How T cell receptors interact with peptide-MHCs: A multiple steered molecular dynamics study

ARTICLE *in* PROTEINS STRUCTURE FUNCTION AND BIOINFORMATICS · NOVEMBER 2011

Impact Factor: 2.63 · DOI: 10.1002/prot.23104 · Source: PubMed

3 AUTHORS:

Michel A Cuendet
Weill Cornell Medical College
23 PUBLICATIONS **609** CITATIONS
SEE PROFILE

Vincent Zoete
Swiss Institute of Bioinformatics
94 PUBLICATIONS **2,633** CITATIONS
SEE PROFILE

Olivier Michielin
University of Lausanne
153 PUBLICATIONS **3,649** CITATIONS
SEE PROFILE

PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

How T cell receptors interact with peptide-MHCs: A multiple steered molecular dynamics study

Michel A. Cuendet,[1,2] Vincent Zoete,[1] and Olivier Michielin[1,3,4]*

[1] Swiss Institute of Bioinformatics, Génopode, 1015 Lausanne, Switzerland
[2] Department of Chemistry, New York University, New York, New York 10003
[3] Ludwig Institute for Cancer Research, Lausanne Branch, 1066 Epalinges, Switzerland
[4] Multidisciplinary Oncology Center, Centre Hospitalier Universitaire Vaudois, Rue du Bugnon 46, Switzerland

ABSTRACT

The T-cell receptor (TCR) interaction with antigenic peptides (p) presented by the major histocompatibility complex (MHC) molecule is a key determinant of immune response. In addition, TCR-pMHC interactions offer examples of features more generally pertaining to protein-protein recognition: subtle specificity and cross-reactivity. Despite their importance, molecular details determining the TCR-pMHC binding remain unsolved. However, molecular simulation provides the opportunity to investigate some of these aspects. In this study, we perform extensive equilibrium and steered molecular dynamics simulations to study the unbinding of three TCR-pMHC complexes. As a function of the dissociation reaction coordinate, we are able to obtain converged H-bond counts and energy decompositions at different levels of detail, ranging from the full proteins, to separate residues and water molecules, down to single atoms at the interface. Many observed features do not support a previously proposed two-step model for TCR recognition. Our results also provide keys to interpret experimental point-mutation results. We highlight the role of water both in terms of interface resolvation and of water molecules trapped in the bound complex. Importantly, we illustrate how two TCRs with similar reactivity and structures can have essentially different binding strategies.

Proteins 2011; 79:3007–3024.

Key words: protein–protein interaction; hydrogen bonds; electrostatic interactions; solvation; TCR; Tax; HLA-A2; A6, B7; Gromacs.

INTRODUCTION

The TCR-pMHC system

The triggering and propagation of a cellular immune response depends on the recognition of an antigenic peptide (p) presented by the major histocompatibility complex (MHC). pMHC complexes are recognized by CD4+ and CD8+ T lymphocytes expressing a repertoire of hypervariable αβ T cell receptors (TCR) negatively selected in the thymus against self antigens. TCRs can be extremely selective: mutations of a single amino acid on the peptide were observed to completely disrupt recognition.[1] On the other hand, TCRs display some degree of cross-reactivity, some TCRs being able to recognize several different pMHC complexes. While the key role of the TCR-pMHC interaction has long been established, the nature of the ensuing molecular events that lead to T cell activation remains largely unknown.[2]

TCR-pMHC recognition has been the focus of intensive research in the last decade. A number of crystal structures of different TCR and pMHC complexes have led to the identification of some general trends in the binding modes.[3] In particular, a diagonal orientation of the TCR seems to be imposed by conserved residues on the MHC α_1 and α_2 helices, although it has been suggested that this orientation might also be due to larger scale protein arrangements.[4] TCR recognition is mediated by its complementary determining regions (CDR), a set of three loops on each of the α and β chains, which constitute the interface with the MHC. The germline encoded CDR1 and CDR2 are shorter and less diverse than the CDR3 loop. CDR1 and CDR2 have been found to form contacts mostly with the MHC, whereas CDR3 is usually interacting primarily with the peptide.[5] CDRs are flexible and can undergo large conformational rearrangements upon pMHC binding,[6,7] especially the longer CDR3 loops. These rearrangements do not propagate away from the antigen binding site[8] and the overall structure of the TCR remains

Additional Supporting Information may be found in the online version of this article.
Grant sponsor: Swiss National Center of Competence in Research; Grant numbers: 3200B0-103173; Grant sponsor: OncoSuisse; Grant numbers: OCS 01381-08-2003, 02555-02-2010.
*Correspondence to: Olivier Michielin, Swiss Institute of Bioinformatics, Génopode, 1015 Lausanne, Switzerland. E-mail: olivier.michielin@unil.ch.
Received 12 November 2010; Revised 18 May 2011; Accepted 20 May 2011
Published online 27 June 2011 in Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/prot.23104

largely unchanged upon binding. CDR conformational plasticity plays a role in the selectivity and cross-reactivity of the TCR.[9]

Upon TCR binding, the MHC retains an invariant overall structure, except for some local rearrangements.[10] The peptide however is able to adapt its conformation upon binding[11] and the shape of the peptide in the unbound state was shown not to correlate with T cell activation.[12] Despite several very detailed studies of particular TCR-pMHC structures, as well as reviews of all known complexes,[3,5,13,14] no universal and well localized structural determinant for T cell activation has been identified so far. Instead, the variability in the pitch, twist, and roll of the TCR and the even greater variability in CDR conformations reflect the fact that for each active TCR an individual solution to pMHC recognition is evolved in the cell by random variation and selection during the TCR maturation process.[15]

However, overall characteristics of the TCR-pMHC complex, such as the binding affinity and kinetics, do correlate with T-cell activation. Reactive TCR binding affinities tend to be in the low micromolar range with relatively slow on-rates and fast off-rates. The question whether T cell activation depends rather on the affinity or the kinetics remains a matter of ongoing debate.[16,17] According to a recent review,[18] there appears to be no general thermodynamic signature for T cell activation, but rather a broad range of entropy/enthalpy compensation patterns.[19,20]

This variety of thermodynamic behaviors in TCR-pMHC complexes poses interesting challenges to standard models for protein-protein association. In the literature on protein-protein binding, two separate models emerged.[21] In the 'induced fit' model,[22] the change to the bound conformation is induced by the binding partner. In the 'conformer selection' model,[23] the bound conformation is selected from a number of conformations pre-existing in the unbound state. Induced fit and conformer selection are in fact not mutually exclusive. Rather, they might represent extremes of a continuum of hybrid mechanisms,[24] or be consecutive steps of more complex binding scenarios, as captured by the flexible protein recognition model (FPRM).[25] The FPRM seems to successfully apply to a wide range of protein-protein interactions.[26]

In the case of TCR-pMHC, binding occurs in two stages with different kinetics. First, the TCR encounters the pMHC with a fast and concentration-dependent rate. In a second slower and concentration-independent step, interface shape complementarity is achieved.[27] Wu *et al.* put forward the two-step model,[28] consistent with these findings.[5] In this model, the MHC α_1 and α_2 helices are first recognized by the CDR1 and 2. Then the highly specific and flexible CDR3 loops sample the peptide surface, and adjust their conformation to it. This two-step model can thus be seen as a form of induced fit.

The model described above has however been challenged by the low proportion of interaction energy mediated by the CDR1 and CDR2 in some complexes.[4,29] In addition, a recent survey of known TCR-pMHC complexes shows that interactions between CDR1/2 and the peptide, or between CDR3 and MHC are common. Moreover, analysis of solvent-accessible surface variations showed that most prominent residues of both the peptide and the MHC are already in contact with the TCR in the transition state, which is incompatible with the two-step model.[30] The same analysis showed that both CDR3 loops are involved in the transition state. In addition, conformer selection has recently been shown to be involved in the binding of the particular TCR and MHC molecules studied here (see below).[10]

Focus on HLA-A2/Tax/A6

Among the most studied pMHCs is the HLA-A2 MHC molecule presenting the viral Tax nonapeptide (LLFGYPVYV). X-ray structures are available for this pMHC in complex with two TCRs, A6 and B7 (see Figure 1), as well as for the V7R, Y8A, and P6A peptide mutants with the A6 TCR. In addition, extensive kinetic and thermodynamic measurements have been reported for these systems. Despite the very different sequences of their CDRs, both A6 and B7 have similar binding affinities and elicit similar immune responses (although the specific functional response differs[32]). In contrast to this, a single mutation of the peptide, such as proline 6 to alanine (P6A) significantly decreases the TCR affinity and completely abrogates the T cell cytotoxic response.[1] In this study, we focus on three TCR-pMHC complexes: Tax-A6, Tax-B7, and P6A-A6. Because all complexes include the same HLA-A2 MHC molecule, we use this short-hand notation throughout the remaining of the text.

The A6 and B7 TCRs bind similarly and cover largely overlapping regions of the pMHC.[20] Figure 1 shows how the CDRs of each TCR are disposed on the surface of the pMHC in the crystal structures. Both TCRs share the same V_α chain, which interacts with the N-terminal part of the peptide. The overall binding modes of the CDRα loops is well conserved, whereas the CDRβ loops take different conformations. Comprehensive mutational scanning of Tax showed that A6 and B7 have different functional responses upon substitutions in the peptide,[32] suggesting that the binding mechanism is very different. In addition, the dissociation rates of A6 and B7 are affected in opposite ways by increased ionic strength,[20] which highlights a different usage of electrostatics.

More importantly, extensive thermodynamic measurements by Baker and coworkers[9,20] showed that A6 and B7 follow two very different entropy/enthalpy compensation patterns, see Table I. In line with the common view of protein–protein interactions, Tax-B7 binding involves

