Large-scale molecular dynamics simulations of HLA-A*0201 complexed with a tumor-specific antigenic peptide: Can the ?3 and ?2m domains be neglected?

ARTICLE *in* JOURNAL OF COMPUTATIONAL CHEMISTRY · NOVEMBER 2004

Impact Factor: 3.59 · DOI: 10.1002/jcc.20100

3 AUTHORS:

Shunzhou Wan
University College London
27 PUBLICATIONS **363** CITATIONS
SEE PROFILE

Peter Coveney
University College London
230 PUBLICATIONS **4,080** CITATIONS
SEE PROFILE

Darren Flower
Aston University
237 PUBLICATIONS **9,174** CITATIONS
SEE PROFILE

Large-Scale Molecular Dynamics Simulations of HLA-A*0201 Complexed with a Tumor-Specific Antigenic Peptide: Can the $\alpha 3$ and β_2m Domains Be Neglected?

SHUNZHOU WAN,[1,2] PETER COVENEY,[1] DARREN R. FLOWER[2]
[1]*Centre for Computational Science, Department of Chemistry, University College London, 20 Gordon Street, London WC1H 0AJ, UK*
[2]*Edward Jenner Institute for Vaccine Research, Compton, Berkshire RG20 7NN, UK*

Received 18 December 2003; Accepted 8 June 2004
DOI 10.1002/jcc.20100
Published online in Wiley InterScience (www.interscience.wiley.com).

Abstract: Large-scale massively parallel molecular dynamics (MD) simulations of the human class I major histocompatibility complex (MHC) protein HLA-A*0201 bound to a decameric tumor-specific antigenic peptide GVYDGREHTV were performed using a scalable MD code on high-performance computing platforms. Such computational capabilities put us in reach of simulations of various scales and complexities. The supercomputing resources available for this study allow us to compare directly differences in the behavior of very large molecular models; in this case, the entire extracellular portion of the peptide–MHC complex vs. the isolated peptide binding domain. Comparison of the results from the partial and the whole system simulations indicates that the peptide is less tightly bound in the partial system than in the whole system. From a detailed study of conformations, solvent-accessible surface area, the nature of the water network structure, and the binding energies, we conclude that, when considering the conformation of the $\alpha 1$–$\alpha 2$ domain, the $\alpha 3$ and β_2m domains cannot be neglected.

 J Comput Chem 25: 1803–1813, 2004

Key words: immunogenic peptide; MHC class I; HLA-A*0201; large-scale molecular dynamics

Introduction

Cellular immunity is mediated by the T cell, whose surface is enriched in T-cell receptors (TCR) and functions by binding to major histocompatibility complex proteins (MHCs) expressed on the surfaces of other cells. MHCs bind, in turn, small peptide fragments derived from both host and pathogen proteins. These short peptides are commonly known as epitopes. It is this ternary molecular recognition event, which lies at the heart of the cellular immune response, that allows T cells to distinguish self from nonself. Without this presentation, other aspects of the immune response could not occur.

MHCs fall into two classes: Class I and Class II. Class I MHCs are expressed by most cells in the body, and are recognized by T cells whose surfaces are rich in CD8 coreceptor protein, while Class II MHCs are only expressed on so-called "professional" antigen presenting cells (APC), and are recognized by cells that express abundant CD4 coreceptors. Class I peptides are largely, but not exclusively, derived from intracellular proteins, which are targeted by the ubiquitinylation pathway to the proteasome, which protealytically cleaves them into short peptides. These peptides are then bound by the transmembrane peptide transporter TAP, which translocates them from the cytoplasm to the endoplasmic reticulum where they can undergo further protealytic trimming. Ultimately peptides of 8 to 11 residues are bound by MHCs, which transport them to the cell surface, where they are "presented" to the immune system. Conversely, Class II peptides, which are 15–20 amino acids in length, and are derived primarily from extracellular proteins, are taken up into APCs either by receptor-mediated endocytosis or fluid phase pinocytosis. Protealytic cleavage of endocytosed proteins follows, mediated by a variety of Cathepsins in special endosomal compartments, before derived peptides are presented by structurally distinct Class II MHCs on the surface of the APC. In what follows, we focus our attention on Class I.

To bind antigenic peptides, there are some conserved interactions between the MHC and peptide, which involve the topographically clustered residues at the very ends of the binding groove (the

Correspondence to: P. Coveney; e-mail: p.v.coveney@ucl.ac.uk

Contract/grant sponsor: The E.P.S.R.C.; contract/grant number: GR/R67699

Contract/grant sponsors: BBSRC, MRC, GlaxoSmithKline, and the UK Department of Health (to the Jenner Institute)

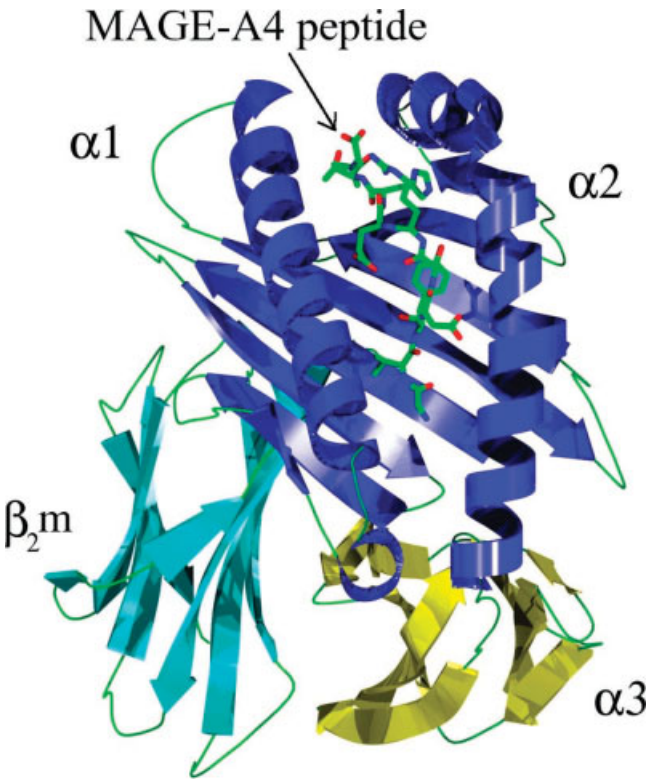

Figure 1. Ribbon representation of the HLA-A*0201:MAGE-A4 complex. The α-helices and β-sheets of the three domains are indicated in different colors: $\alpha1$–$\alpha2$ domain in blue, $\alpha3$ domain in yellow, and β_2m domain in cyan. The loops are shown in green lines. The bound peptide is displayed as a ball-and-stick model. The picture was created using ALTER.[10] [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

A and F pockets) and the terminal groups of the associated peptide.[1] The center part of the peptide bulges out of the peptide-binding groove. The peptide–MHC (p–MHC) interactions result in a strong preference for stable binding of short peptides.[2] It is well known that only peptides possessing high-affinity interactions with MHC molecules are recognized as T-cell epitopes.[3] A broad spectrum of sophisticated bioinformatic/data-centric methods for the prediction of T-cell epitopes has evolved in recent years.[4]

