Monte Carlo study of the effect of β_2 -microglobulin on the binding cleft of the HLA-A2 complex

DJAMAL BOUZIDA,¹ JEAN GARNIER,^{1,3} RICHARD BROWER,¹
JAMES CORNETTE,² AND CHARLES DELISI¹

¹ BioMolecular Engineering Research Center, Department of Biomedical Engineering, Boston University,
Boston, Massachusetts 02215

² Department of Mathematics, Iowa State University, Ames, Iowa 50011

(RECEIVED August 16, 1993; ACCEPTED March 17, 1994)

Abstract

Peptide recognition by class I products of the major histocompatibility complex requires association of the class I heavy chain with β_2 -microglobulin. We present results of Monte Carlo simulations of the β -pleated sheet floor of the human class I MHC molecule, HLA-A2, with and without β_2 -microglobulin. We find a significant effect of β_2 -microglobulin on the side chains of residues near a region that would accommodate the C-terminus of a bound peptide. By modeling simultaneously each loop and its neighboring strand at either end of the class I cleft, we find that β_2 -microglobulin restricts the conformational space of residues that are central to binding peptides. The effect is most pronounced for R97 and H114 and somewhat less important for Y99 and Y116, the latter forming strong hydrogen bonds with neighboring residues in the heavy chain itself.

Keywords: β_2 -microglobulin; dynamically optimized Monte Carlo simulation; HLA; immune system; MHC; Robson–Platt potential

Cytotoxic T cells are generally activated by foreign peptides bound to major histocompatibility complex (MHC) class I products on the surface of infected cells (reviewed by Kourilsky & Claverie, 1989). Class I MHC consists of 2 polypeptide chains: a heavy chain with three 90-amino acid-long extramembranous domains designated α_1 , α_2 , and α_3 , and an associated light chain, β_2 -microglobulin (β_2m), similar in size to each of the heavy chain domains. The central feature of the heavy chain, determined now by a number of crystallographic studies (Bjorkman et al., 1987a, 1987b; Saper et al., 1991; Fremont et al., 1992; Madden et al., 1992), is a cleft formed by the α_1 and α_2 domains, bounded by 2 α -helical walls and a β -sheet floor (Kinemage 1). β_2m is noncovalently associated with the bottom surface of the floor, whereas potentially immunogenic peptides rest above the floor, interacting with its top surface and the walls of the cleft.

Although β_2m does not interact directly with the bound peptide, experiments performed both in vivo (Townsend et al., 1989) and in vitro (Townsend et al., 1990) indicate that it strongly influences peptide presentation by class I molecules and, similarly, proper heavy–light chain assembly is dependent on the peptide

(Elliot & Eisen, 1990; Vitiello et al., 1990; Kozlowski et al., 1991; Silver et al., 1991). Thus, after separation, the 2 chains reassemble more efficiently in the presence of the peptide, while peptide association with class I depends strongly on the presence of excess β_2m in the incubating medium.

Because of the mutual interdependence of the 3 components of the complex, the dissociation rate of β_2m has been used as a measure of how tightly a given peptide binds to class I MHC molecules (Parker et al., 1992). Furthermore, when the heavy and light chains are separated under denaturing conditions, a large portion (50%) of the β -sheet structure is lost when allowed to renature (Lancet et al., 1979). The result has been interpreted as a radical conformational change produced by the absence of β_2m , which may explain the inability of non- β_2m -associated heavy chains to be expressed at the cell surface. Hence, β_2m probably plays a role in stabilizing the structure of the β -pleated sheet floor of the binding cleft.

A detailed secondary structure analysis of the α_1 and α_2 domains, allowing 16 possible states for each residue, indicates that in the absence of β_2m , certain residues in the floor are unlikely to be in their observed crystallographic configuration (Vajda, 1993). Most of these residues occur mainly in segments that are in contact with β_2m , suggesting that they are held in their position by its presence rather than by the specific properties of the local amino acid sequence.

Although both studies suggest the possibility of a radically altered backbone structure in the absence of β_2m , the use of de-

Reprint requests to: Charles DeLisi, Department of Biomedical Engineering, Boston University, 44 Cummington Street, Boston, Massachusetts 02215; e-mail: delisi@buenga.bu.edu.

³ On leave of absence from the Unité d'Ingénierie des Protéines, Bio-technologies, INRA, 78352, Jouy-en-Josas, France.

naturants to disrupt the structure suggests that large scale structural alteration may not occur spontaneously within times of experimental interest. Moreover, secondary structure predictions have only moderate accuracy. In view of these caveats, and because rigorous conformational searching of large structures strains the limits of even massively parallel computing, a full-scale calculation seems unwarranted at present. In this initial study, we therefore focus primarily on exploring the consequences of β_2 m on side-chain rearrangements and correlations in the floor of the binding cleft using equilibrium Monte Carlo (MC) computer simulations. Even with this restriction, the calculations are computer intensive and yield useful insights, as indicated below.