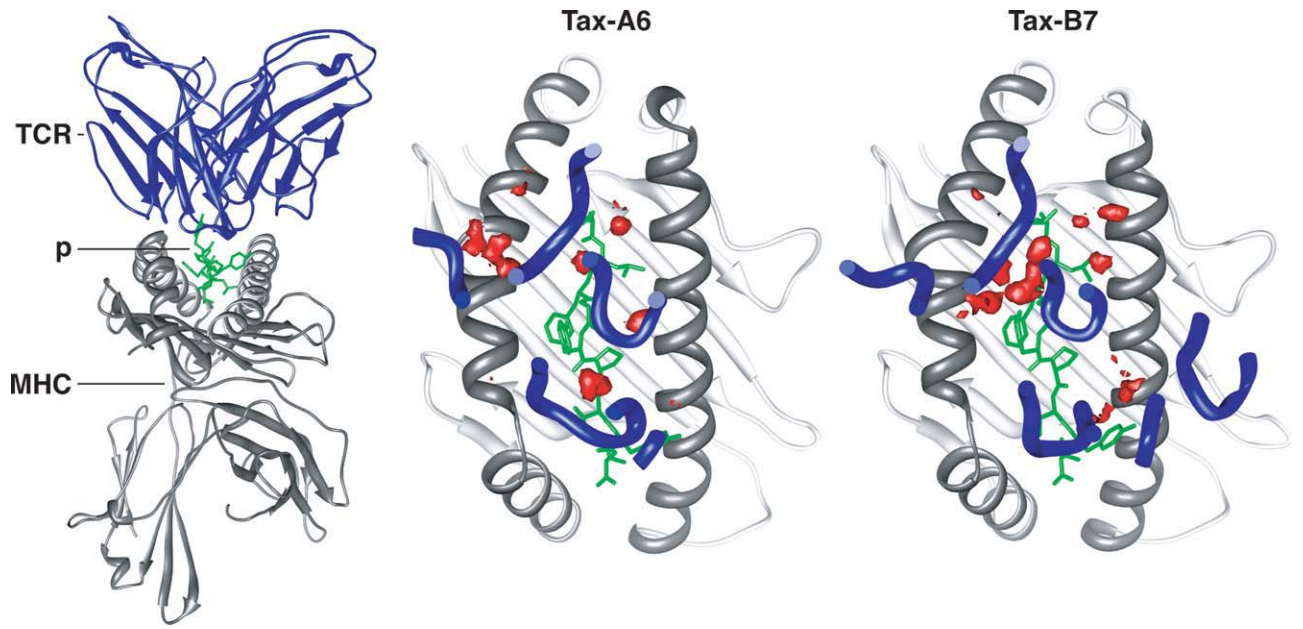

Figure 1

Left panel: the Tax-A6 system. The MHC is depicted in grey ribbons; the peptide is in green; the TCR is in blue ribbons. Middle and right panels: top view of the TCR-pMHC interface of the Tax-A6 and Tax-B7 systems. The N-terminus of the peptide points to the top of the figure. Parts of the CDR loops making contacts with the pMHC are shown as blue tubes, the rest of the TCR was chopped off. CDRs appear as follows: on the upper row, from left to right are CDR2α, CDR1α, and CDR3α. On the lower row are CDR3β, CDR1β, CDR2β. Only a small portion of CDR1β, and none of CDR2β is shown for Tax-A6, because these loops make very few contacts with the pMHC. The density of trapped water molecules (see text for definition) is indicated as red isosurfaces at equal density level for both systems. On average 10.3 water molecules are simultaneously trapped in Tax-A6 and 14.5 in Tax-B7. Rendered with UCSF Chimera.[31]

an entropic cost offset by a large enthalpic gain. On the other hand Tax-A6 binding benefits from a large favorable entropic contribution, combined with a more modest enthalpic gain. The Tax-A6 favorable binding entropy can be attributed to burial of hydrophobic surface area without compensating reductions in TCR dynamics. The question is then why does Tax-B7 behave differently. The difference arises mainly between the bound state and the transition state. In particular, the Tax-B7 complex shows a very large enthalpic activation barrier from the bound state to the transition state, which indicates the need to break short-range interactions. This is consistent with the existence of a significant number of bound water molecules mediating H-bonds, whose incorporation upon binding results in an entropic penalty.[20]

The P6A mutation in the peptide induces only minute changes in the structure of the TCR-pMHC interface[1] with respect to the wild type Tax peptide. The only significant structural difference is a packing defect consisting of an enlarged cavity partially filled with a bound water molecule[1] below the CDR3α of the TCR. However, this single mutation completely abrogates T cell cytotoxicity by incurring a loss of 9.8 kJ/mol in binding free energy (which can be recovered stepwise using side chains of increasing length[34] at position 6). Simulations have shown[36] that this binding free energy difference results partly from the worse shape complementarity with the TCR and the less favorable internal energy in the pMHC, but also from the greater solvation energy of the unbound P6A-MHC complex.

Important residues

In a quest for the molecular determinants of the TCR-pMHC recognition, the role of many residues has been investigated individually. Mutations have been attempted either on the peptide itself, or on the surface of the MHC. Their impact has been measured either at the functional level using cytotoxicity or cytokine release assays, or directly at the molecular level using TCR-pMHC binding affinity and kinetics data. For example,

Table I

Summary of TCR-pMHC Complexes

	X-ray		ΔG^0	$-T\Delta S^0$	
Complex	PDB	Reference	kJ/mol	kJ/mol	Reference
Tax-A6	1ao7	33	−32.3	−14.5	20
P6A-A6	1qrm	1	−22.5	n.d.	34
Tax-B7	1bd2	35	−33.5	+27.4	20

ΔG^0 is the standard binding free energy. ΔS^0 is the standard binding entropy and T is the temperature, here 25°C.

an energetic hot spot for recognition was identified by mutating 15 TCR contact sites on the MHC,[37,38] whose relevance was questioned by a later study.[15]

The decisive role of the peptide sequence in TCR-pMHC binding was understood very early. In the Tax-A6 and Tax-B7 complexes, a comprehensive investigation[32] characterized the effect of all possible single mutations in the Tax peptide (LLFGYPVYV) (no mutation was able to improve A6 or B7 binding). A noticeable feature of the TCR contact surface is a deep pocket that accommodates the prominent Tyr5 side chain of the peptide. In B7, this pocket is highly specific for aromatic residues. In the A6 structure, this pocket is larger, allowing many different residues to be accommodated, including a bulky non-natural side chain.[7] The role of Tax Tyr5 on A6 or B7 TCR binding affinity and kinetics has been further investigated using a variety of mutations at this site, including Ala, Phe, Trp.[30,39]

The case of the Y5F mutation is particularly interesting because it is able to single out the contribution of the phenol OH group (Tyr5-OH) at the tip of the Tyr5 side chain. Suppression of this group was shown to have a greater impact on the Tax-B7 complex (bringing the K_D from 1.35 to 6.49 μM, $\Delta\Delta G = 3.9$ kJ/mol) than on the Tax-A6 complex (K_D from 2.23 to 2.79 μM, $\Delta\Delta G = 0.5$ kJ/mol).[20] By varying denaturant concentrations, the important role of the Tyr5-OH bond in the transition state of Tax-A6 was established, with a ϕ value close to one (0.91 ± 0.63). This is less the case in the Tax-B7 complex ($\phi = 0.25 \pm 0.02$).[30] Side chains of TCR residues interacting with Y5, S31α of A6 and Y104β of B7, were also shown to make stabilizing contacts in the transition state.

Another residue of interest is H151 of the MHC. An advanced calorimetric study on the Tax-A6 complex using buffers with various ionization enthalpies found that H151 gets deprotonated upon binding.[40] 0.28 protons are released to the buffer at pH 7.4. It seems counterintuitive that binding can affect H151, because it does not make any direct contact with the TCR in the crystal structure.

Overview of this work

General models of protein-protein association, such as the conformer selection or induced fit mechanisms, remain controversial. Concurrently, details of the TCR-pMHC recognition mechanism remain anecdotal, and no general model accounting for both specificity and cross-reactivity has been put forward. This calls for a more detailed understanding of the interactions in particular TCR-pMHC complexes, such as Tax-A6 and Tax-B7. Towards this goal, molecular modeling can be a useful complement to experiment. It has already proven its usefulness towards clinical applications, as a tool for the engineering of TCRs to be used in cancer immunotherapy.[41] Other simulation studies have sought to understand the contribution of particular side chains to TCR-pMHC binding. These include a thermodynamic integration approach to the P6A peptide mutation[36] and a computational alanine scanning and per-residue binding free energy decomposition study.[42]

There are further questions on which molecular dynamics (MD) simulation can give interesting insights:

- Structural biologists often refer to (native) H-bonds and contacts observed in a crystal structure as a proxy to characterize protein–protein interfaces[3,5] and attempt to rationalize experimental results. However at room temperature, these contacts are inherently transient and side chain conformations fluctuate a lot. Thus how meaningful is the concept of native H-bond or contact?
- In this fluctuating environment, can we draw maps of interactions H-bonds and contacts between particular residues or atoms along the dissociation pathway?
- Water molecules play a key role in the TCR-pMHC interaction. In particular, water-mediated H-bonds seem to be a key factor to explain the different thermodynamic behaviors of the Tax-A6 and the Tax-B7 complexes. In addition, the effect of the peptide P6A mutation cannot be understood without considering trapped water molecules. Can we draw a global picture of the interplay of solvent in TCR-pMHC binding?
- The nature of the transition state remains largely unknown. Can simulation give insights on the role of the different CDRs or individual residues in the transition state, thus helping to assess binding and recognition models, such as the two-step model?

To address these questions, we have conducted MD simulations of the Tax-A6, P6A-A6, and Tax-B7 complexes in explicit water. In this work, we go beyond simulations of the bound complex at equilibrium used in most previous studies.[36,41,42] We perform steered MD (SMD)[43,44] to gradually force the dissociation of the TCR-pMHC complex. We have achieved extensive sampling by using a multiple trajectory approach, which allowed us to report converged numbers for energetic components or atomic contact statistics at any point along the dissociation pathway. The total simulation time amounts to $\sim 1.35 \mu s$ per complex, collected from around 150 independent trajectories each.

In a previous study, we used the Tax-A6 and P6A-A6 systems to explore the applicability of the Jarzynski identity to calculate binding free energies in large protein-protein complexes.[45] Here, we use the same pulling scheme, which enforces straight unbinding with overall conserved protein structure in each part (see Methods). We keep in mind that the results presented here are subject to the key assumption that the sampled SMD unbinding pathways are realistic.

In the following, the Methods section explains the details of the simulation protocol and of H-bond and contacts accounting. In the Results and Discussion section, we first present interaction and resolvation H-bonds counts along the reaction coordinate. We give special attention to trapped and H-bond mediating water molecules. We then decompose the TCR-pMHC interaction in terms of H-bonds and hydrophobic contacts for individual residues, with a focus on the specific H-bonds involving the phenol OH group of the peptide Tyr 5 residue. We then look at the repartition of the binding energy among CDRs. Finally, we decompose the energy variations upon unbinding into contributions of TCR, pMHC or solvent energies.

METHODS

Simulation setup

Simulations were setup similarly for all three TCR-pMHC complexes. Initial atomic coordinates for the Tax-A6, P6A-A6, and Tax-B7 complexes were taken from the Protein Data Bank[46] (see Table I). Only the variable domain of the TCR was kept in the system. Crystal water molecules were kept in place. Histidine protonation states were decided and hydrogen atoms were constructed using the Gromacs 3.3 simulation package.[47] The centers of mass of the TCR and pMHC were aligned along the z-axis and the proteins were solvated in a rectangular periodic box of water with enough space to allow for a 2 nm dissociation course along the z-axis (see below). Counter ions were added in sufficient number to neutralize the overall charge of the system. This setup yielded system sizes of around 7.5 × 8.0 × 14.5 nm including a total of around 84,000 atoms.

Classical MD simulations were performed using the Gromos 43a1 force field[48,49] within Gromacs 3.3. All bonds involving hydrogen atoms were constrained to their equilibrium distance using SETTLE,[50] a version of the SHAKE[51] algorithm. The rigid SPC[52] water model was used. The Lennard-Jones interactions were evaluated with a smooth switch function bringing the energy and forces to zero at 1.4 nm separation. Electrostatic interactions were calculated using the particle-mesh Ewald (PME) method[53] during the dynamics. However, PME does not allow direct decomposition of the electrostatic energy among different parts of the system. For such decompositions, electrostatic energy components were re-evaluated on saved trajectories using a cut-off radius of 1.4 nm and assuming a uniform dielectric environment beyond.

After energy minimization, random velocities were assigned according to a Maxwell-Boltzmann distribution at 50 K. The system was heated up to 300 K during 300 ps of MD with positional restraints on all backbone heavy atoms and C_β atoms. This was followed by 700 ps of equilibration at constant temperature and pressure using two separate Nosé-Hoover[54] thermostats with time constant 0.1 ps for solvent and protein, as well as a Parrinello-Rahman[55,56] barostat with time constant 0.5 ps. This combination of thermostat and barostat preserves the correct NPT ensemble and was applied in all subsequent simulations.