As a complementary tool to experimental methods, molecular dynamics (MD) simulations have been used to study peptide–MHC complexes. Simulations can provide details, not accessible by experimental techniques, concerning individual particle motions as a function of time. Because of the size of the p–MHC complexes (Fig. 1), approximations are usually introduced within the models to reduce the computational cost. One such approximation is to simulate only the antigen-binding site,[5] based on the assumption that the $\alpha3$ and β_2m domains have negligible effects on the three-dimensional structure of the $\alpha1$–$\alpha2$ domains.[6] Another approximation only includes the $\alpha1$–$\alpha2$ domains, but now constraints are imposed on the backbone atoms of these domains.[7] A third approximation is to include all the complex in the simulation, but only with atoms around the binding site free to move.[8] However, it is known from experiments that the $\alpha3$ and β_2m domains are important in stabilizing the $\alpha1$–$\alpha2$ domains.[1] Their neglect (approximation 1) may affect the conformation of the $\alpha1$–$\alpha2$ domains, while the constraints on the backbone atoms of the $\alpha1$–$\alpha2$ domains (approximation 2) or on the atoms far from the binding site (approximation 3) introduce artificial effects on the dynamics of the protein. It is well known that the properties of a simulated molecular system may change when different models and/or different boundary conditions are applied.[9] Therefore, one may ask whether the $\alpha3$ and β_2m domains can be neglected when simulating a p–MHC complex. To answer this question, a much larger scale of molecular dynamics simulation is required than is commonly used today.

In this article, we shall study the dynamics of molecular models comprising approximately 30,000 and 60,000 atoms for a time interval of 1 ns. The use of scalable MD codes on high-performance computers enables us to tackle such large problems with very short turn around times. LAMMPS (Large-scale Atomic/Molecular Massively Parallel Simulator, see http://www.cs.sandia.gov/~sjplimp/lammps.html)[11] is one such code, which was originally designed for distributed-memory parallel computers but runs on any parallel platform that supports the MPI message-passing library, as well as single-processor workstations. It scales well not only in terms of good speed-up relative to the equivalent serial performance, but also in its ability to handle large problems.[12] The validation and the efficiency of LAMMPS have been established in our previous work.[12,13] The benchmark figures for LAMMPS on our systems are shown in Figure 2. The results show scalability up to 256 processors, with speedups of 205 for the partial system and 210 for the whole system. The scalability performance of the code makes it possible to perform 1-ns simulation of the whole model in 17-h wall clock time on 256 processors of a Cray T3E supercomputer. LAMMPS had been benchmarked against the standard community code AMBER.[14] The serial performance of LAMMPS is comparable to Amber, while its scalability is far better (compare also the benchmarks on the Amber homepage: http://amber.scripps.edu/amber7.bench4.html). The typical behavior of such

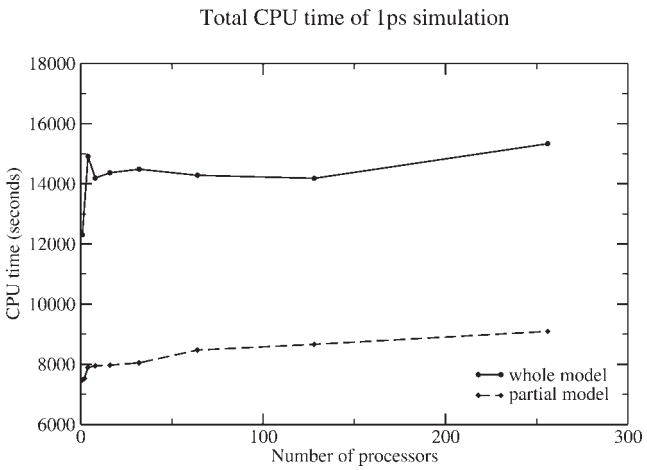

Figure 2. Total CPU time consumed per picosecond of simulations for partial and whole models on varying numbers of processors on a Cray T3E. Time step is 2fs and SHAKE is used. Perfect linear scaling is a horizontal line.

conventional legacy codes is that their performance tapers off rapidly beyond a few tens of processors where scaling remains close to linear for hundreds of processors for LAMMPS on tightly coupled parallel computers, provided the system size is large enough to warrant distributing the system over so many processors, because the interprocessor communication overhead otherwise becomes prohibitive. As demonstrated by our current work, the application of high-performance computing (HPC) technology effects a shift in our thinking about biomolecular simulation, putting us in reach of simulations of various scales and complexities hitherto not thought viable.

In this article, we study the Class I MHC molecule HLA-A*0201 in complex with the decameric peptide GVYDGREHTV, an epitope from the well-studied tumor antigen MAGE-A4.[15] MAGE-A4, which is expressed in 63% of esophageal carcinomas, 51% of lung carcinomas, but not by noncancerous cells, is a promising target for immunotherapy.[16] The HLA-A2:MAGE-A4 complex has also been studied at unusually high resolution, at least for an MHC molecule, using X-ray crystallography.[17] The present study uses two large models with periodic boundary conditions—the complete p–MHC and a partial model of the complex[5]—to investigate the effects of the $\alpha 3$ and β_2m domains on the conformations of the $\alpha 1$–$\alpha 2$ domains and on the binding of the peptide.

Model and Methods

Simulations

Because of the lack of model building and analysis routines in LAMMPS, the AMBER package[18] was used for preparing the initial molecular models and analyzing the final results. The initial atomic coordinates were taken from the X-ray crystal structure of the HLA-A*0201 complex with MAGE-A4, a decameric peptide GVYDGREHTV (Protein Data Bank entry 1I4F).[17] The complex has been determined at 1.4 Å resolution. To evaluate the effects of $\alpha 3$ and β_2m on $\alpha 1$–$\alpha 2$ domains, two simulations were performed: one for the entire complex ("whole"), and one for a partial representation, in which only the antigen-binding site is taken into account ("partial"). The xleap module of AMBER[18] was used to add hydrogen atoms, counterions (Na^+), and then to solvate the two systems using TIP3P water molecules.[19] The partial and whole complexes were each solvated by overlaying a preequilibrated box of water and deleting all water molecules that either overlapped a protein atom or a crystallographic water molecule. A water molecule is regarded as overlapping a protein atom or a crystallographic water molecule if the distance between any solvent atom to the closest solute/crystallographic water atom is less than the sum of the atoms' van der Waals distances. The distance between the edges of the box and the closest atom in the protein was 10 Å. This procedure resulted in a final number of 30,574 atoms for the partial system and 58,825 atoms for the whole system.

The potential parameterization used was the Cornell et al. 1994 force field.[20] Periodic boundary conditions were used. Equilibrations were required to bring the artificially created periodic simulation cell to a reasonable density because van der Waals voids always remain. First, minimizations were run with position restraints on protein atoms to remove any potentially high-energy contacts. Then 20 ps of NVT simulation was performed to equilibrate to the final temperature of 300 K, assigning random velocities based on a Maxwell distribution at 100 K as the initial temperature; heavy atoms within the protein were restrained at their X-ray–defined positions with a force constant 1 kcal/mol/Å^2 for backbone and 0.5 kcal/mol/Å^2 for side-chain atoms. This was followed by 120 ps of NPT simulation at 300 K with no position restraints. All equilibrations were done using the sander module of the AMBER package.[18] Bond lengths to all hydrogen atoms and the internal geometry of all the water molecules were constrained with the SHAKE algorithm.[21] The equations of motion were integrated with a 2-fs time step. The system was maintained at 300 K using the weak-coupling algorithm.[22] The nonbonded cutoff distance was 9 Å, and the Particle Mesh Ewald (PME) method[23] was used to calculate the full electrostatic energy of the periodic simulation cell in an infinite lattice of repeating images. After 30 ps of NPT simulation, the volume reached its equilibrium value within a limited range of fluctuations. The average volume of the final 110-ps equlibrium simulations was employed in our production simulations using LAMMPS on the Cray T3E machine.