The regions of interest are residues in segment 25–33: VGYVDDTQF, which is allowed to interact with its neighboring strand 5–11: MRYFFTS; and residues in segment 113–124: YHQAAYDGGKDYI, which interacts with its neighboring strand 95–101: VQRMVGC.⁴ The stability of these regions is studied subject to the constraint that the remainder of the molecule be in its crystallographic configuration. The problem of extreme local anisotropy and heterogeneity of the environment, which severely limits the computational effectiveness of conventional methods, is taken into account using a recently developed MC technique (Bouzida et al., 1991, 1992).

We find that the effect of β_2 m is greatest on the region located near the C-terminus of the bound peptide, restricting largely the motion of R97 and H114, which are adjacent to pocket E of the binding cleft as defined in Saper et al. (1991). The restriction is caused in part by direct interaction with β_2 m, but its influence appears to be mainly mediated indirectly, as a cooperative effect on the neighboring residues whose side-chain structure is greatly dependent on the presence of β_2 m.

Results

The standardization of the segments of interest has been performed using a simulated annealing process described in the Methods section. During each annealing cycle, the structure with the smallest deviation from the X-ray structure is recorded and used as the initial structure for the next annealing cycle. The lowest all-atom and backbone atom RMS deviations (RMSDs) of the final standardized structures are shown in Table 1. The final structures obtained using this procedure have an all-atom RMSD from the crystal structures less than 0.5 Å, which is within experimental resolution in proteins (Chothia & Lesk, 1986).

Starting from the lowest RMSD standardized structures, each loop and its neighboring strand is relaxed simultaneously into its equilibrium state at room temperature (300 K). The initial maximum angle change for each variable is set to 15°, and the acceptance ratio is tuned to 50% with a calculated standard deviation of about 15%. We generate a total of approximately 300,000 thermally probable conformations in each region studied, of which 300 uncorrelated structures are retained and stored to study individual residue rotamers as explained below. In each case, the internal energy and its fluctuations are computed, and the global RMSD for each generated conformation with respect

Table 1. RMS deviations of the final standardized structures

Sequence	Angle count	RMSA ^a (Å)	RMSB ^b (Å)
L1 V ₂₅ GYVDDTQF ₃₃	34	0.484	0.490
L2 Y ₁₁₃ HQAAYDGGKDYI ₁₂₄	52	0.369	0.330
B1 M ₉₅ RYFFTS ₁₀₁	32	0.481	0.312
B2 V ₉₅ QRMVGC ₁₀₁	29	0.385	0.274

^a RMSA, all-atom RMSD.

^b RMSB, backbone (N, C α , C, O) RMSD.

to the crystal structure is recorded. Some dihedrals are found to be more flexible than others. The maximum changes in angle ranged from 3° to 60°, with an overall average angle change of 20°.

The Monte Carlo methodology described below provides a picture of the flexibility of specified sections of the heavy chain in the presence and absence of β_2 m. The flexibility can be conveniently represented in terms of the distribution of values of the RMSD for the various segments from their crystallographic coordinates. For L1B1, unimodal distributions are obtained (Fig. 1) both in the presence and in the absence of β_2 m, with the modes displaced on average by 1.14 ± 0.13 Å. Consequently, β_2 m appears to have very little effect on the conformation of this region. The L2B2 region, on the other hand, is sensitive to the presence of β_2 m, the distribution being unimodal in its presence and bimodal in its absence. Based on the RMSDs from the X-ray structure, 2 conformers appear in the absence of β_2 m. Conformer 1 of the bimodal distribution corresponds to an average all-atom RMSD of 1.05 ± 0.06 Å. The average all-atom RMSD of conformer 2 is 1.60 ± 0.13 Å. There is no significant internal energy barrier between the two; they both correspond to the same average energy, with a 3 kcal/mol energy fluctuation.

The RMSDs of individual amino acids with respect to the native structure (β_2 m present) for the single conformer of the L1B1 region, and for the 2 conformers obtained in the L2B2 region, are shown in Figure 2. The backbone RMSDs of residues in the L1B1 region are less than 0.2 Å, whereas for residues in the L2B2 region, most of the residues in each conformer have a backbone RMSD greater than 0.2 Å; 2 residues in each conformer (G120 and D122 in conformer 1, and D119 and G120 in conformer 2) have a backbone RMSD of 0.5–0.6 Å. In the second conformer, residues Y118, D119, and G120 undergo backbone deviations as large as their all-atom deviations. A possible explanation is that these residues, representing the turn residues for loop L2, actually form hydrogen bonds with I1, R3, and W60 of β_2 m when present. The removal of β_2 m frees the glycine-aspartate backbone, which will consequently move more. These backbone movements are about 3 times as great as those observed in the L1B1 region.

