PROTEINS: Structure, Function, and Bioinformatics 67:1026–1047 (2007)

Comparison Between Computational Alanine Scanning and Per-Residue Binding Free Energy Decomposition for Protein–Protein Association Using MM-GBSA: Application to the TCR-p-MHC Complex

Vincent Zoete[1]* **and Olivier Michielin**[1,2,3]*
[1]*Swiss Institute of Bioinformatics (SIB), Molecular Modeling Group, Genopode, CH-1015 Lausanne, Switzerland*
[2]*Ludwig Institute for Cancer Research, National Center of Competence in Research (NCCR) Molecular Oncology, 1066 Epalinges, Switzerland*
[3]*Multidisciplinary Oncology Center (CePO), Lausanne University Hospital (CHUV), Switzerland*

ABSTRACT **Recognition by the T-cell receptor (TCR) of immunogenic peptides (p) presented by Class I major histocompatibility complexes (MHC) is the key event in the immune response against virus-infected cells or tumor cells. A study of the 2C TCR/SIYR/H-2K^b system using a computational alanine scanning and a much faster binding free energy decomposition based on the Molecular Mechanics—Generalized Born Surface Area (MM-GBSA) method is presented. The results show that the TCR-p-MHC binding free energy decomposition using this approach and including entropic terms provides a detailed and reliable description of the interactions between the molecules at an atomistic level. Comparison of the decomposition results with experimentally determined activity differences for alanine mutants yields a correlation of 0.67 when the entropy is neglected and 0.72 when the entropy is taken into account. Similarly, comparison of experimental activities with variations in binding free energies determined by computational alanine scanning yields correlations of 0.72 and 0.74 when the entropy is neglected or taken into account, respectively. Some key interactions for the TCR-p-MHC binding are analyzed and some possible side chains replacements are proposed in the context of TCR protein engineering. In addition, a comparison of the two theoretical approaches for estimating the role of each side chain in the complexation is given, and a new ad hoc approach to decompose the vibrational entropy term into atomic contributions, the linear decomposition of the vibrational entropy (LDVE), is introduced. The latter allows the rapid calculation of the entropic contribution of interesting side chains to the binding. This new method is based on the idea that the most important contributions to the vibrational entropy of a molecule originate from residues that contribute most to the vibrational amplitude of the normal modes. The LDVE approach is shown to provide results very similar to those of the exact but highly computationally demanding method. Proteins 2007;67:1026–1047.** © 2007 Wiley-Liss, Inc.

Key words: TCR; MHC; p-MHC; protein association; MM-GBSA; binding free energy; molecular recognition; computational alanine scanning; free energy decomposition; molecular dynamics; continuum model; GB-MV2

INTRODUCTION

The specific cellular immune response is based on the recognition by cytotoxic T lymphocytes of immunogenic peptide (p) presented at the surface of the target cell by a Class I major histocompatibility complex (MHC). Binding of the T-cell receptor (TCR) to the p-MHC complex results in the activation of the CTL and target cell destruction. Because of the central role of the T lymphocyte in the immune response, the molecular basis of the interaction between the TCR and the p-MHC is of general interest in immunology and medicine, and more specifically in cancer immunotherapy. There is, therefore, a need for methods aimed at understanding the role of each residue of the different components of the TCR-p-MHC complex, not only at the structural level, but also from a thermodynamic point of view. Such approaches could be used, in turn, to guide new experimental investigations, and to rationally optimize tumor specific TCR or peptide sequences.

Computational molecular modeling methods provide an insight into the structural and energetic effect of mutations on the association constant between two pro-

Grant sponsor: The Swiss Institute of Bioinformatics; Grant sponsor: Swiss National Science Foundation; Grant numbers: SCORE 3232B0-103172 and 3200B0-103173; Grant sponsor: Oncosuisse; Grant number: OCS 01381-08-2003.

*Correspondence to: Vincent Zoete or Olivier Michielin, Swiss Institute of Bioinformatics (SIB), Molecular Modeling Group, Genopode, CH-1015 Lausanne, Switzerland. E-mail: Olivier.Michielin@unil.ch or Vincent.Zoete@isb-sib.ch

Received 8 March 2006; Revised 20 October 2006; Accepted 22 December 2006

Published online 21 March 2007 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.21395

teins. They describe the binding process in terms of contributions from van der Waals and electrostatic interactions energies, desolvation free energy, and entropy.[1–4] Free energy simulations[5,6] are the most accurate methods for studying the effect of mutations. Thermodynamical integration has been applied to reproduce and rationalize the experimental binding free energy difference of a particular TCR (A6) for a wild type peptide (Tax) and a mutant peptide (Tax-P6A), both presented in HLA A2.[3] However, despite increased computational power, these methods are still quite time consuming and cannot be applied to study the role of a large number of residues. Among the simplified approaches that have been developed to address this issue, the molecular mechanics—Poisson-Boltzmann surface area (MM-PBSA) approach[7,8] is one of the most promising and widely used. The MM-PBSA approach has been applied to estimate the binding energy for protein/protein[9,4,10] or protein/ligand systems[11–13]. In this approach, the binding free energy is estimated as the sum of the gas-phase energies, solvation free energies, and entropic contributions, averaged over a series of snapshots from MD trajectories. The gas-phase terms are obtained by simple single point energy calculations, whereas the electrostatic contribution to the solvation term is calculated by solving the Poisson-Boltzmann equation.[14] In addition, it is possible to use this approach to perform a so-called computational alanine scanning[1,2] (CAS) in which the absolute binding free energy is calculated for the wild type system, as well as for several mutants in which one residue has been replaced by an alanine. The difference in the binding free energy of the wild type and of the mutants may be compared directly to the results of an experimental alanine scanning (AS).[1,2] An approach related to MM-PBSA, where the electrostatic contribution to the solvation free energy is determined using a generalized Born model, has been introduced under the name Molecular Mechanics—Generalized Born Surface Area (MM-GBSA).[4] The GB model makes this variant attractive because it is much faster than the PB approach and allows the decomposition of the electrostatic solvation free energy into atomic contributions in a straightforward manner. This allows an easy and rapid binding free energy decomposition (BFED) for the wild type system[4,10], and offers a faster alternative to the CAS for the detailed study of protein/protein interactions at the residue level. Also, this approach allows to study the contributions arising from nonmutable groups of atoms, such as backbone atoms. In addition, contrary to CAS, the BFED is a nonperturbing approach that does not require introducing a mutation in the system. In the CAS approach, it is questionable whether simply modifying a given side chain to alanine in the MD simulation trajectory of the wild type system can lead to a good representation of the conformational space of the mutant, since no eventual conformational modification induced by the mutation is investigated. However, it is also questionable how the binding free energy contribution of a given side chain in the wild type complexation may be always representative of the change in the binding free energy upon mutation, since the conformational modifications induced by the mutations are not included in the model either and since, for instance, the modification of the solvation free energy of close side chains upon mutation is not evaluated directly. Also, on the contrary to the total free energy, the free energy components are not state functions and the values of these contributions are thus dependent on the decomposition scheme used. Obviously, both methods cannot be expected to provide results exactly comparable to experimental values obtained from an AS. However, as explained later, we have obtained a fair agreement between the experimental and theoretical results in the present study using a BFED based on a single wild type complex MD simulation. A detailed and methodologically important analysis of the question of convergence, and dependence of the results upon the parameters and the physical models used to calculate the energy terms in the MM-PB(GB)SA approach has been recently published.[15] This study uses the Ras-Raf protein–protein complex as an example.

In the present work, we use the MM-GBSA method to explore the molecular basis of the 2C TCR binding to the SIYR/H-2K^b p-MHC complex, and point out the most important residues in this interaction. This system was chosen because of the availability of a 3D structure of the 2C TCR/SIYR/H-2K^b complex[16] (Fig. 1) and of an experimental AS mutagenesis of the 2C TCR.[18] The two *in silico* approaches mentioned earlier were used to study this interaction: a CAS and a BFED for the wild type system. To our knowledge, the present study offers the first systematic comparison of the two approaches. In addition, a new intuitive approach is introduced to decompose approximately the vibrational entropy term into atomic contributions. The theoretical basis of this approach will be the subject of a future paper. This approach is shown to increase significantly the correlation between the calculated and experimentally determined contributions to the binding free energy of the wild type system. The BFED on a per-residue basis showed satisfying consistency with experimentally determined AS data.

COMPUTATIONAL METHODS

This section presents the methods to estimate and decompose the binding free energy, and to perform a CAS of the TCR-p-MHC interaction.

System Setup

The starting coordinates for the MD simulation were taken from the X-ray structure of the complex between 2C TCR, H-2K^b, and the superagonist peptide SIYR at 2.8 Å,[16] corresponding to entry 1G6R in the Protein Data Bank (PDB[19,20]). To decrease the computational requirements, the constant domains of TCR (residues 115–223 for Cα and 117–247 for Cβ) were removed. These domains are far from the interaction surface with the p-MHC and are not expected to make important contributions to the binding. Moreover, it has been reported that the variable domains of

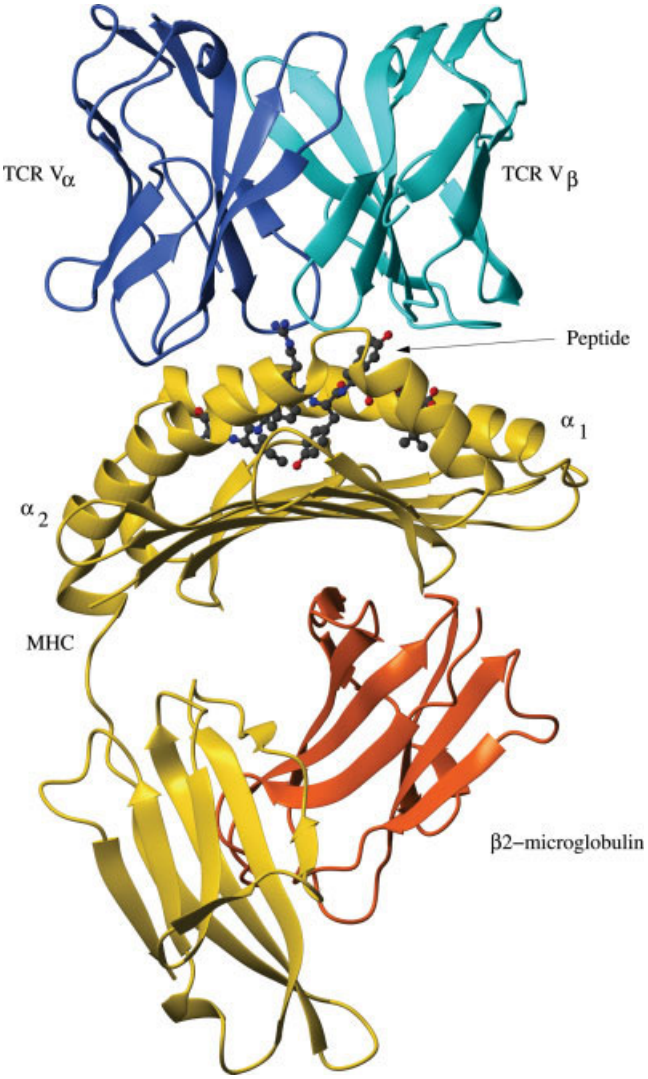

Fig. 1. X-ray structure of the 2C TCR/SIYR/H-2K^b complex, from the 1G6R PDB file. All figures of molecular structures were prepared with the MOLMOL program.[17] [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

TCR, Vα and Vβ, are fully functional in the absence of the constant domains: single chain Vα + Vβ (sFv) have been shown experimentally to have similar affinities to p-MHC.[21] Both Vα and Vβ contain 109 residues. Their new C-termini were aminated to prevent the introduction of an unphysical charge. The entire MHC (i.e., domains α1, α2, and α3) and the β2-microglobulin were included. On the basis of visual inspection, His41 of TCR Vβ was charged, while other histidine residues where set neutral and protonated at N_δ or N_ε depending on their environment. Other titratable groups were taken in their standard protonation state at neutral pH. The total charge was +4 for TCR Vα, +4 for TCR Vβ, −8 for MHC, 0 for the β2-microglobulin, and +1 for the peptide. The overall charge is +1 for the TCR-p-MHC complex. All calculations were carried out using CHARMM[22] (version c31b1) and the molecules were described by the "all-atom" force field CHARMM22.[23]

Molecular Dynamics Simulations

The complex was solvated in a rectangular 139.69 × 93.13 × 93.13 Å^3 box of TIP3P[24] water molecules previously equilibrated at 300 K and 1 atm pressure. Water molecules overlapping the complex were removed, leading to a system with 36903 water molecules and a total of 120118 atoms. The entire system was minimized using 200 steps of the steepest descent (SD) algorithm. The solvent was equilibrated at 300 K during 20 ps in the presence of the fixed complex. Then, a 5 kcal/(mol A^2) harmonic restraint was applied to the backbone of the complex and the entire system was heated to 300 K during 12 ps. The system was equilibrated during 90 ps at 300 K while removing progressively the restraints and during additional 100 ps without any restraint. The MD simulation was conducted with periodic boundary conditions (PBC)[5] during 1 ns, in the NVE ensemble. This duration allowed a global convergence of the log(SIYR/K^b reactivity) versus ΔG^{sc}_{bind} regression (see Binding Free Energy Decomposition for the Wild Type System). The Verlet leapfrog integrator was used for time propagation with a time step of 0.002 ps. Bond lengths involving bonds to hydrogen atoms were constrained using SHAKE.[25] A 12-Å cutoff was applied to the shifted electrostatic and switched van der Waals interactions.

Binding Free Energy

The MM-GBSA approach

The absolute binding free energy for the TCR/p-MHC association was calculated using the MM-GBSA approach.[7,8]

In this approach, the binding free energy, ΔG_{bind}, is written as the sum of the gas phase contribution, ΔG^0_{bind}, the desolvation free energy of the system upon binding, ΔG_{desolv}, and an entropic contribution, $-T\Delta S$.

$$\Delta G_{bind} = \langle \Delta G^0_{bind} \rangle + \langle \Delta G_{desolv} \rangle - \langle T\Delta S \rangle. \quad (1)$$

The brackets, $\langle\rangle$, indicate an average of these energy terms along the MD simulation trajectory (see later).

The gas-phase contribution to the binding free energy is equal to the sum of the van der Waals and electrostatic interaction energies between the TCR and the p-MHC, E_{vdW} and E_{elec}, and ΔE_{intra}, the difference in the internal energy of the TCR and the p-MHC between the complex and the isolated parts.

$$\Delta G^0_{bind} = \Delta E_{intra} + E_{vdW} + E_{elec}, \quad (2)$$

with

$$\Delta E_{intra} = E^{TCR/p\text{-}MHC}_{intra} - (E^{TCR}_{intra} + E^{p\text{-}MHC}_{intra}), \quad (3)$$

and

$$E^X_{intra} = E^X_{intra,bond} + E^X_{intra,vdW} + E^X_{intra,elec}, \quad (4)$$

where $E^X_{intra,bond}$, $E^X_{intra,vdW}$, and $E^X_{intra,elec}$ are the energy of the bonded terms (bonds, angles, dihedral angles, and improper angles) within a given molecule X, and the van

der Waals and electrostatic interactions between the atoms of this molecule, respectively. ΔE_{intra}, E_{elec}, and E_{vdW} are calculated according to the CHARMM22 molecular mechanics force field, with a dielectric constant of 1 and no cutoff.