In the LAMMPS simulations, the final structures from the preliminary AMBER runs were used as the initial structures for the production runs, with SHAKE applied for bonds involving hydrogen atoms, and a time step of 2 fs. The Cornell et al. 1994 force field was used, because the AMBER force field has been implemented in LAMMPS.[13] The system was thermostatted by the Langevin method[22] at 300 K. The nonbond cutoff distance was 9 Å, and the PPPM (Particle-Particle Particle-Mesh) technique[24] was used to treat the long-range interactions. Both simulations were run for a time interval of 1 ns.

Binding Free Energy

The binding free energy was calculated using the MM/GBSA (molecular mechanics/generalized Born surface area) method.[25] This method combines molecular mechanical interaction energies with solvation terms based on implicit solvation models. The free energies for complex, free MHC, and free peptide are calculated for snapshot structures taken from the MD trajectory. The binding free energy is calculated as:

$$\Delta G_b = G(\text{pMHC}) - G(\text{free MHC}) - G(\text{free peptide}) \quad (1)$$

$$G(\text{molecule}) = \langle E_{\text{MM}} \rangle + \langle G_{\text{sol}} \rangle - \text{TS} \quad (2)$$

where ΔG_b is the binding free energy in water, E_{MM} is the molecular mechanical energy, G_{sol} is the solvation energy, and $-TS$ is the entropy contribution to the solvation. $\langle\rangle$ denotes the average for a set of structures along an MD trajectory. The molecular mechanics energy represents the internal bonded energy (E^{bonded}), electrostatic ($E_{\text{int}}^{\text{ele}}$), and van der Waals ($E_{\text{int}}^{\text{wvW}}$) interactions, while the solvation energy G_{sol} is divided into two parts, the electrostatic ($G_{\text{sol}}^{\text{ele}}$) and the hydrophobic ($G_{\text{sol}}^{\text{np}}$) contributions:

$$E_{\text{MM}} = E^{\text{bonded}} + E_{\text{int}}^{\text{ele}} + E_{\text{int}}^{\text{vdW}} \quad (3)$$

$$G_{\text{sol}} = G_{\text{sol}}^{\text{ele}} + G_{\text{sol}}^{\text{np}} \quad (4)$$

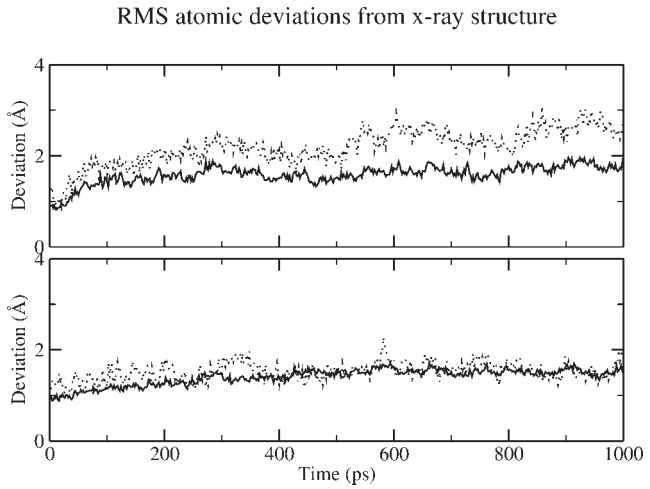

Figure 3. RMS deviation from X-ray structure vs. simulation time (ps). Above: partial MHC–peptide system; Below: whole MHC–peptide system. Solid line: main chain of protein; Dotted line: main chain of peptide.

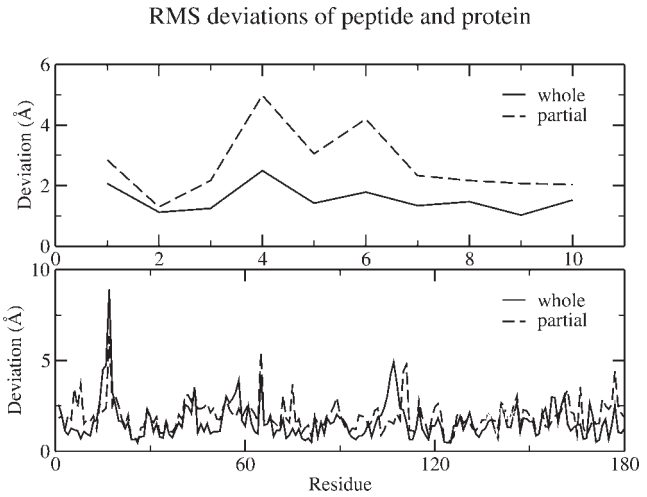

Figure 4. RMS deviations of peptide (top) and antigen-binding site of MHC protein (bottom) from X-ray structure. Solid line: whole system simulation; Dashed line: partial system simulation.

The molecular mechanics energy is calculated using an empirical force field. The electrostatic contribution to the solvation free energy, G_{sol}^{ele} is calculated by the generalized Born (GB) method.[25] Studies have shown that there is a good correspondence between GB and finite-difference Poisson–Boltzmann (PB) calculations,[26,27] although the latter has been used more frequently to calculate the electrostatic solvation free energy. The hydrophobic contribution, G_{sol}^{np}, is estimated by empirical methods based on solvent accessible surface (SAS): G_{sol}^{np} 0.00542 × SAS + 0.92 kcal/mol.[28] We assume that the entropy contributions are similar for the two models because it is the same peptide bound to the same binding site. When we calculate the binding free energy difference between the two simulations, the entropy contribution is therefore assumed to cancel.

Results

In this section, we discuss the simulation results we have obtained in five separate areas: conformational analysis, peptide/protein distances, solvent accessible surface area, the water network structure, and the binding free energy calculation.

Conformation of the Complex

The root-mean-square (RMS) deviations were analyzed after RMS fitting of the main-chain atoms at the antigen-binding site, which consists of two long α-helices and an eight-stranded antiparallel β-sheet.[17] Atoms within the loops are not included in the RMS fitting because of their flexibility. Figure 3 shows the time-dependent RMS deviations of α1–α2 domains from the X-ray structure. For the partial system, about 300 ps were required for equilibration, while the whole system required about 600 ps, equilibration here being a function of the size of the system. It should be noted that, for the whole MHC–peptide system, there is relative motion among three domains during the simulation. Even with the most powerful supercomputers and scalable MD codes that we are utilizing, it is not possible to perform MD simulations on long enough time scales to sample the relative motions of domains. However, *local* motions are well represented within these simulations. In the following analyses, trajectories from the final 400 ps were used for both simulations.