The residues in L1B1 undergo very small side-chain variations — all less than 0.7 Å. The largest RMSDs correspond to R6 (0.7 Å) and F8 (0.4 Å) in strand B1, and Y27 (0.6 Å) in loop L1, as shown in Figure 2. Both F8 and Y27 are in contact with β_2 m. It should be noted that L1B1 is as close to the α_3 domain in the heavy chain as it is to β_2 m, and it is possible that the low relative RMSDs obtained for residues in this region could be attrib-

⁴ Segments 25–33 and 5–11 will be denoted by L1 and B1, respectively, and by L1B1 when referred to collectively. The corresponding abbreviations for segments 113–124 and 95–101 are L2, B2, and L2B2.

Fig. 1. Energy/RMS histograms. A: Region L1B1 in the presence of β_2m . B: Region L1B1 in the absence of β_2m . C: Region L2B2 in the presence of β_2m . D: Region L2B2 in the absence of β_2m . RMS refers to all-atom RMSDs.

uted to the presence of the residues from the α_3 domain that help stabilize it independently of β_2m .

In contrast to L1B1, the all-atom RMSDs for each residue in L2B2 from their native state obtained in the absence of β_2m are relatively large (Fig. 2). The largest all-atom RMSD values are obtained for residues Q96, R97, M98 (conformer 1 only), H114, Q115, K121, and D122 (conformer 1 only). Interestingly, all these residues occur around peptide binding pockets D and E, as described by Saper et al. (1991), although some of them (Q115 and K121) are actually pointing toward β_2m . The primary difference between the conformers is due mainly to the larger all-atom RMSDs of residues Q115 and K121 and the relatively smaller deviations of R97, M98, H114, and D122 in conformer 2. Q115 and K121 side chains are affected more because they are directly pointing toward β_2m and thus interacting with it. However, the variation in the RMSDs in the residues pointing toward the peptide may be a sign of some correlation between residues on both sides of the β -sheet floor.

To gain more insight into the correlation between the conformers, we studied the relative orientations of side chains in the 2 regions. The different rotamers of the residues that are affected by the presence of β_2m are shown in Figure 3. The 2 regions L1B1 and L2B2 contain altogether 66 χ dihedral angles. In the presence of β_2m , 47 of these rotamers are found within

1 standard deviation of the observed X-ray rotamer, and only 12 of the dihedral angles occurring in side chains of residues R6, D30, M98, Q115, and D122 differ by 2 or more standard deviations. On the basis of observed rotamer frequencies (Dunbrack & Karplus, 1993), 7 of the 12 simulated rotamers are found to be the most observed ones, 4 had no preferential rotamer values, and only the χ_1 value of M98 was anomalous (i.e., the simulated rotamer corresponded to a rarely observed conformation). This result underlies the accuracy of the simulation of the side chains when β_2m is present, which allows us to be more confident about the changes occurring when β_2m is removed.

Two rotamers are obtained for residue H114 when the simulation is run without β_2m : one in the vicinity of the X-ray structure and the other corresponding to a flip of the ring thereby interchanging the positions of the nitrogens in the ring with the carbons and vice versa. The histidine conformer in the vicinity of the X-ray structure is found to be more populated (3:1) than the other conformer. Furthermore, the corresponding angular fluctuations around the average angle are 9° higher for the X-ray conformer. Interestingly, the single conformer obtained when β_2m is present does not correspond to the conformation of the histidine in the X-ray structure. As the H114 ring appears to form 1 hydrogen bond in either conformation (Table 2), the crystallographic orientation of the histidine ring given by the ni-

Fig. 2. All-atom and backbone RMSDs from the native state (β_2m present) of individual amino acids in each region simulated. **A:** L1B1 region. **B:** L2B2 region (conformer 1). **C:** L2B2 region (conformer 2).

trogen positions cannot be determined by experimental data alone. Furthermore, the fact that residue H114 can flip from one rotamer to another when β_2m is removed clearly illustrates the increased flexibility of that part of the binding cleft.

Two rotamers are also found for χ_1 of residue Q115 in the absence of β_2m , while only one is found when β_2m is present,

as shown in Figure 3. In this case, the value of χ_1 in the X-ray structure is 6° , which is highly unlikely. The simulated structures move away from this value to avoid steric hindrance between the backbone nitrogen and the C_γ atom. The conformer obtained in the presence of β_2m has an average χ_1 of -76° , corresponding to a position of the C_γ atom between the backbone N and C atoms. The standard deviation around this average is only 6° . The more populated rotamer obtained when β_2m is absent corresponds to the positioning of the C_γ atom between the backbone H and C instead, and the standard deviation is 10° in this case.