For each TCR-pMHC complex, ~150 trajectories were run in parallel, each starting from the same structure but different initial random velocities. Using multiple trajectories was shown to provide better sampling than using a single trajectory of the same total duration.[57] In addition, the multiple trajectory approach is more appropriate to sample the nonequilibrium steered dissociation process. At the end of the equilibration phase described above, each trajectory was subjected to SMD to induce dissociation of the TCR and pMHC. This dissociation process was realized using the individual pulling SMD method previously described by the same authors.[45] In essence, individual pulling potentials are applied to the *z*-coordinate of each heavy atoms of the backbone, such that the internal structure and orientation of each part of the complex (TCR or pMHC) is preserved. SMD is achieved by moving the reference position of all potentials in concert. This method is appropriate for the present system, because both TCR and MHC are rather rigid molecules, which have similar structures in complex and in solution[3,5] (except the peptide and the CDR loops of the TCR, which can undergo conformational changes upon binding and were hence not subjected to pulling).

The direction for the pulling is chosen to be along the vector joining the centers of mass of the TCR and the pMHC, which is aligned along the z-axis of the simulation box. There are two reasons for this choice. First, in the present system, the center of mass axis is perpendicular to the TCR-pMHC molecular interface (see Fig. 1). There is evidence that TCR-pMHC interactions happens between closely packed multimers *in vivo*,[58] both attached to two parallel cellular membranes, which would prevent large sideways or rotatory movements. Second, pulling along the center of mass axis ensures that no force momentum develops in the proteins. Note that we apply the steering potentials along the dissociation axis only, motions in other directions are unrestrained.

After applying the pulling potentials to the bound complex and before starting to pull (i.e., move the reference positions of the pulling potentials), the system is equilibrated for 300 ps at the equilibrium center of mass distance $\xi = 0$. In the SMD phase itself, the center of mass distance is increased by 2.0 nm over 4 ns of simulation time by moving the reference positions of the pulling potentials in opposite directions for TCR and pMHC. At the end of the SMD simulation, the TCR finds itself separated from the pMHC by ~2.0 nm of solvent.

In addition to the SMD trajectories, additional sampling of the bound state was provided by performing

2 ns of equilibrium simulation for each trajectory starting at the end of the equilibration stage described above. Equilibrium sampling of the unbound state was also performed. For this, the steering potential was removed at the end of the SMD trajectories and the system was equilibrated for 1 ns in the unbound state. This was followed by a production run of another 1 ns on each of the independent trajectories. For the three TCR-pMHC complexes covered in this work, the combined simulation time amounts to more than 4 μs.

H-bonds and contacts

Garboczi *et al.*[33] identified 21 H-bonds between TCR and pMHC in the Tax-A6 crystal structure, which we call native H-bonds. For our discussion of native H-bonds, we retain the same set of native H-bonds for the P6A-A6 complex, although only 16 of these can be observed in the P6A-A6 X-ray structure. For the Tax-B7 complex, we define as native H-bonds the 10 H-bonds observed in the crystal structure by Ding *et al.*[35] using the same criterion as Garboczi *et al.*

The criterion of Garboczi *et al.* and Ding *et al.* relies solely on a cutoff on the distance between the H-bond acceptor and donor atoms. A closer look at the crystal structures reveals that this criterion counts H-bonds which cannot possibly exist simultaneously. To account for H-bonds during our dynamics, we choose a more stringent criterion: (i) donor-acceptor heavy-atom distance smaller than 0.35 nm, (ii) angle acceptor-donor-hydrogen smaller than 30°. Using this definition of a H-bond, we find 13 H-bonds in the Tax-A6 crystal structure. Using the same criterion, we find only 6 H-bonds in the P6A-A6 crystal structure, and 4 in the Tax-B7 crystal structure, as shown in Table II. During the dynamics at 300 K, however, all native H-bonds are also observed using our more stringent criterion.

H-bonds were monitored along equilibrium or SMD trajectories using the Gromacs 3.3 tool g_hbond.[47] For each trajectory the output is a binary table expressing the existence of all H-bond between each particular donor-acceptor pairs for all frames of the trajectory. The Gromacs tool g_hbond considers Nitrogen to be always an acceptor, which is wrong in most cases except in unprotonated histidines. We carefully discard occurrences corresponding to this situation, although they represent a low proportion of the total number of H-bonds.

For analysis, we pool together the occurrence tables coming from all trajectories, and sum up the occurrences of pairs according to the desired level of detail: entire molecule, residues, or individual atoms. First, we monitor the variation of the total number of H-bonds between TCR and pMHC, as well as between each of these parts and the solvent (see Fig. 2). We can retain a finer resolution and look at all the H-bonds that a given residue is able to make across TCR and pMHC during the dissociation process (see Fig. 4). Alternatively, we can focus on a given atom, and monitor interactions to all H-bonding partners during the simulation (see Fig. 5).

Table II
H-bond Counts

Source	Description	Tax-A6	P6A-A6	Tax-B7
X-ray	No angle cutoff	21	16	10
	With angle cutoff	13	6	4
MD	Simultaneous	10.9	9.3	8.4
	Total number	300	344	374
	Significant number	89	99	98
	Ratio native	52%	45%	18%
SMD	Δ water-pMHC	10	11	13
	Δ water-TCR	6	3	14

Number of H-bonds between TCR and pMHC in the crystal structures (with or without the angle cutoff) or in the equilibrium MD of the bound complex. Here, "simultaneous" is the average number of H-bonds simultaneously observed in a given conformation of an MD trajectory. The total number of different H-bonds occurring at any point in any equilibrium trajectory is reported under "total number." Those occurring in at least 1% of the frames are under "significant number." The fraction of all occurrences corresponding to native H-bonds observed in the crystal structure (see text) is reported under "ratio native." The "Δ water-pMHC" and "Δ water-TCR" indicates the gain of H-bonds with water molecules observed between the beginning and the end of the SMD dissociation trajectories.

Non-polar contacts are determined as any pair of atoms not consisting of an H-bond donor and an H-bond acceptor sitting within a cutoff radius of 0.535 nm of each other. This distance corresponds to half the depth of the CH_3 - CH_3 Lennard-Jones interaction energy function in the Gromos force field. Non-polar contacts are accounted exactly as described above for H-bonds (see Fig. 4).

H-bond mediating water molecules are defined as those simultaneously engaged in one or more H-bonds with both the TCR and the pMHC. The same criterion was used as for single H-bonds. Counts of water-mediated H-bonds were performed with a modified version of the program g_hbond from the Gromacs 4 software package.[59] Trapped water molecules are defined as sitting simultaneously within a cutoff distance of 0.475 nm of both TCR and pMHC heavy atoms, and not within the same distance of a free water. A free water is one that is not within 0.475 nm of both TCR and pMHC heavy atoms. This cutoff distance is meant to represent a non-polar interaction of a water molecule with the protein, and was determined as half the depth of the OW - CH_3 Lennard-Jones potential energy well. To perform the counts of trapped waters on MD trajectories, we used the program g_select that comes with the Gromacs 4.5 release. The criteria defined above are somewhat arbitrary and can occasionally include some water molecules that are not properly mediating H-bonds or trapped at the interface. However the statistical average of these counts is meaningful and gives insights on the TCR-pMHC binding mechanisms, regardless of the particular cutoff values chosen.

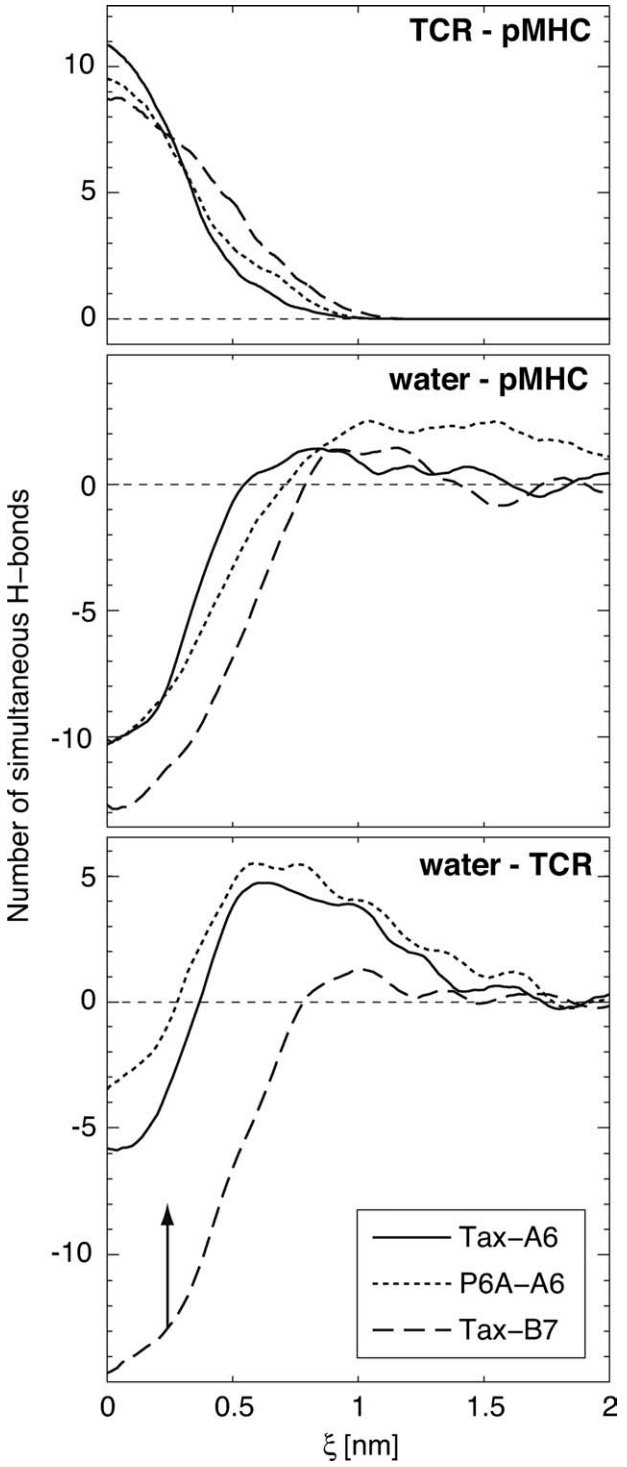

Figure 2

Number of H-bonds along the dissociation pathway between TCR and pMHC and between each part and water molecules. In all panels, the zero of the scale is defined as the unbound state ($1.7 < \xi < 2.0$) of the Tax-A6 complex. In the lower panel the arrow indicates that the B7 curve has been shifted by 25.9 H-bonds to make its unbound state match that of A6 to allow comparison.

RESULTS AND DISCUSSION

TCR-pMHC interaction H-bonds

Important actors in the TCR-pMHC binding mechanism are the H-bonds between the two molecules. We define as native H-bonds those observed in the crystal structure (see the Methods section for definitions). We retain 21 native H-bonds for Tax-A6 and P6A-A6 and 10 for Tax-B7. Native H-bonds observed in crystal structures are often used as proxy to explain biomolecular mechanisms in structural biology. However, we show here that native H-bonds are only part of the picture.

