Current status and future challenges in T-cell receptor/peptide/MHC molecular dynamics simulations

Bernhard Knapp, Samuel Demharter, Reyhaneh Esmaielbeiki and Charlotte M. Deane

Corresponding author. Bernhard Knapp, University of Oxford, Department of Statistics, South Parks Road, Peter Medawar Building for Pathogen Research, Oxford, OX1 3SY, United Kingdom. Tel.: +44 (0) 1865 2818989; Fax: +44 (0) 1865 281333; E-mail: bernhard.knapp@stats.ox.ac.uk

Abstract

The interaction between T-cell receptors (TCRs) and major histocompatibility complex (MHC)-bound epitopes is one of the most important processes in the adaptive human immune response. Several hypotheses on TCR triggering have been proposed. Many of them involve structural and dynamical adjustments in the TCR/peptide/MHC interface. Molecular Dynamics (MD) simulations are a computational technique that is used to investigate structural dynamics at atomic resolution. Such simulations are used to improve understanding of signalling on a structural level. Here we review how MD simulations of the TCR/peptide/MHC complex have given insight into immune system reactions not achievable with current experimental methods. Firstly, we summarize methods of TCR/peptide/MHC complex modelling and TCR/peptide/MHC MD trajectory analysis methods. Then we classify recently published simulations into categories and give an overview of approaches and results. We show that current studies do not come to the same conclusions about TCR/peptide/MHC interactions. This discrepancy might be caused by too small sample sizes or intrinsic differences between each interaction process. As computational power increases future studies will be able to and should have larger sample sizes, longer runtimes and additional parts of the immunological synapse included.

Key words molecular dynamics simulations; T-cell receptor; MHC

Introduction

The interaction between the T-cell receptor (TCR) of T cells and the major histocompatibility complex (MHC) of antigen presenting cells is a fundamental part of the adaptive immune response [1]. MHCs present potentially pathogenic peptides on the cell surface. These peptide/MHC (pMHC) complexes are bound by the highly polymorphic and flexible complementary determining regions (CDRs) of TCRs. Effective extracellular interaction between

TCR and pMHC causes a signalling cascade inside the T-cell membrane. This cascade triggers further downstream pathways within the cell, which finally lead to gene transcription in the nucleus and an immune reaction. If a peptide is not bound by MHCs, or pMHCs are not recognized by TCRs, then no further immune reaction against this peptide can take place [2].

The binding between a TCR and a pMHC is weak (usually in the range of 1–100 μM) and degenerate, i.e. many TCRs bind the same

Bernhard Knapp is a postdoctoral research fellow in Computational Structural Biology at the University of Oxford. His research focuses on the modelling of immune system-related protein structures and their dynamics.
Samuel Demharter is a PhD student in the Department of Computer Science at the University of Oxford. He uses coarse-grained molecular simulations to investigate functional motions in large protein structures.
Reyhaneh Esmaielbeiki is a postdoctoral researcher in Computational Structural Biology at University of Oxford. She has been working on protein interface prediction and modelling of membrane proteins.
Charlotte M. Deane is a Professor in the Department of Statistics, University of Oxford. Her research interests include the areas of protein structure prediction and evaluation, as well as protein interaction networks.

pMHC and many pMHCs are bound by the same TCR. However, not every TCR/pMHC interaction leads to T-cell activation. This prompts the question of what mechanism leads to effective TCR triggering? [3], i.e. TCR/peptide/MHC (TCRpMHC) interact and bind but only in some cases is a signal transmitted from this extracellular interaction to the cytoplasmic signalling network. This is likely to involve subtle structural adjustments. Understanding the basic mechanisms of TCRpMHC interactions is important, as it would aid our ability to treat cancer [4, 5], autoimmune diseases, allergy [6] and other immune system-related health conditions [7] by the rational design of customized peptides and TCRs. Several hypotheses of TCR-triggering mechanisms have been proposed [3, 8, 9]. Many of them are likely to involve subtle conformational adjustments in the TCR/peptide/MHC interface.

In this review we will focus on structural bioinformatics studies investigating conformational changes in the (TCR)pMHC interface and how these changes relate to function. We review Molecular Dynamics (MD) simulations that aim to understand the dynamics in the (TCR)pMHC interaction (Figure 1) and provide a detailed comparison of 25 studies of the over 500 MD simulations published since 2008 (for earlier studies see [10, 11]). Firstly, we describe methods of (TCR)pMHC complex modelling and (TCR)pMHC MD trajectory evaluation as necessary parts of any simulation study. Secondly, we classify simulations into categories and give an overview of approaches and results. Finally, we discuss the future directions and challenges.

Modelling of (TCR)pMHC complexes

The first step in an MD simulation is the generation of an initial structure (a workflow example for a complete pMHC MD simulation project is described in Supplementary Figure A1). Often this structure is a computationally altered model of an X-ray structure. Different modelling approaches can be employed depending on how similar the (TCR)pMHC complex of interest is to the closest X-ray structure.

Change of a small number of residues

Often X-ray structures are similar to the (TCR)pMHC complex of interest. If only one or very few amino acids need to be changed, an overall structure conservation can be assumed. In this case, side-chain substitution tools are a quick and easy way to model the complex of interest. It was shown previously [12, 13] that, for example, SCWRL [14] is a reliable tool for the generation of altered peptide ligands in a high-throughput framework such as PeptX [15].