In order to visualize correlations between the different rotamers in neighboring residues, clusters of conformations that correspond to each doublet of χ angles are plotted in Figure 4. When plotting χ_1 of Q115 versus χ_2 of H114 in the presence of β_2m , only 1 cluster of conformations (cluster 1 in Fig. 4) corresponding to average (χ_1, χ_2) values of $(-77^\circ, -94^\circ)$ appears. However, 3 clusters of conformations (clusters 2, 3, and 4 in Fig. 4) are obtained in the absence of β_2m , and they correspond to (χ_1, χ_2) values of $(-170^\circ, -94^\circ)$, $(-170^\circ, 95^\circ)$, and $(-64^\circ, 95^\circ)$. Similarly, when correlating χ_3 of R97 with χ_2 of H114, 2 clusters appear in the presence of β_2m , one of which disappears ($\chi_3 = 67^\circ$) to give rise to 2 other clusters of H114. A stereo view of 2 coupled conformations of H114 and R97, one in the presence of β_2m and the other in its absence, is shown in Figure 5 and Kinemage 2.

To assess the physical significance of the observed differences in the conformations of residues R97, H114, and Q115, we carried out a systematic study of the hydrogen bonding network involving these 3 residues with the rest of the molecule. We compared the hydrogen bond lengths in the average structure obtained when β_2m is present with those found in the average structure of the second conformer when β_2m is absent. The results are shown in Table 2. The 3 residues form 5 backbone/backbone hydrogen bonds, which are found to be similar in both structures; they are not represented in the table. It can be seen that 3 strong and 2 weak hydrogen bonds are found in the average structure when β_2m is present, whereas only 1 strong and 2 weak bonds are seen when β_2m is absent, allowing more conformational freedom to the side chains. When β_2m is present, R97 is hydrogen bonded to H74 and D77. However, when it is removed, R97 flips around its χ_3 angle, breaking the previous 2 hydrogen bonds to form new ones with T73 and H74 and, more interestingly, with H114. The latter has flipped around its χ_2 angle to form yet another hydrogen bond with its backbone. Furthermore, residue Q115, which forms 1 hydrogen bond with its backbone and a weak hydrogen bond with L126 when β_2m is present, is found completely free when β_2m is removed. From an energetic point of view, the internal energy of the H114 ring has its lowest value (-9.7 kcal/mol) in its native position in the presence of β_2m (cluster 1 of Fig. 4). When β_2m is removed, the internal energy of H114 ring is increased to -8.0 kcal/mol (cluster 2), -6.1 kcal/mol (cluster 3), and -4.6 kcal/mol (cluster 4).

The correlation of rotamers observed between H114 and Q115, as shown in Figure 4, cannot be explained by a specific interaction between the 2 side chains alone. These 2 chains are in a *trans* conformation, and the movement of their backbone is relatively small. This correlation is more likely due to a common effect by the force field originating from the β_2m atoms. It is clear from the data shown in Table 3 and the stereo view

Fig. 3. Rotamer representation for selected residues. The χ angles are measured following the IUPAC-IUB notation (IUPAC-IUB Commission on Biochemical Nomenclature, 1970). We have positioned the N, C, and H backbone atoms around the circle representing χ_1 for R6 to help the reader see the orientation of the side chain; the C_α - C_β bond goes into the paper through the center of the circle. For the other χ angles, only the relative positions of the C_α , C_β , and C_γ atoms are shown. A χ angle is 0 for a horizontal line pointing to the right and positive in the counterclockwise direction. The arrows in the figure correspond to the orientation of the side chain in the X-ray structure, the thick line represents its orientation in the absence of β_2m , and the thinner line that sticks out of the circle is the side-chain orientation in the native case. The angular values in degrees outside the circle represent the standard deviation from the average value of the rotamer. When more than 1 rotamer for the side chain is present, a percentile value of the number of conformations in that rotamer is indicated. Values obtained in the presence of β_2m are preceded by an asterisk.

Table 2. Lengths of hydrogen bonds formed by R97, H114, and Q115 in the presence or absence of β_2m subunit

AA1 ^a	Donor	AA2 ^b	Acceptor	Without β_2m (Å)	With β_2m (Å)
R97	NE-HE	H114	NE2-CD2	2.76	— ^c
R97	NH1-HH2	T73	CB-OG1	2.70	—
R97	NH1-HH1	H74	NE2-CD2	2.28	—
R97	NE-HE	H74	NE2-CD2	—	2.73
R97	NH2-HH3	D77	CG-OD1	—	2.25
H114	ND1-HD	H114	C-O	—	2.16
Q115	N-HN	Q115	CD-OE1	—	2.21
Q115	NE2-HE1	L126	C-O	—	2.89

^a AA1 is the donor amino acid.^b AA2 is the acceptor amino acid.^c A hydrogen bond length greater than 4 Å is represented by a dash.

of the flexible part in Figure 5 and Kinemage 2 that for most residues pointing toward β_2m —Q96, M98, Q115, A117, G120, and D122—the β_2m subunit brings a stabilizing (negative) internal energy that is lost when β_2m is removed. Some of the residues pointing away from β_2m , such as H114 and D119, exhibit the same internal energy behavior. This suggests that upon removal of the β_2m subunit, the stabilizing energy disappears, allowing more freedom and new conformations to the corresponding residues.

