHC bound peptide. Pöhlmann et al. [6] and Starikov et al. [7] performed MD simulations of the same system (PDB-id: 1jge). However, they observed no specifiable conformational substates for the bound peptide. Pöhlmann et al. used the OPLS-AA/L force field and TIP4P water model and Starikov et al. used a 7 Å water shell, both may suffer from a limited configurational sampling.

Further simulation studies can be envisaged to find out whether the existence of "protruding substates" actually correlates to the activation potency of TCRpMHC complexes, as obtained

http://mc.manuscriptcentral.com/tandf/jenmol

For Peer Review Only

from experimental measurements. The present work evaluates the computational tools and prepares their future application in the prediction of “protruding substates” for a given pMHC complex. In addition, single peptide residues could be exchanged and the impact on “protruding substates” computationally evaluated for vaccine design.

For Peer Review Only

5 Acknowledgments

We are indebted to Prof. Dr. Anneliese Schimpl for continuous support and advice in preparing the manuscript.

The computational work for the current investigation has been supported by Prof. Dr. Rudolf Karch, Core Unit for Medical Statistics and Informatics, Medical University of Vienna, Austria.

For Peer Review Only

6 References

[1] M. Karplus and G.A. Petsko, *Molecular dynamics simulations in biology*, Nature 347 (1990), pp. 631-639.

[2] T. Hansson, C. Oostenbrink, and W. van Gunsteren, *Molecular dynamics simulations*, Curr. Opin. Struct. Biol. 12 (2002), pp. 190-196.

[3] D. Rognan, L. Scapozza, G. Folkers, and A. Daser, *Molecular dynamics simulation of MHC-peptide complexes as a tool for predicting potential T cell epitopes*, Biochemistry 33 (1994), pp. 11476-11485.

[4] D. Rognan, S. Krebs, O. Kuonen, J.R. Lamas, J.A. Lopez de Castro, and G. Folkers, *Fine specificity of antigen binding to two class I major histocompatibility proteins (B*2705 and B*2703) differing in a single amino acid residue*, J. Comput. Aided Mol. Des 11 (1997), pp. 463-478.

[5] H. Toh, N. Kamikawaji, T. Tana, T. Sasazuki, and S. Kuhara, *Molecular dynamics simulations of HLA-DR4 (DRB1*0405) complexed with analogue peptide: conformational changes in the putative T-cell receptor binding regions*, Protein Eng 11 (1998), pp. 1027-1032.

[6] T. Pöhlmann, R.A. Bockmann, H. Grubmuller, B. Uchanska-Ziegler, A. Ziegler, and U. Alexiev, *Differential peptide dynamics is linked to major histocompatibility complex polymorphism*, J. Biol. Chem. 279 (2004), pp. 28197-28201.

[7] E.B. Starikov, L. Nilsson, and M. Hulsmeyer, *A single residue exchange between two HLA-B27 alleles triggers increased peptide flexibility*, Eur. Biophys. J. 33 (2004), pp. 651-655.

[8] U. Omasits, M. Neumann, O. Steinhauser, R. Valenta, R. Kobler, and W. Schreiner, *Molecular Dynamics Simulations in Immunology*, in *Proceedings of the 2nd Austrian Grid Symposium*, J. Volkert, T. Fahringer, D. Kranzlmüller, and W. Schreiner, eds., Oesterreichische Computer Gesellschaft, Vienna, 2007, pp. 81-90.

[9] W.S. Meng, H. von Grafenstein, and I.S. Haworth, *A model of water structure inside the HLA-A2 peptide binding groove*, Int. Immunol. 9 (1997), pp. 1339-1346.

[10] P.M. Petrone and A.E. Garcia, *MHC-peptide binding is assisted by bound water molecules*, J. Mol. Biol. 338 (2004), pp. 419-435.

[11] M. Zacharias and S. Springer, *Conformational flexibility of the MHC class I alpha1-alpha2 domain in peptide bound and free states: a molecular dynamics simulation study*, Biophys. J. 87 (2004), pp. 2203-2214.

[12] O. Michielin and M. Karplus, *Binding free energy differences in a TCR-peptide-MHC complex induced by a peptide mutation: a simulation analysis*, J. Mol. Biol. 324 (2002), pp. 547-569.