In our equilibrium and SMD simulations, we observe that a large number of non-native H-bonds also transiently form, following fluctuations of the side chain conformations. For example, a total of 300 different H-bonds (combinations of different acceptors and donors) are observed at least once during the equilibrium simulations of the bound Tax-A6 complex, see Table II. Among these, 89 different H-bonds are significant (defined as occurring on average in more than 1% of the sampled conformations). Native H-bonds represent only a fraction (52%) of all H-bonds occurring in the Tax-A6 complex during the dynamics at 300 K.

Table II also shows that the P6A-A6 and Tax-B7 complexes use a slightly more diverse set of H-bonds than the Tax-A6 complex. The slightly lower proportion (45%) of native H-bonds observed in the P6A-A6 complex is expected, given that native H-bonds were defined in Tax-A6. The very low proportion (18%) of native H-bonds for the Tax-B7 complex partly stems from the fact that there are only 10 native H-bonds. However, the number of simultaneous H-bonds observed in the Tax-B7 simulation is almost as high as in Tax-A6. This implies that the Tax-B7 crystal structure reveals only a fraction of the H-bonds involved at 300 K. The more diverse ensemble of H-bonds found in Tax-B7 also suggests that the Tax-B7 interface is less restrictive for residue conformations than Tax-A6. The important point here is that the number of H-bonds occurring in the dynamics is not related to the number of H-bonds observed in the crystal structures. Native H-bonds do not give a faithful image of the total number and variability of H-bonds participating in the TCR-pMHC interaction at room temperature.

Using our SMD simulations, we are able to monitor the variation in the number of simultaneous H-bonds during the dissociation process, as a function of the center-of-mass distance between the TCR and the pMHC, denoted by the reaction coordinate ξ. This is shown in the upper panel of Figure 2. On each curve, the expected error before smoothing is lower than 0.35 H-bond. After smoothing over intervals of 0.1 nm, the expected error falls below 0.03 H-bonds.

We see that beyond $\xi = 1.0$ nm, H-bonds have statistically disappeared in all complexes (although some can

still scarcely be formed, see Fig. 4). In the bound state, P6A-A6 has on average one interaction H-bond less than the wild type complex. On the other hand, P6A-A6 retains one more H-bond in the second phase of the dissociation process ($0.6 < \xi < 0.8$ nm). The fact that the P6A complex forms on average more H-bonds than the Tax complex at larger distances where the side chains are less tightly packed is consistent with the greater diversity of donors and acceptors reported in Table II for the P6A complex.

Turning to the case of the B7 TCR, the upper panel of Fig. 2 shows that Tax-B7 forms fewer simultaneous H-bonds in the bound state than the A6 complexes. However, at $\xi = 0.5$ nm Tax-B7 is still able to form 5 simultaneous H-bonds, compared to only 2 in Tax-A6. This is a first indication that the dissociation mechanism of the B7 TCR is different from the A6 TCR. This is further illustrated by monitoring the ratio of native H-bonds upon dissociation, see Supporting Information Figure S1. We note that is difficult to identify which particular residues are responsible for the differences in H-bonds mentioned above, because individual H-bonds are transiently formed between a large collection of occasional donors and acceptors, some of which available only in partially unbound states.

Solvation H-bonds

Complementary to the breaking of TCR-pMHC interaction H-bonds is the resolvation of each part of the complex after dissociation. The variation of the solvent H-bonds for TCR or pMHC during the SMD dissociation simulations is plotted in the lower panels of Figure 2. Instead of showing the absolute number of solvent H-bonds, we take as a reference the unbound state ($1.7 < \xi < 2.0$) of the HLA-A2/Tax and A6 molecules. The zero point for the scale of the y-axis in the lower panels of Figure 2 was defined accordingly (around 382 for A6 and 364 for the pMHC). The corresponding value for B7 is around 356, so the B7 curve would fall out of the scale. To nevertheless allow comparison with A6, the B7 curve was shifted up by 25.9 H-bonds, which is indicated by an arrow on Figure 2 (no other curves was shifted). The differences of solvent H-bonds between bound and unbound states in equilibrium simulations are also reported in Table II.

Note that extensive sampling is necessary to converge counts of H-bonds with water. This is due to the fact that this number is strongly fluctuating, because of the transient nature of H-bonds with diffusing water molecules and the flexibility of exposed side chains. On a single trajectory, root mean square fluctuations reach 12.2 H-bonds. A further difficulty is that we are interested in small variations compared to the large total number of simultaneous water-protein H-bonds (e.g., variation of 6 solvent H-bonds over a total of 382 for the A6 TCR) drawn from an even larger number of possible H-bond combinations. By averaging over 150 trajectories, the expected error is around 1.1 H-bond. After smoothing in windows of 0.1 nm, the expected error falls below 0.5 H-bond. Thus, variations and differences between curves on Figure 2 are statistically significant with the number of trajectories available.

We first focus on the effect of the P6A mutation on resolvation H-bonds, as shown on Figure 2. On the pMHC side (middle panel), we see that the P6A mutation allows for one to two additional H-bonds in the unbound state, but does not induce any change in the bound state. This is surprising, because the mutation creates room for an additional interfacial water molecule in place of the Tax proline pyrrole ring, which can make an H-bond with the backbone nitrogen atom of Ala 6 of the P6A peptide. This additional trapped water molecule is resolved in the 1bd2 crystal structure of Tax-P6A,[1] and appears in the counts of trapped water molecules (see Fig. 3, middle panel). The effect of this particular water molecule is probably compensated by indirect variations in the total number of solvent H-bonds arising in other parts of the protein interface.

On the side of the A6 TCR (Fig. 2, lower panel), the counts of solvent H-bonds show a very distinctive transition state around $\xi = 0.5$ nm. In a first phase, the TCR recovers with the solvent the same number of H-bonds it had with the pMHC ($\sim$11). At further distance, the A6 TCR loses 5 solvent H-bonds, which can be assumed to be converted to internal H-bonds as the CDRs rearrange to conformations more favorable to the unbound state. We verify that the P6A mutation has no effect on the solvation of the unbound A6 TCR, as expected. In the bound state, however, we observe that the P6A mutant is able to form 2.5 solvent H-bonds more than the wild type.

We now turn to the resolvation of the Tax-B7 complex, also shown on Figure 2. First focusing on the pMHC side (middle panel), we verify that the pMHC recovers the same amount of solvent H-bonds in the unbound state as for the A6 TCR, as expected. In the bound state however, the B7 TCR covers a slightly different region of the pMHC surface and is able to suppress three solvent H-bonds more than the A6 TCR. On the TCR side (lower panel), the transition state observed with A6 is not present with B7. Instead, B7 recovers over 8 solvent H-bonds more than A6 upon dissociation. The major part of this difference (5 H-bonds) is related to the reorganization of the A6 TCR after the transition state, which does not happen in the B7 TCR. The remaining part is due to a difference of $\sim$2.5 H-bonds formed within the TCR upon dissociation (A6 gains 3.5 internal H-bonds, B7 only one; data not shown). If we admit that forming an internal H-bond implies mobilizing one donor and one acceptor previously exposed to water, this is consistent with the 5 additional solvent H-bonds lost in A6 upon reorganization.

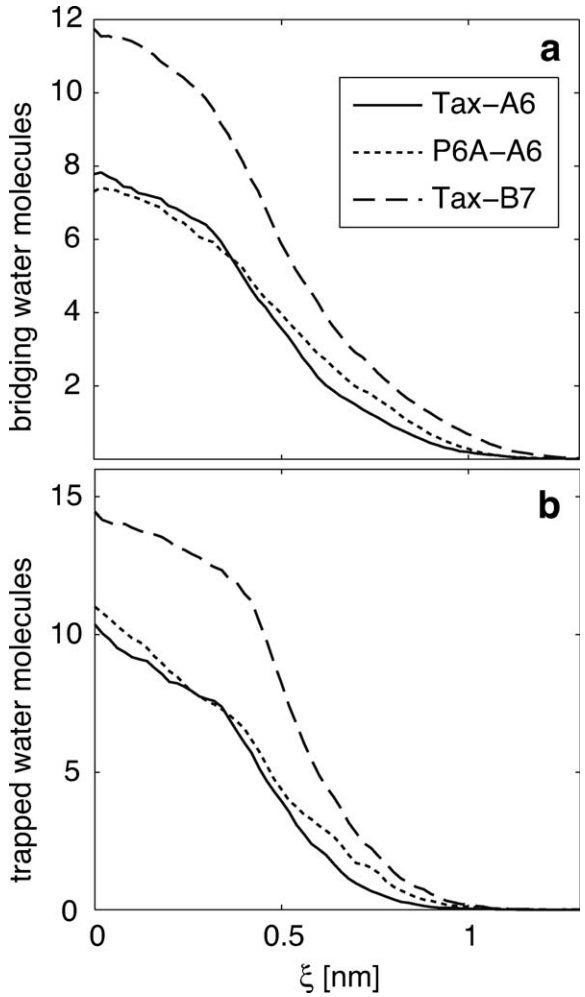

Figure 3

Number of water molecules involved at the interface of the three TCR-pMHC complexes. (**a**) Water molecules engaging simultaneously in at least one H-bond with each of TCR and pMHC. (**b**) Trapped water molecules (see text for definitions).

Overall, the B7 TCR shows a net deficit of H-bonds in the bound TCR-pMHC compared to the A6 TCR (9 vs. 11, see top panel of Fig. 2), together with an excess of solvent H-bonds in the unbound state (3 on the pMHC side, and 3 on the TCR side, not counting the internal TCR reorganization in A6). So we can say that, from the perspective of H-bond counts alone, B7 binding seems less favorable than A6 binding.

Interfacial water

Water molecules can play two distinct roles at the TCR and pMHC interface, by either mediating H-bonds between both partners, or by being trapped in cavities (possibly without H-bonding). H-bond mediating water molecules consolidate the interaction with a favorable enthalpic contribution. On the other hand, water molecules that are trapped at the interface represent an entropic cost related to their restricted mobility. Note that both definitions (see Methods section) are not mutually exclusive and many trapped water molecules will also mediate H-bond bridges. Fig. 3 shows counts of H-bond-mediating and trapped water molecules as a function of ξ the for the three TCR-pMHC complexes.

To make sure that the number of interfacial water molecules found in the simulations were not dependent on the initial number of crystal water molecules resolved at the TCR-pMHC interface, we counted the water molecules trapped in the X-ray structures using the same definition. Table III shows that the number of crystal waters trapped at the interface is uncorrelated to the counts observed on Figure 3. These numbers are themselves independent of the X-ray structure resolution, which is similar for all three complexes.

Counts of trapped and H-bond mediating water molecules in the bound state equilibrium simulations shown in Table III are very consistent with the initial values for $\xi = 0$ on Figure 3. Overall, the water counts varied very little during the course of the equilibrium simulations of the bound complexes, and we can assume convergence within error bars of less than one unit (data not shown).

On Figure 3 we see that both Tax-A6 and P6A-A6 complexes behave very similarly. P6A-A6 has one more trapped water molecule than the wild type complex. This water molecule very plausibly corresponds to the ordered water molecule found By Ding *et al.*[1] in the P6A-A6 crystal structure, occupying the enlarged cavity left vacant by the pyrole ring of Pro 6. This molecule is H-bonded to the backbone amide group liberated by the substitution of Pro for Ala at position 6 of the peptide. We note however that this water molecule is not within H-bonding distance of the TCR in the crystal structure, indicating that it interacts with a second water molecule occupying the same cavity, rather than directly with the TCR. This explains why the additional trapped water molecule does not appear in the counts of H-bond mediating water molecules. As a counter-example, the additional water molecule observed in the P6A-A6 complex between $\xi = 0.5$ and $\xi = 0.9$ nm appears to be both trapped and H-bond mediating.