A closer look at the rotamers of Y99 and Y116 shows that these residues do not exhibit large all-atom motions even when β_2m is removed, which is in agreement with their low RMSDs (Fig. 2). The effect on these 2 residues is mainly on their backbone, not on the side chains. Their backbone fluctuations are not negligible and are twice as much as the highest L1B1 backbone fluctuations in the case of Y99. These particular side chains are known to play active roles in peptide binding. An analysis of their interactions shows that the stability is mainly due to their fixed hydrogen bonding with heavy chain residues that stay fixed during the simulation. Y99 is hydrogen bonded to H70, while Y116 is found buried and hydrogen bonded to D77.

Discussion and conclusions

Bound peptides are often stabilized by amino- and carboxyl-terminal charges bonding polar and/or charged groups in the class I cleft (Madden et al., 1992). Of the 8 MHC residues identified as making important contributions to these interactions, 6 amino acids (Y59, Y84, T143, L146, Y159, and Y171) are on the α -helical walls and hydrogen bond to 2 other amino acids (Y7 and Y123) on the β -floor. All those in the wall but Y159 occur within 1 turn from the end, and none are in proximity of β_2m .

In addition to the charged end groups of the peptide, additional stability is generally provided by the so-called anchor residues, which usually occur close to the ends of the peptide. In this paper, we were interested in heavy-chain residues that simultaneously affect, either directly or indirectly, both the peptide and β_2m . Although the simulations were limited to 2 specific regions of the HLA-A2 complex, thus excluding added conformational changes occurring in the other parts of the binding cleft or excluding possible denaturation of the whole heavy chain when β_2m is removed, they were, however, central to

Fig. 4. Clustering of conformations corresponding to different χ angles of chosen residues. **A:** χ_1 (Q115) versus χ_2 (H114). **B:** χ_3 (R97) versus χ_2 (H114).

peptide binding. The MC simulations that we have undertaken proved to be effective in analyzing the effect of β_2m on the conformation and binding activity of the binding cleft in the α_1 and α_2 domains.

The simulations show clearly that the effect of β_2m is more pronounced on residues near a region that would accommodate the C-terminus of a bound peptide. Simulation of this region (L2B2) in the absence of β_2m indicates a bimodal distribution in the all-atom RMSD. The peak corresponding to a higher RMSD disappears when β_2m is present. No significant effect of β_2m is observed in the simulations of L1B1. This suggests that

Fig. 5. Stereo figure of 2 conformations of the L2B2 region. The structure drawn with a thin line is obtained in the presence of β_2m , whereas the other is obtained in its absence.

β_2m plays the role of indirectly restricting some key residues in the L2B2 region to their native conformation for effective binding of the peptide.

The simulations also indicate an interesting coupling between residues facing toward β_2m , and R97 and H114 facing the bound peptide. R97 and H114 are polymorphic residues known to affect peptide recognition (Latron et al., 1991). These results suggest that residues that do not bind peptide nevertheless exert an influence on binding by affecting the conformation of those residues that do.

Methods

In order to efficiently explore conformational space, we restrict all portions of the molecule other than the loop and β -strand segments (regions L1B1 and L2B2) to their crystallographic coordinates, and use a force field with bond lengths and angles fixed to standard values (Robson & Platt, 1986). The variable dihedral angles are the backbone ϕ and ψ and the side-chain χ 's, as described by Robson and Platt (1986). We use a united-atom representation where only the polar hydrogens that are capable of forming hydrogen bonds are taken into account. Polar hy-

drogen atoms have also been added in the rigid part at the stereochemical standard positions.

Standardization

The class I geometry of the relevant segments was standardized using a simulated annealing process, which was found to be more efficient in attaining smaller standard deviations from the crystal structure than the usual "gradient-based" minimization methods (J.F. Gibrat & J. Garnier, pers. comm.). The procedure consists of finding the set of dihedral angles of each segment, which provides the smallest deviation from its corresponding X-ray structure. It consists of 3 stages: (1) building an extended structure with standard bond lengths and bond angles from the same sequence of residues as the segment; (2) bringing the 3 atoms at the N-terminus of the strand — N', H', and C $_{\alpha}$ — in contact with the corresponding atoms in the X-ray structure using a translation vector and a rotation matrix (see, for example, page 148 of Goldstein, 1981); and (3) minimization using simulated annealing of the harmonic cost function $\mathcal{C}$ defined as

$$\mathcal{C} = \sum_{i=1}^N (\mathbf{r}_i - \mathbf{r}_i^X)^2. \quad (1)$$

In Equation 1, N is the number of atoms in the segment, and $\mathbf{r}_i$ and $\mathbf{r}_i^X$ are the position vectors of the i th atom in the modeled and observed structures, respectively. The cooling process used is exponential, where the temperature at each cycle is decreased by a constant ratio,