ΔG_{desolv} is the difference between the solvation free energy of the complex and that of the isolated parts.

$$\Delta G_{\text{desolv}} = \Delta G_{\text{solv}}^{\text{TCR/p-MHC}} - (\Delta G_{\text{solv}}^{\text{TCR}} + \Delta G_{\text{solv}}^{\text{p-MHC}}) \quad (5)$$

The solvation free energy, $\Delta G_{\text{solv}}^{X}$, where X stands for the TCR-p-MHC complex or the isolated TCR and p-MHC, is divided into the electrostatic, $\Delta G_{\text{elec,solv}}^{X}$, and the nonpolar, $\Delta G_{\text{np,solv}}^{X}$, contributions.

$$\Delta G_{\text{solv}}^{X} = \Delta G_{\text{elec,solv}}^{X} + \Delta G_{\text{np,solv}}^{X} \quad (6)$$

$\Delta G_{\text{np,solv}}^{X}$, which can be considered as the sum of a cavity term and a solute–solvent van der Waals term, is assumed to be proportional to the solvent accessible surface area, (SASA); that is., $\Delta G_{\text{np,solv}}^{X} = \sigma \times \text{SASA}$. This well-known and often used approximation comes from the fact that the solvation free energy of saturated nonpolar hydrocarbons is linearly related to the SASA.[26,27] We used a value of 0.0072 kcal/(mol Å^2) for the parameter σ.[4,28,29] The SASAs were calculated analytically with CHARMM. The electrostatic contribution to the solvation free energy, $\Delta G_{\text{elec,solv}}^{X}$, was calculated using the analytical generalized Born (GB) GB-MV2 model implemented in CHARMM.[30,31] This model was found to reproduce the solvation free energies calculated by solving the Poisson equation with 1% accuracy.[30,31] The use of a GB equation allows one to decompose the electrostatic contribution to the binding free energy on a per-atom basis in a straightforward manner (see later and Refs. 4 and 10).

The entropy term is decomposed into translational (ΔS_{trans}), rotational (ΔS_{rot}), and vibrational (ΔS_{vib}), contributions

$$\Delta S = \Delta S_{\text{vib}} + \Delta S_{\text{trans}} + \Delta S_{\text{rot}} \quad (7)$$

These terms are calculated using standard equations of statistical mechanics.[32,33] $T\Delta S_{\text{trans}}$ and $T\Delta S_{\text{rot}}$ are functions of the mass and moments of inertia of the molecule, respectively. The vibrational entropy term is calculated with the quantum formula from a normal mode analysis:[33]

$$S_{\text{vib}} = \frac{k\alpha}{e^{\alpha} - 1} - \ln(1 - e^{-\alpha}) \quad (8)$$

with ω the frequency of the normal mode, and $\alpha = \hbar\,\omega/kT$. To calculate the vibrational entropy term, the normal mode analysis was preferred to the quasiharmonic analysis of the MD simulations. Actually, in the study of the Ras-Raf complexation using the MM-PBSA or MM-GBSA approach, Gohlke and Case[15] found that the latter method did not yield convergent vibrational entropy contributions within 10 ns of simulation time, and also led to strong deviations from the results obtained with a normal mode analysis, giving an overall unreasonable entropic contribution.

The VIBRAN module of the CHARMM program was used to calculate and diagonalize the force constant matrix for the normal mode vectors and frequencies determination. To cope with the available memory of the computers, the normal modes were only calculated for the residues belonging to the contact surface between the TCR and the p-MHC: residues 24–32 (CDR1), 48–56 (CDR2), 66–72 (HV4), and 92–104 (CDR3) for TCR Vα, residues 25–32 (CDR1), 47–57 (CDR2), 69–76 (HV4) and 93–107 (CDR3) for TCR Vβ, residues 58–80 (α1) and 144–167 (α2) for MHC, and all residues of the peptide. The rest of the complex was fixed and thus its contribution to the vibrational entropy of complexation was neglected. To further limit memory demands, the normal modes were calculated with a cutoff of 8 Å on the electrostatic and van der Waals interactions. A distance dependent dielectric constant of $\varepsilon = 4r$ was used. The normal modes were calculated on minimized structures. Starting from a conformation extracted from the trajectory, the minimization was performed using the Adopted Basis Newton–Raphson minimization algorithm, using the same cutoff scheme and constraints as for the normal mode calculations, until the root mean square of the energy gradient reached a value of 10^{-7} kcal/(mol Å). This gradient has been shown to be satisfactory for calculating normal modes and is expected to yield real frequencies only.[33] Since the normal mode analysis is computationally demanding, $-T\Delta S$ was averaged over only 50 frames of the MD trajectory.

With the exception of the entropy terms (see earlier), all the energy terms were calculated for 500 frames regularly separated by 2 ps along the 1-ns trajectory performed for the complex. We have chosen to calculate all the terms using a single trajectory of the complex, and not additional trajectories for the isolated parts, since considerable cancellation and more stable results are expected this way.[10] With this approach, the difference in the internal energy of the isolated parts upon complexation is neglected, that is $\Delta E_{\text{intra}} = 0$. This is not expected to play a major role in a per-residue decomposition, but could alter the absolute values of ΔG_{bind}.

Computational alanine scanning

The CAS consists of replacing a given side chain by an alanine and recalculating the absolute binding free energy for the mutated system. The difference in the binding free energy of the wild type and alanine mutant, $\Delta\Delta G_{\text{bind}}$, may be compared with the results of an experimental AS:

$$\langle \Delta\Delta G_{\text{bind}} \rangle = \langle \Delta G_{\text{bind}}^{\text{mutant}} \rangle - \langle \Delta G_{\text{bind}}^{\text{wild type}} \rangle \quad (9)$$

and thus

$$\langle \Delta\Delta G_{\text{bind}} \rangle = \langle \Delta E_{\text{vdW}} \rangle + \langle \Delta E_{\text{elec}} \rangle + \langle \Delta\Delta G_{\text{elec,desolv}} \rangle + \langle \Delta\Delta G_{\text{np,desolv}} \rangle - T\langle \Delta\Delta S \rangle \quad (10)$$

The binding free energy of the alanine mutant is calculated using the MM-GBSA approach described earlier,

from the set of snapshots obtained for the wild type complex. The side chain of the residue under investigation is simply truncated, replacing the Cγ atom by a hydrogen. For each frame, the new methyl side chain is minimized in the context of the fixed remaining system using 100 steps of SD minimization before calculating the energy terms. One underlying approximation of the CAS is thus that the structural perturbation of the system upon such a mutation is small enough so that its effect on the binding free energy may be obtained from the trajectory of the wild type system. Owing to the prohibitive computational cost, and in accord with previous studies,[2] the entropy term was neglected. Taking this term into account would have required a normal mode analysis for each of the mutants. Also, the study of a given residue with this approach is computationally demanding, since it requires the complete recalculation of the absolute binding free energy. On the contrary, the BFED for the wild type system described below requires only one absolute binding free energy calculation to estimate the contributions of all residues to the binding. Therefore, because of computational limitations, the AS was only performed for the TCR residues, for which the results of an experimental AS is available, while all the residues of the TCR-p-MHC complex could be studied using the BFED approach.

Binding free energy decomposition

Owing to the linear form of Eq. 1, a BFED at the atomic level is possible. It was used to evaluate the contribution of each residue to the total binding free energy of the wild type system, as well as the contributions of its side chain and backbone. For this purpose, one half of a pairwise electrostatic interaction energy between two atoms, each belonging to a different component of the complex (TCR or p-MHC), is attributed to both of them. Therefore, the contribution of atom i to the total electrostatic interaction energy between the two components is given by

$$E^i_{\text{elec}} = \frac{1}{2}\sum_j \frac{q_i q_j}{r_{ij}}, \tag{11}$$

where j loops over all the atoms of the component to which i does not belong to. r_{ij} is the distance between the two atoms with charges q_i and q_j, respectively. Similarly, one half of the pairwise intercomponent van der Waals interaction energies between TCR and p-MHC is attributed to the atoms that are part of the interaction pairs, which avoids double counting.

In the single trajectory method that is used here, ΔE_{intra} is equal to zero since the internal energies of the complex and the separated parts are calculated from the same trajectory. Thus, there is no need for a per atom decomposition.

The SASA of each atom i in the complex, $\text{SASA}^{i,\text{TCR-p-MHC}}$, and in each of the two parts, $\text{SASA}^{i,\text{TCR}}$ and $\text{SASA}^{i,\text{p-MHC}}$, are calculated by CHARMM. The contribution of this atom to the nonpolar solvation term is $\Delta G^i_{\text{np,solv}} = \sigma \times (\text{SASA}^{i,\text{TCR-p-MHC}} - (\text{SASA}^{i,\text{TCR}} + \text{SASA}^{i,\text{p-MHC}}))$, where $\text{SASA}^{i,\text{TCR}}$ or $\text{SASA}^{i,\text{p-MHC}}$ is equal to zero depending on what component the atom belongs to.

The GB-MV2 approach uses the expression of Still and coworkers[29] to calculate the electrostatic solvation free energy term:

$$\Delta G_{\text{elec,solv}} = k\sum_{i,j} \frac{q_i q_j}{\sqrt{r_{ij}^2 + \alpha_i\alpha_j\ \exp(-r_{ij}^2/K_s\alpha_i\alpha_j)}} \tag{12}$$

where $k = -166.0\ (\varepsilon^{-1}_{\text{solute}} - \varepsilon^{-1}_{\text{solvent}})$, with energies expressed in kcalories per mole and distances in Angstrom. $\varepsilon_{\text{solute}}$ and $\varepsilon_{\text{solvent}}$ are the dielectric constant of the solute and the solvent, that is 1 and 80, respectively. α_i and α_j are the Born radii of atoms i and j, respectively, calculated with the GB-MV2 approach. i and j loop over all the atoms of the system. The constant K_s is equal to 8 in GB-MV2,[31] unlike the original Still equation, where it is equal to 4. This pairwise expression allows us to define the contribution of atom i to $\Delta G_{\text{elec,solv}}$ as

$$\Delta G^i_{\text{elec,solv}} = k\frac{q_i^2}{\alpha_i} + \frac{1}{2}k\sum_{j\neq i} \frac{q_i q_j}{\sqrt{r_{ij}^2 + \alpha_i\alpha_j\ \exp(-r_{ij}^2/K_s\alpha_i\alpha_j)}} \tag{13}$$

A theoretically exact way to calculate the contribution of a given group of atoms to the total vibrational entropy is to zero the mass of these atoms and recalculate the normal modes and the corresponding total entropy. The difference between the vibrational entropy of the wild type system and of the system with some zeroed masses gives the vibrational entropic contribution of the corresponding atoms. Unfortunately, this approach is very time consuming since it requires to recalculate the normal modes for each group of atoms of interest. Therefore, we introduce in this study the linear decomposition of the vibrational entropy (LDVE) approach to approximately describe the decomposition of the vibrational entropy into atomic contributions that necessitates only one normal modes calculation for the wild type system. This approach is thus much faster and makes the estimation of the entropic contribution of each atom feasible. It is based on the idea that the most important contributions to the vibrational entropy term upon binding originate from residues that contribute the more to the vibrational amplitude.

To evaluate the contribution of a given atom to the vibrational entropy of a system, we decompose each normal mode of the system into atomic contributions, and then sum the terms relative to the given atom over all the normal modes. In order to decompose the entropic contribution of a given normal mode into atomic terms, we propose a pragmatic approach that consists in attributing to a given atom a portion of the energy proportional to its part in the total vibrational amplitude of this normal mode. If S_p is the contribution of the pth

normal mode to the total entropy, relative to the complex or a given isolated part, and $(\delta x_i, \delta y_i, \delta z_i)$ the coordinates of the mass weighted Cartesian displacement of this normal mode relative to atom i of mass m_i, then we write S_p^i the contribution of atom i to S_p as

$$S_p^i = S_p \times \frac{m_i(\delta x_i^2 + \delta y_i^2 + \delta z_i^2)}{\sum_{1<j<n} m_j(\delta x_j^2 + \delta y_j^2 + \delta z_j^2)} \quad (14)$$

where n is the total number of atoms in the system. Since the mass weighted Cartesian displacement vectors are normalized, we have $\sum_{1<j<n} m_j(\delta x_j^2 + \delta y_j^2 + \delta z_j^2) = 1$. Then, we write the contribution, S^i, of atom i to the total vibrational entropy of the system as

$$S^i = \sum_{1<p<N} S_p^i \quad (15)$$

where $N = 3n-6$ is the total number of normal modes for the system. By summing the S^i values over groups of atoms, it is possible to obtain the approximate weight of a side chain or of the backbone of a given residue on the entropic term, for instance.

Finally, by summing the atomic contributions E^i_{elec}, E^i_{vdW}, $\Delta G^i_{\text{elec,desolv}}$, $\Delta G^i_{\text{np,desolv}}$, and $-T\Delta S^i$ over the atoms of a given residue, we obtain its contribution to the total binding free energy, $\Delta G^{\text{res}}_{\text{bind}}$. It is also possible to calculate separately the contribution of the backbone, $\Delta G^{\text{bb}}_{\text{bind}}$, or the side chain, $\Delta G^{\text{sc}}_{\text{bind}}$, of a given residue.

RESULTS AND DISCUSSION

We present the assessment of the LDVE method developed for the rapid per-residue decomposition of the vibrational entropy in the first subsection. Then, we examine the MD simulation of the TCR-p-MHC complex in the next subsection. The estimation of the absolute binding free energy is discussed in the third subsection. The contributions of the TCR residues to the binding are presented in the fourth subsection, and the results obtained by CAS and BFED are compared. A detailed discussion of the outliers is provided. The contributions of the MHC and peptide residues to the binding are discussed in the last two subsections.

Assessment of the Vibrational Entropy Decomposition

The LDVE method introduced in the present study (see Binding Free Energy Decomposition) was assessed by comparing its results with those obtained with the exact method for two systems: a globular scorpio protein toxin (64 residues) containing two helices and a β strand (PDB[19,20] code 1AHO), and the human lamin protein (PDB code 1X8Y), which is a 86-residues fully helical protein in which all side chains are solvent accessible. The normal mode analysis of the systems were performed as described in the Methods, and the total vibrational entropy obtained was TS = 710.61 kcal/mol for 1AHO and 985.20 kcal/mol for 1X8Y. The entropy contributions were calculated for each residue (backbone and side chain atoms) from the fully minimized structures, using the LDVE and exact approaches, as described in the Methods. These systems were chosen because their limited size allows the calculation of the entropy contributions of all residues using the exact method in a reasonable amount of time, without the need to fix a part of the system. Also, thanks to their different structures, the impact of the burying of residues on the efficiency of the approximate method may be assessed.