[13] V. Zoete and O. Michielin, *Comparison between computational alanine scanning and per-residue binding free energy decomposition for protein-protein association using*

For Peer Review Only

MM-GBSA: Application to the TCR-p-MHC complex, Proteins 67 (2007), pp. 1026-1047.

[14] S. Wan, P.V. Coveney, and D.R. Flower, *Molecular basis of peptide recognition by the TCR: affinity differences calculated using large scale computing*, J. Immunol. 175 (2005), pp. 1715-1723.

[15] T. Darden, D. York, and L. Pedersen, *Particle mesh Ewald: An N.log(N) method for Ewald sums in large systems*, J. Chem. Phys. 98 (1993), pp. 10089-10092.

[16] T.E. Cheatham and B.R. Brooks, *Recent advances in molecular dynamics simulation towards the realistic representation of biomolecules in solution*, Theor. Chem. Acc. 99 (1998), pp. 279-288.

[17] B.K. Andrews, T. Romo, J.B. Clarage, B.M. Pettitt, and J. Phillips, *Characterizing global substates of myoglobin*, Structure 6 (1998), pp. 587-594.

[18] D. Sengupta, R.N. Behera, J.C. Smith, and G.M. Ullmann, *The alpha helix dipole: Screened out?*, Structure 13 (2005), pp. 849-855.

[19] A. Wada, *The alpha-helix as an electric macro-dipole*, Advances in Biophysics (1976), pp. 1-63.

[20] W. Weber, P.H. Hünenberger, and J.A. McCammon, *Molecular dynamics simulations of a polyalanine octapeptide under Ewald boundary conditions: Influence of artificial periodicity on peptide conformation*, J. Phys. Chem. B 104 (2000), pp. 3668-3675.

[21] G.M. Torrie and J.P. Valleau, *Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling*, J. Comput. Phys. 23 (1977), pp. 187-199.

[22] D. Rognan, N. Zimmermann, G. Jung, and G. Folkers, *Molecular dynamics study of a complex between the human histocompatibility antigen HLA-A2 and the IMP58-66 nonapeptide from influenza virus matrix protein*, Eur. J. Biochem. 208 (1992), pp. 101-113.

[23] S. Wan, P. Coveney, and D.R. Flower, *Large-scale molecular dynamics simulations of HLA-A*0201 complexed with a tumor-specific antigenic peptide: can the alpha3 and beta2m domains be neglected?*, J. Comput. Chem. 25 (2004), pp. 1803-1813.

[24] The Austrian Grid Initiative, http://www.austriangrid.at, cited 2008

[25] H.J.C. Berendsen, D. van der Spoel, and R. van Drunen, *GROMACS: A message-passing parallel molecular dynamics implementation*, Comput. Phys. Commun. 91 (1995), pp. 43-56.

[26] E. Lindahl, B. Hess, and D. van der Spoel, *GROMACS 3.0: a package for molecular simulation and trajectory analysis*, J. Mol. Model. 7 (2001), pp. 306-317.

[27] G. Kingsley and J. Sieper, *Current perspectives in reactive arthritis*, Immunol. Today 14 (1993), pp. 387-391.

For Peer Review Only

[28] H.M. Bergman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, and P.E. Bourne, *The Protein Data Bank*, Nucleic Acids. Res. 28 (2000), pp. 235-242.

[29] D.R. Madden, J.C. Gorga, J.L. Strominger, and D.C. Wiley, *The three-dimensional structure of HLA-B27 at 2.1 A resolution suggests a general mechanism for tight peptide binding to MHC*, Cell 70 (1992), pp. 1035-1048.

[30] M. Hulsmeyer, R.C. Hillig, A. Volz, M. Ruhl, W. Schroder, W. Saenger, A. Ziegler, and B. Uchanska-Ziegler, *HLA-B27 subtypes differentially associated with disease exhibit subtle structural alterations*, J. Biol. Chem. 277 (2002), pp. 47844-47853.

[31] J. Zielkiewicz, *Structural properties of water: comparison of the SPC, SPCE, TIP4P, and TIP5P models of water*, J. Chem. Phys. 123 (2005), pp. 1-6.