$$T_{i+1} = \gamma T_i, \quad \text{where } 0 < \gamma < 1, \quad (2)$$

such that the algorithm runs at each temperature long enough to allow a reasonable sample of possible conformations to take place some specified number of times. The factor γ can be written in terms of the initial and final temperatures, T_i and T_f :

$$\gamma = \left(\frac{T_f}{T_i} \right)^{\frac{1}{Nc-1}}, \quad (3)$$

where Nc is the number of annealing temperatures used in the process. During annealing, only the dihedral angles are allowed to vary, and the step sizes are dynamically optimized at each temperature (Bouzida et al., 1991, 1992). The standardization of each segment in the 2 regions of interest was carried out separately using 100,000 MC steps with an annealing schedule that lowered the temperature from 100 to 0.1 (same units as the cost function — $\text{\AA}^2$ in this case) gradually and exponentially every 2,000 steps.

Table 3. Relative contributions of the heavy chain and β_2m interatomic interactions to the average internal energy of residues in the L2B2 region

Residue	E_{tot}^a (kcal/mol)	E_{hc}^b (kcal/mol)	$E_{\beta_2m}^c$ (kcal/mol)
Q96	-11.25	-4.06	-7.19
R97	6.56	3.58	2.98
M98	-4.50	-2.35	-2.15
Y99	12.50	9.12	3.38
G100	-3.13	-2.97	-0.16
H114	-9.70	-7.15	-2.55
Q115	-5.97	-4.80	-1.17
Y116	11.87	6.99	4.88
A117	-1.02	-0.62	-0.40
Y118	4.42	0.61	3.87
D119	-13.06	-7.15	-5.91
G120	-4.28	0.47	-4.75
K121	2.34	0.73	1.61
D122	-15.83	-9.58	-6.25
Y123	8.08	3.87	4.21

^a Average internal energy of each residue (side-chain and backbone atoms), sum of columns 3 and 4.

^b Average internal energy contribution of heavy chain atoms.

^c Average internal energy contribution of β_2m atoms.

The target function

The standardized structures are used as starting points for the MC simulations. During both equilibration and the subsequent MC updates, we use a united-atom energy function for the system given by the sum of pairwise interactions of nonbonded atoms comprising both electrostatic, van der Waals energies and torsional potentials as defined by Robson and Platt (1986):

$$E = E_{fi} + E_{ff} + E_{fr} + E_{tr}. \quad (4)$$

In Equation 4, E_{fi} and E_{ff} represent the nonbonded interactions within each flexible segment and between segments, respectively; E_{fr} is the sum of the interactions between the flexible part and the rigid part; and E_{tr} is the torsional energy around the side-chain dihedral angles. There are no explicit water molecules in the simulations. The dielectric constant is set equal to 5, which has been used to model loops in proteins with reasonable accuracy (Gibrat et al., 1992; Higo et al., 1992; Collura et al., 1993).

The potential was applied to all pairwise interactions without imposing a cutoff on the interaction distance. The reason for such complete inclusion is that the potential energy of the relaxed L2 loop calculated as a function of cutoff radius in the presence of the entire molecule fluctuates widely until about 25 Å, as shown in Figure 6. Consequently, a cutoff less than 25 Å would be inaccurate. On the other hand, once 25 Å is reached, more than 80% of the total number of nonbonded interactions in the molecule have been included, so the time savings by truncating at this point are insignificant. In addition, to ensure that the interaction of the L1B1 region with the nearby α_3 region is properly taken into account, neighboring residues from the α_3 domain are added to the residues of the α_1 and α_2 domains that constitute the rigid part of the molecule with β_2m . The relevant side chains are D183, A184, S207–I213, T233–F241, H263, Q264, and G265. Distant interactions with the α_3 domain were neglected when simulating the L2B2 region.

Equation 4 is applied to all the atoms of the molecule except the carboxyl terminal atoms C_α , C' , and O' , which are constrained to be near their crystallographic coordinates. The constraint is imposed by adding an energy term that is a quadratic function of distance from the observed location given by

$$E_{xr} = K_x \sum_{C_\alpha, C', O'} (r_i - r_i^x)^2. \quad (5)$$

The bond constant K_x is chosen to be equal to 10 kcal/mol/Å², constraining the 3 atoms to remain within approximately 0.1–0.3 Å of their observed locations at room temperature.

The Monte Carlo method

It has been common practice to carry out standard MC simulations (reviewed by Valleau & Torrie, 1977; Valleau & Whittington, 1977) by performing 2 consecutive sets of simulations. In the first short set, optimized values of the relevant parameters are obtained. These values are then used in a second set of longer simulations to compute the physical quantities of interest. This strategy is effective for homogeneous systems, such as simple fluids. However, most molecular systems are inhomogeneous and anisotropic and therefore require step-size updates as new information is generated.

