Predicting Binding Affinities of Protein Ligands from Three-Dimensional Models: Application to Peptide Binding to Class I Major Histocompatibility Proteins

Didier Rognan,*,† Sanne Lise Lauemøller,‡ Arne Holm,§ Søren Buus,‡ and Vincenzo Tschinke▽

Department of Pharmacy, Swiss Federal Institute of Technology, CH-8057 Zürich, Switzerland, Division of Experimental Immunology, Department of Medical Microbiology and Immunology, University of Copenhagen, DK-2200 Copenhagen, Denmark, Chemistry Department, Royal Veterinary and Agricultural University, DK-2200 Copenhagen, Denmark, and NOVARTIS Pharma, Basel, CH-4002 Basel, Switzerland

Received May 10, 1999

A simple and fast free energy scoring function (Fresno) has been developed to predict the binding free energy of peptides to class I major histocompatibility (MHC) proteins. It differs from existing scoring functions mainly by the explicit treatment of ligand desolvation and of unfavorable protein−ligand contacts. Thus, it may be particularly useful in predicting binding affinities from three-dimensional models of protein−ligand complexes. The Fresno function was independently calibrated for two different training sets: (a) five HLA-A*0201-peptide structures, which had been determined by X-ray crystallography, and (b) three-dimensional models of 37 H-2K^k-peptide structures, which had been obtained by knowledge-based homology modeling. For both training sets, a good cross-validated fit to experimental binding free energies was obtained with predictive errors of 3−3.5 kJ/mol. As expected, lipophilic interactions were found to contribute the most to HLA-A*0201-peptide interactions, whereas H-bonding predominates in H-2K^k recognition. Both cross-validated models were afterward used to predict the binding affinity of a test set of 26 peptides to HLA-A*0204 (an HLA allele closely related to HLA-A*0201) and of a series of 16 peptides to H-2K^k. Predictions were more accurate for HLA-A2-binding peptides as the training set had been built from experimentally determined structures. The average error in predicting the binding free energy of the test peptides was 3.1 kJ/mol. For the homology model-derived equation, the average error in predicting the binding free energy of peptides to K^k was significantly higher (5.4 kJ/mol) but still very acceptable. The present scoring function is thus able to predict with a good accuracy binding free energies from three-dimensional models, at the condition that the backbone coordinates of the MHC-bound peptide have first been determined with an accuracy of about 1−1.5 Å. Furthermore, it may be easily recalibrated for any protein−ligand complex.

Introduction

With the ever increasing number of protein structures available at the three-dimensional level, predicting the binding free energy of a small molecule to the active site of a protein has become one of the major challenges in computational chemistry aimed at lead finding or optimization. Basically, one has to find a compromise between the accuracy of the method to choose and the number of ligands to examine.[1] On one hand, very accurate free energy differences may be obtained by comparing the binding of a pair of ligands to a common protein by free energy perturbation techniques.[2] The price to pay is rather CPU-intensive calculations, the difficulty to compare more than two ligands, the restricted nature of the change to perform (small chemical change), and the difficulty to predict absolute binding free energies[3,4] although some progresses have been recently reported.[5] On the other hand, binding free energies may be estimated from atomic coordinates by means of simple scoring functions.[6−9] These free energy estimates are less accurate (prediction error of about 7−8 kJ/mol, for high-resolution crystal structures) but may be obtained rather quickly for thousands of potential ligands. Between these two extremes, force-field,[10−13] 3D-QSAR,[14−16] and continuum methods[17,18] may be applied to a series of 10−20 related ligands for estimating binding free energy differences.

Free energy scoring functions, although not always very accurate, are particularly interesting in prioritizing hits obtained from de novo design or 3D database searching. However, they face four major problems: (i) correct estimation of the entropic contribution to the absolute binding free energy, (ii) reliable estimate of solvation/desolvation processes, (iii) lack of repulsive terms, and (iv) a questionable application to three-dimensional molecular models.

Up to now, free energy scoring functions have been applied with reasonable success to reproduce experimental binding affinities from X-ray coordinates of protein−ligand complexes.[6−9,16] Very few predictions have indeed been made from three-dimensional models[9] as one has to tackle a two-dimensional problem: docking

* To whom correspondence should be addressed. Phone: +41.1.635 60 36. Fax: +41.1.635 68 84. E-mail: didier@pharma.ethz.ch.
† Swiss Federal Institute of Technology.
‡ University of Copenhagen.
§ Royal Veterinary and Agricultural University.
▽ NOVARTIS Pharma.

10.1021/jm9910775 CCC: $18.00
Published on Web 10/01/1999

Table 1. Training Set A: 5 HLA-A*0201-Peptide Complexes[20]

		ΔG_{bind}, kJ/mol				
peptide	PDB code	exp.[a]	Ludi[b]	Chemscore[c]	Zhang[d]	Fresno[e]
TLTSCNTSV	1hhg	−37.32	−62.36	−33.12	−32.63	−36.85
FLPSDFFPSV	1hhh	−48.45	−72.10	−56.48		−48.56
GILGFVFTL	1hhi	−46.94	−57.40	−42.12	−52.30	−47.03
ILKEPVHGV	1hhj	−37.60	−78.32	−43.09	−42.69	−38.96
LLFGYPVYV	1hhk	−45.48	−77.92	−57.28	−45.10	−45.57

[a] Experimental values have been taken from the literature.[27,68] Reported IC_{50} values were assumed to be similar to equilibrium dissociation constants K_D as the concentration [L] of the radiolabeled ligand in the unbound state is much lower than the equilibrium dissociation constant K_D of the labeled ligand in the binding assay. Thus, $\Delta G^{exp} \approx RT \ln(IC_{50})$. [b] Measured with InsightII, release 970 (MSI, San Diego, CA). [c] Measured from Fresno, using H-bond, lipophilic, and rotatable bond terms as described for Chemscore.[8] [d] Taken from Zhang et al.[32] [e] Predicted values using the cross-validated model (see text).

and scoring. Therefore, there is a need for a flexible and fast scoring function in which desolvation energy of the ligand and unfavorable protein−ligand contacts are taken into account.

Class I MHC-peptide complexes represent a particularly attractive case for docking and free energy scoring algorithms. Determining the complete MHC-bound conformation of an oligomeric peptide (up to nine residues) is not a trivial problem as the ligand is highly flexible and the middle part of the peptide (P_4 to P_{C-2}, P_n standing for position n and P_C for the C-terminus) usually zigzags[19] or bulges[20] out the binding groove using very different paths. Peptide docking using Monte Carlo,[21] molecular dynamics simulations,[22] dynamic programming,[23−25] free energy mapping,[26] or threading[27−29] can rather well reproduce the conformation of the MHC-binding amino acids (rms deviations for backbone atoms lower than 1 Å) but have more difficulties to represent the middle bulging part (rms deviations between 1 and 2 Å). Some attempts to reproduce experimental binding free energy values from MHC-peptide X-ray structures have been described in the last five years. Continuum electrostatic calculations may find the correct ranking but not the absolute binding free energies.[30] Knowledge-based scores based on pairwise contacts are only efficient for apolar peptides.[28] Up to now, the best scores have been found for a training set of four MHC-nonapeptide X-ray structures (1hhg, 1hhi, 1hhj, 1hhk; Table 1) after binding free energy decomposition into structure-derived atomic contact energies, electrostatic interactions and backbone conformational entropy change.[31,32] However, Zhang's equation has only been used for four X-ray structures and its predictive value for molecular models is hypothetical. Herewith, we present the development a new scoring function (Fresno) based on previous works of Böhm[6,7] and of the Proteus group[8,9] as well as its application to a nontrivial problem: predicting, from three-dimensional models, the absolute binding free energy of polar and flexible ligands (octa- to decapeptides) to class I major histocompatibility molecules.