As can be seen in Figure 2(A,B), the correlation between the residue vibrational entropy contributions

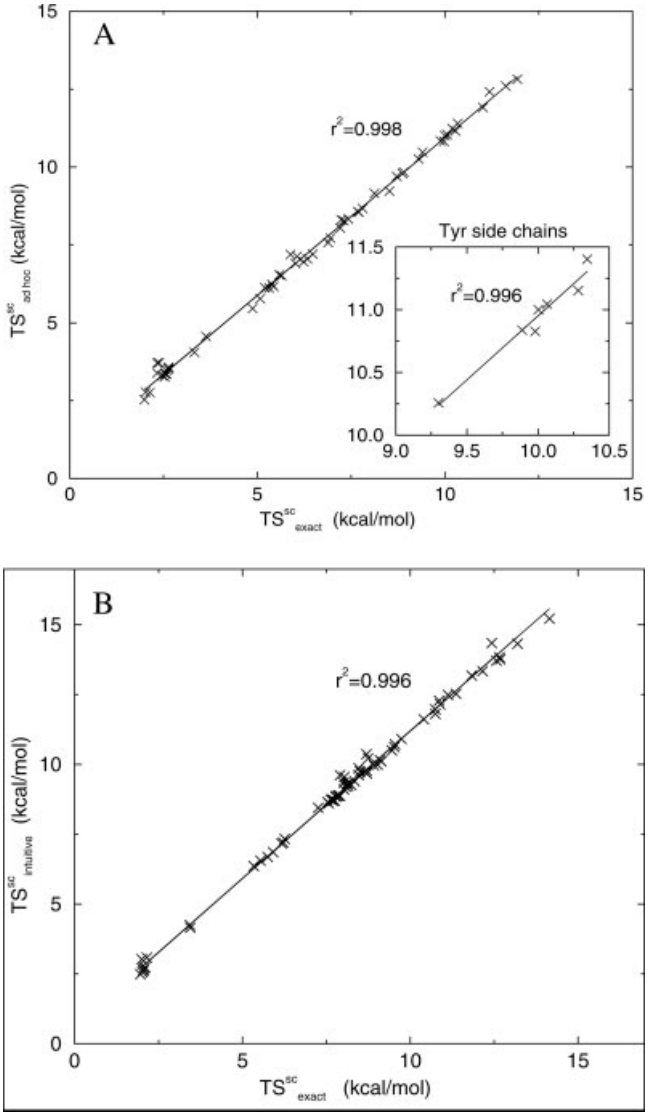

Fig. 2. Comparison between residue vibrational entropy contributions calculated using the exact and the LDVE approaches, for (**A**) the globular scorpio protein toxin, 1AHO, and (**B**) the human lamin protein, 1X8Y.

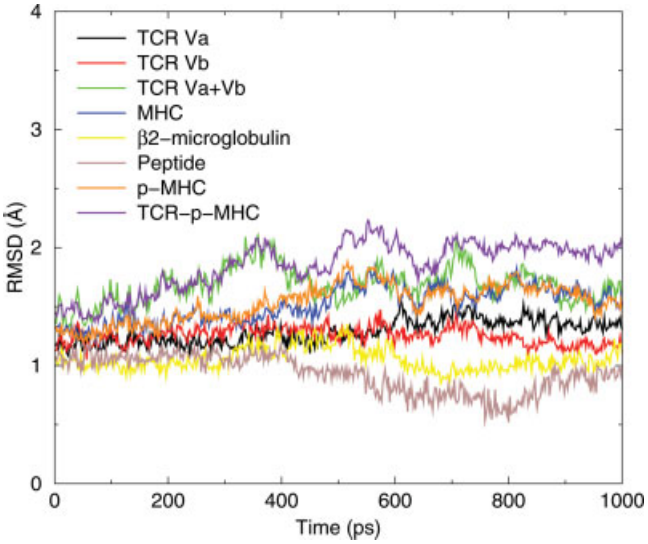

Fig. 3. RMS deviations from the X-ray structure during the MD simulation, calculated for the backbone atoms of the entire TCR-p-MHC complex, and components of the system. The RMSD relative to a given component of the complex was calculated after superimposing the backbone atoms of this component. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

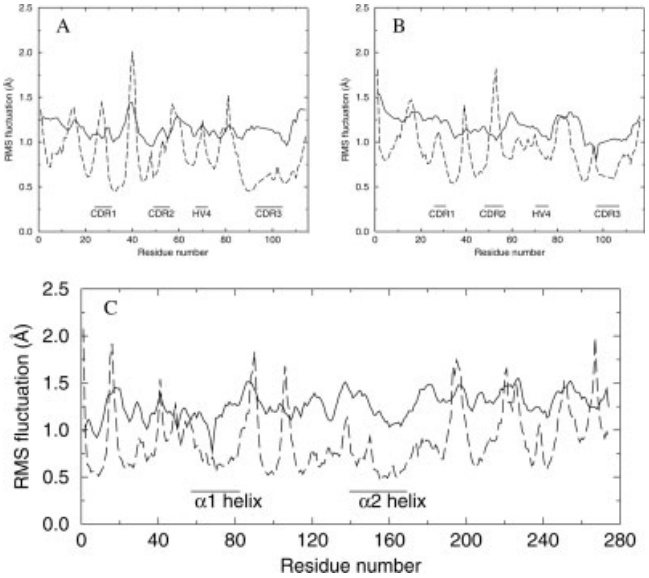

Fig. 4. RMS fluctuations of the backbone atoms calculated from the experimental X-ray B-factors (continuous line) using the formula $B = (8/3)\pi^2 \langle \Delta r \rangle$ and from the MD simulation (dashed line) for TCR Vα (**A**), TCR Vβ (**B**), and MHC (**C**).

calculated using the exact, TS^{sc}_{exact}, and the LDVE method, TS^{sc}_{LDVE}, is extremely good for the two systems:

$$TS^{sc}_{LDVE} = 1.01 \times TS^{sc}_{exact} + 0.81 \text{ for 1AHO},$$

and

$$TS^{sc}_{LDVE} = 1.06 \times TS^{sc}_{exact} + 0.64 \text{ for 1X8Y}.$$

The correlation coefficients are 0.999 and 0.998 for 1AHO and 1X8Y, respectively. The slopes of the regression lines are very close to one, that is 1.01 and 1.06, respectively. Thus, the difference between the contributions calculated using the exact and the LDVE approaches is nearly constant. The shape of the protein (extended or globular) has little influence on the regressions. The regression found for α-helical residues (1X8Y, see earlier) is similar to that for the β-sheet residues of 1AHO (residues 1–5, 32–37, and 44–52): $TS^{sc,\beta\text{-sheet}}_{LDVE} = 1.02 \times TS^{sc,\beta\text{-sheet}}_{exact} + 0.79$ (correlation coefficient, 0.999). The regression found for the 18 buried side chains of 1AHO, defined as having an SASA lower than 20 Å^2, is also very similar to that for the entire protein: $TS^{sc,buried}_{LDVE} = 1.01 \times TS^{sc,buried}_{exact} + 0.72$ (correlation coefficient of 0.999), and to that for solvent exposed residues (1X8Y, see earlier). The approach is sensitive enough to reproduce the small entropy differences between side chains of the same type. For example, the inset of Figure 2(A) shows the regression between the exact and LDVE entropy contributions calculated for the tyrosine side chains of 1AHO. As can be seen, the regression is very good (correlation coefficient of 0.979) and very similar to that found for all side chains: $TS^{sc,Tyr}_{LDVE} = 1.02 \times TS^{sc,Tyr}_{exact} + 0.74$.

Thus, the LDVE provides values that are very similar to those of the exact method. This finding is of major interest for free energy calculations since it allows the full decomposition of the vibrational entropy of a system, which is most of the time computationally inaccessible by the exact method. For example, the LDVE method was 60–70 times faster than the exact method for 1AHO and 1X8Y: respectively 192 and 356 min with a pentium P-IV 3.2 GHz for the exact method, compared to 3 and 5 min with LDVE. The LDVE method will be used in the following.

MD Simulation

To verify the stability of the TCR-p-MHC complex during the MD simulation, the all heavy atoms root mean square deviation (RMSD) from the starting X-ray structure was followed along the trajectory (Fig. 3). As can be seen, the RMSD from the X-ray structure is around 1.9 Å for the entire TCR-p-MHC system, which is quite satisfying in view of the size of the system, that is 599 residues. The RMSD of to each component of the complex along the trajectory is also given in Figure 3. Again, we see that all parts of the complex are stable during the MD simulation, with a RMSD of around 1.6 Å for both TCR and p-MHC. The binding surface, defined as the CDR1, CDR2, HV4, and CDR3 loops of α and β chains of the TCR, the α1 and α2 helices of MHC and the peptide, shows a RMSD of about 1.8 Å at the end of the simulation. Thus, the system shows a satisfying stability during the MD simulation. The transient RMSD decrease observed for the peptide between 400 and 900 ps is not due to a particular conformational change, but simply to thermal fluctuations.

Figure 4 shows the RMS fluctuations of the backbone atoms of TCR and MHC residues along the MD simulation or calculated from the experimental X-ray B-fac-

TABLE I. Binding Free Energy for the Complexation of 2C TCR to H-2K^b/SIYR, and Contributions of Solvation, van der Waals and Electrostatic Interactions, and Entropic Terms

	H-2K^b/SIYR/2C TCR	2C TCR	H-2K^b/SIYR	Δ
$\langle E_{vdW} \rangle$	−105.68 (6.7)	—	—	−105.68 (6.7)
$\langle E_{elec} \rangle$	−593.74 (26.8)	—	—	−593.74 (26.8)
$\langle \Delta G_{elec,solv} \rangle$	−7037.24 (102.34)	−2683.19 (37.9)	−5005.60 (79.3)	651.55 (28.3)
$0.0072 \times \langle SAS \rangle$	198.03 (1.5)	76.40 (0.8)	139.99 (1.1)	−18.37 (0.6)
$\langle \Delta G^0_{bind} \rangle + \langle \Delta G_{desolv} \rangle$				−66.24 (10.6)
$\langle -T\,S_{vib} \rangle$	−1404.50[a] (4.9)	−741.85[a] (3.4)	−678.61[a] (3.0)	15.96 (5.4)
$\langle -T\,S_{trans} \rangle$	−15.83[b]	−14.89[b]	−15.45[b]	14.51[b]
$\langle -T\,S_{rot} \rangle$	−16.39 (<0.01)	−14.73 (<0.01)	−15.80 (<0.01)	14.14 (<0.01)
$\langle \Delta G_{bind} \rangle$				−21.63 (12.1)
$\Delta G_{bind,exp}$				−6.2[16]

Energies are in kcal/mol. Brackets denote an average over the MD simulation. Figures in parentheses are standard deviations.
[a]These values were calculated only for the residues of the contact surface between the TCR and the p-MHC.
[b]The standard deviation of $-T\,S_{trans}$ is not defined since it is only a function of the mass of the system, which is constant.

tors.[5,34] As can be seen, the RMS fluctuations obtained from the MD simulation are lower than those calculated from the B-factors, except for the most flexible part of the proteins where the theoretical and experimental fluctuations are of the same order. Globally, there is a qualitative agreement between the calculated and experimental backbone fluctuations: the most flexible parts are the same. For both Vα and Vβ TCR, the CDR1, CDR2, and HV4 loops are among the most flexible regions of the proteins, while CDR3 that are confined in the center of the interface appear more rigid. The α1 and α2 helices of the MHC protein were found rigid, both experimentally and theoretically. All these data argue that all the microstates sampled by the MD simulation represent a relevant subset of the accessible states.

To address the stability of the results as a function of the length of the trajectory, this MD simulation was prolonged up to 2.2 ns. Also, since several short MD simulations are expected to provide a more thorough and physically relevant sampling of the conformational space in the vicinity of the native structure,[35] three additional MD simulations of the system were performed, each 1 ns in length. The BFED analysis was performed for the prolonged and for the additional trajectories. They provided results very similar to those of the 1-ns MD simulation presented here. In the following, we only report the results corresponding to the first 1-ns trajectory, since we consider that it represents a typical analysis that can be performed for such a large system in a reasonable amount of time. The additional control analysis is presented in the Appendix.

Calculated Absolute Binding Free Energy for the TCR-p-MHC Association

Table I contains the different terms contributing to the calculated absolute binding free energy for the TCR-p-MHC complexation. The sum is compared to the experimental value.

The calculated value of $\langle \Delta G^0_{bind} \rangle + \langle \Delta G_{desolv} \rangle$ averaged over the MD simulation is −66.24 kcal/mol. The contributions to the binding free energy of the van der Waals and electrostatic interactions, and of the electrostatic and nonpolar solvation free energies are −105.68, −593.74, 651.55, and −18.37 kcal/mol, respectively. The favorable contribution of the electrostatic interactions between TCR and p-MHC is more than compensated by the electrostatic desolvation free energy upon complexation, so that the total electrostatic term contributes unfavorably to the binding, which is in agreement with other MM-PBSA and MM-GBSA studies.[1,2,4,9,10] Care should be taken in interpreting absolute values in such calculations[15] (see later). On the contrary, both the van der Waals and the nonpolar contribution to the solvation free energy contribute favorably to the binding. Therefore, the favorable binding free energy for the complexation comes predominantly from the nonpolar terms $\langle E_{vdW} \rangle$ and $\langle \Delta G_{np,desolv} \rangle$, while the polar interactions provide most of all the directional constraint for the complexation, that is the relative positions of the molecules.[4,10,36,37] Table I also gives the statistical uncertainty for the averages along the MD simulation trajectory. We see that the main source for uncertainties are the electrostatic terms, $\langle E_{elec} \rangle$ and $\langle \Delta G_{elec,desolv} \rangle$, which are most sensitive to the atomic coordinates, in agreement with previous studies.[10] Figure 5 shows the instantaneous $\Delta G^0_{bind} + \Delta G_{desolv}$ value along the 1 ns of the MD simulation. As can be seen, no drift in this value can be observed during the MD simulation. The slope of the regression line between time and $\Delta G^0_{bind} + \Delta G_{desolv}$ is only 2.10^{-3} kcal/(mol ps).