Table III
Interfacial Water Counts

Source	Description	Tax-A6	P6A-A6	Tax-B7
X-ray	Resolution (Å)	2.6	2.8	2.5
	Trapped H_2O	1	8	5
MD	H-bond mediating H_2O	8.0	7.9	11.3
	Trapped H_2O	10.0	10.9	14.0

The first line indicates the resolution of the X-ray crystal structures (see references in Table I). Other lines report the number of trapped or H-bond mediating water molecules at the TCR-pMHC interface, observed either in the crystal structure or in the equilibrium MD simulations.

The most striking feature of Figure 3 is the large difference between the Tax-B7 complex and the Tax-A6 complex. There is an excess of 4 trapped molecules and 4 H-bond mediating water molecules (they need not be the same) in the bound Tax-B7, with respect to Tax-A6. This seems to imply that B7 retains more solvent H-bonds than A6 at the interface in the bound complex. All other things equal, this in turn would imply that B7 would gain less solvent H-bond than A6 upon unbinding. However, the counts of TCR-solvent H-bonds in Figure 2 show the opposite: B7 in fact gains 9 solvent H-bonds more than A6 upon unbinding. This apparent paradox could be explained if the B7 interface contained more H-bond donors or acceptors than the A6 interface. However, both A6 and B7 bury approximately the same surface area (1042 vs. 968 Å^2, respectively), of which roughly the same amount is from polar residues (407 vs. 397 Å^2, respectively).[20] We also verified that B7 in solution forms slightly fewer simultaneous H-bonds with water than A6 (59.5 vs. 62.7 for residues closer than 0.5 nm from the pMHC in the crystal structure).

Further analysis shows that solvent H-bond counts behave differently at the core of the interface, at the periphery of the interface, or in the rest of the TCR (see Supporting Information Fig. S2). It turns out that the variation at the core of the interface is almost identical for both TCRs. At the interface periphery however, B7 gains 5 solvent H-bonds upon unbinding, whereas no variation is observed for A6. This can be reconciled with the results of Figure 3 if we assume that H-bond mediating water molecules at the surface of the Tax-B7 complex are constrained in a conformation that hinders optimal solvation of each binding partner by preventing H-bonds between nearby donors or acceptors and other water molecules.

What Figures 2 and 3 undoubtedly show, is that the solvent plays radically different roles in the binding mechanisms of A6 and B7. The complex interactions of water molecules at a protein-protein interface are difficult to rationalize from global observables, because it strongly depends on local details of the molecular surfaces.

Theoretical studies have shown that water molecules trapped at biomolecular binding interfaces generally induce both an entropic penalty and a favorable binding enthalpy contribution due to H-bonding.[60] In general, the net result of this enthalpy/entropy compensation depends a lot on the surrounding of the particular water binding site, but contributes on average only very modestly to complex stability.[61] This is consistent with the marginal difference in affinity observed between the Tax-A6 and Tax-B7 complexes. It was however shown that a confined water molecule can account for between 4 and 12.5 kJ/mol of entropic penalty.[62] Thus, 4 additional trapped water molecules could account for a large part of the 42 kJ/mol difference of binding entropy between the Tax-A6 and Tax-B7 complexes (see Table I).

Trapped water molecules can be found at many different locations in particular snapshots of the bound state MD trajectory. However, if the density of those water molecules is measured, distinct maxima appear, indicating the most favored positions. This is represented by red isosurfaces on Figure 1. For the Tax-A6 complex, 9 clearly separated such maxima emerge. The tenth trapped water molecule present on average (see Table III) can be located at various regions of lower density. In the Tax-B7 complex, the distribution of trapped water molecules is entirely different, although the V_α chain of the TCR is the same and the CDRα loops take similar conformations. In particular, a cavity filled with trapped water molecules appears between CDR1α, CDR3α, and the peptide in Tax-B7, whereas this same region hosts only one trapped water molecule in Tax-A6. The little overlap between both distributions of interfacial water molecules is another indication that the binding modes of the A6 and B7 TCRs are very different.

H-bonds and hydrophobic contacts for individual residues

Figure 4 shows occurrence counts of H-bonds and hydrophobic contacts (see Methods section) between TCR and pMHC, as a function of ξ, pooled at the level of individual residues. Note that contributions of the backbone are also included. Only residues that make an interaction in at least 2.5% of the conformations from dissociation MD trajectories are reported. Looking at the upper panels of Figure 4, we can analyze the contributions of the various CDRs to the TCR-pMHC binding. First, we note that for A6, the contribution of CDR1β and CDR2β is very limited, with the modest exception of E30β. For B7, CDR2β interacts much more markedly with the MHC. In all complexes, CDR2α and CDR2β interact almost exclusively with the MHC, but CDR1α, CDR3α, and CDR3β interact equally with the peptide and the pMHC. This is not in line with the general view according to which CDR1 and 2 interact primarily with MHC and CDR3 with the peptide.

Lower panels on Figure 4 show interactions of residues in the MHC and peptide. Among the major players in all three systems are residues, R65, K66 and A69, which were identified as an energetic footprint for TCR binding.[37] Also involved is A158, which, together with A69 has been identified as an essential component of a conserved anchoring cup for the TCR.[13] A more recent study[14] identified positions 65, 69 and 155 as a "restriction triad" essential for TCR binding in most known complexes; these residues show important contributions on Figure 4. Interestingly, some MHC residues participating in a significant number of H-bonds or contacts were shown by alanine scanning experiments not to be essential for T cell activation,[37] such as H151 or Q155. This shows that the penalty associated with the

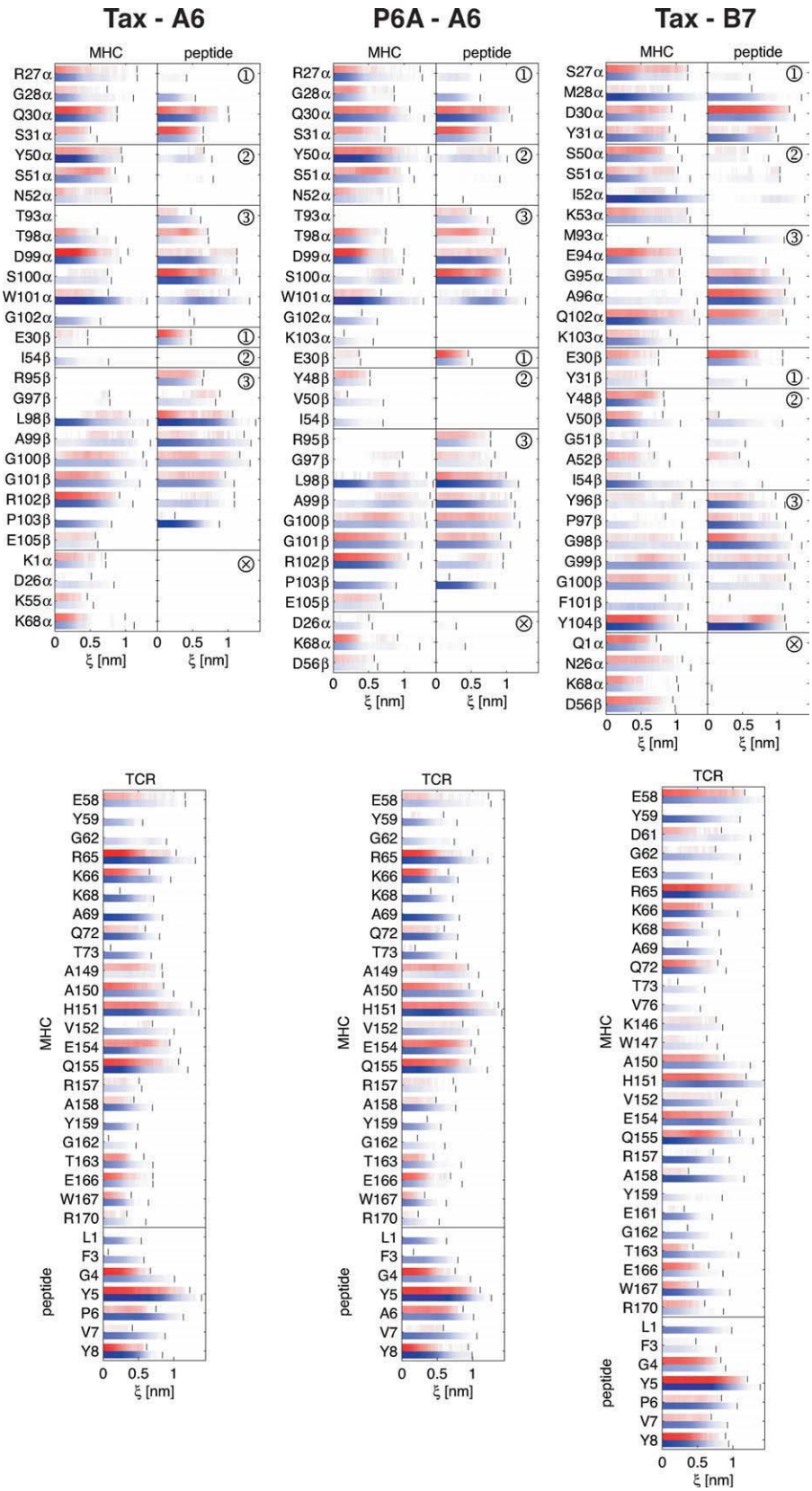

Figure 4

Occurrence of H-bonds (red) and hydrophobic contacts (blue) between TCR and pMHC as a function of the distance between centers of mass. See text for definitions of these two interactions. Bright colors indicate a high occurrence of the interaction between the given residue and the other part of the complex. Colors fading to white indicate lower occurrences. The color scales are arbitrary, but consistent over all nine panels. A vertical line indicates the position of the last interaction recorded for each given residue. The circled numbers in the upper panels indicate the CDR to which the listed residues belong. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

desolvation of particular side chains can offset the enthalpic gain expected from protein-protein contacts. Note the case of H151, which has been put forward as the residue responsible for the pH dependence of the Tax-A6 binding.[40] Here, we see how His151 can indeed influence binding although it does not interact with the TCR in the crystal structure. His151 interacts significantly with the TCR at room temperature and is even one of the last residues interacting with the TCR at large ξ.

Averaged at the residue level, the interaction data presented in Figure 4 does not show a clean-cut transition state between the bound state and the unbound state. Even less can the various states involved in the FPRM model[25] be discerned. If we however assume that the transition state (or the recognition complex of the FPRM) is located between $\xi = 0.5$ nm and $\xi = 1.0$ nm, we can analyze which residues are involved. Some TCR residues such as W101α or Y104β show a regain of interactions with the peptide after the initial dissociation phase. Alternatively, some residues of CDR3β, such as L98β, A99β and G100β in A6, or G98β and G99β in B7 show some transient contacts with the MHC, which are not present in the bound state. For most residues, however, the H-bond and contact pattern in this range of ξ is essentially the same as in the bound state. In particular, CDR1α interacts largely with the peptide in the transition state, and CDR3s have mostly the same proportion of interactions with the peptide as in the bound state. This is not consistent with the two-step model,[28] in which the final contacts of CDR3 with the peptide should form after the main contacts of CDR1 and 2 with the MHC have been secured.