Results and Discussion

Binding to HLA-A2 Alleles: Predicting Binding Affinities from a Set of Five X-ray Structures. As class I MHC proteins are structurally unstable in their free form and need a peptide in order to properly fold,[33] there are rather few MHC-peptide complexes for which a crystal structure is available and where the binding affinity of the bound peptide is known with accuracy. The HLA-A*0201 human class I protein is the best characterized protein as it has been cocrystallized with at least 10 different peptides, of which the binding free energy of five of them has been determined experimentally (Table 1). Furthermore, we wished to study the potential benefit of two novel terms (buried-polar interactions and ligand desolvation) in our binding free energy estimation with respect to two universal scoring functions, that described by Böhm[6] which we will call Ludi and the one published by Eldridge et al.[8] which we will call Chemscore.

Using default coefficients[6] for each term of the equation, Ludi largely overestimated the binding affinity of all five peptides (Table 1) by 11−40 kJ/mol (about 2 to 7 orders of magnitude in binding affinity). This poor performance ($r^2_{pred} = 0.010$, $s = 6.09$ kJ/mol) has three main reasons: (i) As the Fresno estimate should be applicable to 3D-models, all protein−ligand structures have been here energy-minimized whereas the procedure described by Böhm for calibrating his scoring function only involved addition of hydrogen atoms without any energy relaxation; (ii) the lack of an explicit treatment of desolvation effects is particularly detrimental to the oligomeric peptides studied here; (iii) the free energy contribution of the two salt bridges, involving both terminal ends of the peptides with the MHC binding groove is certainly overestimated. The first Ludi scoring function that we tested[6] cannot distinguish between buried and water-accessible ionic interactions. Hence, MHC-peptide salt bridges in the five tested complexes are all accessible to water.[20] It should be pointed out that the last two points have been recently addressed by a new version of the Böhm's function.[7] The first point (minimized vs experimental structures) is still a matter of debate. However, we feel that energy-minimized structures are necessary for reliable 3D-models, especially for the correct positioning of hydrogen atoms usually missing in X-ray structures but whose coordinates strongly influence the hydrogen-bonding score. Furthermore, it reduces the user's influence on the starting positions of these hydrogen atoms, even if a minimized structure may remain close to the starting one.

Using Chemscore-based default coefficients and parameters[8] (H-bonding, lipophilic, and rotational terms), a better correlation is found ($r^2_{pred} = 0.522$, $s = 4.32$ kJ/mol) because of a better description of electrostatic and rotational terms for our data set (Table 1). Recalibrating the Chemscore scoring equation by fitting computed to experimental data afforded a predictive model ($q^2 = 0.645$, $s_{press} = 4.47$ kJ/mol) that could be improved by addition of either the buried-polar term ($q^2 = 0.865$, s_{press}

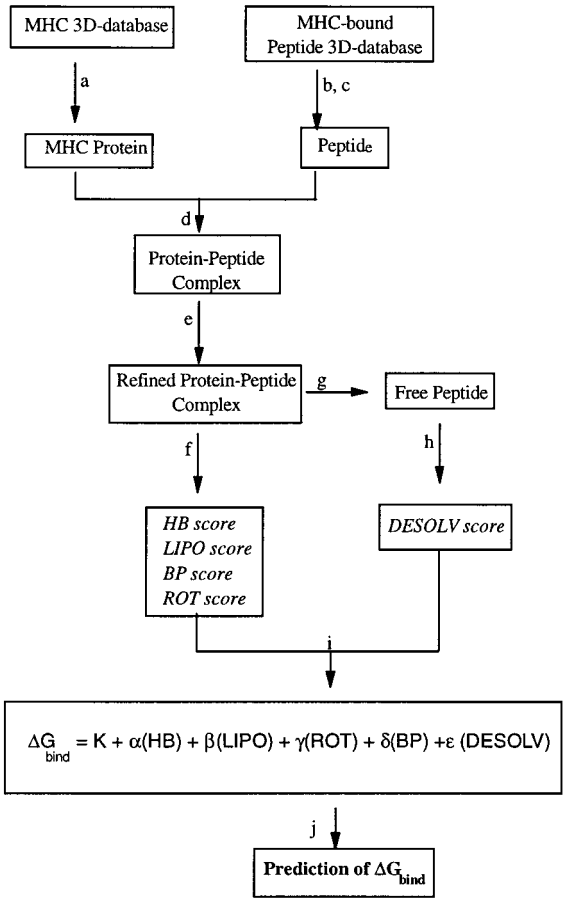

Figure 1. Flowchart of the methods used for quantifying and predicting the free energy of binding of peptide ligands to a class I MHC protein (see Experimental Section for computing protocols): (a) homology modeling of the MHC protein; (b) knowledge-based building of MHC-anchor residues (P_1-P_3, $P_{C-1}-P_C$); (c) knowledge-based Loop search—building the P_4-P_{C-2} sequence; (d) merging the peptide into the MHC binding groove; (e) energy-minimization (EM) of the peptide, EM of the MHC-peptide complex, 100 ps simulated annealing (SA) of the MHC-bound peptide, EM of the last SA conformer; (f) Fresno; (g) extracting coordinates of the peptide from the MHC-peptide complex; (h) DelPhi; (i) multiple linear regression; (j) prediction of binding free energies for test sets.

$= 2.78$) or the desolvation score ($q^2 = 0.918$, $s_{press} = 2.14$ kJ/mol). When both terms are added to the final regression equation (eq 1), a still good predictive model is obtained with a correct ranking of all five peptides (Figure 2a).

$$\Delta G_{bind} = -33.614 - 0.014\text{HB} - 0.076\text{LIPO} + 0.017\text{ROT} + 0.021\text{BP} + 0.026\text{DESOLV} \quad (1)$$

$$(q^2 = 0.895,\ s_{press} = 3.448,\ n = 5)$$

This scoring scheme (see eq 1 and Table 1) was selected for further predictions. Interestingly, the contribution of each term to the predicted binding free energy is physically reasonable. The highest contributions to the binding free energies were given for the lipophilic (from −20 to −30 kJ/mol) and the buried-polar (about +10 kJ/mol) terms. This observation is in agreement with previous reports suggesting the predominant role of hydrophobic interactions in HLA-A2-peptide interactions.[30,34] In contradiction to a previous report,[6] the inclusion of explicit desolvation energies was shown

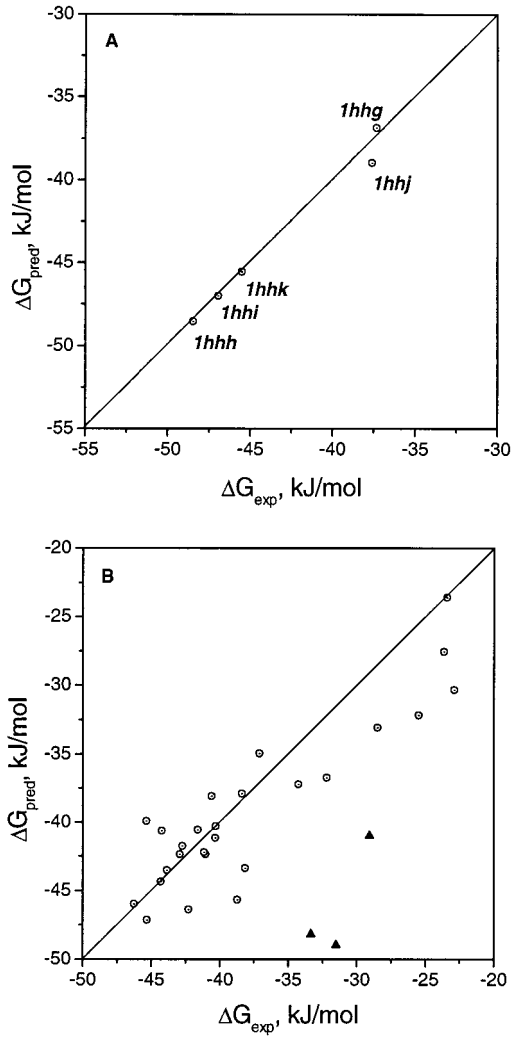

Figure 2. Predicted versus experimental binding free energies for HLA-A2-binding peptides: (a) cross-validated fit to experimental data (see Table 1); (b) prediction of a test set of 26 peptides. Outliers, indicated by up triangles, have been discarded from the regression analysis.