In Figure 4, the heavy atoms' RMS deviation is shown graphically for each residue. The loop regions have large deviations from the X-ray structure, while the two long helices and the β-sheets have relatively small deviations. For the peptide, the largest deviations occur in the central region, between residues pAsp4 and pGlu7, due to the flexibility of these residues. We find that the deviation of the peptide heavy atoms from the X-ray structure is substantially greater for the partial system (2.97 ± 0.24

Table 1. RMS Deviation (in Å) of Backbone Atoms of the Secondary Structures from X-ray Structure.[a]

MHC region	Partial model	Whole model
α 49–53	2.08 ± 0.42	1.19 ± 0.20
α 56–85	1.59 ± 0.19	1.37 ± 0.16
α 137–150	1.85 ± 0.33	1.38 ± 0.20
α 151–161	1.62 ± 0.19	1.71 ± 0.27
α 162–175	1.70 ± 0.22	0.78 ± 0.19
β 46–47	2.04 ± 0.36	0.88 ± 0.19
β 31–37	1.06 ± 0.20	0.84 ± 0.13
β 21–28	1.06 ± 0.15	1.01 ± 0.13
β 3–12	1.37 ± 0.14	0.86 ± 0.11
β 94–103	1.59 ± 0.14	1.04 ± 0.10
β 109–118	1.67 ± 0.16	1.40 ± 0.14
β 121–126	1.17 ± 0.18	0.70 ± 0.11
β 133–135	1.31 ± 0.31	0.89 ± 0.22

[a]Averaged over the last 400 ps of simulation.

Table 2. Distance Differences (in Å) between Centers of Geometry of Each Secondary Structure, Compared with X-ray Structure.[a]

	α1	α2	β1	β2	β3	β4	β5	β6	β7	β8
α1		0.93	0.36	0.60	0.05	0.54	0.62	0.95	0.21	0.78
α2	0.71		0.45	0.44	0.16	0.15	−0.01	0.06	−0.31	−0.43
β1	0.78	1.22		−0.17	−0.24	−0.24	−0.16	0.02	−0.32	−0.16
β2	0.12	0.39	−0.15		0.23	−0.03	0.01	0.10	−0.16	−0.12
β3	−0.71	0.08	−0.10	0.18		0.08	0.05	0.18	−0.19	−0.09
β4	−0.65	−0.63	0.13	0.14	0.29		0.02	0.13	−0.10	−0.14
β5	−0.89	−1.01	−0.23	−0.29	−0.18	−0.43		0.13	−0.06	−0.20
β6	−0.06	−1.69	0.47	0.27	0.44	0.13	0.60		0.09	−0.32
β7	−0.29	−0.79	−0.55	−0.87	−0.61	−0.80	−0.05	0.03		0.14
β8	0.70	−0.45	0.21	−0.34	−0.10	−0.50	0.08	−0.41	0.26	

[a]α1: 56–85, α2: 137–175; β1: 46–47, β2: 31–37, β3: 21–28, β4: 3–12, β5: 94–103, β6: 109–118, β7: 121–126, β8: 133–135. The upper triangle of numerical data is from the whole system simulation, while the lower triangle pertains to the partial system simulation.

Å) than for the whole MHC–peptide system (1.62 ± 0.16 Å), while the deviation of the protein heavy atoms is a little larger (2.23 ± 0.12 vs. 2.02 ± 0.07 Å). The larger deviations observed with the peptide in the partial system indicate that the peptide is considerably less tightly bound in the partial system than in the whole system.

The mass-weighted RMS deviations of the backbone atoms of the secondary structures from the X-ray structure are listed in Table 1. The deviations for atoms within the α-helices are larger than those within the β-sheets. The secondary structures retain their integrity well in both simulations, with the exception of the α-helix of residues comprising from 162 to 175 (residue numbering refers to the heavy chain if not otherwise stated) in the partial system simulation, owing to the lack of the α3 domain in this model. However, their relative positions do exhibit some differences from the X-ray structure. In Table 2, the differences of distances between centers of geometry of each secondary structural motif with those in the X-ray structure are listed: the upper triangle is for the whole system simulation, while the lower triangle corresponds to the partial system simulation. The β1 to β8 sheets are juxtaposed from the N-terminal to the C-terminal of the peptide. The increase of the distance between α1 and α2 compared with the X-ray structure reflects the increased width of the groove in our simulations. In the partial system simulation, the middle sheets (β4, β5) at the bottom of groove bulge towards the peptide, while β1, β2, and β8 turn aside from it, whereas the whole system simulation does not exhibit these effects obviously (see Fig. 5). Similarly, the differences of the relative distances among β-sheets are larger in the partial system simulation than those in the whole system simulation. The averaged values of the absolute differences are 0.24 Å for the whole model system and 0.44 Å for the partial system. Presumably these deviations in the partial system simulation are due to the lack of α3 and β_2m domains.

The RMS fluctuations were calculated for each residue by averaging over all heavy atoms in each residue. The averaged structures from the last 400 ps of the two simulations were used as

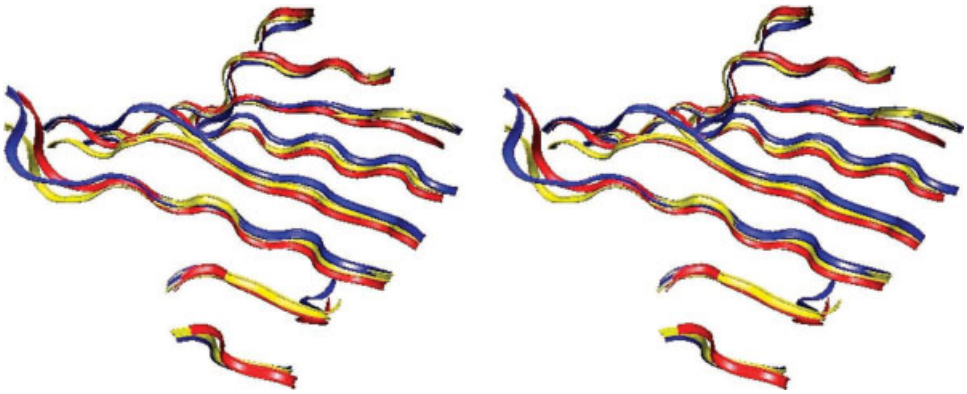

Figure 5. Stereoview of the β-sheets of the average structures from the partial system simulation (blue) and the whole system simulation (yellow), compared with the X-ray structure (red). From top to bottom, the sheets are juxtaposed from the N-terminal to the C-terminal of the peptide. The view is directly onto the peptide-binding side.