Is the data on Figure 4 able to point out the influence of the P6A mutation on TCR interactions? We see that the overall binding patterns of Tax-A6 and P6A-A6 are extremely similar. Slight differences appear in which residues of low importance are selected by the 2.5% threshold. Computing differences in the interaction counts (data not shown) allows identifying the residues with highest variations. For example, in terms of hydrophobic contacts on the TCR side, L98α is most affected, because it sits just on top of peptide Pro 6. In terms of H-bonds, the strongest feature is a weakening of the MHC R65 and TCR D99α interactions upon the P6A mutation. This could be due to a modification of CDR3α packing due to its proximity to the peptide Pro 6 side chain. Note that these differences are barely visible on Figure 4, which indicates that their influence must be minor. Indeed, according to a previous simulation study,[36] direct TCR-pMHC interactions account for less of the total binding free energy difference than solvation or internal reorganization effects. In that same study, D99α was also identified as the TCR residue with the most unfavorable energy contribution upon P6A mutation. The resemblance of H-bond and contact counts for residues not directly impacted by the P6A mutation is a confirmation that the results on Figure 4 are reliable, because they are reproducible from totally independent simulations.

A close-up on Tax Y5

The approach described above to represent per-residue counts of H-bonds or contacts can be refined to the level of single atoms. To illustrate this, we focus on the Tyr5 side chain of the Tax peptide, whose phenol OH group (Tyr5-OH) makes different H-bonds with the A6 and B7 TCRs. These can be related to the experimental effects of the Y5F mutation, which essentially corresponds to replacing Tyr5-OH with an apolar hydrogen atom.

Figure 5 shows the occurrence of significant H-bonds with Tyr5-OH as a function of ξ. In the bound Tax-A6 complex, Tyr5-OH essentially makes one very frequent H-bond with S31α of the TCR. The H-bond scores indicated in Figure 5 show that all other H-bonding partners in A6 play marginal roles. With the B7 TCR, the Tyr5-OH H-bond pattern is very different. Tyr5-OH acts as a donor in two moderately frequent H-bonds with the carboxylate oxygens of D30α. In addition, Tyr5-OH is an acceptor for an H-bond with the backbone nitrogen of A96α. The cumulated score of Tyr5-OH H-bonds with B7 is around 25% higher than with A6. Also, it might be entropically more favorable for Tyr5-OH to have several H-bonding partners with B7 than to be limited to a single conformation tightly bound with A6. This is consistent with the higher $\Delta\Delta G$ observed experimentally in the Tax-B7 complex for the Y5F mutation.

In addition to the Y5F mutation on the Tax peptide, two alanine mutations have been experimentally tested on the TCR at positions involved in H-bonds with Tyr5-OH. These are S31α in A6 and Y104β in B7. Both mutations have a similar impact of around 2 kJ/mol on the binding free energy.[30] The measured ϕ value for the mutation of Y104β to Ala in B7 indicates that the side chain of this residue is strongly involved in the transition state.[30] In Figure 5 we can see that Tyr5-OH interacts with B7-Y104β in the transition state, although this interaction never occurs in the bound state. The moderate ϕ value of the Y5F mutation with the B7 TCR, together with the fact that Y104β has other possible binding partners in the transition state (see Fig. 2) discourages further discussion of this residue.

The case of S31α in the A6 TCR is more conclusive. The ϕ value associated with the S31α to Ala mutation indicates that the side chain of this residue is significantly involved in the transition state. We see on Figure 2 that A6-S31α interacts mainly with the peptide. Given its quite buried location in the TCR pocket, we can identify Tyr5-OH as the major H-bonding partner of the side chain of S31α, as confirmed in Figure 5 by the strong occurrence of this H-bond up to $\xi \approx 0.5$. Given the high ϕ value associated with the Y5F mutation as well, we can

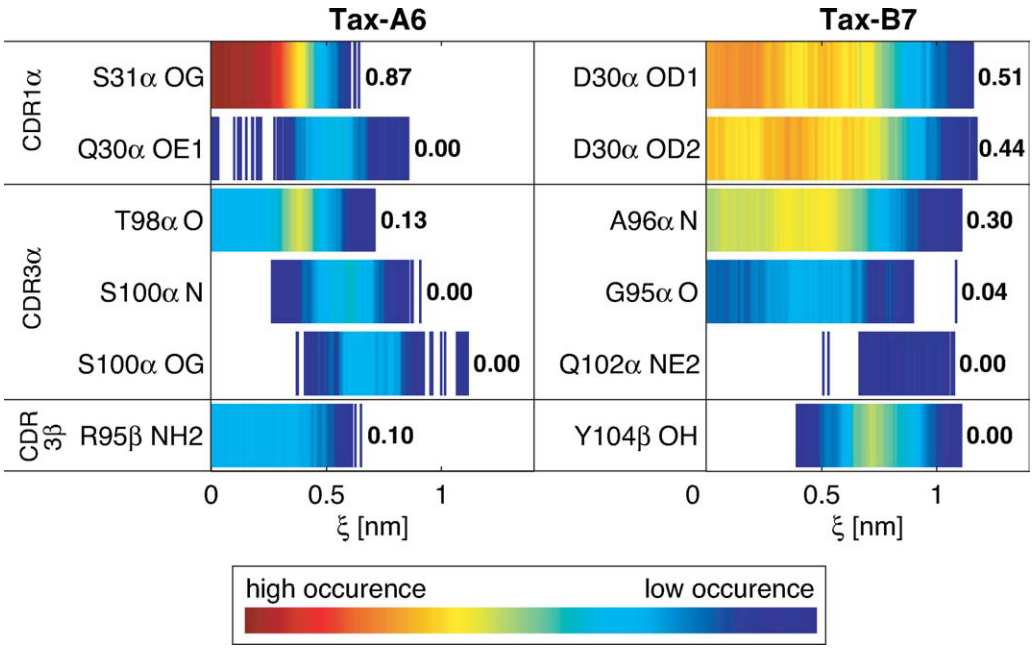

Figure 5

Occurrence of H-bonds between the hydroxyl oxygen atom of Tyr5 in the Tax peptide and the A6 (left) or the B7 (right) TCR, as a function of the TCR and pMHC center of mass distance ξ. The color represents the number of occurrences of a particular H-bond in a small window of ξ. The scale is the same for all H-bonds in both panels. The darkest shade of red corresponds to an occurrence in 90% of conformations around ξ. The number at the right of each color bar is the occurrence of the particular H-bond in the bound state (ratio of conformations in which the H-bond is present in the interval $0.0 < \xi < 0.1$ nm). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

conclude that the H-bond between A6-S31α and Tyr5-OH must be present in the transition state.

Following a combination of arguments from simulation and experiment, we have shown two examples of particular H-bonds between the Tax peptide and the TCR that are formed during the transition state. This sort of interaction is not consistent with the two-step model proposed by Wu *et al.*,[28] in which the peptide engages the TCR only in the bound state. The simulation results point more to the mechanism suggested by Baker and coworkers,[30] in which the transition state already involves interactions with the most salient side chains of the peptide.

CDR interaction energies with peptide or MHC

To understand the role of the different molecular regions involved in the TCR-pMHC interface, it is useful to decompose the interaction energy. The left panel of Figure 6 shows the interaction energy of each CDR loop with either the peptide or the MHC, averaged over the equilibrium bound state trajectories. The total contribution of the peptide is very similar in all three complexes (23%, 25%, and 26%, for Tax-A6, P6A-A6, and Tax-B7, respectively). The contribution of the highly specific CDR3s is much larger for A6 (66% on Tax and 61% on P6A) than for B7 (only 45%), with respect to CDR1 and 2. With the A6 TCR, the CDR3α has a much stronger contribution than CDR3β, although CDR3β buries a larger proportion of the total buried surface (33%) than CDR3α (24%).[3] There is a similar but weaker trend with B7 (22% and 21% of the buried surface for CDR3α and β). This underlines the fact that considering the buried surface is not a good indicator of the importance of protein domains for binding.

Focusing on the effect of the P6A mutation, we see that it has a limited impact on the repartition of binding energy over the different CDRs. The contribution of CDR3β, which is closest to the mutation site, is barely affected. The only significant difference occurs in the interaction of CDR3α with the MHC. This difference can be attributed to the disturbance of the salt bridge between D99α of the TCR and R65 of the MHC, already apparent in the study of H-bonds above. Note that CDR3α was also found to be the preponderant factor of binding free energy difference between Tax-A6 and P6A-A6 in the study of Michielin and Karplus.[36]

The right panels on Figure 6 show the binding energy contributions of the CDRs as a function of ξ. For the Tax-A6 complex, we see that between $\xi = 0.5$ nm and $\xi = 1.0$ nm, both CDR3 are responsible for almost all the interaction energy. In the Tax-B7 complex, the binding energy is more evenly spread across the different CDRs

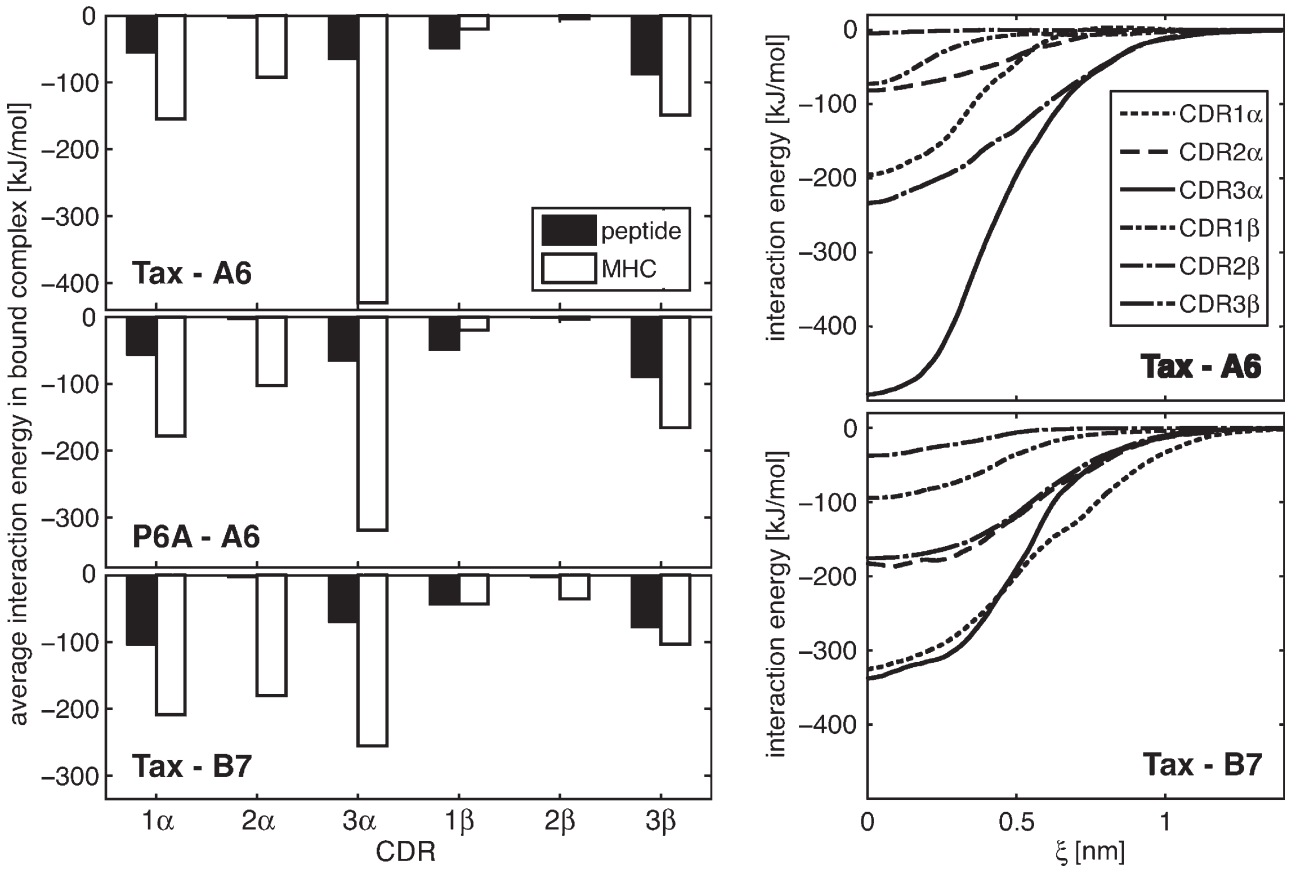

Figure 6

Left : interaction energy between each CDR and the peptide or the MHC, calculated in the bound state of all TCR-pMHC complexes. Right : interaction energy between each CDR and the pMHC as a function of the TCR-pMHC separation ξ, for the Tax-A6 and Tax-B7 complexes.