to significantly improve the model. The desolvation term contributes here to about +5 kJ/mol to the total binding free energy. H-Bonding and rotational contributions were in the present equation very weak (±0.5 kJ/mol). Accordingly, error limits on the regression coefficients were higher for these two terms (0.012 for α, 0.011 for γ) than for LIPO, BP, and DESOLV coefficients (0.04 for β, 0.08 for δ, 0.014 for ϵ). However, it is somewhat surprising that one should be able to derive a predictive model of so high quality based on such a limited training set. Therefore, its real predictive value was checked on a test set of 26 nonapeptides binding to the class I MHC molecule, HLA-A*0204. The binding free energy to the HLA-A*0204 allele was calculated by Fresno and in parallel tested by direct experimentation. We reasoned that it should be possible to use the HLA-A*0201 training set in the prediction of binding to HLA-A*0204 as the latter only varies from the HLA-A*0201 allele by a single point mutation in the binding groove.[35] This is supported by the finding that self-peptides eluted from the two related alleles are very similar.[36] However, calculating reliable free energy scores from 3D-models requires first an accurate docking of the ligand to its host protein. For class I MHC-binding ligands, the

Table 2: Predicting the Backbone Conformation and the Binding Free Energy of 5 HLA-A*0201-Binding Peptides

peptide	PDB code	loop[a]	rmsd, Å[b]	ΔG_{bind}, kJ/mol exp	ΔG_{bind}, kJ/mol model
TLTSCNTSV	1hhg	1prt: 90−93	1.04, 1.27	−37.32	−36.37
FLPSDFFPSV	1hhh	1ajs: 362−366	1.59, 1.82	−48.45	−45.09
GILGFVFTL	1hhi	1gig: 10−13	1.14, 0.46	−46.94	−44.55
ILKEPVHGV	1hhj	1prt: 90−93	0.93, 0.87	−37.60	−39.82
LLFGYPVYV	1hhk	1bnc: 3−6	0.78, 1.44	−45.48	−45.92

[a] Sequence used to build the P_4-P_{C-2} loop: 1prt, D-3-phosphoglycerate deshydrogenase; 1ajs, aspartate aminotransferase; 1gig, hc19 Igg1 Fab fragment; 1bnc, biotin carboxylase. [b] Root-mean-square deviations of peptide backbone atoms from the X-ray coordinates, before and after simulated annealing of the MHC-bound peptide (see Experimental Section).

docking step may be particularly difficult for two reasons: (i) the ligand is a highly flexible oligomeric peptide, (ii) one-half of the bound ligand bulges out of the MHC binding site in order to further interact with a T-cell receptor.[37−39] To validate our docking procedure (Figure 1, see Experimental Section), we tried first to reproduce the MHC-bound backbone conformation of all five reference peptides in their X-ray structure (Table 2). Except for the decapeptide, rms deviations of the initial docked conformation from the X-ray structure varies between 0.78 and 1.14 Å (Table 2). Refinement the protein-bound peptide by simulated annealing did not dramatically alter the conformation of the bulging loop (compare the values before and after simulated annealing refinement, Table 2). The models obtained for the five complexes were then used to derive the binding free energy of the reference five ligands, using the regression derived from the original training set composed of X-ray structures. Again, a very good agreement was obtained between experimental and predicted binding free energies ($r^2_{pred} = 0.846$, $s = 2.40$ kJ/mol). This good fit demonstrates that the docking and refinement procedures are able to propose very reliable three-dimensional models of MHC-peptide complexes. It should be noted that the bulging part of the bound peptide only contains auxiliary anchors to the MHC binding groove[40] and that a poorer score is obtained by taking only into account MHC anchoring amino acids for intermolecular interactions while giving a mean value for the rotational term (data not shown). However, it is important that docking does not overemphasize protein−ligand interactions.[41] In none of the five complexes remodeled in this study was the binding free energy of the corresponding peptide significantly overestimated.

Our docking procedure being validated, we tried to predict the binding free energy of a series of 26 peptides to HLA-A*0204 (Figure 2b). The very low standard error of prediction of this new test set (3.49 kJ/mol) illustrates the robustness of the training model ($r^2_{pred} = 0.774$). Interestingly, the scoring function was able to correctly rank the most potent and the weakest binders of the series. It should be noted that many of the weakest binders were overestimated by about 5 kJ/mol (Figure 2b). All these peptides bear a negatively charged residue at the auxiliary anchor position P_3. Asp and Glu amino acids at P_3 are known to disfavor binding,[40] by a probable electrostatic repulsion with the aromatic upper part of the pocket D in the MHC binding cleft.[42] The 100 ps simulated annealing refinement of these peptides was, however, unable to totally expel the P_3 side chain from the binding groove, thus overestimating the binding free energy of this small subset. Totally hydrophobic peptides were also correctly ranked among the most potent binders. To check whether Fresno was able to correctly score very low binding peptides presenting another HLA specificity, we tried to rank a HLA-B*2705-binding peptide (RRIKAITLK) whose sequence does not share any identity with known HLA-A2-binding ligands[44] and whose structure has been modeled in the electron density map of the HLA-B*2705 allele.[43] The scoring function was indeed able to predict a very poor binding for this peptide ($\Delta G_{bind} = -18.51$ kJ/mol, $K_i = 560$ μM).

As a maximal error limit, we chose a value of 7.5 kJ/mol, close to the standard deviation of two known scoring functions.[6−8] If the discrepancy between experimental and predicted binding free energies was higher than this cutoff, the corresponding ligand was defined as an outlier. Up to now, only three peptides could not have been properly handled by our docking/scoring protocol (Figure 2b). Their binding free energies were overestimated by 10−18 kJ/mol. The binding affinity of these three outliers is unexpectedly weak, with regard to their sequence that perfectly fulfills the HLA-A*0204 binding motif.[36] They all share a small hydrophobic side chain at P_2 and P_9 and an aromatic amino acid at P_3, three favored features for a proper binding to HLA-A*0204. Therefore, it is very difficult to explain the weak binding affinity of these three ligands. It should be noticed that scoring the three outliers from residue-dependent positional coefficients derived from experimental binding affinities was also unsuccessful.[47] Intramolecular repulsions between adjacent bulky side chains (P_3, P_5, and P_7) that have already been reported to affect the bound conformation of MHC ligands,[20,45] could be an issue for the three outliers. However, one of the reference peptide (1hhk, Table 1) also presents bulky side chains at these three positions but binds as expected to HLA-A*0201. The only way to reproduce with less discrepancies the experimental binding affinity values for this restricted set was to expel the P_3 side chain from its binding pocket D, although there is no experimental support for this hypothesis. As the ligand strain energy was not taken into account here, it is also possible that the observed discrepancies may be related to distortion of the peptide 3D-structure upon binding. Thus, we feel that wrong predictions are probably not attributable to the free energy scoring function itself but to the inability to correctly predict reliable three-dimensional structures for the corresponding complexes.