The entropic contribution of the solvent is included in the free energy of solvation estimated from the the GB-MV2 continuum solvation model. In addition, it is necessary to estimate and add the entropy change of

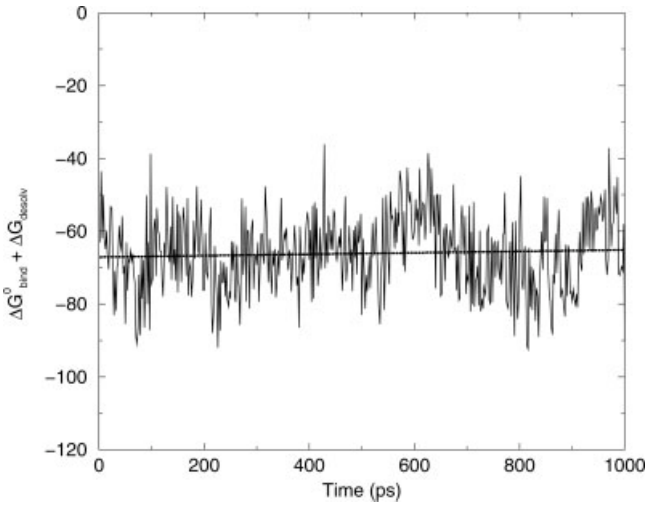

Fig. 5. $\Delta G^0_{bind} + \Delta G_{desolv}$ for 500 snapshots extracted at 2 ps intervals from MD simulation. Energies are in kcalories per mole. The regression line is shown in dotted line.

the solute molecules upon binding to determine the absolute binding free energy of complexation. The entropy contributions of the solute have been calculated from the standard equations of statistical mechanics (see Methods). $\langle -T\ S_{vib} \rangle$, $\langle -T\ S_{trans} \rangle$, and $\langle -T\ S_{rot} \rangle$ make a +44.61 kcal/mol unfavorable contribution to the binding free energy, leading to a total binding free energy of about -21 ± 12 kcal/mol according to this approach. Thus, the approach overestimates largely the experimental binding free energy of −6.2 kcal/mol.[16] As previously noticed,[10] the calculation of an absolute binding free energy with this approach is obtained through differences of large numbers leading to a small final value. Therefore, relatively small errors in the first ones can lead to large errors in the absolute binding free energy. In this respect, and taking account of the approximations of the model, the use of a single trajectory of the complex to estimate all energy terms, the neglect of the reorganization energy upon binding, the approximations in the calculation of the entropy terms, the absence of ionic strength, and the uncertainties in the calculated values, it is not surprising to find a noticeable error in the absolute binding free energy. Moreover, this error is not problematic since we are mostly interested in the relative contributions of the residues to the binding.

As can be seen in Table II, about 60% of the calculated binding free energy, including the vibrational entropy, is attributed to the backbone. Although the interactions involving backbone atoms might be specific in some cases, they are not responsible for the main diversity and specificity associated with TCR-p-MHC recognition. Therefore, the important role played by the backbone atoms in the binding may in part be responsible for TCR cross reactivity, that is the possible recognition of several different p-MHC counterparts by a given TCR. It could be of interest to perform BFED studies for other TCR-p-MHC complexes to determine whether this finding can be generalized. An AS does not address the role of interactions between the backbone of the TCR and p-MHC, but the influence of the side chains only. It is therefore the calculated contribution of the side chains to the binding free energy that needs to be compared to the experimental log(SIYR/K^b reactivity) values.

Contribution to the Binding Free Energy From TCR Residues

In the experimental AS study,[17] the authors expressed the residue contribution to the binding as the log of the reactivity relative to the wild type scTCR, log(SIYR/K^b reactivity). This reactivity was examined in a capture ELISA experiment: the scTCR was added to immobilized F23.1 mAb, and then various concentrations of the SIYR/K^b-streptavidin-horseadish peroxidase (SAv-HRP) tetramers were added while measuring the absorbance variation. Since this binding assay involved multivalent interactions, the intrinsic binding constant for each scTCR mutant was not possible to obtain. Therefore, the authors used the double reciprocal plots absorbance^{-1} versus concentration^{-1} to obtain a reliable measure of binding. Experimental calculations of binding contributions were made relative to the wild type scTCR. The range of log(SIYR/K^b reactivity) values for the different mutants was found to be very similar to the range of $\Delta\Delta G_{bind}$ measured in a 2C TCR/QL9/L^d AS study.[38]

Computational alanine scanning

Results for the CAS are given in Tables III and IV for the TCR Vα and Vβ residues, respectively. A positive $\Delta\Delta G_{bind}$ value corresponds to a residue for which the wild type side chain is more favorable to the binding than an alanine side chain. Owing to the inaccessible amount of computational time required to include it, the vibrational entropy term was neglected in the calculation of $\Delta\Delta G_{bind}$ values (see Methods).

For the 17 alanine residues present in the wild type system, the calculated $\Delta\Delta G_{bind}$ value ranges from −0.11 to +0.06 kcal/mol. These values are very close to the expected value of 0 (the alanine scan of an alanine residue is a null operation), and show that the uncertainty on the calculated values is in the range of ±0.1 kcal/mol.

The $\Delta\Delta G_{bind}$ values were calculated for all the residues of the TCR molecule, whatever their distance from the contact surface. However, as expected, all the residues side chains of TCR making a significant contribution to the binding energy belong to CDR1, CDR2, HV4, and CDR3 loops of Vα and Vβ, with the exception of TCRα Phe33, and Gln 1. In the relation between the calculated $\Delta\Delta G_{bind}$ values and the experimental log(SIYR/K^b reactivity), we consider TCRα Ser27, Lys68, Ser93, Phe100 and TCRβ Asn30, Asn31, Tyr48 and Glu56 as outliers. The same outliers were found in the relation between the experimental log(SIYR/K^b reactivity) and the ΔG^{sc}_{bind}

TABLE II. Contributions From the Side Chains and the Backbone Atoms in $\langle \Delta G^0_{\mathrm{bind}} \rangle + \langle \Delta G_{\mathrm{desolv}} \rangle - \langle T\, S_{\mathrm{vib}} \rangle$

	$\langle E_{\mathrm{vdW}} \rangle$	$\langle E_{\mathrm{elec}} \rangle$	$\langle \Delta G_{\mathrm{elec,desolv}} \rangle$	$\langle \Delta G_{\mathrm{np,desolv}} \rangle$	$-\langle T\, S_{\mathrm{vib}} \rangle$	$\langle \Delta G^0_{\mathrm{bind}} \rangle + \langle \Delta G_{\mathrm{desolv}} \rangle - \langle T\, S_{\mathrm{vib}} \rangle$
Backbone	−44.79	−135.62	150.45	−3.76	2.34	−31.38
Side chains	−60.89	−458.12	501.10	−14.61	13.62	−18.90

Energies are in kcal/mol.

TABLE III. Results of the CAS for the 2C TCRα Residues

Residue	Loop	$\langle \Delta E_{\mathrm{vdW}} \rangle$	$\langle \Delta E_{\mathrm{elec}} \rangle$	$\langle \Delta\Delta G_{\mathrm{elec,desolv}} \rangle$	$\langle \Delta\Delta G_{\mathrm{np,desolv}} \rangle$	$\langle \Delta\Delta G_{\mathrm{bind}} \rangle$
Ser93*	CDR3	−2.04	18.39	−4.74	−0.03	11.57
Tyr31	CDR1	6.53	5.20	−3.48	0.59	8.83
Lys48	CDR2	−1.41	130.47	−122.87	0.34	6.53
Lys68*	HV4	−1.79	112.07	−105.31	0.62	5.59
Phe100*	CDR3	5.64	0.00	−1.58	0.45	4.52
Tyr50	CDR2	5.94	7.85	−10.36	0.15	3.58
Ser27*	CDR1	−0.07	9.91	−6.39	0.02	3.46
Ser51	CDR2	−0.55	2.86	0.78	0.08	3.17
Tyr26	CDR1	2.03	3.45	−3.43	0.03	2.08
Phe33	—	0.41	−0.27	1.03	0.00	1.17
Tyr49	CDR2	0.37	−0.14	0.38	0.00	0.62
Gln1	—	0.24	1.48	−1.26	0.00	0.46
Thr29	CDR1	0.46	2.61	−4.48	0.12	−1.28
Asp53	CDR2	0.13	−57.00	55.12	0.02	−1.72

Only residues making a significant favorable or unfavorable contribution are given ($|\Delta\Delta G_{\mathrm{bind}}| \geq 0.4$ kcal/mol). Energies are in kcal/mol. Outliers are marked with an asterisk.

TABLE IV. Same as Table III, but for TCRβ Residues

Residue	Loop	$\langle \Delta E_{\mathrm{vdW}} \rangle$	$\langle \Delta E_{\mathrm{elec}} \rangle$	$\langle \Delta\Delta G_{\mathrm{elec,desolv}} \rangle$	$\langle \Delta\Delta G_{\mathrm{np,desolv}} \rangle$	$\langle \Delta\Delta G_{\mathrm{bind}} \rangle$
Asn30*	CDR1	1.86	27.98	−17.03	0.28	13.09
Tyr50	CDR2	6.30	1.08	−2.01	0.23	5.61
Tyr107	CDR3	1.67	4.33	−4.44	0.23	1.80
His29	CDR1	3.71	1.11	−3.86	0.08	1.04
Phe75	HV4	0.04	−0.02	0.68	0.00	0.71
Ser94	—	0.04	−0.82	0.35	0.00	−0.42
Asn31*	CDR1	1.33	1.20	−2.97	0.00	−0.44
Glu73	—	0.03	−48.32	47.82	0.00	−0.47
Met32	—	0.06	−0.06	−0.49	0.00	−0.49
Glu1	—	0.04	−51.19	50.47	0.01	−0.67
Gln72	HV4	0.74	−0.56	−1.49	0.12	−1.18
Gln25	—	0.14	0.09	−2.56	0.00	−2.32
Glu56*	CDR2	0.49	−37.39	32.28	0.15	−4.47

values obtained by BFED for the wild type system. They are defined as having $|\log(\mathrm{SIYR/K}^b \text{ reactivity}) + \Delta G^{\mathrm{sc}}_{\mathrm{bind}}| > 1.25$. A detailed discussion of these outliers is given below. Figure 6 shows the relation between $\log(\mathrm{SIYR/K}^b$ reactivity) and $\Delta\Delta G_{\mathrm{bind}}$, excluding the outliers. We see that there is a satisfying correlation between these two quantities; the correlation coefficient is 0.722:

$$\Delta\Delta G_{\mathrm{bind}} = -0.36 + 2.66 \times \log(\mathrm{SIY\,R/K}^b \mathrm{reactivity}) \quad (16)$$

A comparable correlation was also found by P. Kollman and coworkers in the study of the AS of the 1:1 human growth factor hormone-receptor.[2] The authors also found several outliers in this relationship.

To get further insight into the contributions of the different energy terms on the binding free energy differences, the change in the van der Waals and electrostatic interaction energies, as well as the electrostatic and nonpolar desolvation free energies upon mutation are reported in Tables III and IV. With the exception of Vα Ser51 and Vβ Met32, the change in the electrostatic desolvation free energy is strongly anticorrelated to the change in the electrostatic energy, as already noticed in other studies.[2] A gain in electrostatic interaction energy upon mutation is paid by a change in electrostatic desolvation energy of the same amplitude but opposite sign (see Vα Asp53, for example), and vice versa (see Vα Tyr26, for example). This suggests a good balance between the molecular mechanics terms and the solvation free energy terms.[2] We can also

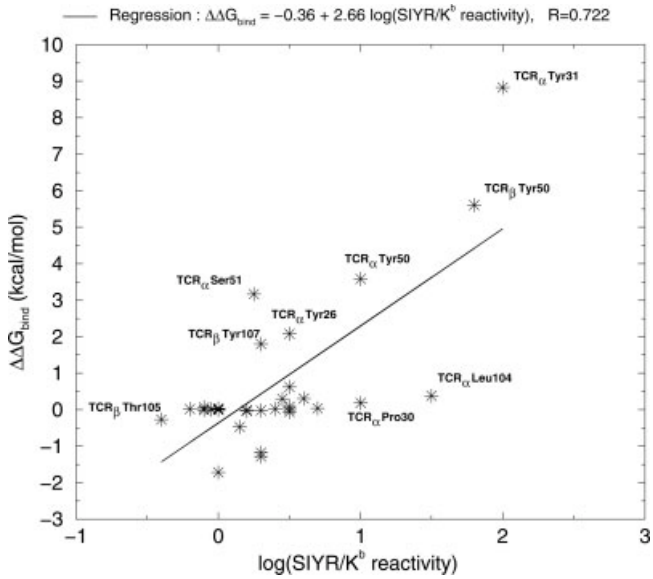

Fig. 6. Regression between the experimental log(SIYR/K^b reactivity) and the calculated $\Delta\Delta G_{bind}$ values obtained by CAS. Energies are in kcalories per mole etc. The vibrational entropy term is not included (see Methods and Discussion). The correlation is shown as a thick black line. Noticeable residues are labeled.

notice that in the case of the outliers listed earlier, the balance between the change in the electrostatic interaction energy and the electrostatic desolvation free energy upon complexation is not achieved.

We also tried to take account of the vibrational entropy in the CAS approach. The exact approach would require the computationally very expensive calculation of the $-T\Delta S_{vib}$ values for all the mutants. However, to keep computing time manageable, we used the LDVE approach. Assuming that the vibrational entropy contribution of an alanine side chain to the binding is constant relative to its position in the sequence, we simply removed from the $\Delta\Delta G_{bind}$ value relative to a given mutation the vibrational entropy contribution of the corresponding wild type side chain calculated from the BFED scheme. This vibrational entropy correction gives satisfying results: after removing the outliers described earlier, the regression coefficient between log(SIYR/K^b reactivity) and $\Delta\Delta G_{bind}$ becomes 0.748, with $\Delta\Delta G_{bind} = -0.37 + 2.53$ log(SIYR/K^b reactivity).

Binding free energy decomposition for the wild type system

As explained earlier, the contributions of the rotational and translational entropies on a residue basis are expected to be negligible or constant. Therefore, we will focus on $\Delta G^0_{bind} + \Delta G_{desolv} - T\Delta S_{vib}$, and decompose it into contributions arising from side chains, ΔG^{sc}_{bind}, and backbones, ΔG^{bb}_{bind}. As for the CAS, the fast LDVE method was used to compute vibrational entropies (see Assessment of the Vibrational Entropy Decomposition).

Tables V–VIII give the side chains contributions, ΔG^{sc}_{bind}, of residues belonging to TCR Vα, TCR Vβ, MHC, and the peptide.

With the only limitation that the vibrational entropy contributions were only estimated for the residues at the contact surface, the ΔG^{sc}_{bind} values were calculated for all the residues of the system, whatever their distance from the contact surface. However, as in the case of the CAS, nearly all the residues side chains of TCR making a significant contribution to the binding energy belong to CDR1, CDR2, HV4, and CDR3 loops. The list of the most important residues for the binding is very similar to that obtained from the CAS (see Tables III and IV).

Figure 7 shows the relation between the experimental and theoretical values, excluding the outliers discussed later: Vα Ser27, Lys68, Ser93, Phe100 and Vβ Asn30, Asn31, Tyr48 and Glu56. In this figure, the ΔG^{sc}_{bind} values were calculated including (grey stars) or excluding (black diamonds) the contribution of the vibrational entropy. Without the entropy term the relation between log(SIYR/K^b reactivity) and ΔG^{sc}_{bind} is satisfying. The correlation coefficient is −0.672:

$$\Delta G^{sc}_{bind} = 0.10 - 1.13 \times \log(\text{SIYR/K}^b\text{reactivity}) \quad (17)$$

The slope is negative since a more favorable (i.e. more negative) binding free energy contribution corresponds to a larger loss in activity upon mutation, characterized by a larger (i.e. more positive) value of log(SIYR/K^b reactivity). Similar correlations were found by Gohlke et al. in the study of the AS of the H-Ras/C-Raf1 and H-Ras/RalGDS systems using the MM-GBSA approach.[4] The authors also found several outliers in these relationships.