and CDR3 contributions are still largely present in this range of ξ. In both cases, Figure 6 does not support the two-step model of Wu *et al.*, in which only CDR1 and CDR2 would interact with the MHC at larger separation.

Protein and solvent energies upon dissociation

Monitoring the direct interaction energy between parts of the TCR and pMHC interface tells only a part of the story, and does not capture the most important energetic differences between TCR-pMHC complexes. To illustrate this, we recorded not only the total TCR-pMHC interaction energy, but also the the internal energy in each of the binding partners. In addition, the large number of trajectories at hand enabled us to extract converged averages of the solvent internal energy and of the solvent-protein interaction energy. Figure 7 shows the evolution of these energies as a function of ξ. For all energy curves, the zero energy reference was taken as the unbound state, defined as the average over the range ξ = 1.5 nm to ξ = 2.0 nm. To allow comparison, the same scale is used across all panels of Figure 7, including the inset.

For all complexes, large-scale energy compensations occur between the different parts of the system. Upon dissociation, favorable protein-protein interactions are broken for the benefit of solvent-protein interactions formed as the interfaces are solvated. This gain in solvation energy comes at the expense of the solvent internal energy. Conversely, both TCR and pMHC are able to relax internal strains caused by induced fit mechanisms. Except for the TCR-pMHC interaction energy, the energy contributions do not reach a strictly constant value at ξ = 2.0. At large ξ, internal energies in the protein and the solvent tend to slowly decrease, whereas the solvent-protein interaction energy increases. These drifts are due to slow reorganization processes continuing after the unbinding.

The variation of the total energy of the system over the dissociation process is eventually much smaller than any of the separate contributions, as shown in the inset of Figure 7. The total system energies are subject to large uncertainties due to the summation over large and opposed contributions, as well as the slow relaxation

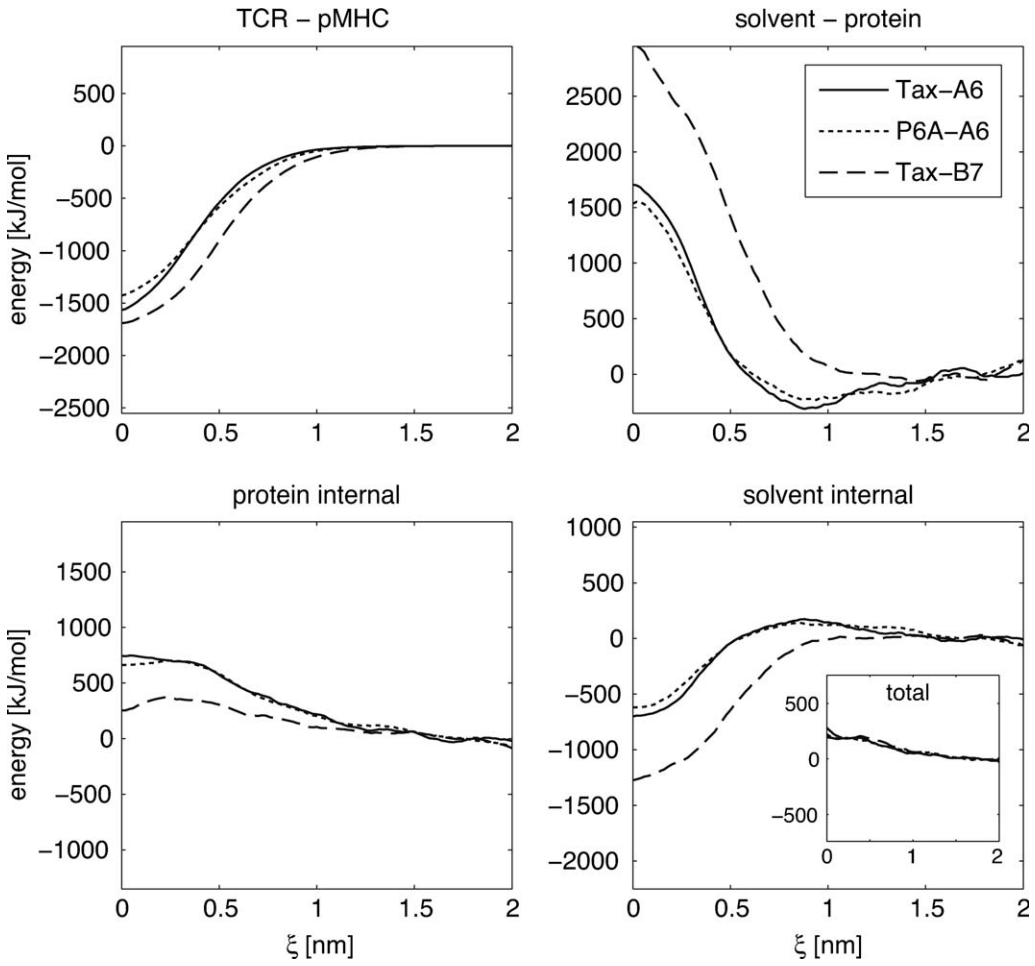

Figure 7

Variation of energy components as a function of the unbinding reaction coordinate ξ: interaction energy between TCR and pMHC, interaction energy between proteins and solvent, protein internal energy, solvent internal energy. For each complex, the sum of energy components from all four panels represents the total energy of the system, which is shown in the inset. For each curve, the reference is the unbound state, defined as the average value over the last 0.5 nm of the dissociation process. To allow comparison, all panels (including the inset) use the same energy scale.

phenomenon discussed above. However, comparing energetic contributions in Figure 7 brings valuable insights, which we discuss next.

We start by focussing on the influence of the P6A mutation. Expectedly, this small alteration leaves the energy profiles on Figures 7 largely unchanged, except for small variations close to the bound state. In the P6A-A6 complex, lost contacts with the proline pyrrole ring make the interaction energy less favorable at short range. However, this effect is partly compensated by the fact that there is less internal strain in the P6A-A6 complex, whose interface is less packed than with the bulkier proline side chain. The solvent-protein energy seems to favor the P6A-A6 complex, whereas the solvent internal energy seems to favor the wild type. Subtle differences in solvation contributions are difficult to assess from Figure 7, because of large cancellation effects and missing entropic contributions. Michielin and Karplus[36] found that the MHC-bound P6A peptide has a better solvation free energy, which destabilizes the P6A-A6 complex by $\sim$ 2.5 kJ/mol with respect to the Tax-A6 complex.

The behavior of the Tax-B7 complex clearly diverges from that of both complexes involving the A6 TCR, as evident on Figure 7. With B7, the direct TCR-pMHC interaction is only slightly stronger than with A6, but it is not compensated by the large internal reorganization energy found with A6. Both solvent internal and solvent-protein contributions are of much larger amplitude with the B7 TCR, which is consistent with the larger variation in solvent H-bonds observed in Figure 2. However, this is not foreseeable based on the crystal structures, because the Tax-B7 complex buries slightly less surface area than Tax-A6, with a similar ratio of polar residues.[20]

On the two right panels of Figure 7, The Tax-A6 complex shows a sort of transition state around $\xi = 0.9$ nm, in which the solvent-protein interaction is more favorable

(and solvent internal energy less favorable) than in the unbound state. In contrast, this transition state is completely absent in the Tax-B7 energy profiles. This is consistent with the behavior for H-bonds between the TCR and the solvent observed in the lower panel of Figure 2. The fact that energetic contributions on Figure 7 correlate with H-bond counts is a hint that the electrostatics must play a prominent role. Supporting Information Figure S3 is identical to Figure 7, but shows only the electrostatic contribution to the energy in each part of the systems. We verify that indeed, most of the features of Figure 7 are reproduced in Supporting Information Figure S3, indicating that most of the differences between A6 and B7 have their origin in the electrostatic interaction, especially within the solvent.

CONCLUSION

Detailed aspects of protein-protein interactions can be characterized by SMD simulations as a function of the distance between binding partners. This requires that enough trajectories are available to converge ensemble averages at any given protein-protein separation, including for highly fluctuating observables involving solvent molecules. In the case of the TCR-pMHC complex, ~150 independent 4 ns trajectories in which the center of mass separation is increased of 2 nm appear to be enough to produce converged dissociation profiles for observables whose level of detail ranges from whole proteins down to a single atom. Once the SMD trajectories have been produced, a wealth of information can be extracted. The methods proposed here can be further applied to extract detailed H-bond or contact maps for any residue or atom in the system, or interaction energy profiles in and between any parts of the system, including the solvent. Thus the existing SMD data can be further exploited with limited effort as new biological questions arise or new experimental results call for interpretation.

The three TCR-pMHC systems studied here are good illustrations of both specificity and cross reactivity at a molecular recognition interface. Although the P6A-A6 complex has a very different affinity compared to the wild type Tax-A6, we see that both complexes share very similar features in terms of specific H-bonds or energy contributions. On the other hand, the Tax-B7 complex has a binding affinity similar to the Tax-A6, but a completely different binding mechanism. The B7 TCR creates a very different set of H-bonds and hydrophobic contacts to the pMHC, inducing a different repartition of the binding energy among its CDRs. Focussing on the phenol OH group of the Tyr5 residue of the Tax peptide, we have shown that A6 and B7 exploit very distinct H-bonding modes.

Another major difference between the A6 and B7 TCRs comes from their interactions with the solvent, which plays a major role. In particular, B7 traps more water molecules than A6 at different regions of the interface. Furthermore, B7 does not show the transition state observed in A6 in terms of reorganization energy and solvent interactions. We were able to show that the differences between A6 and B7 result from a different usage of electrostatics for binding, especially in protein-solvent or solvent-solvent interactions. This is consistent with experimental observations that A6 and B7 binding is differently affected by changes in ionic strength.[20] In retrospect, as noted previously by Baker and coworkers,[15] it is not so surprising to observe different behaviors for A6 and B7, given the fact that TCRs are produced by random CDR sequence variations followed by negative selection in the thymus. We can thus imagine that each TCR arrives at a unique solution of how to bind a given pMHC, if the only selection criteria are affinity and binding kinetics.