[32] D. van der Spoel, E. Lindahl, B. Hess, A.R. van Buuren, E. Apol, P.J. Meulenhoff, D.P. Tieleman, A.L.T.M. Sijbers, K.A. Feenstra, R. van Drunen, and H.J.C. Berendsen, *Gromacs User Manual version 3.2*, 2004.

[33] C. Oostenbrink, A. Villa, A.E. Mark, and W.F. Van Gunsteren, *A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6*, J. Comput. Chem. 25 (2004), pp. 1656-1676.

[34] G.A. Kaminski, R.A. Friesner, J. Tirado-Rives, and W.L. Jorgensen, *Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides*, J. Phys. Chem. B 105 (2001), pp. 6474-6487.

[35] H.J.C. Berendsen, J.P.M. Postma, W.F. Van Gunsteren, A. DiNola, and J.R. Haak, *Molecular dynamics with coupling to an external bath*, J. Chem. Phys. 81 (1984), pp. 3684-3690.

[36] J.P. Ryckaert, G. Ciccotti, and H.J.C. Berendsen, *Numerical integration of the cartesian equations of motion of a system with constraints; molecular dynamics of n-alkanes.*, J. Comput. Phys. 23 (1977), pp. 327-341.

[37] F. Eisenhaber, P. Lijnzaad, P. Argos, C. Sander, and M. Scharf, *The Double Cube Lattice Method: Efficient Approaches to Numerical Integration of Surface Area and Volume and to Dot Surface Contouring of Molecular Assemblies*, J. Comput. Chem. 16 (1995), pp. 273-284.

[38] H. Frauenfelder, F. Parak, and R.D. Young, *Conformational substates in proteins*, Annual review of biophysics and biophysical chemistry 17 (1988), pp. 451-479.

[39] M. Neumann, *Dielectric relaxation in water. Computer simulations with the TIP4P potential*, J. Chem. Phys. 85 (1986), pp. 1567-1580.

[40] M.A. Cuendet, *The Jarzynski identity derived from general Hamiltonian or non-Hamiltonian dynamics reproducing NVT or NPT ensembles*, J. Chem. Phys. 125 (2006).

[41] S. Wan, P.V. Coveney, and D.R. Flower, *Peptide recognition by the T cell receptor: comparison of binding free energies from thermodynamic integration, Poisson-*

For Peer Review Only

Boltzmann and linear interaction energy approximations, Philos. Transact. A Math. Phys. Eng Sci. 363 (2005), pp. 2037-2053.

[42] M. Krogsgaard, N. Prado, E.J. Adams, X.L. He, D.C. Chow, D.B. Wilson, K.C. Garcia, and M.M. Davis, *Evidence that structural rearrangements and/or flexibility during TCR binding can contribute to T cell activation*, Mol. Cell 12 (2003), pp. 1367-1378.

[43] J.B. Reiser, C. Gregoire, C. Darnault, T. Mosser, A. Guimezanes, A.M. Schmitt-Verhulst, J.C. Fontecilla-Camps, G. Mazza, B. Malissen, and D. Housset, *A T cell receptor CDR3beta loop undergoes conformational changes of unprecedented magnitude upon binding to a peptide/MHC class I complex*, Immunity. 16 (2002), pp. 345-354.

[44] M.A. Saper, P.J. Bjorkman, and D.C. Wiley, *Refined structure of the human histocompatibility antigen HLA-A2 at 2.6 A resolution*, J. Mol. Biol. 219 (1991), pp. 277-319.

[45] W. Humphrey, A. Dalke, and K. Schulten, *VMD: visual molecular dynamics*, J. Mol. Graph. 14 (1996), pp. 33-38.