Binding to the H-2K^k Protein: Predicting Binding Affinities from Three-Dimensional Models. Relating binding affinities to X-ray coordinates by a free energy decomposition and using the derived free energy equation for quantitative predictions is an absolute prerequisite for any reliable free energy scoring function. However, in many cases, experimentally determined 3D structures of protein−ligand complexes are missing. Thus, the scoring function should also perform well when the training set is composed of three-dimensional molecular models. As class I MHC proteins share a common three-dimensional fold and the peptide binding mode is highly conserved,[46] homology modeling of any MHC-peptide complex is feasible. To probe our

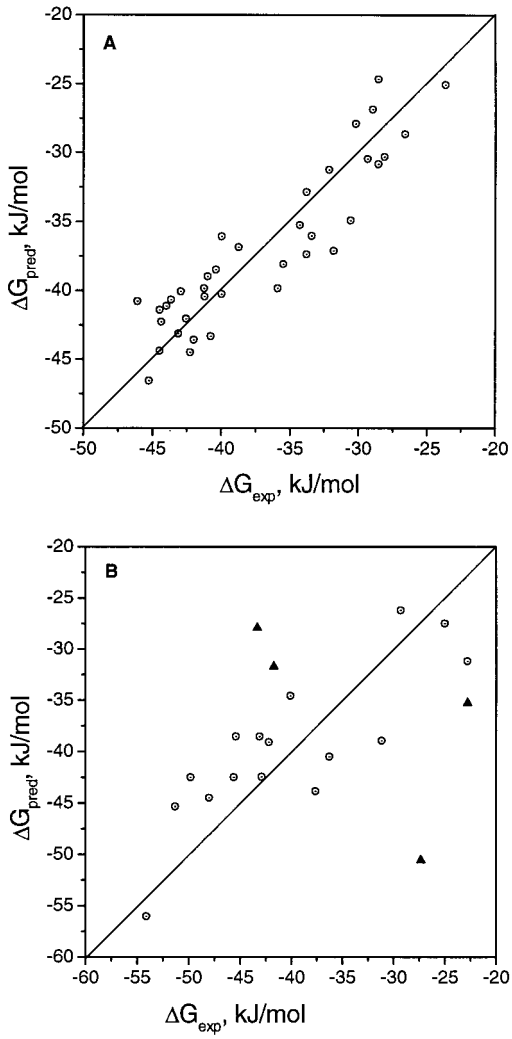

Figure 3. Predicted versus experimental binding free energies for H-2K^k-binding peptides: (a) cross-validated fit to experimental data; (b) prediction of a test set of 16 peptides. Outliers, indicated by up triangles, have been discarded from the regression analysis.

scoring function on 3D-models, we then used a training set of 37 related H-2K^k-binding octapeptides (Figure 3a). H-2K^k is a murine class I MHC protein with a well-defined peptide binding motif[47] but for which no X-ray structure currently exists. The training set is composed of 37 peptides (sequence: Phe^1-Xaa^2-Ser^3-Thr^4-Gly^5-Asn^6-Leu^7-Xaa^8) where both main anchor positions (P2 and P8) to K^k have been systematically varied.[48] All models were built by homology modeling, and new coefficients were derived by multiple regression (eq 2).

$$\Delta G_{bind} = -35.596 - 5.440\text{HB} + 0.344\text{ROT} + 0.063\text{BP} + 0.021\text{DESOLV} \quad (2)$$

$$(q^2 = 0.777,\ s_{press} = 3.157,\ n = 37)$$

An excellent cross-validated fit to experimental binding free energy values could again be obtained (Figure 3a). Surprisingly, it was necessary to discard the lipophilic interaction score from the regression analysis in order to get a physically reasonable coefficient for other terms. In fact, the addition of lipophilic interactions not only decreased the q^2 value but was also found to disfavor binding ($\beta = +0.021$). The best combination was found when these lipophilic interactions were discarded (Figure 4). The suggestion that electrostatic interactions drive the binding of peptides to H-2K^k (H-bond contribution up to −60 kJ/mol) is supported by experimental positional-scanning combinatorial libraries[47] which shows that negatively charged residues are preferred at the major anchoring residue (P_2). However, it should be pointed out that the H-bonding score, alone, was not able to give any reliable correlation (Figure 4). The H-bond coefficient ($\alpha = -5.44$ kJ/mol) has a much higher absolute value than that of the previous equation (Table 1) and is closer to that reported by Böhm and Eldridge for totally different training sets.[6,8] The buried-polar term contributed to about +35 kJ/mol to the total binding free energy whereas rotational and desolvation terms had a lower impact (about +5 kJ/mol) on the final score.

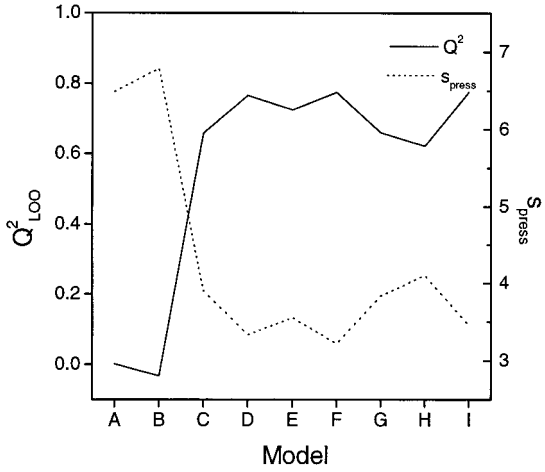

Figure 4. Influence of the regression terms on the predictivity of the H-2K^{k}training model: (A) H-bonding only; (B) H-bonding and lipophilic scores; (C) H-bonding, lipophilic, and rotational scores; (D) H-bonding, lipophilic, rotational, and buried-polar scores; (E) H-bonding, lipophilic, rotational, and desolvation scores; (F) H-bonding, lipophilic, rotational, buried-polar, and desolvation scores; (G) H-bonding, rotational, and buried-polar scores; (H) H-bonding, rotational, and desolvation scores; (I) H-bonding, rotational, buried-polar, and desolvation scores. q^2 and s_{press} values were obtained by a leave-one-out (LOO) cross-validation analysis.

Using the equation derived from the HLA-A2 training set, no correlation could be obtained, thus showing the necessity to recalibrate the scoring function for each new application. The predictivity of the H-2K^k model was further examined for a test set of 16 peptides, a mixture of octa- and nonameric ligands (Figure 3b). After building the three-dimensional models as previously described, a good correlation was found for the test set ($r^2_{pred} = 0.698$, $s = 5.33$ kJ/mol, Figure 3b). The error of prediction is higher than that obtained from the HLA-A2 test set ($s = 3.49$ kJ/mol) and probably reflects more uncertainties in the three-dimensional coordinates of the K^k-peptide models. Furthermore, the path followed by peptides in the K^k binding groove is restrained by the presence of two bulky side chains in the middle of the binding cleft (Arg^{97}, Arg^{155}). Therefore, the predicted free energies are much more sensitive to the sequence and the conformation of the bulging part of the peptide than for HLA-A2-binding peptides. This structural consideration may explain why more outliers (compounds for which $|\Delta G_{exp} - \Delta G_{pred}| > 7.5$ kJ/mol) were

observed in the K^k series (Figure 3b). As described for HLA-A2-binding ligands, it is likely that errors describing the bound conformation of the middle part (P_4 to P_{C-1}) of the K^k-bound peptides are partly responsible for the observed discrepancies.

Conclusion

The present report is to the best of our knowledge one of the very first aimed at a quantitative prediction of absolute binding free energies from three-dimensional homology models. It should be stated that we did not want to develop a universal scoring function but to optimize a precise equation for class I MHC-binding peptides. The explicit treatment of unfavorable protein−ligand contacts by a simple distance-based function is particularly well suited for molecular models where repulsive forces between the host protein and its ligand can be exerted. Estimation of ligand desolvation by a continuum electrostatic method allows the study of polar ligands of peptidic nature. Docking, refinement, and binding free energy estimation can be partly automated and are fast enough (about 30 min per MHC-nonapeptide complex on a R5000 SGI O2 workstation) to be used on a much larger scale for scanning potential MHC-binders and T-cell epitopes from protein amino acid sequences.[49] It has the main advantage over neural networks[50,51] or experimentally derived position-dependent coefficients[47,52] to require very few experimental data for training and is applicable to any protein−ligand complex once the scoring equation has been recalibrated.

Experimental Section

Analysis of Peptide Binding. This was performed essentially as previously described.[47] Briefly, MHC molecules were affinity-purified from appropriate cell lines. The binding of a radioiodinated indicator peptide could be determined by incubating the MHC molecules with the labeled peptide for 48 h at 18 °C followed by a gel filtration size separation of free and MHC-bound peptide and measuring free vs bound peptide by gamma spectrometry. Preliminary experiments determined the concentration of MHC needed to obtain 15−25% binding, thus avoiding ligand depletion. The binding of all other test peptides was subsequently determined by adding increasing concentrations of the peptide in question to the reaction mixture thereby competing with the labeled peptide for binding. The concentration of test peptide needed to effect 50% inhibition (the IC_{50}, an approximation of the K_D) was interpolated. Peptides were synthesized by a standard FMOC chemistry strategy, and peptide purity and identity were ascertained by HPLC and mass spectrometry.