Our simulations do not support the two-step model for TCR engagement, in which the CDR1 and CDR2 preferentially contact the MHC at large distance, while the CDR3 establishes final contacts with the peptide mainly at short distances. Instead, our H-bond occurrence maps and interaction energy decompositions indicate that, while CDR2 interacts mostly with the MHC, both CDR1 and CDR3 interact equally with peptide and MHC, even at large distances. The model proposed by Baker and coworkers[30] is more consistent with this study. In this model, the transition state is characterized by contacts by the most protruding side chains and expulsion of the solvent from 65 to 70% of the final buried surface, including a large contribution of the peptide. Despite the evidence that a number of contacts and H-bonds exist only at intermediate TCR-pMHC separation, as well as a signature in the resolvation of A6, no clean-cut transition state can be characterized from this study. Such a continuum of transition states is compatible with the FPRM binding model,[25] or other hybrid models between induced fit and conformer selection.[24]

Lessons learned for general protein–protein interaction include the fact that solvent plays a key role in two different ways. First, variations of solvation energies (protein–solvent and solvent internal) exceed contributions from the proteins themselves (interaction energy, or variation of internal energy) upon binding. Second, specific water molecules trapped at the interface can influence the binding mechanism and thermodynamics. Here, the shape of the surface recognized by the TCR is not given by the pMHC alone, but it is modulated by the trapped water molecules. Different patterns of trapped water molecules can facilitate the cross-reactivity observed with various TCRs.[3] We have also learned that the crystal structure does not give a faithful representation of the population of H-bonds involved in the TCR-pMHC complex at room temperature. In addition, the number of water molecules trapped at the protein–protein interface cannot be inferred from the number of water molecules resolved in the crystal

structure. Furthermore, counting H-bonds or measuring buried solvent accessible surface are poor predictors for the actual binding thermodynamics.

This study shows how energy contributions and H-bond or contact statistics for protein–protein unbinding (including solvent) can be obtained from multiple SMD simulations. We however need to keep in mind the main assumption underlying this work: we assume that the pathway followed by the system in our SMD dissociation simulations is representative of the real unbinding mechanism. We have seen that this assumption is supported by *a priori* arguments and by the fact that the results we obtain allow meaningful interpretation of experiments. Longer unbinding simulations with less restraints on the TCR orientation, as well as reverse binding simulations, would be necessary to further validate the results shown in this study.

ACKNOWLEDGMENTS

The authors are grateful to the Vital-IT computing facility of the Swiss Institute of Bioinformatics, and the Hewlett-Packard Labs in Bristol for supplying the computing power necessary for this work. They Specially thank R. Fabbretti and V. Flegel for their technical assistance. They also thank B. Baker and K. Piepenbrink for their insightful discussions.

REFERENCES

1. Ding YH, Baker BM, Garboczi DN, Biddison WE, Wiley DC. Immunity 1999;11:45.
2. Choudhuri K, van der Merwe PA. Sem Immunol 2007;19: 255–261.
3. Rudolph MG, Stanfield RL, Wilson IA. Annu Rev Immunol 2006;24:419.
4. Garcia KC, Adams EJ. Cell 2005;122:333–336.
5. Rudolph MG, Luz JG, Wilson IA. Annu Rev Biophys Biomol Struct 2002;31:121.
6. Reiser JB, Darnault C, Grégoire C, Mosser T, Mazza G, Kearney A, van der Merwe PA, Fontecilla-Camps JC, Housset D, Malissen B. Nat Immunol 2003;4:241.
7. Gagnon SJ, Borbulevych OY, Davis-Harrison RL, Turner RV, Damirjian M, Wojnarowicz A, Biddison WE, Baker BM. J Mol Biol 2006;363:228–243.
8. Reiser JB, Gregoire C, Darnault C, Mosser T, Guimezanes A, Schmitt-Verhulst AM, Fontecilla-Camps JC, Mazza G, Malissen B, Housset D. Immunity 2002;16:345–354.
9. Armstrong KM, Piepenbrink KH, Baker BM. Biochem J 2008;415: 183–196.
10. Borbulevych OY, Piepenbrink KH, Gloor BE, Scott DR, Sommese RF, Cole DK, Sewell AK, Baker BM. Immunity 2009;31:885–896.
11. Tynan FE, Reid HH, Kjer-Nielsen L, Miles JJ, Wilce MCJ, Kostenko L, Borg NA, Williamson NA, Beddoe T, Purcell AW, Burrows SR, McCluskey J, Rossjohn J. Nat Immunol 2007;8:268–276.
12. Borbulevych OY, Insaidoo FK, Baxter TK Jr, DJP, Johnson LA, Restifo NP, Baker BM. J Mol Biol 2007;372:1123–1136.
13. Marrack P, Scott-Browne JP, Dai S, Gapin L, Kappler J. Annu Rev Immunol 2008;171–203.
14. Burrows S, Chen Z, Archbold J, Tynan F, Beddoe T, Kjer-Nielsen L, Miles J, Khanna R, Moss D, Liu Y, Gras S, Kostenko L, Brennan RM, Clements CS, Brooks AG, Purcell AW, McCluskey J, Rossjohn J. Proc Natl Acad Sci USA 2010;107:10608.
15. Gagnon SJ, Borbulevych OY, Davis-Harrison RL, Baxter TK, Clemens JR, Armstrong KM, Turner RV, Damirjian M, Biddison WE, Baker BM. J Mol Biol 2005;353:556–573.
16. Rosette C, Werlen G, Daniels MA, Holman PO, Alam SM, Travers PJ, Gascoigne NRJ, Palmer E, Jameson SC. Immunity 2001;15:59.
17. Tian S, Maile R, Collins EJ, Frelinger JA. J Immunol 2007;179: 2952–2960.
18. Armstrong KM, Insaidoo FK, Baker BM. J Mol Recognit 2008;21:275–287.
19. Ely L, Beddoe T, Clements C, Matthews J, Purcell A, Kjer-Nielsen L, McCluskey J, Rossjohn J. Proc Natl Acad Sci USA 2006;103: 6641–6646.
20. Davis-Harrison RL, Armstrong KM, Baker BM. J Mol Biol 2005;346:533.
21. Boehr DD, Wright PE. Science 2008;320:1429.
22. Koshland DE. Proc Natl Acad Sci USA 1958;44:98–104.
23. Kumar S, Ma B, Tsai C, Sinha N, Nussinov R. Protein Sci 2000; 9:10–19.
24. Zhou H. Biophys J 2010;98:L15–L17.
25. Grünberg R, Leckner J, Nilges M. Structure 2004;12:2125–2136.
26. Teilum K, Olsen JG, Kragelund BB. Cell Mol Life Sci 2009;66: 2231–47.
27. Gakamsky DM, Lewitzki E, Grell E, Saulquin X, Malissen B, Montero-Julian F, Bonneville M, Pecht I. Proc Natl Acad Sci USA 2007;104:16639–16644.
28. Wu LC, Tuot DS, Lyons DS, Garcia KC, Davis MM. Nature 2002;418:552.
29. Borg N, Ely L, Beddoe T, Macdonald W, Reid H, Clements C, Purcell A, Kjer-Nielsen L, Miles J, Burrows S, McCluskey J, Rossjohn J. Nat Immunol 2005;6:171–180.
30. Davis-Harrison RL, Insaidoo FK, Baker BM. Biochemistry 2007; 46:1840–1850.
31. Pettersen E, Goddard T, Huang C, Couch G, Greenblatt D, Meng E, Ferrin T. J Comput Chem 2004;25:1605–1612.
32. Hausmann S, Biddison WE, Smith KJ, Ding Y-H, Garboczi DN, Utz U, Wiley DC, Wucherpfennig KW. J Immunol 1999;162: 5389–5397.
33. Garboczi DN, Ghosh P, Utz U, Fan QR, Biddison WE, Wiley DC. Nature 1996;384:134.
34. Baker BM, Gagnon SJ, Biddison WE, Wiley DC. Immunity 2000; 13:475.
35. Ding Y-H, Smith KJ, Garboczi DN, Utz U, Biddison WE, Wiley DC. Immunity 1998;8:403–411.
36. Michielin O, Karplus M. J Mol Biol 2002;324:547–569.
37. Baker BM, Turner RV, Gagnon SJ, Wiley DC, Biddison WE. J Exp Med 2001;193:551–562.
38. Baxter TK, Gagnon SJ, Davis-Harrison RL, Beck JC, Binz AK, Turner RV, Biddison WE, Baker BM. J Biol Chem 2004;279: 29175.
39. Piepenbrink KH, Borbulevych OY, Sommese RF, Clemens J, Armstrong KM, Desmond C, Do P, Baker BM. Biochem J 2009; 423:353–361.
40. Armstrong KM, Baker BM. Biophys J 2007;93:597–609.
41. Zoete V, Irving MB, Michielin O. J Mol Recognit 2009;23: 142–152.
42. Zoete V, Michielin O. Proteins 2007;67:1026.
43. Grubmüller H, Heymann B, Tavan P. Science 1996;271:997.
44. Park S, Schulten K. J Chem Phys 2004;120:5946–5961.
45. Cuendet MA, Michielin O. Biophys J 2008;95:1–16.
46. Berman H, Westbrook J, Feng Z, Gilliland G, Bhat T, Weissig H, Shindyalov I, Bourne P. Nucleic Acids Res 2000;28:235–242.
47. Lindahl E, Hess B, van der Spoel D. J Mol Mod 2001;7:306.
48. van Gunsteren WF, Billeter SR, Eising AA, Hünenberger PH, Krüger P, Mark AE, Scott WRP, Tironi IG. Biomolecular simulation: the GROMOS96 manual and user guide. Zürich: Vdf Hochschulverlag AG an der ETH Zürich; 1996.

49. Daura X, Mark AE, Gunsteren WFV. J Comput Chem 1998;19: 535–547.
50. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. J Comput Chem 1997;18:1463.
51. Ryckaert JP, Ciccotti G, Berendsen HJC. J Comput Phys 1977; 23:327–341.
52. Berendsen HJC, Postma JPM, van Gunsteren WF, Hermans J. Intermolecular Forces. Dordrecht: Reidel; 1981. p 331.
53. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. J Chem Phys 1995;103:8577.
54. Hoover WG. Phys Rev A 1985;31:1695.
55. Parrinello M, Rahman A. J Appl Phys 1981;52:7182.
56. Nosé S, Klein ML. Mol Phys 1983;50:1055.
57. Caves LSD, Evanseck JD, Karplus M. Protein Sci 1998;7:649.
58. Cebecauer M, Guillaume P, Mark S, Michielin O, Boucheron N, Bezard M, Meye BH, Segura JM, Vogel H, Luescher IF. J Biol Chem 2005;280:23820.
59. Hess B, Kutzner C, van der Spoel D, Lindahl E. J Chem Theory Comput 2008;4:435–447.
60. Li Z, Lazaridis T. Phys Chem Chem Phys 2007;9:573–581.
61. Reichmann D, Phillip Y, Carmi A, Schreiber G. Biochemistry 2008; 47:1051.
62. Young T, Abel R, Kim B, Berne BJ, Friesner RA. Proc Natl Acad Sci USA 2007;104:808–813.