In pure side-chain substitution approaches, the average root mean square deviation (RMSD) between the average peptide
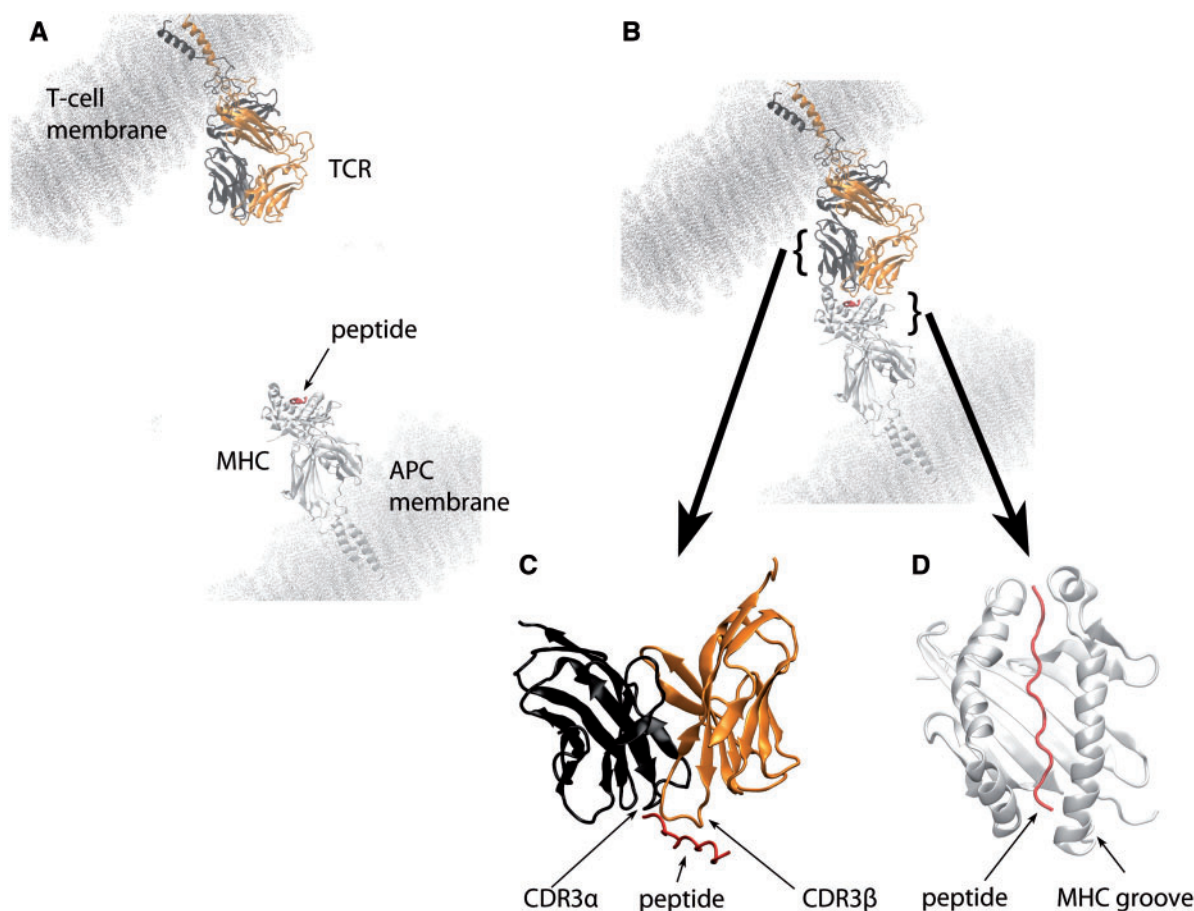

Figure 1. Interaction between TCR, peptide and MHC. White cartoon: MHC; black cartoon: TCRα chain; orange cartoon: TCRβ chain; grey dots: cell membranes. (**A**) A TCR bound to the membrane of a T cell and an MHC bound to the membrane of an antigen presenting cell (APC). TCR and MHC are not bound. (**B**) Same as in (**A**) but in the bound configuration. (**C**) Zoomed and rotated view of the variable TCR regions. The main peptide interaction loops of the TCR (CDR3α and CDR3β) are labelled. For clarity the MHC is not shown. (**D**) Zoomed and rotated view of the peptide bound in the MHC binding groove. For clarity the TCR is not shown. Experimental X-ray structures are only available for the ectodomains. Therefore, this rendering is based on the modelled structure published by Wan *et al.* [64]. A colour version of this figure is available at BIB online: http://bib.oxfordjournals.org.

backbone and the individual peptides is about 0.93 Å [12]. If side-chain substitutions slightly affect the backbone configuration, then the energy minimization and warming-up procedure during MD should take care of the necessary adjustments.

A typical example where this protocol would be used would be where the X-ray structure closest to the system of interest contains the correct MHC but the peptide differs by one amino acid from the system of interest.

Change of multiple peptide residues with expected impact on the backbone

If a whole peptide needs to be re-modelled and substantial changes in the peptide backbone conformation are expected, then specific tools can be applied. Recent examples include:

Bui *et al.* [16] presented PePSSI that predicts a peptide's structure within the MHC binding groove using a van der Waals contacts-based scoring system. Their method takes explicit water at the binding interface into account. In a benchmark of eight structures, they achieved all-atom RMSDs ranging from 1.3 Å to 2.5 Å when comparing predicted and X-ray conformations.

Bordner *et al.* [17] developed a docking protocol that optimizes a physical energy function of a multitude of conformational peptide variables. Their Internal Coordinate Mechanics tool-based [18] method achieved a backbone RMSD of 0.75 Å when cross-docking a flexible decapeptide. Furthermore, they showed the method's utility by achieving an average backbone RMSD of <1 Å when applied to homology models of two different allotypes.

Khan *et al.* [19] published the docking protocol pDOCK taking into account flexibility along the whole peptide. For the majority of test cases (159 of 186), their method achieved a Cα RMSD of <1 Å.

Alternatively, general ligand–protein docking tools (reviewed in [20]) can be used to model the peptide structure in the MHC binding groove.

This type of approach would be taken when the X-ray structure closest to the system of interest contains the correct MHC but the peptide has no similarities to the one of interest.

Change of multiple residues in TCR and MHC with expected impact on the backbone

If larger sequence changes from the most similar X-ray structure are necessary in the MHC and/or TCR, then standard homology modelling approaches might be most appropriate. It is assumed that the overall fold of a TCR and an MHC is conserved; however, the orientation of secondary structure elements can change by a few degrees. Also, conformation and length of several loops can differ (for details see the IMGT database [21]). A large number of homology modelling tools such as MODELLER [22] are available.

High-accuracy homology models are based on high sequence identity templates. The main chain RMSD error of these high-accuracy homology models can be around 1 Å or lower, which makes them roughly comparable with the accuracy of medium-resolution nuclear magnetic resonance (NMR) structures or low-resolution X-ray structures (for a review on homology modelling, see [23]). In the case of TCRpMHC modelling, the sequence identity is likely to be in the range of 80–90%, which would normally introduce only minor errors in side-chain positions that can be accommodated by the energy

minimization and warming-up procedure of the MD simulation protocol.

Homology modelling would be used when the X-ray structure closest to the system of interest contains the correct pMHC but some regions of the TCR do not match and have to be modelled.

Relative orientation between TCR and pMHC

TCRs and pMHCs bind each other in a relatively conserved way. However, the detailed binding angle differs based on MHC allele and TCR type [1]. This conserved binding pattern could be due to co-evolution of key amino acids to preserve TCR/pMHC binding or that the diagonal orientation of the TCR over pMHC allows for the strongest binding affinity and signal generation [24].

Several computational approaches for identifying the exact TCR/pMHC binding orientation have been reported. For example, Ferber *et al.* [25] used long-range intra-molecular coulomb interactions with implicit solvation to predict the TCR/pMHC binding mode. Roomp and Domingues [26] predicted the contacts between the TCR and MHC using a rule-based approach and Leimgruber *et al.* [27] developed a TCRpMHC structural modelling pipeline using sequence- and structure-based alignment.

Alternative methods to determine TCR orientation over pMHC include protein–protein docking algorithms (reviewed in [28]) and template-based superimposition on closely related and known TCRpMHC X-ray structures.

This type of procedure is needed when there is an X-ray structure of the pMHC of interest and a second structure of the TCR of interest but there is no combined experimental structure.

MD simulation protocols

MD simulations solve Newton's equation of motion for a given system of particles over time. In contrast to the static picture of a single structure, this gives insight into how particles move and behave over time. A set of rules, often called a force field, describes the interaction between the particles. MD simulations save the structural configuration at different time points to a trajectory file that can later be used for analysis.