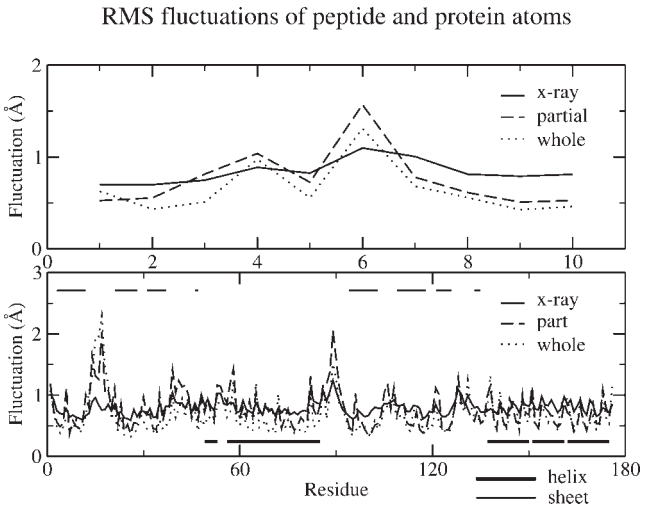

Figure 6. RMS fluctuations of peptide (top) and antigen-binding site of protein (bottom). Solid line: X-ray structure; dashed line: partial system simulation; dotted line: whole system simulation. The α-helix and β-sheet domains are indicated using thick and thin lines, respectively.

the reference configurations. The fluctuations in the X-ray structure were calculated from the experimental temperature factors.[17] They are compared in Figure 6. The general pattern in the variance of residues from the simulations matches very well with that of the X-ray crystal structure, significant enhancements around residues 17 and 89 being caused by the fluctuations of those loops. The loops are expected to readjust to adopt a favorable configuration in the course of the dynamics run. The major thermal fluctuations occur, as expected, in the turn residues and in the random coil, while the fluctuations within the two long α-helices and the eight-stranded antiparallel β-sheets are small. For the peptide, pGly1, pVal2, and pVal10 are the MHC anchors and have small fluctuations. Residue pArg6 is a candidate TCR anchor,[17] and the large observed fluctuation for this residue is indicative of its flexibility. This observation is consistent with its capacity to interact with the TCR.

Peptide and Protein Distances

In Table 3 we present the average distances between the centers of geometry of the peptide anchors and the centers of geometry of the relevant pockets from the molecular dynamics simulations and from the p–MHC complex X-ray structure. The results from our simulations are obtained by averaging the distances from each time frame (every 2 ps) of the last 400 ps simulations. The definition of the pockets was made according to Saper.[29] Most of the average distances from the two simulations are comparable with the crystal structure values. The distance between pGly1 and pocket A exhibits some deviation from the X-ray structure in the partial system simulation. From Table 4, it can been seen that the RMS deviations of the residues around pGly1 from the partial system simulation are larger than those from the whole system simulation. These deviations are calculated in a residue-based average. In the partial system simulation, the N-terminal of the peptide shifts to the α2 domain and the nitrogen of the pGly1 form hydrogen bonds with Thr163, Cys164, and Cys101. The bulky side chain of Trp167 covers the pGly1 and shields it from the solvent. Hence, although there is a shift of the residue, the solvent accessible surface area is still comparable with the X-ray structure (see below).

Table 3. Anchor–Pocket Distances in the Peptide–MHC Complex.

Anchor	Pocket	X-ray (Å)	Part (Å)	Whole (Å)
PGly1	A	3.16	5.99 ± 0.29	3.37 ± 0.22
PVal2	B	3.66	3.51 ± 0.23	3.40 ± 0.20
pTyr3	C	4.92	6.31 ± 0.52	4.91 ± 0.26
pHis8	E	7.99	9.14 ± 0.32	8.13 ± 0.31
pVal10	F	2.25	2.84 ± 0.20	2.90 ± 0.19

Pocket A: residue Met5, Tyr7, Tyr59, Glu63, Tyr159, Thr163, Trp167, Tyr171.
Pocket B: residue Phe9, Ala24, Met45, Lys66, Val67, Tyr99.
Pocket C: residue Phe9, His70, Thr73, His74, Arg97.
Pocket E: residue His114, Trp133, Trp147, Val152, Leu156.
Pocket F: residue Asp77, Thr80, Leu81, Tyr116, Tyr123, Thr143, Lys146, Trp147.

Solvent-Accessible Surface Area (SASA) of Ligand

Table 5 lists the solvent-accessible surface area (SASA) of ligand, and the relative accessibility of each residue compared to the accessibility of that residue type (x) in an extended ALA-x-ALA tripeptide.[30] The simulation data from the whole system are closer to the experimental data than that of the partial system. Residues pGly1, pVal2, and pTyr3 bind into pockets A, B, and C, respectively; they are buried away from the solvent. The next three residues (pAsp4, pGlu5, and pArg6) are solvent accessible. Residues pGlu7, pHis8, and pThr9 are partly solvent-exposed, and make simultaneous contacts with the MHC protein. The side chain

Table 4. RMS Deviation (in Å) of Heavy Atoms around Pocket A from X-ray Structure.[a]

Residue	Partial	Whole
Met5	1.58 ± 0.28	1.22 ± 0.31
Tyr7	2.22 ± 0.26	1.05 ± 0.33
Tyr59	1.85 ± 0.38	1.64 ± 0.36
Glu63	1.96 ± 0.59	1.05 ± 0.24
Cys101	2.16 ± 0.28	1.27 ± 0.28
Tyr159	2.00 ± 0.25	1.51 ± 0.28
Thr163	2.07 ± 0.32	1.10 ± 0.52
Cys164	1.87 ± 0.21	1.24 ± 0.29
Trp167	1.25 ± 0.36	0.65 ± 0.24
Tyr171	1.48 ± 0.36	0.49 ± 0.21
Average[b]	1.81 ± 0.35	1.10 ± 0.31

[a]Fit to the backbone atoms at secondary structures; the last 400-ps simulations are used.
[b]Averaged over all heavy atoms in these residues.

Table 5. Solvent-Accessible Surface Area (SASA, in Å^2) of Ligand.

	X-ray		Partial[a]		Whole[a]	
Residue	SASA	%[b]	SASA	%[b]	SASA	%[b]
Gly1	0.00	0.00	4.15 ± 2.01	5.18	2.55 ± 1.84	3.18
Val2	0.00	0.00	20.02 ± 4.01	13.22	3.42 ± 1.83	2.25
Tyr3	11.22	5.30	52.79 ± 17.30	24.81	14.63 ± 3.83	6.88
Asp4	72.38	51.60	118.04 ± 7.64	84.08	104.93 ± 8.33	74.74
Gly5	38.29	47.80	39.57 ± 7.13	49.40	33.99 ± 5.77	42.43
Arg6	165.81	69.40	197.61 ± 22.84	82.77	164.92 ± 18.83	69.07
Glu7	43.11	25.00	32.24 ± 5.29	18.72	39.93 ± 7.61	23.18
His8	27.09	14.80	66.78 ± 9.21	36.52	42.62 ± 12.05	23.30
Thr9	36.20	26.00	55.19 ± 7.00	39.63	52.76 ± 5.93	37.88
Val10	3.86	2.50	11.73 ± 3.52	7.75	14.06 ± 5.65	9.28

[a]Averaged over the last 400 ps.
[b]Relative accessibility of each residue calculated as the % accessibility compared to the accessibility of that residue type in an extended ALA-x-ALA tripeptide.

of pVal10 is completely packed into pocket F. The simulation of the whole system agrees very well with the experimental results, while the partial system shows significant deviation.

Water Network Structure

The conserved water molecules in the binding groove of the complex are shown in the average structures of the last 400-ps simulations in Figure 7, in comparison with their positions in the X-ray structure. A water molecule is regarded as being conserved when its fluctuation is smaller than 3 Å and the maximum deviation from the average position is less than 7 Å for the duration of these 400-ps simulations. These water molecules are also listed in Table 6. The water numbers in the pdb file[17] are used in Table 6, except for the water molecules added by xleap of AMBER.