Including the vibrational entropy term increases significantly the correlation. The coefficient correlation becomes −0.716.

$$\Delta G^{sc}_{bind} = 0.10 - 0.99 \times \log(\text{SIYR/K}^b\text{reactivity}). \quad (18)$$

This enhancement of the correlation between log(SIYR/K^b reactivity) and ΔG^{sc}_{bind} is due to the fact that several residues exhibit significant favorable (Vα Leu104 and Vβ Asn28) or unfavorable (Vα Tyr31, Tyr50 and Ser51, and Vβ Tyr50 and Tyr107) vibrational entropy contributions, which bring the ΔG^{sc}_{bind} value closer to the ideal regression when the vibrational entropy contribution is included (see Fig. 7). With the exception of Vβ Tyr107, all the residues exhibiting an unfavorable vibrational entropy contribution in this list are in contact with the p-MHC system, that is the accessible vibrations are limited upon complexation. On the contrary, all the residues exhibiting a favorable vibrational entropy contribution in this list are not in contact with the p-MHC system.

Figure 8 shows that the log(SIYR/K^b reactivity) versus ΔG^{sc}_{bind} regression for all residues except the outliers has globally converged after 400–500 ps. Moreover, as can be seen in Figure 9, the individual ΔG^{sc}_{bind} values have converged for most of the important residues at that time. The

TABLE V. Contributions of the 2C TCRα Side Chains to $\langle \Delta G^{0}_{bind} \rangle + \langle \Delta G_{desolv} \rangle - \langle T\, S_{vib} \rangle$, From the BFED Approach

Residue	Loop	$\langle E^{sc}_{vdW} \rangle$	$\langle E^{sc}_{elec} \rangle$	$\langle \Delta G^{sc}_{elec,desolv} \rangle$	$\langle \Delta G^{sc}_{np,desolv} \rangle$	$-\langle T\, S^{sc}_{vib} \rangle$	$\langle \Delta G^{sc}_{bind} \rangle$
Ser93*	CDR3	0.63	−9.06	3.04	−0.01	0.17	−5.23
Phe100*	CDR3	−3.59	−0.74	1.36	−0.72	0.65	−3.04
Tyr31	CDR1	−3.46	−2.37	3.31	−0.61	0.75	−2.38
Tyr50	CDR2	−3.70	−4.02	5.63	−0.57	0.58	−2.08
Lys68*	HV4	0.87	−56.18	53.34	−0.34	0.59	−1.72
Ser27*	CDR1	−0.52	−5.14	4.32	−0.32	0.06	−1.60
Lys48	CDR2	0.63	−65.57	62.44	−0.26	1.25	−1.51
Tyr26	CDR1	−1.06	−1.46	1.79	−0.11	−0.11	−0.95
Ala28	CDR1	−0.73	0.49	−0.37	−0.33	0.13	−0.81
Ala101	CDR3	−0.40	−0.38	0.24	−0.08	0.02	−0.60
Leu104	CDR3	−0.10	−0.15	0.06	−0.00	−0.33	−0.52
Ser51	CDR2	−0.57	−1.58	1.77	−0.46	0.47	−0.37
Gln1	—	−0.16	−0.44	0.28	−0.00	0.00	−0.32
Ala103	CDR3	−0.04	−0.17	0.11	0.00	−0.19	−0.29
Pro30	CDR1	−0.10	−3.20	3.14	0.00	−0.12	−0.28
Phe66	—	−0.08	−0.10	0.18	0.00	−0.26	−0.26
Tyr49	CDR2	−0.21	0.15	−0.05	−0.00	−0.13	−0.24
Ser102	CDR3	−0.96	−4.31	5.08	−0.17	0.13	−0.23
Thr29	CDR1	−0.37	−1.55	3.03	−0.18	−0.12	0.81
Asp53	CDR2	−0.10	28.56	−27.32	−0.01	0.02	1.15

Only side chains making a significant favorable or unfavorable contribution are given ($|\Delta G^{sc}_{bind}| \geq 0.2$ kcal/mol). Energies are in kcal/mol. Outliers are marked with an asterisk.

TABLE VI. Same as Table V but for TCRβ

Residue	Loop	$\langle E^{sc}_{vdW} \rangle$	$\langle E^{sc}_{elec} \rangle$	$\langle \Delta G^{sc}_{elec,desolv} \rangle$	$\langle \Delta G^{sc}_{np,desolv} \rangle$	$-\langle T\, S^{sc}_{vib} \rangle$	$\langle \Delta G^{sc}_{bind} \rangle$
Asn30*	CDR1	−1.63	−14.42	10.83	−0.67	0.99	−4.90
Tyr50	CDR2	−3.85	−0.79	2.52	−0.71	0.25	−2.58
Arg69	HV4	−0.10	−23.63	22.05	0.00	−0.09	−1.77
Thr105	CDR3	−0.73	0.97	−0.58	−0.22	−0.07	−0.63
Asn27	CDR1	−0.33	−0.20	−0.05	−0.02	0.07	−0.53
Tyr107	CDR3	−0.88	−2.26	2.08	−0.27	0.91	−0.42
Asn28	CDR1	−0.39	−0.05	0.36	−0.12	−0.19	−0.39
Ala2	—	−0.25	−0.11	0.08	−0.08	0.00	−0.36
Ala52	CDR2	−0.14	−0.06	0.12	−0.05	−0.10	−0.23
Leu106	CDR3	−0.10	0.14	−0.12	−0.00	−0.14	−0.22
Phe75	HV4	−0.03	−0.03	0.04	0.00	−0.19	−0.21
Ser54	CDR2	−0.12	0.24	0.04	−0.05	0.10	0.21
Ser94	—	−0.05	0.38	−0.09	0.00	0.03	0.27
Gln72	HV4	−0.42	0.17	0.67	−0.08	−0.01	0.33
Glu73	HV4	−0.02	24.18	−23.87	0.00	0.05	0.34
Glu1	—	−0.08	25.64	−25.18	−0.02	0.00	0.36
Asn31*	CDR1	−0.81	−0.55	1.96	−0.07	0.04	0.57
His29	CDR1	−2.24	−0.64	4.01	−0.37	0.68	1.44
Glu56*	CDR2	−0.31	18.66	−14.54	−0.12	−0.26	3.43

convergence has not been achieved for Vα Ser27 and Ser102 and Vβ Glu56. Except Vα Ser102, these residues correspond to outliers in the relation between log(SIYR/K^b reactivity) and ΔG^{sc}_{bind}. A more detailed analysis and systematization of the outliers is given later.

Comparison between alanine scanning and binding free energy decomposition

Figure 10 shows the regression between the calculated $\Delta\Delta G_{bind}$ values obtained by CAS and the ΔG^{sc}_{bind} values obtained by BFED of the wild type system. To obtain a relevant comparison, the entropy term was excluded from the ΔG^{sc}_{bind} values. As can be seen, the correlation between the two theoretical approaches is very good. Thus, it appears that for the system under investigation, the BFED of the wild type system provides informations similar to that of the CAS. However, the BFED offers two advantages over CAS. First, it is much faster, since it requires only one calculation of an absolute binding free energy of complexation, while this term must be recalculated for all mutants in the CAS approach. Sec-

TABLE VII. Same as Table V but for MHC

Residue	$\langle E^{sc}_{vdW}\rangle$	$\langle E^{sc}_{elec}\rangle$	$\langle \Delta G^{sc}_{elec,desolv}\rangle$	$\langle \Delta G^{sc}_{np,desolv}\rangle$	$-\langle T\, S^{sc}_{vib}\rangle$	$\langle \Delta G^{sc}_{bind}\rangle$
Arg155	−4.67	46.05	−44.50	−0.62	0.54	−3.20
Thr163	0.12	−3.79	1.88	−0.26	0.18	−1.87
Val76	−1.53	0.28	0.18	−0.49	0.24	−1.32
Glu166	0.89	−80.18	77.89	−0.35	0.64	−1.11
Ala158	−0.80	0.32	−0.20	−0.30	0.16	−0.82
Ala150	−0.65	−0.19	0.33	−0.20	−0.01	−0.72
Lys66	−1.36	39.20	−38.32	−0.15	−0.09	−0.72
Trp167	−0.60	−0.24	0.55	−0.19	−0.07	−0.55
Lys146	−1.28	32.01	−31.48	−0.38	0.64	−0.49
Trp147	−0.25	−0.10	−0.03	0.00	−0.08	−0.46
Glu55	−0.03	−40.50	40.20	0.00	0.00	−0.33
Tyr159	−0.39	0.34	−0.28	−0.06	0.08	−0.31
Phe74	−0.11	−0.21	0.21	0.00	−0.19	−0.30
Arg169	−0.04	33.69	−33.35	−0.00	0.00	0.30
Lys68	−0.15	37.35	−36.80	−0.01	−0.09	0.30
Lys131	−0.02	33.12	−32.79	0.00	0.00	0.31
Lys89	−0.02	33.66	−33.31	0.00	0.00	0.33
Arg44	−0.01	35.26	−34.91	0.00	0.00	0.34
Arg170	−0.09	43.57	−43.07	−0.00	0.00	0.41
Glu61	−0.25	−53.99	54.52	−0.05	0.18	0.41
Gln72	−1.44	0.05	1.81	−0.41	0.46	0.47
Glu161	−0.15	−27.70	28.40	−0.04	−0.03	0.48
Glu58	−0.28	−49.06	49.88	−0.10	0.05	0.49
Glu154	0.14	−88.94	88.52	−0.42	1.31	0.61
Ser73	−0.66	−1.36	2.81	−0.16	0.01	0.64
Arg79	−0.48	40.70	−38.87	−0.13	0.10	1.32
Gln149	−2.02	0.06	2.80	−0.57	1.09	1.36
Glu24	−0.03	−36.73	38.19	0.00	0.00	1.43
Arg62	−2.14	46.97	−43.45	−0.48	0.66	1.56
Glu152	−0.45	−46.82	49.02	−0.01	−0.13	1.61
Asp77	−0.36	−49.67	58.70	−0.11	0.02	8.58

Only side chains with $|\Delta G^{sc}_{bind}| \geq 0.3$ kcal/mol) are given.

TABLE VIII. Same as Table V but for the Peptide

Residue	$\langle E^{sc}_{vdW}\rangle$	$\langle E^{sc}_{elec}\rangle$	$\langle \Delta G^{sc}_{elec,desolv}\rangle$	$\langle \Delta G^{sc}_{np,desolv}\rangle$	$-\langle T\, S^{sc}_{vib}\rangle$	$\langle \Delta G^{sc}_{bind}\rangle$
Tyr6	−4.46	−1.57	3.95	−0.63	0.95	−1.76
Tyr3	−0.63	−0.07	0.33	−0.02	−0.02	−0.41
Ser1	−0.14	−0.42	0.28	−0.00	0.02	−0.26
Ile2	−0.06	−0.06	0.12	0.00	−0.19	−0.19
Tyr5	−0.14	0.03	0.02	0.00	−0.09	−0.18
Gly7	0.00	0.00	0.00	0.00	0.00	0.00
Leu8	−0.10	−0.14	0.25	0.00	0.04	0.05
Arg4	−3.81	18.93	−12.00	−1.03	2.50	4.59

All residues are given.

ond, as described in this paper, it is possible to take account of the entropy term in a fast and straightforward manner in the case of the BFED approach. This additional term has been shown in this study to increase significantly the correlation between the experimental and the theoretical results. It is interesting to note that adding the entropy contributions calculated in the BFED scheme to the binding free energy differences calculated using the CAS approach, thus defining a new hybrid approach, enhances the relation between the experimental and calculated activities.

Analysis of some key interactions

In the following, we analyze some TCR residues that are successfully predicted to be important for the binding to p-MHC, not only in terms of interactions with the p-MHC complex but also with the TCR itself. In some cases, possible side chain replacements are proposed in the context of TCR protein engineering. Discussions are based on Tables V and VI, that is on the results of the BFED.

*TCR*α *Tyr31.* This side chain makes van der Waals contacts with the MHC Arg155, the peptide Arg4, and

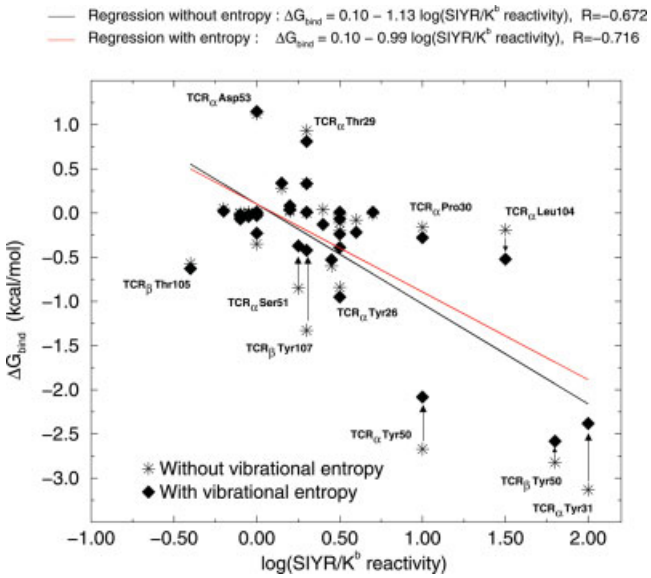

Fig. 7. Regression between the experimental log(SIYR/K^b reactivity) and the calculated ΔG^{sc}_{bind} values. Energies are in kcalories per mole. ΔG^{sc}_{bind} is reported excluding (gray stars) and including (black diamonds) the vibrational entropy. The corresponding correlations are shown. Gray arrows show the effect of including the most important vibrational entropy terms. For clarity, the outliers are not represented. Noticeable residues are labeled. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

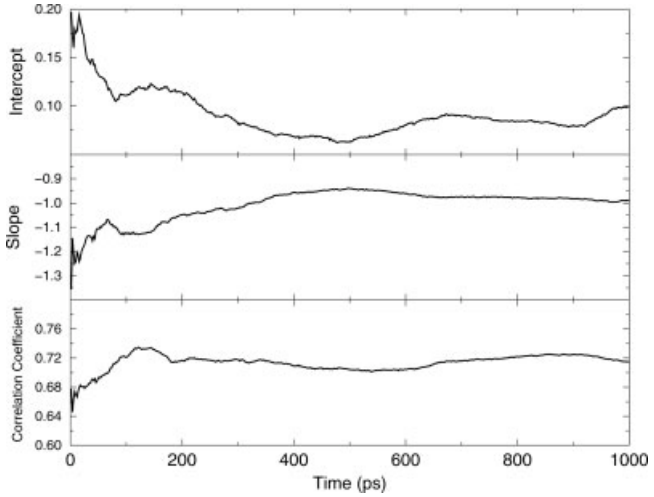

Fig. 8. Regression between the experimental log(SIYR/K^b reactivity) and the calculated ΔG^{sc}_{bind} values. The slope, intercept and regression coefficient values are plotted as a function of the MD simulation time t up to which the averages of the energy terms are calculated. ΔG^{sc}_{bind} values are calculated using the BFED approach, excluding the outliers described in the text.

with residues Phe33, Tyr50, and Ser93 of TCRα and Gly98 of TCRβ [see Fig. 11(A)]. The first two are responsible for the −3.5 kcal/mol favorable contribution to the binding made by the van der Waals interactions for this side chain (See Table V). Also, TCRα Tyr31 exchanges a hydrogen bond with the MHC Arg155 residue that explains the −2.4 kcal/mol favorable contribution provided by the electrostatic interactions. However, the latter is more than compensated by the unfavorable electrostatic desolvation free energy term (+3.3 kcal/mol). The BFED approach thus suggests that it is the aromatic part of TCRα Tyr31 that is essentially responsible for the favorable contribution played by this residue in the binding. On the contrary, the OH function, which makes a hydrogen bond with MHC Arg155, does not provide a favorable contribution in itself. However, removing it would be unfavorable, since it would increase the desolvation energy of MHC Arg155 and thus decrease the binding affinity.