A broad range of MD software packages are available. Well known examples are GROMACS [29], NAMD [30], AMBER [31] and CHARMM [32]. Supplementary Table A1 lists the MD package, force field and further protocol details about all of simulations that we discuss below. A parameter study for peptide/MHC MD simulations was published by Omasits *et al.* [33].

All atom MD simulations are computationally expensive. Even if strong parallel computing servers are available, overall runtimes for simulation projects are likely to be in the range of weeks if not months.

A number of ways to accelerate MD simulations exist. A frequently used way is coarse graining. Examples include bond length constraints [34], increased masses [35], virtual sites instead of hydrogens [36], n-bead models (e.g. amino acids described by three points [37]) or the movement of whole rigid protein regions [38]. Such coarse graining techniques save runtime but neglect details of the simulation. Which level of coarse graining is appropriate always depends on the system and the biological challenge under investigation. Another way to accelerate simulations is steered MD simulations [39]. In these, artificially applied forces pull or push selected parts of

the system under investigation. These simulations quickly reach the desired state but run the risk of being biased.

Methods for (TCR)pMHC trajectory analysis

After an MD simulation has been carried out, it is then necessary to analyze the output. A vast range of methods has been used for the analysis of (TCR)pMHC interactions (see Supplementary Table A1 for the studies and Supplementary Table A2 for a description of the methods). They include RMSD, root mean square fluctuation (RMSF), solvent accessible surface area (SASA), principal component analysis, normal mode analysis, hydrogen bonds, distance measurements, contact maps, backbone dihedral angles, hydrophobic contacts and radius of gyration. Tools to calculate these are implemented in several MD packages. The VMD [40] plug-in vmdICE [41] was developed to give a fast and easy representation of RMSD, RMSF and SASA directly on the three-dimensional structure of the molecule. Other approaches include the import of trajectory data into statistics packages such as Matlab [42].

Specific geometric approaches have also been developed. For example, Dunbar *et al.* [43] developed a method to analyse the relative orientation between the variable domains of the heavy and light chain of antibodies. These regions are highly homologous to the variable domains of the TCR α- and β-chain. Therefore, the method was recently adopted for TCRs [44, 45]. Hischenhuber *et al.* [46, 47] developed a curve and spline method to characterize the MHC α-helices based on differential geometric parameters such as curvature, torsion, ruled surface area, distribution parameter and conical curvature.

Free energy and entropy calculations can be used for (TCR)pMHC characterization. Energetic analysis can give important insights into the TCRpMHC interaction but are less frequently seen in the literature. Examples of free energy calculations are thermodynamics integration [48] or Molecular Mechanics Poisson−Boltzmann Surface Area (B. Knapp *et al.*, "Accurate peptide-MHC binding affinity predictions from in silico molecular dynamics" in review). Free energy calculations are reviewed in detail in [5].

Classification of (TCR)pMHC studies

We have classified (TCR)pMHC MD simulation studies into eight categories. The categories used are: 'simulations of the same MHC binding different peptides', 'simulations of different MHCs binding the same peptide', 'simulations of the same pMHC complex bound by different TCRs', 'simulations of TCRpMHC against pMHC', 'simulations of MHCs without bound peptide', 'simulations including trans-membrane regions and further receptors', 'steered TCR/pMHC simulations' and 'MD for refinement of MHC homology models'. Each study is summarized in the context of the category it fits best but many could be placed in multiple categories. The technical aspects of all studies are summarized in Supplementary Table A1.

Simulations of the same MHC binding different peptides

In this type of study, a particular MHC allele is selected and simulated in complex with two or more different peptides. These studies may include TCRs or consist only of peptide/MHC interactions. Usually the immunological background of the different peptides is known. For example, one peptide is highly immunogenic and the other a single-point mutated version, which does not induce an immune reaction.

A recent study of this type is of Alvarez-Navarro *et al.* [49] who used mass spectrometry to identify novel HLA-B27-restricted epitopes from the bacterium *Chlamydia trachomatis* that can trigger reactive arthritis. The authors performed MD simulations of HLA-B27 in complex with the two bacterial peptides DNA primase $_{211–221}$ (RRFKEGGRGGK) and DNA primase $_{211–223}$ (RRFKEGGRGGKYI) as well as a naturally occurring ligand created by the endogenous processing of the HLA-B27 heavy $_{309–320}$ (RRKSSGGKGGSY) and the original X-ray structure peptide pVIPR$_{400-408}$ (RRKWRRWHL) of PDB accession code 1OGT. On the basis of their MD simulations, the authors concluded that molecular mimicry between bacterial and self-derived HLA-B27 ligands might contribute to the pathology of reactive arthritis.

In Knapp *et al.* [50], we investigated the effect of peptide flanking regions on MHC binding affinity, T-cell activation and pMHC conformation. We simulated HLA-DRB1*01:01 in complex with the 12-mer (KCIEWEKAQHGA) and the 18-mer (NKKCDKKCIEWEKAQHGA) version of the major mug pollen allergen Art v 1. Experimentally we found that the potency to activate Art v 1-specific T-cells and the MHC binding affinity were increased for the 18-mer. MD simulations indicated that the N-terminal peptide flanking region of the 18-mer binds outside the MHC binding groove, which could explain the higher potency of the 18-mer in contrast to the 12-mer.

In a second study of a similar type [51], we investigated the experimental allergic encephalomyelitis (EAE)-associated myelin basic protein (MBP) AcN1-11 (ASQKRPSQRHG) and its mutant peptides AcN1-11[4A] (ASQARPSQRHG) and AcN1-11[4M] (ASQMRPSQRHG) in complex with the MHC class II allele I-A^{u} by experimental and MD techniques. MD indicated that K4 of the wild-type peptide leaves MHC binding pocket [1] number 6 and becomes available for TCR interaction, which was not the case for the two mutant peptides. This was in agreement with experimental binding affinity data, which showed AcN1-11 ($IC_{50} = 7.4$ mM) < AcN1-11[4A] ($IC_{50} = 0.019$ mM) < AcN1-11[4M] ($IC_{50} = 0.00064$ mM). However, we also found that these three altered peptide ligands differ in severity of peptide-induced EAE in mice (AcN1-11[4A] (0% incidence) < AcN1-11 (80% incidence) < AcN1-11[4M] (100% incidence)). As binding affinity did not correlate with *in vivo* disease incidence, we suggested that spatial re-arrangements in the binding surface could offer another level of T-cell regulation.

In a similar study, Laimou *et al.* [52] investigated I-A^{u} bound AcN1-11, AcN1-11[4A] and AcN1-11[4Y] by NMR and MD simulations. They found that the main MHC contact residues S2, P6 and S7 showed a similar configuration for all three peptides, while the TCR contact residue Q3 is buried in the MHC for AcN1-11[4A]. The authors suggested that this could hamper Q3 from forming interactions with the TCR. This blocked interaction might explain why AcN1-11[4A] did not induce EAE in mice [51].