Preparation of 3D-Binary Databases. Three-dimensional binary databases were produced from X-ray structures of class I MHC proteins and MHC-bound peptides using the *mkprodat* utility of the SYBYL 6.4 package (TRIPOS Assoc., Inc., St. Louis, MO). Altogether, 11 class I MHC proteins (HLA-A2.1, HLA-Aw68, HLA-B8, HLA-B27, HLA-B35, HLA-B53, H-2D^b, H-2D^d, H-2K^b, H-2L^d, H-2M3) and 29 class I MHC-bound peptides (7 HLA-A2.1-binding, 1 HLA-Aw68-binding, 5 HLA-B8-binding, 1 HLA-B27-binding, 1 HLA-B35-binding, 2HLA-B53-binding, 2D^b-binding, 1 D^d-binding, 5 K^b-binding, 3 L^d-binding, 1 H-2M3-binding) were extracted from the Protein Data Bank[53] and encoded in two separate databases which will be further used to build by homology all MHC-peptide models studied here.

Coordinates Setup: X-ray Structures. X-ray structures of HLA-A*0201 in complex with 5 different peptides (Table 1) were taken in the first training set. After extracting the crystal coordinates from the Protein Data Bank, hydrogen atoms were added using the SYBYL BIOPOLYMER module. Only the antigen-binding domain (residues 1−182) was used in the calculations.[54] Caution was given to polar hydrogen atoms in order to optimize intra- and intermolecular hydrogen bonds. Hydrogen atoms were then energy-minimized by 100 steps of steepest descent using the Cornell force-field[55] of the AMBER5.0 package.[56] As the minimizations were performed in a vacuum, a distance-dependent dielectric constant ($\epsilon = 4r$) was used, with a twin-cutoff (10−15 Å).

Coordinates Setup: 3D-Models. A 3D-model of the HLA-A*0204 protein was obtained from the X-ray structure of HLA-A*0201 (PDB code: 3hla) by mutating a single residue (Arg^{97} to Met) in the antigen-binding domain. The rotameric state of Met97 was assigned according to HLA-B*5301 (PDB code: 1a1m) after searching the MHC protein binary database for this side chain at position 97. Hydrogen atoms were then added and energy-minimized following the above-described procedure. The H-2K^k protein was built by homology to the X-ray structure of H-2K^b under a previously described procedure.[57]

Whatever the MHC protein, the bound peptide was built in a two-step procedure. Main MHC anchors (P_1, P_2, P_3, P_{C-1}, P_C) were first constructed by homology to the most similar HLA-A*0201-bound peptides for which a X-ray structure has been described (PDB codes: 1hhg, 1hhh, 1hhi, 1hhj, 1hhk, 1a9k, 2clr) or to the K^k-bound conformation of the Influenza Hemagglutinin peptide, $Ha_{255-262}$ (FESTGNLI, one-letter code), for which an experimentally validated model has already been described.[57] Rotameric states of these side anchoring chains were assigned from the peptide binary database. For example, the rotameric state of an arginine at P1 was similar to that found for the same amino acid located at the same position in the PDB entry 2vaa. As the conformation of the MHC-binding part is independent of the peptide sequence,[20] we are pretty confident about the conformation of these MHC anchors. The remaining middle part of the peptide (P_4 to P_{C-2}) may zigzag[19] or bulge[20] out the binding groove using very different paths within the binding groove. Thus, this part was built as a loop connecting positions P_3 to P_{C-1}, using the *loopsearch* utility of SYBYL. In this procedure, a set of 837 high-resolution X-ray structures were searched for a loop of similar length and presenting a similar distance between Cα atoms of the residues delimiting the loop window. Twenty-five loops were selected and clustered into families using a 1.0 Å rms deviation cutoff for backbone coordinates. The loop from the most populated family that shows the highest identity to the target sequence was further selected for insertion. Caution was particularly given to the presence or not of proline residues in the loop sequence. If the target sequence did not contain any proline, proline-containing loops were excluded from the loop search. If the target sequence contained a proline at position P_x, only Pro_x-containing loops were extracted. After loop insertion, side chains were automatically added assuming an extended conformation and quickly refined to the closest minimum by a short conformational scan in order to eliminate steric bumps with the binding groove.

The MHC-bound peptide and the whole MHC-peptide complex were successively relaxed by 1000 steps of AMBER conjugate gradient energy minimization. While fixing protein atoms, the peptide alone was then submitted to a 100 ps simulated annealing (SA) protocol in order to sample the broadest possible conformational space. Starting with random velocities assigned at a temperature of 1000 K, the peptide was first coupled to a heat bath at 1000 K using a temperature coupling constant T_τ of 0.2 ps and then linearly cooled to 50 K for the next 50 ps, while strengthening T_τ to a value of 0.05 ps. During these 100 ps, no protein atom was allowed to move. As the simulated annealing was performed in vacuo, a distance-dependent dielectric function ($\epsilon = 4r$)[58] similar to that used for molecular docking[59] was used. A twin cutoff (10.0, 15.0 Å) was used to calculate nonbonded interactions at every minimization step and every nonbonded pair list update (10 steps), respectively. The last SA conformer was then finally relaxed by 100 steps of conjugate gradient energy refinement.

Regression Terms of the Free Energy Function. As a first step in the investigation of binding interactions, it is useful to assign general atom types to all ligand and receptor atoms. In the present work we adopt the scheme by Eldridge et al.[8] consisting of the four following atom types: lipophilic, H-bond donors, H-bond acceptors, and non-H-bonding polar atoms.

For the full definitions of these atom types we refer to Eldridge et al.[8] It should be noted that the metal atoms discussed by Eldridge et al. have not been applied here. Also, the H-bond donor/acceptor type is assimilated in the present investigation to the H-bond donor type because we define H-bond donor atoms as explicit H-atoms on oxygen or nitrogen.

The terms used in the Fresno scoring function were derived from the work by Eldridge et al.[8] and Böhm,[6] namely H-bond (HB), lipophilic (LIPO), and rotational (ROT) terms. In addition we considered two extra terms, which we defined as the buried-polar (BP) and the desolvation (DESOLV) terms (eq 3).

$$\Delta G_{\text{binding}} = K + \alpha(\text{HB}) + \beta(\text{LIPO}) + \gamma(\text{ROT}) + \delta(\text{BP}) + \gamma(\text{DESOLV}) \quad (3)$$

The number of individual terms tried in the present scoring function was limited to five as we did not want to produce complex equations with several terms that might be collinear. The constant K as well as the regression coefficients $\alpha-\epsilon$ are unknown and will be optimized for each protein−ligand series by multiple linear regression.

The H-bond term (HB) estimates the favorable contribution from hydrogen bonds between the ligand and the receptor. We follow the definition by Böhm,[6] which was also adopted by Eldridge et al.[8] (eq 4) where the summation involves all

$$\text{HB} = \sum_{\text{HB}} g_1(\Delta r)\, g_2(\Delta\alpha) \quad (4)$$

hydrogen bonds between the ligand and the receptor. In eq 4 the functions g_1 and g_2 assign full score to H-bonds of ideal geometry and lower scores to H-bonds deviating from it. g_1 and g_2 are defined as

$$g_1(\Delta r) = \begin{cases} 1 & \text{if } \Delta r \le 0.25 \text{ Å} \\ 1 - (\Delta r - 0.25)/0.4 & \text{if } 0.25 \text{ Å} < \Delta r \le 0.65 \text{ Å} \\ 0 & \text{if } \Delta r > 0.65 \text{ Å} \end{cases} \quad (5)$$

$$g_2(\Delta\alpha) = \begin{cases} 1 & \text{if } \Delta\alpha \le 30° \\ 1 - (\Delta\alpha - 30)/50 & \text{if } 30° < \Delta\alpha \le 80° \\ 0 & \text{if } \Delta\alpha > 80° \end{cases} \quad (6)$$

In eq 4, Δr is the deviation of H-bond length H···X from the ideal value of 1.85 Å. Similarly, $\Delta\alpha$ in eq 6 is the deviation of the H-bond angle X′−H···X from its ideal value of 180°. Following Eldridge et al., we make no distinction between ionic and nonionic hydrogen bonds. In our study, we score H-bonds between the ligand and crystal water molecules as if they were H-bonds between the ligand and the receptor.