TCRα Tyr50. This side chain makes van der Waals contacts with the Arg155 and Ala158 of MHC, as well as with residues Phe33, Tyr31 and Lys48 of TCRα and Gly98 and Thr105 of TCRβ [see Fig. 11(A)]. The OH group of this side chain does not make any hydrogen bond either in the X-ray structure or during the MD simulation. The absence of interactions made by the OH group is in part responsible for the +1.6 kcal/mol unfavorable total electrostatic contribution to the binding made by this side chain (See Table V). Thus, replacing it by a Phe side chain, that is removing the OH group, could be a valuable experiment in a protein engineering study of TCR and can be expected to increase its affinity for the p-MHC complex by decreasing the unfavorable electrostatic contribution. This mutation should keep the numerous van der Waals interactions made by the phenyl ring that are favorable to both the binding to p-MHC and the structural stability of the TCR molecule.

TCRα Tyr26. This side chain makes van der Waals contacts with residues Gln1, Ala28, Pro30, Gly99, and Phe100 of TCRα, as well as with the Arg62 residue of MHC [see Fig. 11(B)]. No hydrogen bond is made by the OH group in the X-ray structure. It only makes a transient hydrogen bond with the TCRα Gly99 backbone during the MD simulation. As for TCRα Tyr50, it could be worthwhile to examine the replacement of this side chain by a phenylalanine.

TCRβ Tyr50. This side chain makes van der Waals contacts with the Ala101 and Ser102 residues of TCRα, with the Asn31 and Tyr48 residues of TCRβ, and with the MHC Ser73, Gln72, and Gly69 [see Fig. 11(C)]. The latter explain the large favorable van der Waals term, −3.8 kcal/mol, in the contribution of this side chain to the binding (See Table V). This side chain does not exchange any hydrogen bond in the X-ray structure, and only transient hydrogen bonds with TCRβ Tyr48 and TCRβ Glu56 side chains are observed during the MD simulation. The total electrostatic contribution to the binding for this side chain is +1.7 kcal/mol. As for TCRα Tyr50 and Tyr26, the BFED approach thus suggests that replacing TCRβ Tyr50 by a Phe side chain could increase the affinity of the TCR for the p-MHC complex by decreasing the unfavorable electrostatic contribution. The impact on the TCR stability should however be investigated.

TCRα Lys48. TCRα Lys48 was not studied experimentally, but was found theoretically to make a strong favor-

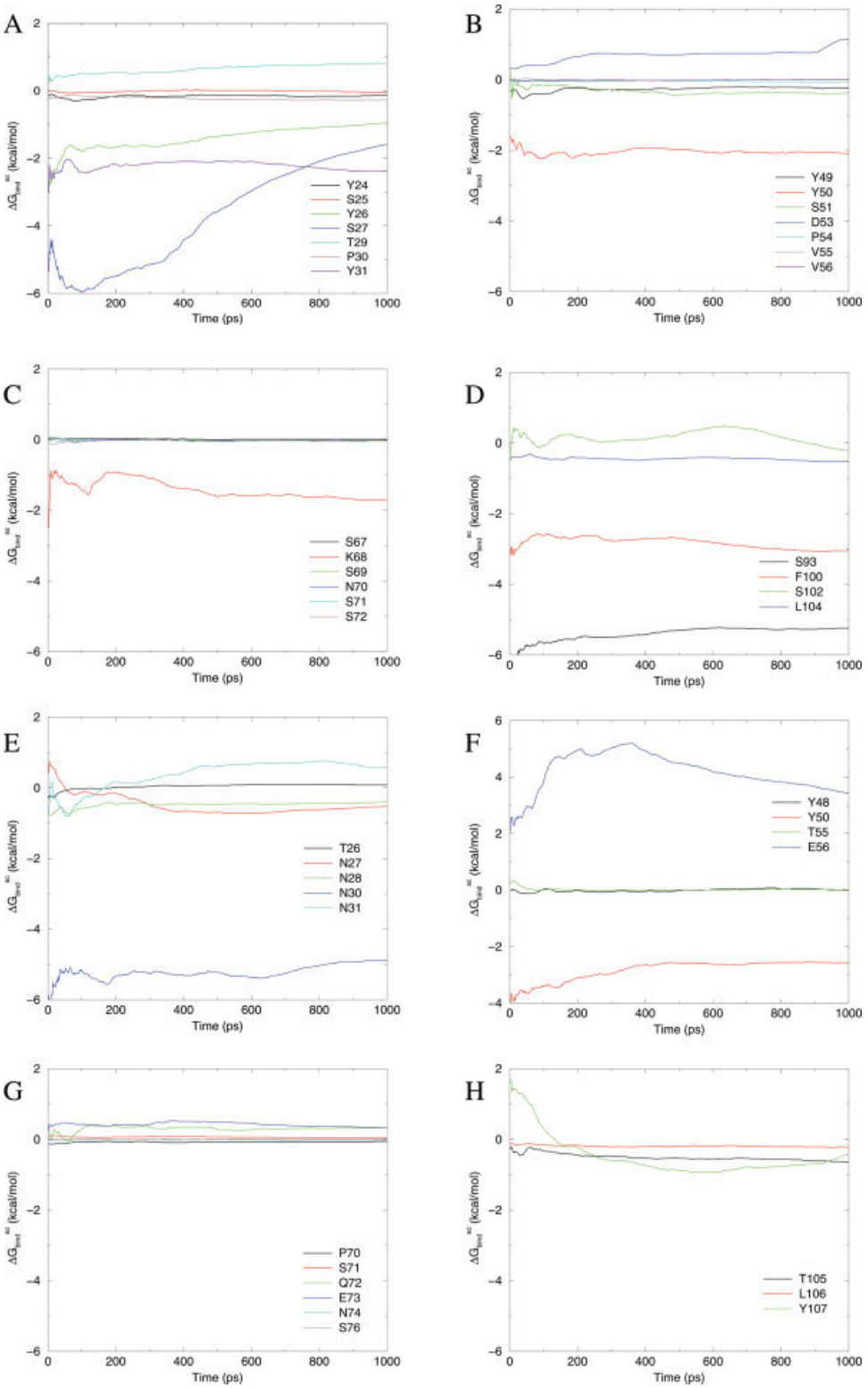

Fig. 9. Averaged ΔG^{sc}_{bind} up to time t for the important residues of Vα CDR1 (**A**), CDR2 (**B**), HV4 (**C**), and CDR3 (**D**) and of Vβ CDR1 (**E**), CDR2 (**F**), HV4 (**G**), and CDR3 (**H**), which were experimentally studied by AS.[18]

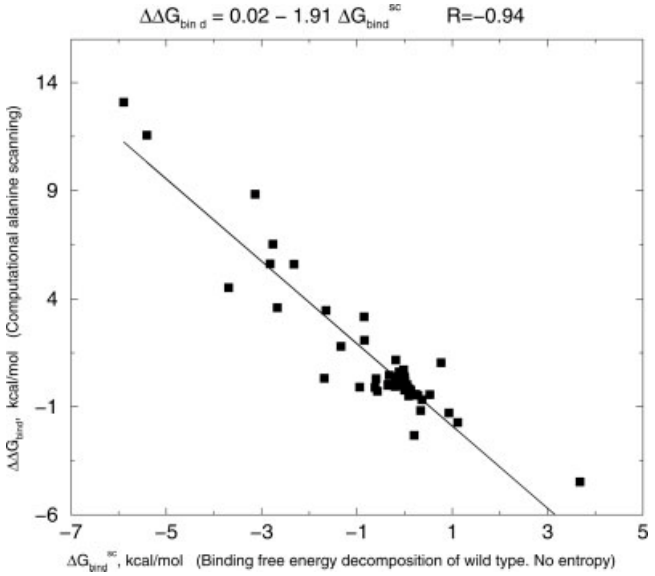

Fig. 10. Regression between the calculated $\Delta\Delta G_{bind}$ values obtained by CAS and the ΔG^{sc}_{bind} values obtained by BFED for the wild type system. For both terms the vibrational entropy is not included.

able contribution to the binding of TCR to p-MHC and is in contact with the MHC molecule. Therefore, it could be of interest for further experiment. This side chain makes van der Waals contacts with residues Phe33, Leu45, and Tyr50 of TCRα and Thr105 of TCRβ [see Fig. 11(D)]. Also, it exchanges a hydrogen bond with the MHC Glu154, which is responsible for the favorable −3.1 kcal/mol contribution of the electrostatic terms to the binding. According to the BFED approach, it could be interesting to investigate experimentally the possible replacement of the side chain with an arginine, since the latter could keep the van der Waals contacts listed earlier and the ionic interaction with the MHC Glu154.

Analysis of the outliers

Several residues were found to be outliers in the relation between log(SIYR/K^b reactivity) and ΔG^{sc}_{bind}. They are described one by one in the Appendix. In a real case of protein engineering, where a computational BFED is performed to supply unavailable experimental AS data, it is of great importance to be able to point out a priori the possible outliers. The discussion of the outliers in the Appendix shows that the residue side chains for which the calculated contribution to the binding is not in agreement with the experimental results are characterized by one or several of the following points. First, some outliers exhibit an unconverged calculated contribution to the binding free energy (TCRα Ser27 and TCRβ Glu56). Some outliers are characterized by the fact that their conformation in the MD simulation differs from that in the crystal or that their contact pattern is changing during the course of the simulation (TCRα Ser27 and Lys68, and TCRβ Asn30 and Asn31). Other outliers are found theoretically to make important contributions to the binding, in agreement with the interactions they make with the peptide or the MHC molecule in the crystal structure and during the MD simulation, but in contradiction with the experimental AS (TCRα Ser93 and Phe100). In such cases, the reliability of the experimental data might be questioned. In the case of TCRα Phe100, our theoretical result is in qualitative agreement with another AS study of the same TCR complexed with QL9-L^d, in which it is expected to make similar contacts with the MHC molecule. For these residues, it could be worthwhile to make further experimental investigations. We can also notice from Tables V and VI that a calculated absolute contribution to the binding free energy larger than the $\Delta\Delta G_{bind}$ amplitude generally observed in experimental AS studies (about 2–3 kcal/mol) may be used to point out possible false positives (TCRα Ser93 and Phe100, and TCRβ Asn30 and Glu56). Similarly, the fact that the electrostatic interaction energy, E^{sc}_{elec}, is not compensating the electrostatic desolvation free energy, $\Delta G^{sc}_{elec,desolv}$, (i.e. if the absolute value of the sum is larger than a threshold of about 10 kcal/mol) may be used as an indicator of a potential outlier. Finally, the present approach is not able to point out residues that play a role in the binding through the stabilization of the active conformation of one partner. This may lead to false negatives, like TCRβ Tyr48.

MHC Residues Contribution to the Binding Free Energy

Table IX gives the known experimental mutation effect of MHC residues on the activity of K^b restricted or alloreactive T cell clones.[39] Among these mutations, five correspond to residues interacting directly with TCR (Glu58, Arg79, Ala158, Glu166, and Trp167). Three of them (Ala158, Glu166, and Trp167) were found experimentally to affect the T-cell activity. For these three residues, the present MM-GBSA study estimates that the corresponding side chain makes a significant favorable contribution to the TCR-p-MHC binding. The two other residues were experimentally not found to affect the T-cell activity, in agreement with the present theoretical approach, which predicts an unfavorable contribution to the binding free energy. Again, the present theoretical approach is not expected to reproduce the mutational data of the MHC residues remote from the TCR since the effect of mutations possibly acting through reorganization of the system is not accessible. Altogether, there is a good agreement between the experimental and theoretical results concerning the MHC. The present approach has also pointed out the possible importance of residues Lys66, Val76, Ala150, Arg155, and Thr163. These residues are in close contact with the TCR and are worth a further experimental investigation.

Peptide Residues Contribution to the Binding Free Energy

As can be seen in Table VIII, the MHC anchoring residues of the peptide, that is Ile2, Tyr5, and Leu8, that

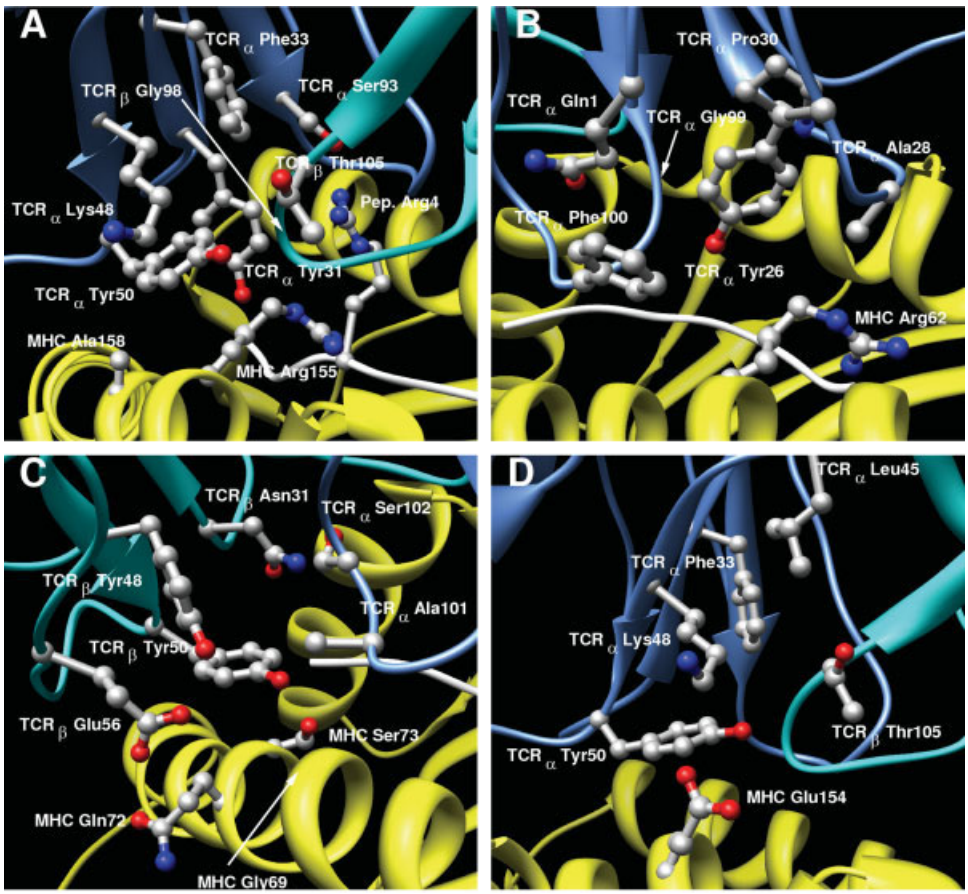

Fig. 11. Interactions between TCRα Tyr31 (**A**), Tyr50 (A), Tyr26 (**B**), TCRβ Tyr50 (**C**) and TCRα Lys48 (**D**), and their respective neighboring residues. The ribbon of TCRα (blue), TCRβ (cyan), MHC (yellow), and peptide (gray) is shown. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