Omasits *et al.* [33] published an MD parameter study using the peptides GRFAAAIAK and ARAAAAAAA bound by HLA-B*27:05. Initial conditions, size of water shell, water model/force field combination, simulation time and simplifications in the system were investigated. The authors found, for example, that the minimal distance between simulation box boundary and pMHC should be >10 Å to avoid artificial stabilization of the non-binder ARAAAAAAA in the MHC binding groove.

In one of the most recent studies of this type [44] (following the proof-of-concept in [53]), we performed 100 ns MD simulations of the LC13 TCR and HLA-B*08:01 in complexes with the Epstein Barr Virus (EBV) peptide FLRGRAYGL and all possible 171 single-point mutations of this peptide. For each of these 172 TCRpMHC MD simulations the peptide immunogenicity is

known. We found that more and less immunogenic peptides induce different changes in the TCRpMHC interface. All these differences were, while being statistically significant, not strong enough to predict peptide immunogenicity. In terms of absolute values, only minor differences in the distribution of interface distances, hydrogen binding footprints of the MHC helices and the orientation between the TCRα and TCRβ chain could be found. In this study, we also demonstrated that the comparison between single simulations might be misleading. We showed that the wild-type peptide FLRGRAYGL and the non-immunogenic mutant FLRGRAAGL differ dramatically in terms of hydrogen bonds, solvent-accessible surface area of CDRs and flexibility of CDRs and peptide. One might be tempted to draw the conclusion that these three properties are responsible for the immune reaction against the wild-type only and not against the mutant. However, if all 172 simulations are taken into account, no such general trends are seen.

The current MD studies do not proof a conserved casual factor for peptide immunogenicity. A large-scale study [44] showed a weak link between peptide immunogenicity and TCR chain orientation or hydrogen binding patterns in the TCRpMHC interface.

Simulations of different MHCs binding the same peptide

In this type of MD study, different MHC alleles are used and simulated with one or more peptides. Usually one of the MHC alleles is associated with a certain type of disease, while the other allele is not.

In a classical study of this type, Narzi *et al.* [54] simulated the ankylosing spondylitis-associated HLA-B*27:05 and the non-ankylosing spondylitis-associated HLA-B*27:09 in complex with one viral (RRRWRRLTV) and three self peptides (RRLPIFSRL, RRKWRRWHL and RRRWHRWRL). Simulations showed an increased entropy for the viral peptide when bound by the disease-associated HLA-B*27:05. The disease-associated MHC allele also showed an increased flexibility of its α1-helix. The authors concluded that conformational flexibility is affected by both the bound peptide and the MHC allele and that conformational flexibility could play a role in the recognition of MHCs by TCRs.

Reboul *et al.* [55] performed MD simulations of HLA-B*35:01 and HLA-B*35:08 in complex with the EBV peptide LPEPLPQGQLTAY. HLA-B*35:01 and HLA-B*35:08 differ only by a single-point mutation (L156R) in the MHC α2-helix. However, only LPEPLPQGQLTAY/HLA-B*35:08 induces a TCR SB27-dominated CTL response. The authors found a higher flexibility for LPEPLPQGQLTAY if bound by HLA-B*35:01 than if bound by HLA-B*35:08. They conclude that increased flexibility might hamper productive TCR interaction and might be responsible for the lack of CTL response.

In Knapp *et al.* [56] we investigated the molecular background of the major mug pollen allergen Art v 1$_{25\text{-}36}$ (KCIEWEKAQHGA) binding to HLA-DR1 and HLA-DR4. Both alleles bind the Art v1 peptide. However, only the frequency of HLA-DR1 is strongly increased in allergic humans. Experimentally we found that DR1 was more effective in T-cell stimulation at low peptide concentrations than DR4. Also, the minimal necessary epitope was shorter for DR1. MD simulations indicated an increased flexibility of both peptide termini when binding to DR1 and an altered SASA of the peptide. On this basis, we concluded that these differences in the peptide/MHC surface between DR1 and DR4 might have an influence on efficient T-cell activation in mug pollen allergy.

Kumar *et al.* [57] performed MD simulations of the multiple sclerosis predisposing HLA-DRB1*15:01 and the protective HLA-DRB1*16:01 in the unbound configuration as well as bound by the MBP peptide ENPVVHFFKNIVTP and the EBV peptide PGRRPFFHPVGEAD. The disease predisposing allele formed a stable complex with both peptides while the non-predisposing allele formed a stable complex with the MBP peptide only. The authors concluded that multiple sclerosis could partly be linked to different interaction and binding mechanisms between peptides and predisposing/non-predisposing MHC alleles.

The overall picture given by this type of simulation is somewhat unclear. For example, two studies [54, 56] found that the disease-associated MHC interface has a higher flexibility, while two other studies [55, 57] found that the disease-associated MHC interface has a lower flexibility. A larger scale analysis will be needed to determine when, if and how interface flexibility is related to an altered immune response (see section 'The future').

Simulations of the same pMHC complex bound by different TCRs

In this type of study, authors use different TCRs on the same or multiple peptide/MHC complexes. These TCRs usually differ from each other in terms of binding affinity or specificity.

A canonical example is Wolfson *et al.* [58] who investigated the effect of two different TCRs as well as four different peptides on the binding interface between TCR and the MHC L^d using eight independent simulations. They simulated the alloreactive 2C TCR as well as the 100-fold affinity increased 2C mutant m6. These two TCRs differ only in the residues 99–103 of CDR3α being GFASA in the case of 2C and HQGRY in the case of m6. The authors found that the binding footprint of the TCR on the pMHC is relatively insensitive to mutations in CDR3α. In contrast, mutations in the peptide induced several small changes in the TCR/MHC interface. The authors concluded that subtle, but important, structural and dynamical characteristics in the TCR/pMHC interface might play an important role in alloreactivity.

A second study by Cuendet *et al.* [39] found a similar binding configuration but a different binding procedure for two different TCRs. This study was based on steered MD simulations and is therefore described in the corresponding section in more detail.

Owing to the limited number of studies of this type, clear conclusions cannot be drawn. Additional large-scale studies will be needed to understand the influence of different TCRs on the peptide/MHC configuration.

Simulations of TCRpMHC against pMHC

This type of study compares the dynamics of the peptide and MHC in the TCR bound and unbound state. Experimentally it was shown that pMHC ligandation of the TCR can induce a conformational change in the constant TCR domain [59].

The only recent study of this type was performed by Stavrakoudis [60] who simulated the DM1 TCR and HLA-B*44:05 in complex with the peptide EENLLDFVRF. From X-ray crystallography it is known that the L5-D6 peptide bond is in the *cis* conformation for the TCR bound state (PDB-ID 3DXA) and in the *trans* conformation for TCR unbound state (PDB-ID 3DX8). The author simulated the TCRpMHC complex in the peptide *cis* conformation (from the TCR bound X-ray structure) as well as a superimposition of the peptide *trans* state (from the TCR

unbound X-ray structure) on the TCRpMHC complex. Energetic as well as solvent-exposed surface analysis showed that the TCR prefers the peptide in the *cis* conformation as shown by the X-ray structure.