The retained water molecules can be grouped into two categories, fixed and mobile, with respect to their contacts with the protein over the duration of the 400-ps simulation period. The "fixed" category is defined as those water molecules that maintain contact with a specific site throughout the simulation: 18, 42, 174, 231, 294, and 432 are considered to be in this category for the whole model simulation; with 18, 42, 174, 186, 231, and 432 for the partial model simulation. They fulfill the hydrogen bonding requirements in their respective sites and maintain contact with the latter during the 400-ps simulation period. In addition to the direct interaction between peptide and protein, there are also some water-mediated interactions between them. The N-terminus of the peptide interacts with the highly conserved residues Tyr7, Tyr59, and Tyr171 directly and indirectly via water 18. The pHis8 forms water-mediated hydrogen bonds to helix α1 and the β-sheet floor via the waters 42 and 432, and also via a water 42-water 294 network in the whole model simulation. The side chain of pGly7 makes contacts with the β-sheet floor (Arg97) via water 174. There is also a water-mediated interaction between the side chain of pGly7 and α1 via water 186 in the partial model simulation. The C-terminal of peptide is attached to Thr80 via the conserved water 231. The "fixed" water molecules retain their contact with the peptide and specific protein residues. Most of them remain near their crystal positions during the 400-ps simulations, within small fluctuations that are comparable with the heavy atoms in the protein. These tightly bound water molecules hold the peptide in place by hydrogen bonds to the MHC residues. The "mobile" category includes all the other water molecules retained in the groove. They do not interact with a specific site on the protein or the peptide. Instead, they exhibit exchange among these sites. They assemble over the pockets C and E, and fill the gap in the peptide-binding cleft. Via these "mobile" water molecules, the peptide establishes interactions indirectly with the α-helices and the β-sheet floor. The large fluctuations and the overlap of these water molecules (see Fig. 7) in the average structures indicate an exchange of mobility of these water molecules. These exchanges occur even in the relatively short 400-ps periods within the production simulations. There is a complex network of hydrogen bonds among five (whole model) or six (partial model) water molecules at the bottom of the groove.

For the whole model simulation, the number of water molecules retained in the groove is in accord with simulations of other MHC complexes[31] and with the X-ray structure. In both simulations, there are other water molecules that occupy the groove at any given time as a result of exchange between the bulk solvent and the water in the binding groove. In certain regions, there are different water molecules at different points in the trajectories: multiple water molecules exchange in and out of these locations. The major differences in the conformation of the peptide for partial and whole model simulations occur in the central region of the peptide, from residues pAsp4 to pGlu7, and this variation is mirrored by the average deviations from the X-ray structure. The large deviations and fluctuations of the region are, for the most part, determined by the surrounding water molecules.

The presence of a water network has previously been established not only by simulation[7] but also by experimental data.[32] Although the water molecules in the groove site play an important role in maintaining the conformation and orientation of the pep-

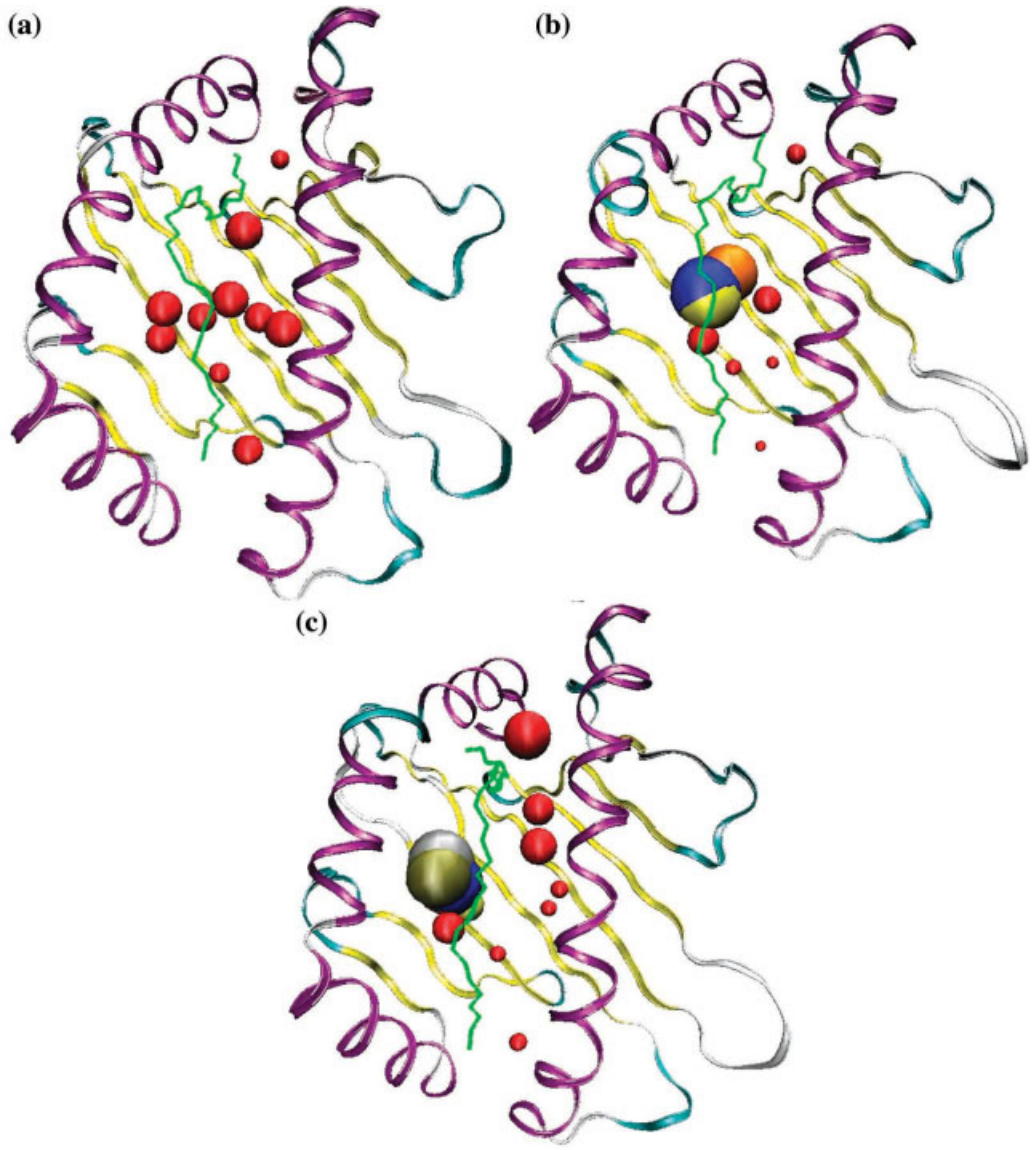

Figure 7. A representation of the groove water molecules in HLA-A*0201 complex: (a) X-ray structure; (b) whole model simulation; and (c) partial model simulation. The radii of the water molecules are their fluctuations from B-factors or from simulations. The overlapped water molecules are shown in different colors for clarity. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

tide,[7] the flexible nature of the water network in the binding groove is believed to play a crucial role in binding a range of ligands.[31,33]

Binding Free Energy

The binding free energy was estimated qualitatively using the standard MM/GBSA method, which is implemented in AMBER program.[18] Because this is a postprocessing technique, we can easily estimate the free energy differences between partial and whole models. Binding free energies were calculated using single p–MHC trajectory for both partial and whole models. This means that the structures for the energy calculations of the MHC and peptide complex were taken from the snapshots from the MD simulations of the p–MHC complex. In the analysis of the binding energies, the water molecules were replaced with implicit solvation models. To make the analysis of the two models comparable, the $\alpha 3$ and β_2m domains in the whole model were also replaced with implicit solvation models. The exterior dielectric constant was set to that of water (80). The dielectric boundary was taken as the solvent-accessible surface defined by a 1.4 Å probe sphere.