TABLE IX. Effect of K^b Mutations on Allo-reactive and Antigen-Specific CTL Clones, in Relation to ΔG^{sc}_{bind}, ΔG^{bb}_{bind}, and Contacts With TCR Residues According to Ref. 39

Residue	Impact	ΔG^{sc}_{bind}	ΔG^{bb}_{bind}	Contact with TCR
Gly56	No	0.00	−0.02	None
Glu58	No	+0.49	−0.10	vdW with sc of Vα Ser27
Arg75	Yes	+0.23	−0.59	None
Arg79	No	+1.32	−0.21	Possible H-bonds with bb of Vβ Ala52 and Gly53
Thr80	Yes	+0.04	−0.37	None
Leu82	Yes	+0.03	−0.03	None
Gly90	Yes	0.00	0.00	None
Met138	Yes	−0.01	0.00	None
Leu141	No	+0.01	+0.06	None
Ala158	Yes	−0.82	+0.04	vdW with sc of Vα Tyr50
Gly162	Yes	0.00	−0.89	None
Glu166	Yes	−1.11	−0.05	HB with sc of Vα Ser51
Trp167	Yes	−0.55	−0.07	vdW with bb and sc of Vα Ala28
Asn174	Yes	0.00	0.00	None

Energies are in kcal/mol.

interact with the MHC molecule but not with the TCR, are making little contribution to the TCR binding to p-MHC. On the contrary, the peptide residues in contact with the TCR molecule, that is Tyr6 and Arg4, are making the most important absolute contributions to the TCR-p-MHC binding free energy.

CONCLUSION

The study of the TCR-p-MHC system using the CAS and the BFED approaches has shown that the two theoretical methods provide very similar results. This is very interesting since the CAS is much more time consuming than the BFED approach. In the latter, the absolute binding free energy of the system is calculated once, and then decomposed on a per-residue or per-atom level. On the contrary, in the CAS approach, the absolute binding free energy must be recalculated for each mutant complex corresponding to a residue of interest.

We have applied the two methods to the study of the 2C TCR/SIYR/H-2K^b complex. Our results suggest that the correlation between the calculated contribution of a given side chain to the binding free energy of the system and the corresponding experimental value is higher when the vibrational entropy term is included. In the present study, we have introduced a new ad hoc approach, called LDVE, to decompose approximately and rapidly the vibrational entropy term into atomic contributions, allowing the calculation of the entropic contribution to the binding of interesting side chains in a straightforward manner. This method is based on the idea that the most important contributions to the vibrational entropy of a molecule originate from residues that contribute most to the vibrational amplitude of the normal modes. This approach has been found to enhance drastically the relation between the experimental log(SIYR/K^b reactivity) values and the ΔG^{sc}_{bind} values obtained by BFED for the wild type system, in the case of the 2C TCR/SIYR/H-2K^b complex.

Owing to the approximations in the model, the MM-GBSA approach does not work very well at reproducing the absolute value of the binding free energy. Indeed, the absolute binding free energy calculated using this approach, that is -21 kcal/mol, overestimates the -6.2 kcal/mol experimental value for the 2C TCR/SIYR/H-2K^{b}complex. However, the interest of the MM-GBSA approach for this study resides in its ability to provide the contribution of each residue to the binding process. Indeed, the two deconvolution approaches, that is the CAS and the BFED, have both provided a reasonable correlation between the experimental and calculated influence of the studied side chains on the binding between the TCR and the p-MHC molecules. This confirms their efficiency to evaluate the contribution of groups of atoms to an overall binding free energy. It is important to note that the limited quality of these correlations may be explained partly by the small number of residues that were studied and found important experimentally. Thus the correlation between theoretical and experimental results is based on a rather limited number of points.

According to the MM-GBSA approach, we have found that the backbone of this TCR-p-MHC complex is responsible for $\sim$60% of the calculated binding free energy. This important contribution may be seen as the source for the basal amount of binding free energy needed for T-cell activation, the rest being brought by sequence-specific interactions.

Altogether, these results suggest that the BFED for the TCR-p-MHC system, including entropic terms, provides a detailed and reliable enough description of the interactions between the molecules in the complex at an atomistic level, and may further be used as an *in silico* investigation tool in TCR protein engineering, for example, in the optimization of specific anti-tumor clones. By providing a list of important or detrimental residues for the binding, and indications regarding the origin of their favorable or unfavorable role, the BFED approach not only helps selecting residues that are worth investigating, but also suggests some possible mutations.

ACKNOWLEDGMENTS

We thank the Cluster versus Cancer Project of the Lausanne University and the VITAL-IT project of the Swiss Institute of Bioinformatics, for providing the computational resources. This work was also supported by the National Center of Competence in Research (NCCR) Molecular Oncology, a research instrument of the Swiss National Science Foundation. We thank Ernest Feytmans (Swiss Institute of Bioinformatics) and Jean-Charles Cerottini (Ludwig Institute for Cancer Research) for helpful discussions.

REFERENCES

1. Massova I, Kollman PA. Computational alanine scanning to probe protein–protein interactions: a novel approach to evaluate binding free energies. J Am Chem Soc 1999;121:8133–8143.
2. Huo S, Massova I, Kollman PA. Computational alanine scanning of the 1:1 human growth hormone-receptor complex. J Comput Chem 2002;23:15–27.
3. 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;324:547–569.
4. Gohlke H, Kiel C, Case DA. Insights into protein–protein binding by binding free energy calculation and free energy decomposition for the Ras-Raf and Ras-RalGDS complexes. J Mol Biol 2003;330:891–913.
5. Brooks CL, III, Karplus M, Pettitt BM. Protein. A theoritical perspective of dynamics, structure, and thermodynamics (Advances in chemical physics, Vol. 71) New York: Wiley-Interscience; 1988.
6. Simonson T, Archontis G, Karplus M. Free energy simulations come of age: protein–ligand recognition. Acc Chem Res 2002; 35:430–437.
7. Srinivasan J, Cheatham TE, III, Cieplak P, Kollman PA, Case DA. Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate-DNA helices. J Am Chem Soc 1998;120: 9401–9409.
8. Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, Lee M, Lee T, Duan Y, Wang W, Donini O, Cieplak P, Srinivasan J, Case DA, Cheatham TE, III. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. Acc Chem Res 2000;33:889–897.
9. Wang W, Kollman PA. Free energy calculations on dimer stability of the HIV protease using molecular dynamics and a continuum solvent model. J Mol Biol 2000;303:567–582.
10. Zoete V, Meuwly M, Karplus M. Study of the insulin dimerization from binding free energy calculations and per-residue free energy decomposition. Proteins 2005;61:79–93.
11. Kuhn B, Kollman PA. A ligand that is predicted to bind better to avidin than biotin: insights from computational fluorine scanning. J Am Chem Soc 2000;122:3909–3916.

12. Lee TS, Kollman PA. Theoretical studies suggest a new antifolate as a more potent inhibitor of thymidylate synthase. J Am Chem Soc 2000;122:4385–4393.
13. Wang J, Morin P, Wang W, Kollman PA. Use of MM-PBSA in reproducing the binding free energies to HIV-1 RT of TIBO derivatives and predicting the binding mode to HIV-1 RT of efavirenz by docking and MM-PBSA. J Am Chem Soc 2001;123: 5221–5230.
14. Gilson MK, Honig BH. Energetics of charge–charge interactions in proteins. Proteins: Struct Func Genet 1988;3:32–52.
15. Gohlke H, Case DA. Converging free energy estimates: MM-PB(GB)SA studies on the protein–protein complex Ras-Raf. J Comput Chem 2004;25:238–250.
16. Degano M, Garcia KC, Apostolopoulos V, Rudolph MG, Teyton L, Wilson IA. A functional hot spot for antigen recognition in a superagonist TCR/MHC complex. Immunity 2000;12:251–261.
17. Koradi R, Billeter M, Wüthrich K. MOLMOL: a program for display and analysis of macromolecular structures. J Mol Graphics 1996;14:51–55.
18. Lee PU, Churchill HR, Daniels M, Jameson SC, Kranz DM. Role of 2C T cell receptor residues in the binding of self- and allo-major histocompatibility complexes. J Exp Med 2000;191: 1355–1364.
19. Bernstein FC, Koetzle TF, Williams GJ, Meyer EE, Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. The protein data bank: a computer-based archival file for macromolecular structures. J Mol Biol 1977;112:535–542.
20. Berman HM, Battistuz T, Bhat TN, Bluhm WF, Bourne PE, Burkhardt K, Feng Z, Gilliland GL, Iype L, Jain S, Fagan P, Marvin J, Padilla D, Ravichandran V, Schneider B, Thanki N, Weissig H, Westbrook JD, Zardecki C. The Protein Data Bank. Acta Cryst 2002;D58:899–907.
21. Gregoire C, Lin SY, Mazza G, Rebai N, Luescher IF, Malissen B. Covalent assembly of a soluble T cell receptor-peptide-major histocompatibility class I complex. Proc Natl Acad Sci USA 1996;93:7184–7189.
22. Brooks BR, Bruccoleri R, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: a program for macromolecular energy, minimization and dynamics calculations. J Comput Chem 1983;4:187–217.
23. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 1998;102:3586–3616.
24. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys 1983;79:926–935.
25. Ryckaert J-P, Ciccotti G, Berendsen HJC. Numerical integration of the Cartesian equations of motion in a system with constraints: molecular dynamics of *n*-alkanes. J Comput Phys 1977;23:327–341.
26. Hermann RB. Theory of hydrophobic bonding. II. Correlation of hydrocarbon solubility in water with solvent cavity surface area. J Phys Chem 1972;76:2754–2759.
27. Amidon GL, Yalkowsky SH, Anik ST, Valvani SC. Solubility of nonelectrolytes in polar solvents. V. Estimation of the solubility of aliphatic monofunctional compounds in water using a molecular surface area approach. J Phys Chem 1975;79:2239–2246.
28. Hasel W, Hendrikson TF, Still WC. A rapid approximation to the solvent accessible surface areas of atoms. Tetrahedron Comput Methodol 1988;1:103–116.
29. Still WC, Tempczyk A, Hawley RC, Hendrickson T. Semianalytical treatment of solvation for molecular mechanics and dynamics. J Am Chem Soc 1990;112:6127–6129.
30. Lee MS, Salsbury FR, Jr, Brooks CL, III. Novel generalized born methods. J Chem Phys 2002;116:10606–10614.
31. Lee MS, Feig M, Salsbury FR, Jr., Brooks CL, III. New analytic approximation to the standard molecular volume definition and its application to generalized born calculations. J Comput Chem 2003;24:1348–1356.
32. McQuarrie DA. Statistical mechanics. New York: Harper & Row; 1976.
33. Tidor B, Karplus M. The contribution of vibrational entropy to molecular association. J Mol Biol 1994;238:405–414.
34. van Gunsteren WF, Mark AE. Validation of molecular dynamics simulation. J Chem Phys 1998;108:6109–6116.
35. Caves LSD, Evanseck JD, Karplus M. Locally accessible conformations of proteins: multiple molecular dynamics simulations of crambin. Prot Sci 1998;7:7649–7666.
36. Kangas E, Tidor B. Electrostatic specificity in molecular ligand design. J Chem Phys 2000;112:9120–9131.
37. Archontis G, Simonson T, Moras D, Karplus M. Specific amino acid recognition by aspartyl-tRNA synthetase studied by free energy simulations. J Mol Biol 1998;275:823–846.
38. Manning TC, Schlueter CJ, Brodnicki TC, Parke EA, Speir JA, Garcia KC, Teyton L, Wilson IA, Kranz DM. Alanine scanning mutagenesis of an alphabeta T cell receptor: mapping the energy of antigen recognition. Immunity 1998;8:413–425.
39. Sun R, Shepherd SE, Geier SS, Thomson CT, Sheil JM, Nathenson SG. Evidence that the antigen receptors of cytotoxic T lymphocytes interact with a common recognition pattern on the H-2Kb molecule. Immunity 1995;3:573–582.

APPENDIX

Analysis of the Outliers

Vα Ser27. According to the experimental AS, the mutation of Vα Ser27 to alanine has little influence on the binding affinity. However, the contribution to the MM-GBSA binding free energy provided by its side chain has been estimated to −1.55 kcal/mol, one half originating from the electrostatic terms and the other from the apolar terms. In the X-ray structure, the Ser27 side chain makes a hydrogen bond with the MHC Arg62 side chain and a van der Waals interaction with the apolar part of the MHC Glu58 side chain. The hydrogen bond is present throughout the first half of the simulation during which it is formed and broken continuously, but is absent thereafter because of a small conformational change of the system in this region. The van der Waals interaction with the MHC Glu58 side chain is also lost for the same reason (see Fig. A1). The −1.55 kcal/mol contribution to the binding free energy made by the Vα Ser27 side chain is thus mainly originating from the first half of the trajectory when the conformation of this region is closer to the starting X-ray conformation. Therefore, a slow conformational fluctuation of this region in solution may be responsible for the apparent contradiction between the experimental AS result for Vα Ser27 and the existence of important interactions in the X-ray structure. The limited time accessible by MD simulation does not allow a sufficient conformational sampling of this region, resulting in a false hot spot for Vα Ser27. Interestingly, the average of ΔG^{sc}_{bind} of Vα Ser27 up to time t has not converged after 1 ns (Fig. 9), whereas the log(SIYR/K^b reactivity) versus ΔG^{sc}_{bind} regression for all residues except the outliers has globally converged after 400–500 ps (Fig. 8). Another possible explanation for this contradictory result would be a compensatory involvement of water molecule(s) in the alanine mutant, proposed in Ref. 17, whose effect is not taken into account by the present theoretical approach.