The lipophilic term (LIPO) estimates the favorable contribution to binding given by the contacts of lipophilic atoms of the ligand with lipophilic atoms of the receptor. According to the definition by Eldridge et al.,[8] we calculate the contacts between all lipophilic atoms l of the ligand, and lipophilic atoms L of the receptor:

$$\text{LIPO} = \sum_{l,L} f(r_{lL}) \quad (7)$$

where r_{lL} is the distance between atom l and atom L, and the function $f(r)$ assigns a score according to the definition

$$f(r) = \begin{cases} 1 & \text{if } r \le \text{R1} \\ 1 - (r - \text{R1})/3 & \text{if R1} < r \le \text{R2} \\ 0 & \text{if } r > \text{R2} \end{cases} \quad (8)$$

Here, R1 is defined as the sum of the van der Waals radii[60,61] of atoms l and L, plus 0.5, and R2 = R1 + 3, with all distances in angströms.

A buried-polar term (BP) was added to the set of parameters used by Eldridge[8] and Böhm,[6] in an attempt to describe the unfavorable interactions arising from the contact of polar atoms with lipophilic atoms between the ligand and the receptor. Here the polar atoms include H-bond donor and acceptor atoms, as well as non-H-bonding polar atoms. The buried-polar term is fashioned after the lipophilic term of eq 7; however, the contacts are calculated between all lipophilic atoms l of the ligand, and all polar atoms P of the receptor, as well as between all polar atoms p of the ligand, and all lipophilic atoms L of the receptor. The term assumes thus the form

$$\text{BP} = \sum_{l,P} f(r_{lP}) + \sum_{p,L} f(r_{pL}) \quad (9)$$

where the function $f(r)$ is defined in eq 8, as it is for the lipophilic term. Note that this term involves explicitly the non-H-bonding polar atoms, as well as the H-bonding atoms. As a consequence, the buried-polar term expands the contribution of nonlipophilic atoms in this study compared to the previous description of Eldridge.[8] In the latter contribution, the H-bonding atoms are used in the H-bond term and in the rotational term (vide infra), and the non-H-bonding polar atoms are used only in the rotational term.

The rotational term (ROT) estimates the loss of entropy due to the freezing of rotatable bonds of the ligand upon binding. A rotatable bond is any bond between sp^3-sp^3 and sp^3-sp^2 atoms, excluding bonds in rings and bonds to terminal CH_3, CF_3, NH_2, or NH_3 groups. In the present study we tested the term proposed by Eldridge et al.,[8] as well as the earlier expression by Böhm[6] which is a simple count of rotatable bonds (acyclic sp^3-sp^3 and sp^3-sp^2 bonds). Eldridge's term is defined as

$$\text{ROT} = 1 + (1 - 1/N_{\text{rot}}) \sum_{r} (P_{\text{p}}(r) + P'_{\text{p}}(r))/2 \quad (10)$$

where the summation is carried out over the rotatable bonds r that are frozen upon binding. A bond is defined as "frozen" if at least one atom on both sides of the bond is in contact with an atom of the receptor, i.e., at a distance of less than the sum of the relevant van der Waals radii plus 0.5 Å. N_{rot} is the total number of rotatable bonds (frozen and unfrozen), and the terms $P_{\text{p}}(r)$ and $P'_{\text{p}}(r)$ define the fraction (ranging from 0 to 1) of nonlipophilic heavy atoms on either side of the bond r. The terms $P_{\text{p}}(r)$ and $P'_{\text{p}}(r)$ penalize less heavily the lipophilic fragments on each side of the bond. This feature is essential to represent realistically the binding of large lipophilic ligands such as the peptides presented here. We refer the reader to a lengthy discussion on this subject in the relevant paper.[8]

The original rotational term proposed by Böhm,[6] which is simply defined as

$$\text{ROT} = N_{\text{rot}} \quad (11)$$

where N_{rot}, as in eq 10, is the total number of rotatable bonds. In our study the two rotational terms of eqs 10 and 11 are expected to give sizably different estimates, since we are investigating the binding of peptides with many rotatable bonds as well as side chains of various lipophilic character.

The desolvation term (DESOLV) is derived by solving the linear form of the Poisson−Boltzmann equation using the finite-difference method.[62,63] ΔG^0_{reac} is the corrected self-reaction field component of the protein−ligand electrostatic inter-

action energy, meaning the energy required to transfer a molecule from a continuum dielectric (vacuum) to another (water).

$$\mathrm{DESOLV} = \Delta G^0_{\mathrm{reac}} = \Delta G^0_{\mathrm{reac(P-L)}} - \Delta G^0_{\mathrm{reac(P)}} - \Delta G^0_{\mathrm{reac(L)}} \tag{12}$$

Due to the very different sizes of the protein and its ligands, the contribution of the protein−ligand complexes ($\Delta G^0_{\mathrm{reac(P-L)}}$) and of the free protein ($\Delta G^0_{\mathrm{reac(P)}}$) could be omitted from the calculation without affecting the reliability of the results.[64] Thus, this component corresponds here to the free energy of peptide desolvation ($-\Delta G^0_{\mathrm{reac(L)}}$).

Caculation of the Different Energy Scores. After transforming the energy-minimized time-averaged conformation of each complex in the SYBYL (TRIPOS Assoc. Inc.) mol2 format, Fresno directly calculates the HB, LIPO, ROT, and BP scores for the corresponding ligand.

Calculation of Desolvation Energies. Desolvation energies of all ligands were computed using the DelPhi program.[65,66] Peptides were first extracted from the refined MHC-peptide final structure and centered in three-dimensional boxes of resolution 2.0, 1.20, and 1.10 grid points $Å^{-1}$, respectively. For each calculation, 90% of the box was filled with the corresponding molecule. Atomic radii and charges were taken from the AMBER 5.0 parameter set.[55] Inner and outer dielectrics were assigned values of 2.0 and 1.0 (vacuum) or 80 (water environment). An ionic strength of 0.145 M and a ion exclusion radius (Stern layer) of 2.0 Å were used according to previously reported solvent calculations.[66] A probe radius of 1.8 Å was utilized for computing the surface at which the electrostatic potential was extrapolated.

Statistical Analyses. For each MHC-peptide complex, all contributions to the binding free energies (HB, LIPO, ROT, BP, and DESOLV) scores were imported into a SYBYL spreadsheet with the first column figuring the experimental binding free energy used as the dependent column of a PLS analysis.[67] Regression coefficients were derived for each column and optimized for each training set. Leave-one-out cross-validation led to q^2 (cross-validated correlation) coefficient and s_{press} (standard error of prediction) as indicators of the predictivity of the training model

$$q^2 = 1 - \frac{\sum (y_{\mathrm{pred}} - y_{\mathrm{obs}})^2}{\sum (y_{\mathrm{obs}} - y_{\mathrm{mean}})^2} \tag{13}$$

where y_{pred} is a predicted value, y_{obs} is the experimental value, y_{mean} is the best estimate of the mean of all values that might be predicted

$$s_{\mathrm{press}} = \sqrt{\sum (y_{\mathrm{pred}} - y_{\mathrm{obs}})^2/(N - k - 1)} \tag{14}$$

where N is the number of objects and k the number of variables.