For Peer Review Only

7 Tables

Table 1. Overview of molecular dynamics simulations performed. Peptide main anchor residues are printed bold.

peptide	MHC	water model / force field combination	size of water box	parallel simulations	simulation time
ARAAAAAAA ("single site binding peptide")	B*2705 whole	SPC / GROMACS	20Å	4x	1ns
	B*2705 whole	SPC / GROMOS 53a6	20Å	4x	1ns
	B*2705 simplified	SPC / GROMACS	2.5Å	4x	1ns
	B*2705 simplified	SPC / GROMACS	5Å	4x	1ns
	B*2705 simplified	SPC / GROMACS	10Å	4x	1ns
	B*2705 simplified	SPC / GROMACS	10Å	1x	10ns
	B*2705 simplified	SPC / GROMACS	20Å	4x	1ns
	B*2705 simplified	SPC / GROMACS	40Å	4x	1ns
	B*2705 simplified	SPC / GROMOS 53a6	20Å	4x	1ns
	B*2705 simplified	TIP4P / OPLS-AA	20Å	4x	1ns
GRFAAAIAK ("double site binding peptide")	B*2705 simplified	SPC / GROMACS	12Å	1x	12ns
	B*2705 simplified	SPC / GROMACS	17Å	1x	12ns
	B*2705 simplified	SPC / GROMACS	27Å	1x	16ns

For Peer Review Only

Table 2. Characteristics of the "double site binding" peptide in crystal structure and in different substates occurring during the 16ns simulation. SASA is taken relative to SASA of the peptide in water.

configuration	time (ps)	mainly affected residues	SASA (relative)	interaction energy (kJ/mol)		RMSF (nm)
				peptide-water	peptide-MHC	
crystal	0		29.7%	$-3.8\ 10^{7}$	$-8.8\ 10^{7}$	0.037
substate #1	2,000 - 10,000		32.7%	$-4.4\ 10^{7}$	$-8.3\ 10^{7}$	0.037
substate #2	890 - 1,550	5A, 6A, 7I	36.8%	$-4.3\ 10^{7}$	$-8.5\ 10^{7}$	0.045
substate #3	11,050 - 12,500	8A, 9K	39.9%	$-5.1\ 10^{7}$	$-7.5\ 10^{7}$	0.038
substate #4	13,550 - 14,100	5A, 6A, 7I, 8A, 9K	42.3%	$-5.2\ 10^{7}$	$-7.0\ 10^{7}$	0.045

For Peer Review Only

8 Figure captions

Figure 1. HLA-B*2705 molecular surface and its binding pockets.

The pockets definition is according to [44]. Pocket A (M5, Y7, Y59, E63, Y159, E163, W167, Y171), pocket B (H9, T24, E45, L66, C67, Y99), pocket C (H9, K70, T73, D74, R97), pocket D (Y99, H114, L156, Y159, L160), pocket E (H114, W133, W147, V152, L156), pocket F (D77, T80, L81, Y84, D116, Y123, T143, K146, W147). The peptide's primary anchors P2 (conserved, Arg) and P9 form salt bridges and hydrogen bonds to the complementary negatively charged MHC pockets B and F, respectively. P3 interacts with the upper part of the hydrophobic pocket D. This one and all subsequent graphical representations of molecules were produced using the VMD package [45].

Figure 2. Crystal structure and simulated system.

*a) Crystal structure of a pMHC complex (m9 peptide complexed to HLA-B*2705 [30]). The peptide binding domain (α1 and α2 domain, white), the α3 domain (grey) and the associated β-2-microglobulin (black) are represented as "cartoons". The bound peptide is shown as "licorice". The dimensions of the MHC's peptide binding domain are given in Ångström.*

b) A simplified pMHC complex placed in the centre of a simulation box allowing for a 20 Ångström solvation shell resulting in altogether 99,101 atoms (water molecules in the front of the cube are carved out in order to make the pMHC complex visible). This system could be used as input for a MD simulation "as is".

Figure 3. RMSF of peptide residues' backbone atoms.

For Peer Review Only

We performed four independent 1ns runs (same starting structure, different starting velocities) for each simulation box size. The RMSF values shown are averages for the time interval from 600ps to 1000ps, which represents the equilibrated part of a stable run. Results for the individual runs are indicated by different line types. Low RMSF values indicate low positional fluctuations of the backbone atoms during the simulation, which means that these residues are likely to be tightly bound by the MHC. A residue displaying high RMSF values may be loosely bound or not bound at all.