Vα Lys68. Vα Lys68 was found experimentally to have no influence on the binding. However, it has been found theoretically to make a favorable −1.98 kcal/mol contri-

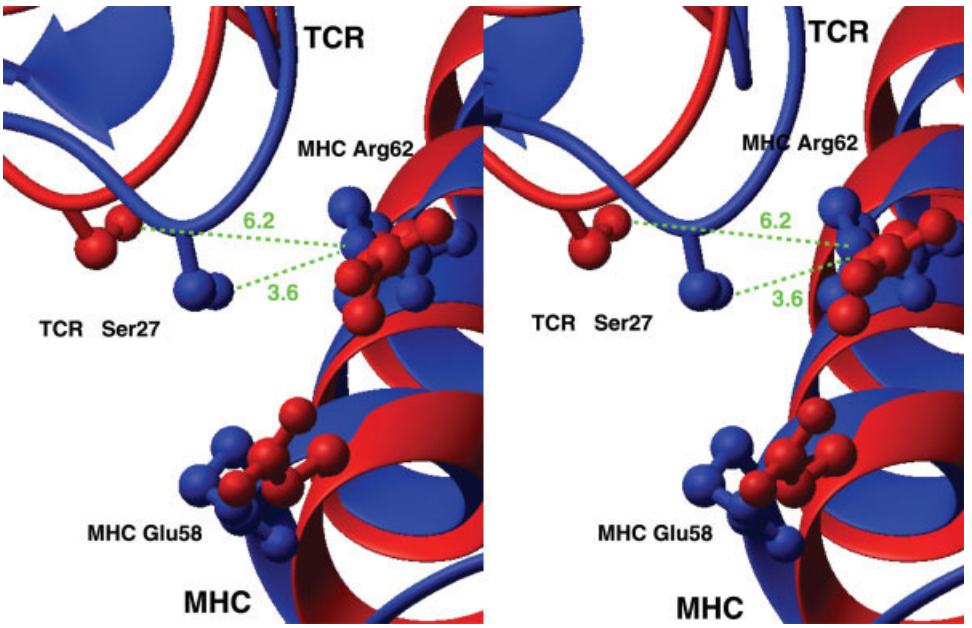

Fig. A1. Cross-eye stereoview of the TCR Vα CDR1 loop and MHC α1 helix conformations in the X-ray structure (blue) and at the end of the 1-ns MD simulation (red). The side chains of TCR Vα Ser27, and MHC Glu58 and Arg62 are shown in ball and stick representation. The distance between TCR Vα Ser27 and MHC Arg62 (green dotted lines) is given in Å. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

bution to the binding free energy, coming from the electrostatic terms. This is explained by the stable ionic interaction between the TCRα Lys68 and MHC Glu166 throughout the simulation. This ionic interaction is not important in the X-ray structure, since the distance between the two side chains is around 6 Å in the X-ray structure while it is around 3 Å in the MD simulation. This favorable contribution from the electrostatic term is not overcome by a large entropic penalty.

Vα Ser93. Vα Ser 93 has been found to make no contribution to the binding experimentally but a strong favorable contribution to the computed binding free energy thanks to the favorable hydrogen bond with the peptide Arg4 side chain. This interaction exists in the X-ray structure and is kept during the entire MD simulation. It is questionable whether mutating the Vα Ser93 to alanine should have no influence on the binding to SIYR-K^b since this mutation is intuitively expected to increase the unfavorable electrostatic desolvation free energy of the peptide Arg4.

Vα Phe100. Whereas Vα Phe100 has been found experimentally to play no role in the binding, it has been found by MM-GBSA to make a larger contribution to the binding free energy thanks to favorable nonpolar contacts. In the X-ray structure, this side chain makes van der Waals interactions with the MHC Arg62, Gln65, and Lys66 side chains and backbones. These interactions are conserved during the MD simulation. Our theoretical result is in qualitative agreement with another experimental AS of the 2C TCR/QL9/L^d, that found TCRα Phe100 to make an important contribution to the binding ($\Delta\Delta G_{\text{bind}} = 0.76$ kcal/mol).[38] In the modeled complex of 2C TCR/QL9/L^d, the TCRα Phe100 side chain also contacts the MHC Arg62 and Gln65 residues, as in the 2C TCR/SIYR/K^d X-ray structure.

Vβ Asn30. This side chain has been found experimentally to make a favorable contribution to the binding, with a log(SIYR/K^d reactivity) around 0.80. Vβ Asn30 has also been found theoretically to make an important contribution to the binding thanks to favorable electrostatic interactions. However, the −4.51 kcal/mol value for $\Delta G^{\text{sc}}_{\text{bind}}$ seems to be largely overestimated, so that this residue falls out of the correlation between log(SIYR/K^d reactivity) and $\Delta G^{\text{sc}}_{\text{bind}}$. In the X-ray structure, this side chain exchanges a hydrogen bond with the MHC Lys146 side chain. During the equilibration, an additional hydrogen bond takes place with the peptide C-terminus carboxylate. The hydrogen bond with MHC Lys146 is conserved during the entire MD simulation. On the contrary, the interaction with the peptide C-terminus remains only during the first 700 ps of the simulation and is then replaced by hydrogen bonds with MHC Ser73 and Asp77.

Vβ Asn31. This residue has been experimentally reported to be favorable to the binding, whereas it has been found to make an unfavorable contribution to the binding free energy, since the favorable electrostatic interaction is more than compensated by the electrostatic desolvation free energy. The van der Waals interaction with the peptide Tyr6 side chain is lost during the MD simulation, and is not compensated by any other interaction with the peptide or the MHC.

Vβ Tyr48. This residue was found experimentally to play an important favorable role in the binding (log(SIYR/K^d

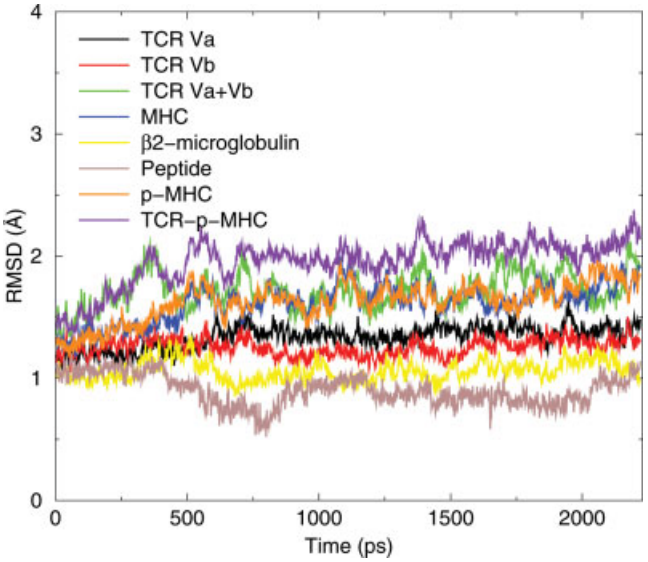

Fig. A2. RMS deviations from the X-ray structure during the MD simulation prolonged up to 2.2 ns (see Fig. 3). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

reactivity) = 1.75), while the MM-GBSA approach estimated its contribution to the binding free energy to be only −0.17 kcal/mol. No interaction between this side chain and the peptide or the MHC exists in the X-ray structure or in the MD simulation. In such a case, one possible explanation is that this residue is playing a structural role by stabilizing adjacent residues. Thus, the MM-GBSA approach cannot be expected to estimate the influence of the tyrosine to alanine mutation as proposed by Lee et al.[18]

Vβ Glu56. The mutation of Vβ Glu56 to alanine was found experimentally to have no influence on the binding while it has been found theoretically to make an unfavorable contribution of $\Delta G^{sc}_{bind} = +3.58$ kcal/mol, because of a large unfavorable electrostatic term. Only a transient hydrogen bond was found between this side chain and the MHC Gln72 side chain during the equilibration and the first 60 ps of the MD simulation. No other contact exists between this residue and the p-MHC system.

Collectively, half of the outliers are situated in the periphery of the TCR-p-MHC complex: TCRα Ser27, Lys68 and Phe100, and TCRβ Glu56.

Additional MD Simulations

To address the stability of the results as a function of the length of the trajectory, the MD simulation was prolonged up to 2.2 ns. Figure A2 shows the RMSD calculated from the starting conformation. As can be seen, the system is stable up to 2.2 ns. Also, three additional MD simulations of the system were performed, each 1 ns in length. The BFED analysis were performed for the prolonged and additional trajectories.

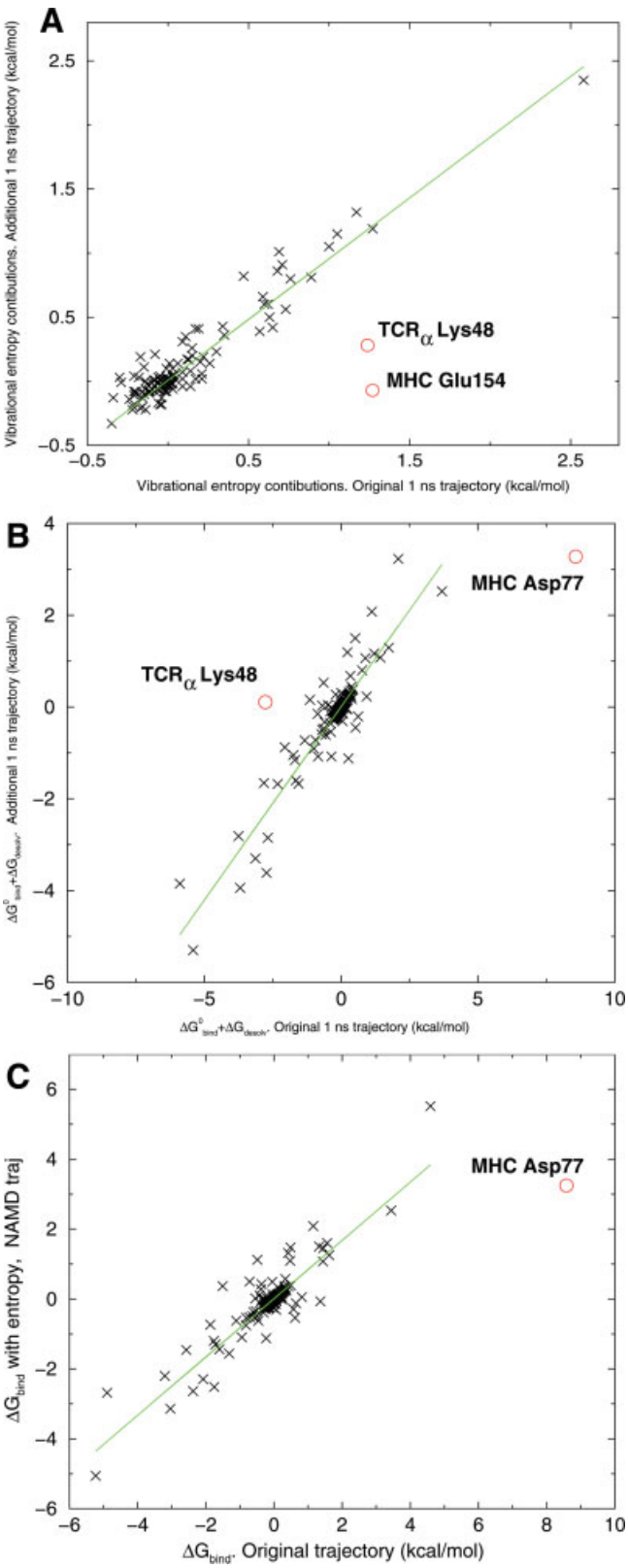

Fig. A3. **A**. Regression between side chains entropy contributions calculated from the original 1-ns MD simulation, and an additional 1-ns MD simulation. **B**. Same as A but for contributions to $\langle \Delta G^{0}_{bind} \rangle + \langle \Delta G_{desolv} \rangle$. **C**. Same as A, but for ΔG^{sc}_{bind}. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

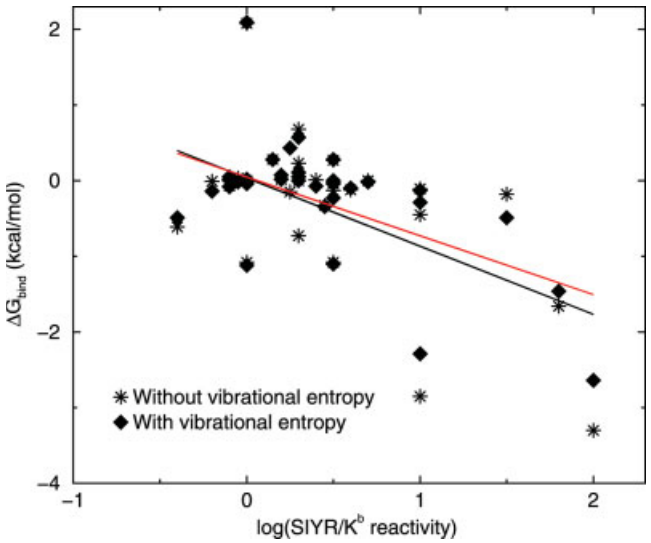

Fig. A4. Regression between the experimental log(SIYR/K^b reactivity) and the calculated ΔG^{sc}_{bind} values for one of the additional 1-ns MD simulations. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

The results obtained from those additional conformational sampling schemes are essentially similar to those described in the article. This confirms that the energy terms calculated from the first nanosecond of the original MD simulation were already converged (see Fig. 8), except for some residues showing long time scale motions that can hardly be addressed in a reasonable amount of CPU time, such as TCRα Lys48 (see below). For example, Figure A3(A) shows the regression between the side chains entropy contributions calculated from the original 1-ns MD simulation, and one of the additional 1-ns MD simulation. As can be seen, the correlation is very satisfying, with a correlation coefficient of 0.95. Two contributions do not follow the correlation: TCRα Lys48 and MHC Glu154. These two residues make a salt bridge in the original 1-ns MD simulation and in the X-ray structure, but not in the new trajectory. Figure A3(B) shows the regression between the side chains contributions to $\langle \Delta G^0_{bind} \rangle + \langle \Delta G_{desolv} \rangle$ calculated from the same two trajectories. As can be seen, the correlation is again satisfying ($R = 0.93$), except for two residues: TCRα Lys48 (see earlier) and MHC Asp77 that does not make any interaction with the TCR. Figure A3(C) shows the regression between the side chains ΔG^{sc}_{bind} values ($R = 0.90$). The outlier of this regression is MHC Asp77. It is interesting to note that TCRα Lys48 and MHC Glu154, which are outliers in the vibrational entropy regression, are not outliers for the $\Delta G_{bind}{}^{sc}$ regression. This is due to a compensation between the "enthalpic" and "entropic" terms: due to the loss of their salt bridge type interaction, their contribution to $\langle \Delta G^0_{bind} \rangle + \langle \Delta G_{desolv} \rangle$ is less favorable in the additional trajectory. However, the absence of this interaction also makes the vibrational entropy contribution less unfavorable. Figure A4 shows the correlation between the experimental log(SIYR/K^d reactivity) and the ΔG^{sc}_{bind} values calculated from this additional 1-ns MD simulation. As can be seen, adding the vibrational entropy term enhances the correlation, as observed and discussed in this article.

Thus, these additional conformational sampling schemes showed similar results to the first one developed in this article. Interestingly, however, changes in some energy terms of a given residue may be observed between different trajectories, as a function of the particular interactions that were sampled. However, a compensation mechanism between these energy terms takes place, so that the total contribution of the residue to the binding free energy remains similar, as mentioned earlier for TCRα Lys48 and MHC Glu154.