Acknowledgment. This work is supported by the Schweizerischer Nationalfonds zur Förderung der wissenschaftlichen Forschung (Project # 31-45504.95) and by the Danish Medical Research Council (Projects # 9501891, 9601615, and 9700901). Authors are grateful to Prof. G. Folkers and Dr. W. Schilling for this collaborative work. D.R. thanks the calculation center of the ETHZ for allocation of computing time on the CRAY-J90 cluster and the Intel Paragon machine.

Supporting Information Available: The sequence and binding free energies (experimental versus predicted) of all peptides studied in the present paper are available free of charge via the Internet at http://pub.acs.org.

References

(1) Kollman, P. Recent advances in structure-based ligand design using molecular dynamics and Monte Carlo methods. *Pharm. Res.* **1998**, *15*, 368−370.
(2) Kollman, P. A. Advances and continuing challenges in achieving realistic and predictive simulations of the properties of organic and biological molecules. *Acc. Chem. Res.* **1996**, *29*, 461−469.
(3) Miyamoto, S.; Kollman, P. A. Absolute and relative binding free energy calculations of the interaction of biotin and its analogues with streptavidin using molecular dynamics/free energy perturbation approaches. *Proteins* **1993**, *16*, 226−245.
(4) Miyamoto, S.; Kollman, P. A. What determines the strength of noncovalent association of ligands to proteins in aqueous solution? *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 8402−8406.
(5) Helms, V.; Wade, R. Computational alchemy to calculate absolute protein−ligand binding free energy. *J. Am. Chem. Soc.* **1998**, *120*, 2710−2713.
(6) Böhm, H. J. The development of a simple empirical scoring function to estimate the binding constant for a protein−ligand complex of known three-dimensional structure. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 243−256.
(7) Böhm, H. J. Prediction of binding constants of protein ligands: A fast method for the prioritization of hits obtained from de novo design or 3D database search programs. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 309−323.
(8) Eldridge, M.; Murray, C. W.; Auton, T. A.; Paolini, G. V.; Lee, R. P. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 425−445.
(9) Murray, C. W.; Auton, T. A.; Eldridge, M. D. Empirical scoring functions. II. The testing of an empirical scoring function for the prediction of ligand−receptor binding affinities and the use of Bayesian regression to improve the quality of the model. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 503−519.
(10) Åqvist, J. A new method for predicting binding affinity in computer-aided drug design. *Protein Eng.* **1994**, *7*, 385−391.
(11) Holloway, M. K.; Wai, J. M.; Halgren, T. A.; Fitzgerald, P. M.; Vacca, J. P.; Dorsey, B. D.; Levin, R. B.; Thompson, W. J.; Chen, L. J.; deSolms, S. J. A priori prediction of activity for HIV-1 protease inhibitors employing energy minimization in the active site *J. Med. Chem.* **1995**, *38*, 305−317.
(12) Krystek, S.; Stouch, T.; Novotny, J. Affinity and specificity of serine endopeptidase-protein inhibitor interactions. Empirical free energy calculations based on X-ray crystallographic structures. *J. Mol. Biol.* **1993**, *234*, 661−679.
(13) Pitera, J.; Kollman, P. A. Designing an optimal guest for a host using multimolecule free energy calculations: Predicting the best ligand for Rebek's "Tennis Ball" *J. Am. Chem. Soc.* **1998**, *120*, 7557−7567.
(14) Vedani, A.; Zbinden, P.; Snyder, J. P.; Greenidge, P. Pseudoreceptor modeling: A new concept for the three-dimensional construction of receptor binding sites. *J. Am. Chem. Soc.* **1995**, *117*, 4987−4994.
(15) Cramer, R. D. I.; Patterson, D. E.; Bunce, J. D. Comparative molecular field analysis (CoMFA). I. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959−5967.
(16) Head, H. J.; Smythe, M. L.; Oprea, T. I.; Waller, C. L.; Green, S. M.; Marshall, G. A. VALIDATE: A new method for the receptor-based prediction of binding affinities of novel ligands *J. Am. Chem. Soc.* **1996**, *118*, 3959−3969.
(17) Zhang, T.; Koshland, D. E. J. Computational method for relative binding energies of enzyme−substrate complexes. *Protein Sci.* **1996**, *5*, 348−356.
(18) Shen, J.; Quiocho, F. A. Calculation of binding energy differences for receptor−ligand systems using the Poisson−Boltzmann method. *J. Comput. Chem.* **1995**, *16*, 445−448.
(19) Guo, H. C.; Jardetzky, T. S.; Garrett, T. P. J.; Lane, W. S.; Strominger, J. L.; Wiley, D. C. Different length peptides bind to HLA-Aw68 similarly at their ends but bulge out in the middle. *Nature* **1992**, *360*, 364−366.
(20) Madden, D. R.; Garboczi, D. N.; Wiley: D. C. The antigenic identity of peptide/MHC complexes: a comparison of the conformations of five viral peptides presented by HLA-A2. *Cell* **1993**, *75*, 693−708.
(21) Caflisch, A.; Niederer, P.; Anliker, P. Monte Carlo docking of oligopeptides to proteins. *Proteins* **1992**, *13*, 223−230.
(22) Rognan, D.; Zimmermann, A.; Jung, J.; Folkers, G. Molecular dynamics study of a complex between the human histocompatibility antigen HLA-A2 and the IMP58-66 nonapeptide from influenza virus matrix protein. *Eur. J. Biochem.* **1992**, *208*, 101−113.
(23) Sezerman, U.; Vajda, S.; Cornette, J.; DeLisi, C. Toward computational determination of peptide-receptor structure. *Protein Sci.* **1993**, *2*, 1827−1843.
(24) Gulukota, K.; Vajda, S.; DeLisi, C. Peptide docking using dynamic programming. *J. Comput. Chem.* **1996**, *17*, 418−428.
(25) Rosenfeld, R.; Zheng, Q.; Vajda, S.; DeLisi, C. Computing the structure of bound peptides. Application to antigen recognition by class I major histocompatibility complex receptors. *J. Mol. Biol.* **1993**, *234*, 515−521.