Simulations of MHCs without bound peptide

To date, no experimental structure of an empty MHC binding groove exists. This might be owing to the instability of such a complex. Therefore, several theoretical groups have performed computational studies on an empty MHC binding site.

In an early study, Painter *et al.* [61] performed 60 ns MD simulations of the peptide-free structure of HLA-DR1. They observed an unfolding of the MHC helix region α50-59. This region adopted a peptide-like conformation binding to MHC pockets P1 to P4. The authors concluded that this might be the peptide-free HLA-DR1 conformation. The authors experimentally support their findings by MHC-conformation-specific superantigen and antibody probes for open/closed configurations of the groove.

Rupp *et al.* [62] also investigated the transition from the open binding groove state of HLA-DR1 to the closed state. Using 30 ns MD simulations they found a straightening of the β1-domain helix. This straightening resulted in a narrowing of the MHC binding groove. The new closed conformation is stabilized by a hydrogen bond formed between β-chain-N82 of the MHC β1domain-α-helix and α-chain-Q9 close to the P1 pocket of the MHC binding groove floor. Experimental mutagenesis supported the importance of these residues. The authors concluded that their simulation result might be the closed HLA-DR1 conformation.

Yaneva *et al.* [63] performed 25 ns MD simulations of HLA-DR3 with and without the invariant chain-associated protein peptide bound. While the simulations with the empty groove resulted in larger fluctuations of the helix domains (α34-40, α51-55, β78-86, β18-24 and β58-70), an overall rearrangement or closing of the groove was not observed. The authors also found that filling the P1 pocket of the MHC with a short peptide or an MHC β-chain V86Y mutation is sufficient to increase the overall stability of the complex. The authors state that this is in agreement with experiments showing that antibodies specific for the peptide-free MHC fail to bind to MHCs where just the P1 pocket is filled. They conclude that filling the P1 pocket of the MHC might be sufficient to stabilize the whole MHC binding groove in its empty state.

The three studies described above draw different conclusions as to the conformation of the empty structure and all three studies support their model by indirect experimental data. A possible conclusion could be that a sole empty MHC configuration does not exist and that an empty MHC is a constantly rearranging structure. This would be in agreement with the fact that to date no experimental X-ray structure of an empty MHC exists.

Simulations including trans-membrane regions and further receptors

Most MD approaches focus on the simulation of the extracellular parts of the (TCR)pMHC. However, for a productive immune reaction, more than these parts are needed. Additional components to be included in the simulations would be the membranes, the trans-membrane regions, the intracellular signalling motifs, CD4/CD8, CD3, CD45, etc. However, with each component the simulation runtime of the system increases

dramatically. Therefore, only few studies including more than the TCRpMHC structure have been reported.

To date, only Wan *et al.* [64] published an MD simulation of membrane-bound TCR, MHC and CD4. Consisting of only 10 ns this simulation is rather short but due to the large system (TCRpMHC, membrane, CD4) still highly resource consuming. This study was the first step in the direction of an all-atom simulation of the immunological synapse. In a subsequent review [10], the authors announced that they will take the next step by simulating a 1 million atom system of four membrane-bound TCRpMHC/CD4 complexes with explicit water. This study will bring the field even closer to an atomistic insight in the immunological synapse.

Bello and Correa-Basurto [65] performed MD simulations of HLA-DRB1*04:01 loaded with the peptide ARAMCSL or VNSDTVGWSWPDGAELPFTIDK as well as in the peptide-free configuration. These simulations were carried out for 100 ns without membranes and in independent runs for 150 ns in the membrane-bound environment. No striking differences between the simulations with and without membranes were found but the authors concluded that membrane-bound simulations were more stable and might provide more reliable insights in peptide/MHC interaction. Thus, including membranes in MHC simulations could benefit the field.

With increasingly available computational resources, the future of (TCR)pMHC simulations will move towards membrane-bound settings with additional receptors. This might be the most appropriate way to understand complex formation, signal creation and transduction.

Steered TCR/pMHC simulations

In steered MD simulations (also reviewed in [5]), authors apply artificial forces to parts of a system to move these parts in a specific direction. These techniques can be applied to obtain insights into molecular rearrangement processes that would take several seconds of real time while current MD simulations are at the micro second scale.