Fig. 6. Variation of the potential energy (in kcal/mol) of the relaxed L2 loop as a function of cutoff radius (in Å).

The most popular procedure is the modified algorithm by Noguti and Gō (1985), which uses scaled collective variables that change in anisotropic random steps. The collective variables are obtained from the eigenvectors and eigenvalues of the second derivative matrix (Hessian) of the energy function with respect to the dihedral angles. Qualitatively, an eigenvalue of the Hessian gives a measure of the “softness” of the energy landscape in some direction specified by the corresponding eigenvector. Hard directions will correspond to eigenvectors whose corresponding eigenvalues are large. This method has been successfully applied to predict the 3-dimensional structure of hypervariable loops in complementary determining regions (CDRs) of antibodies (Gibrat et al., 1992; Higo et al., 1992). However, calculation of the second derivative matrix and its eigenvalues and eigenvectors can become computationally expensive, $\mathcal{O}(N^3)$, when the number of variables becomes large, and as a consequence the matrix is computed only once every thousands of steps (Higo et al., 1992). Furthermore, the information obtained depends only on a single conformation rather than an equilibrium set of conformations.

The MC approach used in this study differs from other MC methods in that it takes account of inhomogeneity by using conformational data collected in real time during the simulation to optimize step size (Bouzida et al., 1991, 1992). A key parameter is the acceptance ratio, that is, the ratio of accepted conformations to the total number of trial conformations. During every cycle of the simulation, each dihedral angle can jump randomly throughout some prespecified range $[-\theta, \theta]$ determined by the acceptance ratio $\langle P \rangle$ obtained during the previous cycle. Because of the inhomogeneity of the molecular system, the range varies from one dihedral angle to another. At the end of each cycle, the maximum angle change θ_{new} for each dihedral is computed, and is given by

$$\theta_{new} = \theta_{old} \frac{\ln(aP_i + b)}{\ln(a\langle P \rangle + b)}, \quad (6)$$

where θ_{old} is the maximum angle used during the previous cycle, P_i represents the ideal acceptance ratio that is input at the beginning of the simulation, and a and b are real parameters chosen such that θ_{old} is either multiplied or divided by a convenient scale factor (around 5) whenever P is 0 or 1 (Corana et al., 1987). If the value of θ_{new} obtained exceeds 180° , meaning that the torsional angle is very flexible, it is reset to 180° .

In the present implementation of this adaptive approach, we employed uncorrelated single angle moves. For the dihedrals on the side chains this resulted in large changes varying from 20° to 60° , but due to the constraining force placed on the ends of the backbone segment, smaller dihedral changes, on the order of 4° , occurred in the backbone itself. The extra computer time needed to update the maximum step sizes in the present single angle update scheme is completely negligible (less than 0.1% of the total CPU time).

Using this procedure, we have sampled equilibrium conformations with and without β_2m . The maximum step size for each variable is updated every cycle of 25 MC-sweeps through each of the 2 regions; 1 MC-step corresponds to moving 1 dihedral angle, whereas 1 MC-sweep corresponds to moving all the dihedral angles of the flexible part once. For each accepted conformation, the RMS position, defined as the square root of the mean of the square of the distances between the atoms in the flexible region and the corresponding ones in the X-ray structure, is calculated. We have found that less than 200 MC-sweeps are needed to bring the energy of each region within the equilibrium fluctuations. In addition, we monitored the autocorrelation time, which is found to be less than 2 MC-sweeps. Thus, most conformations generated in dihedral space using this method are totally uncorrelated. Using an Alpha Decstation, the computer time needed to perform 1 MC-step is about 0.4 s in the absence of β_2m and around 0.6 s in its presence.

Acknowledgments

This work was supported by grant AI3053 awarded to C.D. by the NIAID, NIH. We thank Prof. Temple F. Smith, Prof. S. Vajda, Dr. Yechiel Becker, and Dr. George Vasmatzis for helpful comments and valuable discussions.