Figure 4. Demonstration of the minimum distance between epitope C-terminus and MHC β-sheet.

a) Minimum distances of the epitope's C-terminal alanine to the MHC's β-sheet during 1ns simulations. In the simulation of the complex with a 20Å surrounding water shell (dashed line) the distance first rises gradually from 0.37nm to 0.7nm. Then the C-terminus totally detaches and moves as far as 1.4nm away. Although there are large fluctuations for this distance, in the simulation employing a 5Å water shell (continuous line) no detachment occurs during the 1ns run.

b) The dashed line marks the minimum distance of all the atoms of the epitope's C-terminal alanine to all the atoms of the MHC's β-sheet. The orange (light grey) peptide shows the distance in the crystal structure. The blue (dark grey) peptide depicts the increased distance after C-terminal detachment. For an animated display of the simulation please refer to (supplementary material S1).

Figure 5. Minimum distance between peptide C-terminus and MHC β-sheet during different simulations of the "single site binding" peptide.

For Peer Review Only

Three different force fields and two different water models were used for simulations of the simplified pMHC complex, and two different force fields with the same water model were used for the whole pMHC complex. Each combination was simulated in four independent runs for 1ns within a 20Å hydration shell (black, dark grey, grey, and light grey line).

a) Simulating the simplified pMHC complex enclosed in a 20Å hydration shell results for two out of the four simulations (black and dark grey) in the previously described partial detachment of the peptide (see Figure 4) when the GROMACS force field and the SPC water model are used. The rising distance of the peptide's C-terminal alanine residue to 1nm and above indicates total detachment from the MHC binding cleft.

b) Application of the GROMOS96 53a6 force field on the simplified system leads to peptide detachment in one out of four simulations (black).

c) When the OPLS-AA force field and the TIP4P water model are used on the simplified system, for all four simulations the peptide calmly resides in the binding groove and the minimum distance measure retains its initial low value for the entire length of the 1ns simulation.

d) Simulating the whole pMHC complex using the GROMACS force field and the SPC water model results in an onset of peptide detachment in two out of four simulations (black and dark grey).

e) Using the GROMOS96 53a6 force field on the whole pMHC complex leads to peptide detachment in one out of four simulations (black).

Figure 6. Orientational relaxation (RACF) of the pMHC complex for different sizes of the simulation box.

All the pMHC systems lose their orientational "memory" over time, which is shown by a decay of the RACF $C_1(t)$. The larger the surrounding water shell the faster the loss of memory. All three systems have been simulated for at least 12ns.

For Peer Review Only

Figure 7. Model for system size influencing rotational diffusion.

Both illustrations show a simulation box containing a protein solved in water using periodic boundaries but with different simulation box sizes. The darker the water molecules, the more structured is the water network. A small box hinders rotation by forcing a relatively tightly bound water layer to move in opposite directions.

Figure 8. Analysis of the long time simulation (16ns) of the "double site binding" peptide complexed to HLA-B*2705 (1jge) within a 27Å water box.

a) RMSD of the peptide's backbone (red line) and the MHC's backbone (black line) relative to the initial crystal structure. The peptide spends most of the time in a state (characterized by low RMSD values) that is very similar to the crystal structure, but seems to flip to a different conformational substates around simulation time t≈1ns and between t≈10ns and t≈14ns.

b) RMSD of specific peptide backbone atoms. The Cα atom of the primary anchor arginine (green line) remains close to its initial position throughout the simulation (low RSMD values). The Cα atom of the central alanine (position 5, blue line) temporarily moves farther off its position in the crystal structure after t≈1ns (higher RMSD values) while the Cα atom of the secondary anchor lysine (black line) undergoes an even larger dislocation (large RMSD values) between t≈10ns and t≈14ns. Therefore the peptide's residues differently contribute to the substates. Individual RMSD curves for all the peptide's Cα atoms can be found in (supplementary material S2).

c) Interaction energies of the peptide with the MHC (black line) and with water (grey line). Energies are the sum of Coulomb and Lennard Jones potential.

d) RMSD of each epitope's backbone atom relative to substate #1. This plot shows the substates and backbone atoms involved.

e) Conformational substates defined using above graphs.

For Peer Review Only

Figure 9. Average structures of peptide's conformational substates.

While the white peptide representations depict the substate #1, the grey representations depict the average structures of corresponding substates. The MHC structure shown is an average structure over the whole simulation. Arrows indicate major structural deviations to substate #1

For Peer Review Only

9 Word count

34 pages

8180 words

44563 characters

For Peer Review Only