In a classic study, Cuendet and Michielin [66] performed steered MD simulations of the A6 TCR in complex with HLA-A*02:01. The HLA was loaded with the agonistic HTLV-1 virus Tax nonapeptide LLFGYPVYV or the antagonistic mutant LLFGYAVYV. The authors used a pulling code, which removed the TCR from the peptide/MHC at a speed of 5×10^{-4} nm/ps, i.e. in the 4 ns of simulation a distance of 20 Å. This code applied individual pulling potentials to each heavy atom of the backbone, allowing for a preserved internal structure and orientation of the peptide/MHC and the TCR. The authors performed 152 replica simulations of the agonistic TCRpMHC complex and 162 of the antagonistic complex. In this methodological study, the authors attempt to gain insight into the dissociation mechanism of a TCR from an agonistic and an antagonistic peptide/MHC complex. In a follow-up study [39], the authors extended their approach (steered MD, 4 ns, ~150 replica) from the same systems as in the previous study (A6 TCR / LLFGYPVYV / HLA-A*02:01 and A6 TCR / LLFGYAVYV / HLA-A*02:01) to the B7 TCR / LLFGYPVYV / HLA-A*02:01 system. Additionally they performed non-steered 2 ns simulation of the bound (TCRpMHC) as well as unbound (TCR/pMHC) state of each replica. They found that, while the A6 and B7 TCR have a similar bound X-ray configuration with HLA-A*02:01, their binding procedure differs. In contrast, the binding procedure for the A6 TCR to the agonistic and antagonistic pMHC is highly similar. These types of steered MD studies show that while two X-ray structures of
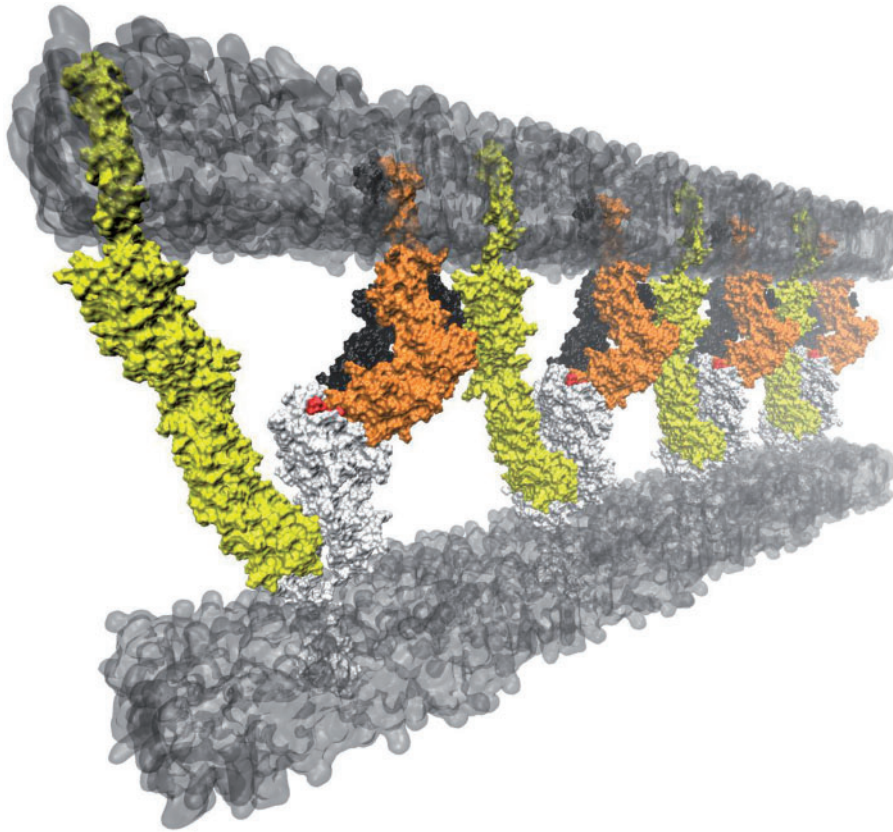
Figure 2. Model of multiple TCRs (α = black, β = orange), MHCs (white), peptides (red), CD4s (yellow), membrane (transparent grey) interactions. Rendering is based on the coordinates of [10]. A colour version of this figure is available at BIB online: http://bib.oxfordjournals.org.
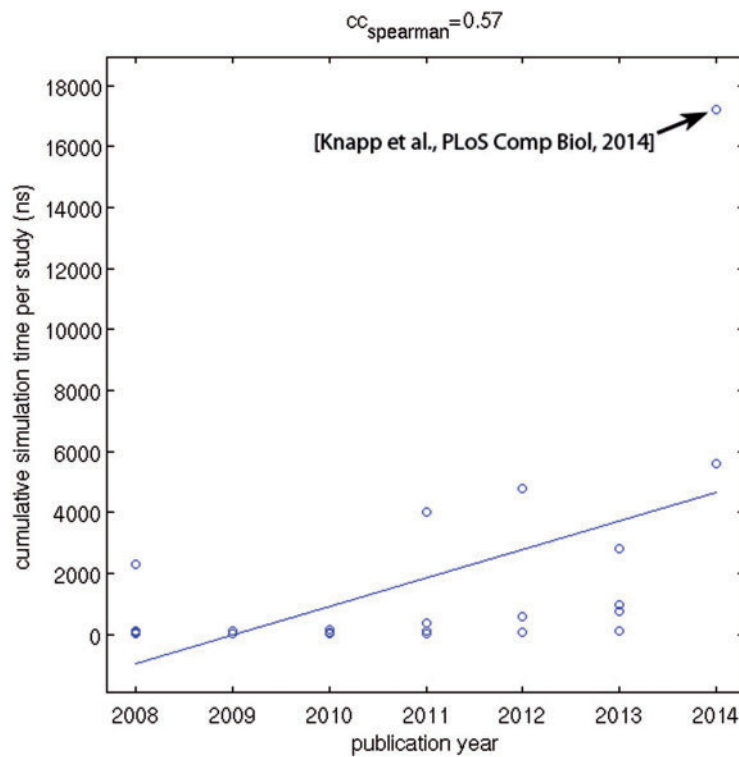

Figure 3. Correlation between publication year and overall simulation time per published study. The simulation time has constantly increased. The largest study containing 172 different 100 ns simulations was recently published by our group (marked in the plot above).

TCRpMHC are highly similar, their trajectory to complex formation might differ.

The long-term aim of such steered simulations could be to monitor the whole engagement and dis-engagement of a T-cell surface to/from an antigen presenting cell surface. However, due to computational resources, it may be some time before this becomes feasible.

MD for refinement of MHC homology models

Several authors have used MD simulations of (TCR)pMHC to refine and validate homology models. While MD simulations are unlikely to reduce the RMSD of a homology model to a (known) X-ray structure [67], such simulations can still check the consistency of a model, that is, does the homology model immediately fall apart or is a stable simulation possible?

Cardenas *et al.* [68] predicted Atlantic salmon MHC structures on the basis of homology modelling and subsequent MD refinement. In the absence of experimental structures for salmon MHC alleles, this study provides insights into residue differences between human and fish MHCs at specific positions in the MHC binding groove.

De Rosa *et al.* [69] homology modelled and MD refined the TCR-Vbeta/collagenII(261-273)/HLA-DR4 complex associated with rheumatoid arthritis to provide insights on molecular mechanisms of self reactivity.

The future

A way to improve the reliability of conclusions: sample size

Many current MD and X-ray studies test one or more hypotheses with a sample size of one complex per group. If the number of investigated properties is large and the number of complexes per group low, then one will always find a difference among them (for an explained example see [44]). Therefore, we suggest that future MD simulations in the (TCR)pMHC system should be planned more like clinical studies rather than explorative investigations. Each study should have one or more clear and testable hypotheses and a corresponding profound statistical power analysis determining the necessary sample size. This could, for example, be: 'We hypothesise that the predisposing MHC allele X will bind viral peptides (n = N) with a *P* percent higher RMSF then the non-predisposing MHC allele Y' where *N* and *P* are determined before the first simulation is run or analysed. On this basis, readers of the article will know (with a certain chance for Type I and Type II error) that the found differences between the two MHC alleles are more than coincidence.

A way to improve the reliability of simulations: perform replica MD simulations

If the runtime is finite, two identical simulations with different random seeds will differ in some aspects. Therefore, multiple replicas of the same simulation should be run. This is especially true if the sample size cannot be increased (see the previous section). Mean values of several replicas are far more reliable than single simulations (B. Knapp *et al., in review*) [70] whose convergence is questionable [71]. The number of replicas can be determined by a boot-strap analysis of the standard deviation between the simulations. For example, in a previous study, 50 replicas were needed for a reliable free energy prediction approach on HIV drugs [70]. In a current study (B. Knapp *et al.,* "Hierarchical Natural Move Monte Carlo gives insight into peptide/MHC detachment processes" submitted for publication), we are investigating the detachment of peptides from MHCs and found that 30–60 replicas are required for a reproducible comparison with experimental data.

Towards simulation of the immunological synapse

Understanding the interaction between MHC, peptide and the variable domains of the TCR is of high interest to the community. However, in future the interest will be in systems that are closer to the natural setting for signal transduction. Including membranes, co-receptors and the CD3 signalling motifs could be first steps. Also, multiple TCRpMHC receptor interactions (Figure 2) could be more realistic and provide additional insights.

Simulation times have considerably increased over the last years (Figure 3). However, there is still a long way to go, as the time necessary for the formation of the immunological synapse is in the range of minutes [72] and not the nanoseconds or microseconds of current simulations.

Supplementary data

Supplementary data are available online at http://bib.oxfordjournals.org/.