References

- Bjorkman PJ, Saper MA, Samraoui B, Bennett WS, Strominger JL, Wiley DC. 1987a. Structure of the human class I histocompatibility antigen, HLA-A2. *Nature* 329:506-512.
- Bjorkman PJ, Saper MA, Samraoui B, Bennett WS, Strominger JL, Wiley DC. 1987b. The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. *Nature* 329:512-518.
- Bouzida D, Kumar S, Swendsen RH. 1991. Almost Markov processes. In: Landau DP, Mon KK, Schüttler HB, eds. *Computer simulation studies in condensed matter physics III. Springer Proceedings in Physics*, vol 53. Berlin/Heidelberg: Springer-Verlag. pp 193-196.
- Bouzida D, Kumar S, Swendsen RH. 1992. Efficient Monte Carlo methods for the computer simulation of biological molecules. *Phys Rev A* 45: 8894-8901.
- Chothia C, Lesk A. 1986. The relation between the divergence of sequence and structure in proteins. *EMBO J* 5:823-826.
- Collura V, Higo J, Garnier J. 1993. Modeling of protein loops by simulated annealing. *Protein Sci* 2:1502-1510.
- Corana A, Marchesi M, Martini C, Ridella S. 1987. Minimizing multimodal functions of continuous variables with the 'simulated annealing' algorithm. *ACM Trans Math Software* 13(3):262-280.
- Dunbrack RL Jr, Karplus M. 1993. Backbone-dependent rotamer library for proteins: Application to side-chain prediction. *J Mol Biol* 230:543-574.
- Elliot TJ, Eisen HN. 1990. Cytotoxic T lymphocytes recognize a reconstituted class I histocompatibility antigen (HLA-A2) as an allogeneic target molecule. *Proc Natl Acad Sci USA* 87:5213-5217.
- Fremont DH, Masazumi M, Stura EA, Peterson PA, Wilson IA. 1992. Crystal structures of two viral peptides in complex with murine MHC class I H-2K^b. *Science* 257:919-927.
- Gibrat JF, Higo J, Collura V, Garnier J. 1992. A simulated annealing method for modeling the antigen combining site of antibodies. *ImmunoMethods* 1:107-125.
- Goldstein H. 1981. *Classical mechanics. Addison-Wesley Series in Physics*. St. Louis: Addison-Wesley.
- Higo J, Collura V, Garnier J. 1992. Development of an extended simulated annealing method: Application to the modeling of complementary determining regions of immunoglobulins. *Biopolymers* 32:33-43.
- IUPAC-IUB Commission on Biochemical Nomenclature. 1970. Abbreviations and symbols for the description of the conformation of the polypeptide chains. *Eur J Biochem* 17:193-201.
- Kourilsky P, Claverie JM. 1989. MHC-antigen interaction: What does the T cell receptor see? *Adv Immunol* 45:107-193.
- Kozlowski S, Takeshita T, Boehncke WH, Takahashi H, Boyd LF, Germain RN, Berzofsky JA, Margulies DH. 1991. Excess β_2 -microglobulin promoting functional peptide association with purified soluble class I MHC molecules. *Nature* 349:74-77.
- Lancet D, Parham P, Strominger JL. 1979. Heavy chain of HLA-A and HLA-B antigens is conformationally labile: A possible role for the β_2 -microglobulin. *Proc Natl Acad Sci USA* 76:3844-3848.
- Latron F, Moots R, Rothbard J, Garrett TPJ, Strominger JL, McMichael AJ. 1991. Positioning of a peptide in the cleft of HLA-A2 by complementing amino acid changes. *Proc Natl Acad Sci USA* 88:11325-11329.
- Madden DR, Gorga JC, Strominger JL, Wiley DC. 1992. The three-dimensional structure of HLA-B27 at 2.1 Å resolution suggests a general mechanism for tight binding to MHC. *Cell* 70:1035-1048.
- Noguti T, Gō N. 1985. Efficient Monte Carlo method for simulation of fluctuating conformations of native proteins. *Biopolymers* 24:527-546.
- Parker KC, DiBrino M, Hull L, Coligan JE. 1992. The β_2 -microglobulin dissociation rate is an accurate measure of the stability of MHC class I heterotrimers and depends on which peptide is bound. *J Immunol* 149:1896-1904.
- Robson B, Platt E. 1986. Refined models for computer calculations in protein engineering. *J Mol Biol* 188:259-281.
- Saper MA, Bjorkman PJ, Wiley DC. 1991. Refined structure of the human histocompatibility antigen HLA-A2 at 2.6 Å resolution. *J Mol Biol* 219:277-319.
- Silver ML, Parker KC, Wiley DC. 1991. Reconstitution by MHC-restricted peptides of HLA-A2 heavy chain with β_2 -microglobulin, in vitro. *Nature* 350:619-622.
- Townsend ARM, Elliot T, Cerundolo V, Foster L, Tse A. 1990. Assembly of MHC class I molecules analyzed in vitro. *Cell* 62:285-295.
- Townsend ARM, Öhlén C, Bastin J, Ljunggren HG, Foster L, Kärre K. 1989. Association of class I major histocompatibility heavy and light chains induced by viral peptides. *Nature* 340:443-448.
- Vajda S. 1993. Conformational filtering in polypeptides and proteins. *J Mol Biol* 229:125-145.
- Valleau JP, Torrie GM. 1977. A guide to Monte Carlo for statistical mechanics, 2. Byways. *Mod Theor Chem* 5:169-194.
- Valleau JP, Whittington SG. 1977. A guide to Monte Carlo for statistical mechanics, 1. Highways. *Mod Theor Chem* 5:137-168.
- Vitiello A, Potter TA, Sherman LA. 1990. The role of β_2 -microglobulin in peptide binding by class I molecules. *Science* 250:1423-1426.