The binding free energies and their components for two models are shown in Table 7. The entropy terms are not included. As mentioned in the Methods section, it is assumed that the entropy

Table 6. Water Molecules Conserved over the Duration of the 400 ps Simulations.

		X-ray	Whole model		Partial model	
Pocket[a]	Number[b]	Fluctuation[c]	Deviation[d]	Fluctuation[e]	Deviation[d]	Fluctuation[e]
A	**18**	0.72	1.05	0.97	1.96	2.07
F	**42**	0.83	1.01	0.64	1.19	0.60
c,e	**66**[f]	0.93	9.17	2.85	—	—
c,e	**104**	1.26	2.44	2.51	—	—
c,e	**118**	1.04	2.16	2.64	2.71	2.61
C	**174**	1.38	2.18	1.22	5.05	0.66
C	**186**	1.20	—	—	3.25	0.66
F	**231**	1.04	1.29	0.44	2.08	0.64
C	**294**[g]	1.22	4.72	0.52	6.33	2.59
C	**432**	1.19	2.30	1.24	1.90	1.32
c,e	1352	—	—	—	—	1.47
c	2089	—	—	—	—	1.35
c,e	4219	—	—	—	—	2.43
c	5351	—	—	—	—	1.13
c,e	8901	—	—	—	—	2.56

[a]Upper case letters denote the water molecule is in the pocket, while lower case means the water molecule is above the pocket(s).
[b]Bold-face number means these are the original water molecules in X-ray structure.
[c]Calculated from B-factors.
[d]Deviations of the average structure from the X-ray positions after fitting the α-helices and β-sheets of protein.
[e]Fluctuations of the water molecules around their average positions.
[f]This water molecule is at the surface of the complex in the X-ray structure.
[g]This water molecule is above pockets C and E in the partial system simulation.

terms are similar for the two models. The whole model has a more favorable binding free energy, which is −72.9 kcal/mol, about 34 kcal/mol more negative than the partial model. The molecular mechanical energy ΔE_{MM} and the solvation energy ΔG_{sol} are very different, which means that in the two models the energetic complementarity between the peptide and protein is quite different. The largest differences between the two models are in the electrostatic contributions, ΔE_{int}^{ele} and ΔG_{sol}^{ele}, that balance each other to a certain extent. The tighter binding in the whole model is controlled by the more favorable electrostatic interactions that dominate over the desolvation penalty. The favorable van der Waals interaction energies and hydrophobic contributions to the solvation enhance the tighter binding in the whole model.

Table 7. MM/GBSA Energies and Standard Deviations Calculated from the Trajectories.[a]

Energy[b]	Partial	Whole
ΔE_{int}^{ele}	−229.6 ± 7.6	−304.7 ± 11.0
ΔE_{int}^{vdW}	−68.6 ± 3.7	−78.8 ± 1.7
ΔE_{MM}	−298.2 ± 10.8	−383.5 ± 12.1
ΔG_{sol}^{np}	−5.6 ± 0.2	−6.5 ± 0.1
ΔG_{sol}^{ele}	264.9 ± 6.9	317.1 ± 10.4
ΔG_{sol}	259.3 ± 6.8	310.6 ± 10.3
ΔG^{ele}	35.4 ± 2.6	12.4 ± 1.8
ΔG_b	−38.8 ± 5.1	−72.9 ± 2.0

[a]Mean energies are averaged over last the 400-ps trajectories. Standard deviations are of four batched means, each averaged over 100-ps trajectories. All units are in kcal/mol.
[b]ΔE_{int}^{ele}: electrostatic molecular mechanical energy; ΔE_{int}^{vdW}: van der Waals molecular mechanical energy; ΔE_{MM}: $\Delta E_{int}^{ele} + \Delta E_{int}^{vdW}$; ΔG_{sol}^{np}: hydrophobic contribution to solvation free energy; ΔG_{sol}^{ele}: reaction field energy calculated by GB; ΔG_{sol}: $\Delta G_{sol}^{ele} + \Delta G_{sol}^{np}$; ΔG^{ele}: $\Delta G_{sol}^{ele} + \Delta E_{int}^{ele}$; ΔG_b: $\Delta G_{sol} + \Delta E_{MM}$.

Conclusions

Large-scale molecular dynamics simulations have been performed for HLA-A*0201 with the decameric peptide GVYDGREHTV, on both partial and full models of the hydrated complex. Periodic boundary conditions were used to avoid edge effects, and the particle–particle particle–mesh (PPPM)[24] summation method was used to compute the long-range electrostatic interactions. The comparison was done to analyze the different behavior of the complex in the presence and absence of the α3 and β_2m domains. Our simulations show that the α3 and β_2m domains have a significant influence on the structural, dynamical, and energetic features of the complex, which is very important for determining the binding efficiencies of epitopes. Detailed comparisons indicate that the peptide is less tightly bound in the partial than in the full system.

Although molecular dynamics simulations have been applied to MHC–peptide interactions before,[4] these have, in general, been severely constrained by the size of the system, necessitating the imposition of various simplifying approximations. Several constrained simulations[5,7,8] have been performed to study the conformational and energetic properties of p–MHC complexes. Constraints have the apparent benefit of preventing simulated structures from drifting too far from observed crystal structures. They are also invaluable to dramatically decrease the system size and/or simulation time required to produce "reliable" results, a particularly attractive feature if MD codes scale poorly and massively parallel computing resources are scarce or unavailable. However, when we compare the RMS deviations and fluctuations of our simulations with those reported in a previous study,[8] it is evident that the imposed constraints restrict the flexibility of the protein. With position constraints either on backbone atoms or on atoms outside a specific sphere, large-scale dynamics manifestly cannot be observed. In our simulations, we find that the all-important binding behavior between peptide and MHC protein is sensitive to such large-scale motions, and so large simulations are to be preferred, rather than ignored. Other published work has already established that the calculated binding free energies are sensitive to the constraints placed on the structure.[34] We conclude that constraints must be used *very carefully*, and preferably justified in conjunction with available experimental data, so as to reduce any inaccuracies that might arise from lack of adequate conformational sampling. Indeed, the imposition of constraints is really rather *ad hoc*, its use invariably being justified retrodictively by agreement between simulation and preexisting experimental data. An elaborate exercise to determine the appropriate form of constraints to impose would itself be a very time-consuming and computationally demanding undertaking and, moreover, would only provide useful results for cases where the answers (e.g., binding energies) are known in advance. The improvement in simulation methodology brought about by the use of high-performance computers, as discussed in this article, leads to a much reduced need for constraint imposition in protein dynamics simulations.