Key Points

- MD simulations can be used to investigate the dynamics of TCRpMHC complexes and have not yet been used for the whole immunological synapse. This might be due to the massive computational power needed.
- Here we presented a review of the methods for TCRpMHC complex modelling and TCRpMHC MD trajectory evaluation as necessary parts of any simulation. We then classified simulations into categories and gave an overview of approaches and results.
- We found that current studies often compare a small number (<10) MD simulations of different complexes. Given that a TCRpMHC complex has roughly 800 amino acids these studies run the risk of reporting differences between simulations that do not hold true for larger sample sizes.
- Future studies should have considerably larger numbers of simulations per group and contain additional parts of the immunological synapse that are not present in current simulations.

Reference

1. Rudolph MG, Stanfield RL, Wilson IA. How TCRs bind MHCs, peptides, and coreceptors. *Annu Rev Immunol* 2006;**24**:419–66.
2. Janeway CA, Travers P, Walport M. *et al*. Immuno Biology. *Garland Sci* 2005;1–888.

3. van der Merwe PA, Dushek O. Mechanisms for T cell receptor triggering. *Nat Rev Immunol* 2011;**11**:47–55.
4. Slingluff CL, Jr. The present and future of peptide vaccines for cancer: single or multiple, long or short, alone or in combination? *Cancer J* 2011;**17**:343–50.
5. Zoete V, Irving M, Ferber M, *et al.* Structure-based, rational design of t cell receptors. *Front Immunol* 2013;**4**:268.
6. Larche M. T cell epitope-based allergy vaccines. *Curr Top Microbiol Immunol* 2011;**352**:107–19.
7. Purcell AW, McCluskey J, Rossjohn J. More than one reason to rethink the use of peptides in vaccine design. *Nat Rev Drug Discov* 2007;**6**:404–14.
8. Gras S, Burrows SR, Turner SJ, *et al.* A structural voyage toward an understanding of the MHC-I-restricted immune response: lessons learned and much to be learned. *Immunol Rev* 2012;**250**:61–81.
9. Wang JH, Reinherz EL. The structural basis of alphabeta T-lineage immune recognition: TCR docking topologies, mechanotransduction, and co-receptor function. *Immunol Rev* 2012;**250**:102–19.
10. Flower DR, Phadwal K, Macdonald IK, *et al.* T-cell epitope prediction and immune complex simulation using molecular dynamics: state of the art and persisting challenges. *Immunome Res* 2010;**6**(Suppl 2):S4.
11. Stavrakoudis A. Conformational flexibility in designing peptides for immunology: the molecular dynamics approach. *Curr Comput Aided Drug Des* 2010;**6**:207–22.
12. Knapp B, Omasits U, Schreiner W. Side chain substitution benchmark for peptide/MHC interaction. *Protein Sci* 2008;**17**: 977–82.
13. Knapp B, Omasits U, Frantal S, *et al.* A critical cross-validation of high throughput structural binding prediction methods for pMHC. *J Comput Aided Mol Des* 2009;**23**:301–7.
14. Krivov GG, Shapovalov MV, Dunbrack RL, Jr. Improved prediction of protein side-chain conformations with SCWRL4. *Proteins* 2009;**77**:778–95.
15. Knapp B, Giczi V, Ribarics R, *et al.* PeptX: using genetic algorithms to optimize peptides for MHC binding. *BMC Bioinformatics* 2011;**12**:241.
16. Bui HH, Schiewe AJ, von Grafenstein H, *et al.* Structural prediction of peptides binding to MHC class I molecules. *Proteins* 2006;**63**:43–52.
17. Bordner AJ, Abagyan R. Ab initio prediction of peptide-MHC binding geometry for diverse class I MHC allotypes. *Proteins* 2006;**63**:512–26.
18. Molsoft LLC. *ICM-Docking.* 2006. http://www.molsoft.com/docking.html.
19. Khan JM, Ranganathan S. pDOCK: a new technique for rapid and accurate docking of peptide ligands to major histocompatibility complexes. *Immunome Res* 2010;**6**(Suppl 1):S2.
20. Sousa SF, Fernades P, Ramos MJ. Protein-ligand docking current status and future challanges. *Proteins* 2006; **65**:15–26.
21. Lefranc MP, Giudicelli V, Ginestoux C, *et al.* IMGT, the international ImMunoGeneTics information system. *Nucleic Acids Res* 2009;**37**:D1006–12.
22. Webb B, Sali A. Comparative protein structure modeling using MODELLER. *Curr Protoc Bioinformatics* 2014;**47**:5.
23. Baker D, Sali A. Protein structure prediction and structural genomics. *Science* 2001;**294**:93–6.
24. Collins EJ, Riddle DS. TCR-MHC docking orientation: natural selection, or thymic selection? *Immunol Res* 2008;**41**: 267–94.
25. Ferber M, Zoete V, Michielin O. T-cell receptors binding orientation over peptide/MHC class I is driven by long-range interactions. *PLoS ONE* 2012;**7**:e51943.
26. Roomp K, Domingues FS. Predicting interactions between T cell receptors and MHC-peptide complexes. *Mol Immunol* 2011;**48**:553–62.
27. Leimgruber A, Ferber M, Irving M, *et al.* TCRep 3D: an automated in silico approach to study the structural properties of TCR repertoires. *PLoS ONE* 2011;**6**:e26301.
28. Janin J. Protein-protein docking tested in blind predictions: the CAPRI experiment. *Mol Biosyst* 2010;**6**: 2351-2362.
29. Pronk S, Pall S, Schulz R, *et al.* GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* 2013;**29**:845–54.
30. Phillips JC, Braun R, Wang W, *et al.* Scalable molecular dynamics with NAMD. *J Comput Chem* 2005;**26**:1781–802.
31. Case DA, Cheatham TE, Darden T, *et al.* The amber biomolecular simulation programs. *J Comput Chem* 2005;**26**:1668–88.
32. Brooks B, Bruccoleri RE, Olafson BD, *et al.* CHARMM—a program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem* 1983;**4**:187–217.
33. Omasits U, Knapp B, Neumann M, *et al.* Analysis of key parameters for molecular dynamics of pMHC molecules. *Mol Simul* 2008;**34**:781–93.
34. Mazur AK. Hierarchy of fast motions in protein dynamics. *J Phys Chem B* 1998;**102**:473–9.
35. Feenstra KA, Hess B, Berendsen HJ. Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. *J Comput Chem* 1999;**20**:768–98.
36. Hess B, Kutzner C, vanderSpoel D, *et al.* GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory Comput* 2008;**4**:435–47.
37. Minary P, Levitt M. Probing protein fold space with a simplified model. *J Mol Biol* 2008;**375**:920–33.
38. Sim AY, Levitt M, Minary P. Modeling and design by hierarchical natural moves. *Proc Natl Acad Sci USA* 2012;**109**:2890–5.
39. Cuendet MA, Zoete V, Michielin O. How T cell receptors interact with peptide-MHCs: a multiple steered molecular dynamics study. *Proteins* 2011;**79**:3007–24.
40. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graph* 1996;**14**:33–8.
41. Knapp B, Lederer N, Omasits U, *et al.* vmdICE: a plug-in for rapid evaluation of molecular dynamics simulations using VMD. *J Comput Chem* 2010;**31**:2868–73.
42. Dien H, Deane CM, Knapp B. Gro2mat: a package to efficiently read gromacs output in MATLAB.. *J Computat Chem* 2014;**35**: 1528–31.
43. Dunbar J, Fuchs A, Shi J, *et al.* ABangle: characterising the VH-VL orientation in antibodies. *Protein Eng Des Sel* 2013;**26**: 611–620.
44. Knapp B, Dunbar J, Deane CM. Large scale characterization of the LC13 TCR and HLA-B8 Structural landscape in reaction to 172 altered peptide ligands: a molecular dynamics simulation study. *PLoS Comput Biol* 2014;**10**:e1003748.
45. Dunbar J, Knapp B, Fuchs A, *et al.* Examining variable domain orientations in antigen receptors gives insight into TCR-like antibody design. *PLoS Comput Biol* 2014;**10**:e1003852.
46. Hischenhuber B, Frommlet F, Schreiner W, *et al.* MH2c: Characterization of major histocompatibility—helices—an information criterion approach. *Comput Phys Commun* 2012; **183**:1481–90.
47. Hischenhuber B, Havlicek H, Todoric J, *et al.* Differential geometric analysis of alterations in MH alpha-helices. *J Comput Chem* 2013;**34**:1862–79.

48. Michielin O, Karplus M. Binding free energy differences in a TCR-peptide-MHC complex induced by a peptide mutation: a simulation analysis. *J Mol Biol* 2002;**324**:547–69.

49. Alvarez-Navarro C, Cragnolini JJ, Dos Santos HG, *et al.* Novel HLA-B27-restricted epitopes from *Chlamydia trachomatis* generated upon endogenous processing of bacterial proteins suggest a role of molecular mimicry in reactive arthritis. *J Biol Chem* 2013;**288**:25810–25.

50. Knapp B, Omasits U, Bohle B, *et al.* 3-Layer-based analysis of peptide-MHC-interaction: in silico prediction, peptide binding affinity and T cell activation in a relevant allergen-specific model. *Mol Immunol* 2009;**46**:1839–44.

51. Knapp B, Omasits U, Schreiner W, *et al.* A comparative approach linking molecular dynamics of altered peptide ligands and MHC with in vivo immune responses. *PLoS ONE* 2010;**5**: e11653.

52. Laimou D, Lazoura E, Troganis AN, *et al.* Conformational studies of immunodominant myelin basic protein 1-11 analogues using NMR and molecular modeling. *J Comput Aided Mol Des* 2011;**25**:1019–32.

53. Knapp B, Dorffner G, Schreiner W. Early relaxation dynamics in the LC 13 T cell receptor in reaction to 172 altered peptide ligands: a molecular dynamics simulation study. *PLoS ONE* 2013;**8**:e64464.

54. Narzi D, Becker CM, Fiorillo MT, *et al.* Dynamical characterization of two differentially disease associated MHC class I proteins in complex with viral and self-peptides. *J Mol Biol* 2012; **415**:429–42.

55. Reboul CF, Meyer GR, Porebski BT, *et al.* Epitope flexibility and dynamic footprint revealed by molecular dynamics of a pMHC-TCR complex. *PLoS Comput Biol* 2012;**8**: e1002404.

56. Knapp B, Fischer G, Van HD, *et al.* Association of HLA-DR1 with the allergic response to the major mugwort pollen allergen: molecular background. *BMC Immunol* 2012; **13**:43.

57. Kumar A, Cocco E, Atzori L, *et al.* Structural and dynamical insights on HLA-DR2 complexes that confer susceptibility to multiple sclerosis in Sardinia: a molecular dynamics simulation study. *PLoS ONE* 2013;**8**:e59711.

58. Wolfson MY, Nam K, Chakraborty AK. The effect of mutations on the alloreactive T cell receptor/peptide-MHC interface structure: a molecular dynamics study. *J Phys Chem B* 2011; **115**:8317–27.

59. Beddoe T, Chen Z, Clements CS *et al.* Antigen ligation triggers a conformational change within the constant domain of the alphabeta T cell receptor. *Immunity* 2009;**30**:777–88.

60. Stavrakoudis A. Cis-trans isomerization of the Epstein-Barr virus determinant peptide EENLLDFVRF after the DM1 TCR recognition of the HLA-B*4405/peptide complex. *FEBS Lett* 2011;**585**:485–91.

61. Painter CA, Cruz A, Lopez GE, *et al.* Model for the peptide-free conformation of class II MHC proteins. *PLoS ONE* 2008;**3**:e2403.

62. Rupp B, Gunther S, Makhmoor T, *et al.* Characterization of structural features controlling the receptiveness of empty class II MHC molecules. *PLoS ONE* 2011;**6**:e18662.

63. Yaneva R, Springer S, Zacharias M. Flexibility of the MHC class II peptide binding cleft in the bound, partially filled, and empty states: a molecular dynamics simulation study. *Biopolymers* 2009;**91**:14–27.

64. Wan S, Flower DR, Coveney PV. Toward an atomistic understanding of the immune synapse: Large-scale molecular dynamics simulation of a membrane-embedded TCR-pMHC-CD4 complex. *Mol Immunol* 2008;**45**:1221–30.

65. Bello M, Correa-Basurto J. Molecular dynamics simulations to provide insights into epitopes coupled to the soluble and membrane-bound MHC-II complexes. *PLoS ONE* 2013;**8**:e72575.

66. Cuendet MA, Michielin O. Protein-protein interaction investigated by steered molecular dynamics: the TCR-pMHC complex. *Biophys J* 2008;**95**:3575–90.

67. Koehl P, Levitt M. A brighter future for protein structure prediction. *Nat Struct Biol* 1999;**6**:108–11.

68. Cardenas C, Bidon-Chanal A, Conejeros P, *et al.* Molecular modeling of class I and II alleles of the major histocompatibility complex in Salmo salar. *J Comput Aided Mol Des* 2010;**24**: 1035–51.

69. De Rosa MC, Giardina B, Bianchi C, *et al.* Modeling the ternary complex TCR-Vbeta/CollagenII(261-273)/HLA-DR4 associated with rheumatoid arthritis. *PLoS ONE* 2010;**5**:e11550.

70. Wright DW, Hall BA, Kenway OA, *et al.* Computing clinically relevant binding free energies of HIV-1 protease inhibitors. *J Chem Theory Comput* 2014;**10**:1228–41.

71. Knapp B, Frantal S, Cibena M, *et al.* Is an intuitive convergence definition of molecular dynamics simulations solely based on the root mean square deviation possible? *J Comput Biol* 2011; **18**:997–1005.

72. Dustin ML, Groves JT. Receptor signaling clusters in the immune synapse. *Annu Rev Biophys* 2012;**41**:543–56.