By using massively parallel molecular dynamics, we have circumvented the need for such simplifications. Previously, we have examined the binding of peptides to MHCs using data driven techniques, namely 2D-QSAR[35] and 3D-QSAR.[36] These techniques are reliant upon the existence of substantial quantities of measured binding data for a given MHC allele. Molecular dynamics methods offer the opportunity to simulate affinities directly without the need for such data.To exploit further the potential of high-performance computing in biomolecular simulation, we are currently performing an extensive series of TCR–p–MHC complex simulations (with models comprising around 100,000 atoms) and free energy calculations to analyze the binding free energy difference of TCR for different peptides presented in HLA-A*0201, which we plan to report in future publications.

The prediction of peptide–MHC affinity is a prerequisite for the prediction of T-cell epitopes,[4] which are, in turn, the focus of modern, rationally designed poly-epitope vaccines. Our biomolecular simulations show orders-of-magnitude speedup relative to single processor runs, reducing the time needed to simulate a nanosecond from months to a few hours. We intend to utilize these substantial performance gains so that atomistic simulation of peptide–MHC complexes ultimately becomes a realistic tool in epitope prediction. Methods such as these, allowing us to predict individual epitopes accurately, are set to become important tools for tomorrow's vaccinologist. In summary, therefore, we believe our approach has potentially important implications for computer-aided vaccine design: we have described a computational methodology that may eventually be able to assist in the rapid *in silico* prediction of the complex dynamic behavior and binding phenomena that underlie the immune response.

Acknowledgments

We thank the E.P.S.R.C., which provided us with access to the 512-processor SGI O3800 and the 816-processor Cray T3E supercomputers at the UK National Supercomputing Service CSAR, on which many of our simulations were performed. We are also grateful to HEFCE for funding our local SGI Onyx2 computer. The Edward Jenner Institute for Vaccine Research thanks its sponsors: BBSRC, MRC, GlaxoSmithKline, and the UK Department of Health.

References

1. Germain, R. N. Cell 1994, 76, 287.
2. Falk, K.; Rotzschke, O.; Stevanovic, S.; Jung, G.; Rammensee, H. G. Nature 1991, 351, 290.
3. Sette, A.; Alexander, J.; Ruppert, J.; Snoke, K.; Franco, A.; Ishioka, G.; Grey, H. M. Annu Rev Immunol 1994, 12, 413.
4. Flower, D. R.; Doytchinova, I. A.; Paine, K.; Taylor, P.; Blythe, M. J.; Lamponi, D.; Zygouri, C.; Guan, P.; McSparron, H.; Kirkbride, H. In Drug Design: Cutting Edge Approaches; D. R. Flower, Ed.; Royal Society of Chemistry, Cambridge, 2002, p. 136.
5. Toh, H.; Kamikawaji, N.; Tana, T.; Muta, S.; Sasazuki, T.; Kuhara, S. Protein Eng 2000, 13, 423.
6. Rognan, D.; Reddehase, M. J.; Koszinowski, U. H.; Folkers, G. Proteins Struct Funct Genet 1992, 13, 70.
7. Meng, W. S.; von Grafenstein, H.; Haworth, I. S. Int Immunol 1997, 9, 1339.
8. Michielin, O.; Karplus, M. J Mol Biol 2002, 324, 547.
9. Hansson, T.; Oostenbrink, C.; van Gunsteren, W. F. Curr Opin Struct Biol 2002, 12, 190.
10. Flower, D. R. J Mol Graph Model 1997, 15, 161.
11. Plimpton, S. J.; Pollock, R.; Stevens, M. In Proc of the Eighth SIAM Conference on Parallel Processing for Scientific Computing, Minneapolis, MN, March 1997, p 8.
12. Giordanetto, F.; Fowler, P.; Gough, A.; Saqi, M.; Coveney, P. V. "Large scale molecular dynamics simulation of native and mutant dihydropteroate synthase-sulfanilamide complexes suggests the molecular basis for dihydropteroate synthase drug resistance" Preprint, 2004.
13. Grindon, C. R.; Harris, S. A.; Evans, T. J.; Novik, K. E.; Coveney, P. V.; Laughton, C. A. CSAR Focus 2001, 7, 5.
14. Grindon, C.; Harris, S.; Evans, T.; Novik, K.; Coveney P.; Laughton, C. Philos Trans R Soc Lond A 2004, 362, 1373.
15. Blythe, M. J.; Doytchinova, I. A.; Flower, D. R. Bioinformatics 2002, 18, 434.
16. Duffour, M. T.; Chaux, P.; Lurquin, C.; Cornelis, G.; Boon, T.; van der Bruggen, P. Eur J Immunol 1999, 29, 3329.

17. Hillig, R. C.; Coulie, P. G.; Stroobant, V.; Saenger, W.; Ziegler, A.; Hülsmeyer, M. J Mol Biol 2001, 310, 1167.
18. Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham, T. E., III; Wang, J.; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Tsui, V.; Gohlke, H.; Radmer, R. J.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. AMBER 7; University of California: San Francisco, 2002.
19. Jorgensen, W. L.; Chandrasekar, J.; Madura, J.; Impey, R.; Klein, M. J Chem Phys 1983, 79, 926.
20. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell J. W.; Kollman, P. A. J Am Chem Soc 1995, 117, 5179.
21. Ryckaert, J.; Ciccotti, G.; Berendsen, H. J. Comput Phys 1977, 23, 327.
22. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. J Chem Phys 1984, 81, 3684.
23. Darden, T.; York, D.; Pedersen, L. J Chem Phys 1993, 98, 10089.
24. Toukmaji, A. Y.; Board, J. A., Jr. Comput Phys Commun 1996, 95, 73.
25. Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. J Am Chem Soc 1990, 112, 6127.
26. Srinivasan, J.; Cheatham, T. E., III; Cieplak, P.; Kollman, P. A.; Case, D. A. J Am Chem Soc 1998, 120, 9401.
27. Zhang, W.; Hou, T.; Qiao, X.; Xu, X. J Phys Chem B 2003, 107, 9071.
28. Sitkoff, D.; Sharp, K. A.; Honig, B. J Phys Chem 1994, 98, 1978.
29. Saper, M. A.; Bjorkman, P. J.; Wiley, D. C. J Mol Biol 1991, 219, 277.
30. Flower, D. R. J Mol Graph Model 1997, 15, 238.
31. Meng, W. S.; von Grafenstein, H.; Haworth, I. S. Int Immunol 2000, 12, 949.
32. Gurlo, T.; Meng, W. S.; Bui, H.-H.; Haworth, I. S.; von Grafenstein, H. Immunol Lett 1999, 70, 139.
33. Ogata, K.; Wodak, S. J. Protein Eng 2002, 15, 697.
34. Veenstra, D. L.; Kollman, P. A. Protein Eng 1997, 10, 789.
35. Doytchinova, I. A.; Blythe, M. J.; Flower, D. R. J Proteome Res 2002, 1, 263.
36. Doytchinova, I. A.; Flower, D. R. J Med Chem 2001, 44, 3572.