(26) Sezerman, U.; Vajda, S.; DeLisi, C. Free energy mapping of class I MHC molecules and structural determination of bound peptides. *Protein Sci*. **1996**, *5*, 1271−1281.
(27) Altuvia, Y.; Schueler, O.; Margalit, H. Ranking potential binding peptides to MHC molecules by a computational threading approach. *J. Mol. Biol*. **1995**, *249*, 244−250.
(28) Altuvia, Y.; Sette, A.; Sidney, J.; Southwood, S.; Margalit, H. A structure-based algorithm to predict potential binding peptides to MHC molecules with hydrophobic binding pockets. *Hum. Immunol*. **1997**, *58*, 1−11.
(29) Schueler-Furman, O.; Elber, R.; Margalit, H. Knowledge-based structure prediction of MHC class I bound peptides: a study of 23 complexes. *Folding Des*. **1998**, *3*, 549−564.
(30) Froloff, N.; Windemuth, A.; Honig, B. On the calculation of binding free energies using continuum methods: Application to MHC class I protein-peptide complexes. *Protein Sci*. **1997**, *6*, 1293−1301.
(31) Weng, Z.; DeLisi, C. Toward a predictive understanding of molecular recognition. *Immunol. Rev*. **1998**, *163*, 251−266.
(32) Zhang, C.; Cornette, J. L.; DeLisi, C. Consistency in structural energetics of protein folding and recognition. *Protein Sci*. **1997**, *6*, 1057−1064.
(33) Townsend, A.; Ohlen, C.; Bastin, J.; Ljunggren, H. G.; Foster, L.; Karre, K. Association of class I major histocompatibility heavy and light chains induced by viral peptides. *Nature* **1989**, *340*, 443−448.
(34) Bouvier, M.; Wiley: D. C. Importance of peptide amino and carboxy termini to the stability of MHC class I molecules. *Science* **1994**, *265*, 398−402.
(35) Arnett, K. L.; Parham, P. HLA class I nucleotide sequences, 1995. *Tissue Antigens* **1995**, *46*, 217−257.
(36) Sudo, T.; Kamikawaji, N.; Kimura, A.; Date, Y.; Savoie, C. J.; Nakashima, H.; Furuichi, E.; Kuhara, S.; Sasazuki, T. Differences in MHC class I self-peptide repertoires among HLA-A2 subtypes. *J. Immunol*. **1995**, *155*, 4749−4756.
(37) Garcia, K. C.; Degano, M.; Stanfield, R. L.; Brunmark, A.; Jackson, M. R.; Peterson, P. A.; Teyton, L.; Wilson, I. A. An $\alpha\beta$ T cell receptor structure at 2.5 Å and its orientation in the TCR-MHC complex. *Science* **1996**, *274*, 209−219.
(38) Garboczi, D. N.; Ghosh, P.; Utz, U.; Fan, Q. R.; Biddison, W. E.; Wiley, D. C. Structure of the complex between human T-cell receptor, viral peptide and HLA-A2. *Nature* **1996**, *384*, 134−141.
(39) Garcia, K. C.; Degano, M.; Pease, L. R.; Huang, M.; Peterson, P. A.; Teyton, L.; Wilson, I. A. Structural basis of plasticity in T cell receptor recognition of a self-peptide-MHC antigen. *Science* **1998**, *279*, 1666−1672.
(40) Ruppert, J.; Sidney, J.; Celis, E.; Kubo, R. T.; Grey, H. M.; Sette, A. Prominent role of secondary anchor residues in peptide binding to HLA-A2.1 molecules. *Cell* **1993**, *74*, 929−37.
(41) Levitt, M. Competitive assessment of protein fold recognition and alignment accuracy. *Proteins* **1997**, *1*, 92−104.
(42) Saper, M. A.; Bjorkman, P. J.; Wiley, D. C. Refined structure of the human histocompatibility antigen HLA-A2 at 2.6 Å resolution. *J. Mol. Biol*. **1991**, *75*, 693−708.
(43) Madden, D. R.; Gorga, J. C.; Strominger, J. L.; Wiley: D. C. The three-dimensional structure of HLA-B27 at 2.1 Å resolution suggests a general mechanism for tight peptide binding to MHC. *Cell* **1992**, *70*, 1035−1048.
(44) Falk, K.; Rötzschke, O.; Stevanovic, S.; Jung, G.; Rammensee, H.-G. Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. *Nature* **1991**, *351*, 290−296.
(45) Fremont, D. H.; Stura, E. A.; Matsumura, M.; Peterson, P. A.; Wilson, I. A. Crystal structure of an H-2Kb-ovalbumin peptide complex reveals the interplay of primary and secondary anchor positions in the major histocompatibility complex binding groove *Proc. Natl. Acad. Sci. U.S.A*. **1995**, *92*, 2479−83.
(46) Batalia, M. A.; Collins, E. J. Peptide binding by class I and class II MHC molecules. *Biopolymers* **1997**, *43*, 281−302.
(47) Stryhn, A.; Pedersen, L. O.; Romme, T.; Holm, C. B.; Holm, A.; Buus, S. Peptide binding specificity of major histocompatibility complex class I resolved into an array of apparently independent subspecificities. quantitation by peptide libraries and improved prediction of binding. *Eur. J. Immunol*. **1996**, *26*, 1911−1918.
(48) Stryhn, A.; Andersen, P. S.; Pedersen, L. O.; Svejgaard, A.; Holm, A.; Thorpe, C. J.; Fugger, L.; Buus, S.; Engberg, J. Shared fine specificity between T-cell receptors and an antibody recognizing a peptide/major histocompatibility class I complex. *Proc. Natl. Acad. Sci. U.S.A*. **1996**, *93*, 10338−10342.
(49) Buus, S. Description and prediction of MHC-binding motifs: the "human MHC project". *Curr. Opin. Immunol*. **1999**, *11*, 209−213.
(50) Gulukota, K.; Sidney, J.; Sette, A.; DeLisi, C. Two complementary methods for predicting peptides binding major histocompatibility complex molecules. *J. Mol. Biol*. **1997**, *267*, 1258−1267.
(51) Honeyman, M. C.; Brusic, V.; Stone, N. L.; Harrison, L. C. Neural network-based prediction of candidate T-cell epitopes. *Nature Biotechnol*. **1998**, *16*, 966−969
(52) Parker, K. C, Bednarek, M. A, Coligan, J. E. Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. *J. Immunol*. **1994**, *152*, 163−175.
(53) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F., Jr.; Brice, M. D.; Rodgers, J. M.; Kennard, O.; Shimanouchi, T. M. T. The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol*. **1977**, *112*, 537−542.
(54) Rognan, D.; Scapozza, L.; Folkers, G.; Daser, A. Molecular dynamics simulation of MHC-peptide complexes as a tool for predicting potential T cell epitopes. *Biochemistry* **1994**, *33*, 11476−11485.
(55) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, J., K. M.; Ferguson, D. M.; Spellmeyer, D. M.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids and organic molecules. *J. Am. Chem. Soc*. **1995**, *117*, 5179−5197.
(56) Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham III, T. E.; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Ferguson, D. M.; Radmer, R. J.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. *AMBER 5.0*; University of California: San Francisco, 1997.
(57) Rognan, D.; Stryhn, A.; Fugger, L.; Lynbaeck, S.; Engberg, J.; Andersen, P. S.; Buus, S. Modelling the interactions of a peptide-major histocompatibility class I ligand with its receptors. I. Recognition by two $\alpha\beta$ T cell receptors. *J. Comput.-Aided Mol. Des*. **1999**, in press.
(58) Whitlow, M.; Teeter, M. M. An empirical examination of potential energy minimization using the well-determined structure of the protein crambin. *J. Am. Chem. Soc*. **1986**, *108*, 71−64−7172.
(59) Meng, E. C., Stoichet, B. K., Kuntz, I. D. Molecular docking using shape descritors. *J. Comput. Chem*. **1992**, *13*, 380−397.
(60) Bondi, A. van der Waals volumes and radii. *J. Phys. Chem*. **1964**, *68*, 441−451.
(61) The choice of the van der Waals radii influences the value of the lipophilic, buried-polar, and rotational terms. In the present work we use the following van der Waals radii in angstroms: H, 1.20; C, 1.70; N, 1.55; O, 1.52; F, 1.47; P, 1.80; S, 1.80; Cl, 1.75; Br, 1.85; I, 1.80.
(62) Warwicker, J.; Watson, H. C. Calculation of the electric potential in the active site cleft due to alpha-helix dipoles. *J. Mol. Biol*. **1982**, *157*, 671.
(63) Klapper, I.; Hagstrom, R.; Fine, R.; Sharp, K.; Honig, B. Focusing of electric fields in the active site of Cu−Zn superoxide dismutase: effects of ionic strength and amino acid modification. *Proteins* **1986**, *1*, 47−59.
(64) Taylor, N. R.; von Iztein, M. A structural and energetics analysis of the binding of a series of *N*-acetylneuramic-acid-based inhibitors to influenza virus sialidase. *J. Comput.-Aided Mol. Des*. **1996**, *10*, 233−246.
(65) Gilson, M. K.; Honig, B. Calculation of the total electrostatic energy of a macromolecular system: solvation energies, binding energies, and conformational analysis. *Proteins* **1988**, *4*, 7−18.
(66) Gilson, M. K.; Sharp, K.; Honig, B. Calculating electrostatic interactions in bio-molecules: Method and error assessment. *J. Comput. Chem*. **1988**, *9*, 327−335.
(67) Wold, S.; Johansson, E.; Cocchi, C. PLS − Partial least-squares projections to latent structures. In *3D QSAR in Drug Design*; Kubinyi, H., Ed.; ESCOM Science Publishers: Leiden, 1993; pp 523−550.
(68) Sette, A.; Sidney, J.; Del Guercio, M. F.; Southwood, S.; Ruppert, J.; Dahlberg, J.; Grey, H. M.; Kubo, R. T. Peptide binding to the most frequent HLA. A class I alleles measured by quantitative molecular binding assays. *Mol. Immunol*. **1994**, *31*, 813−822.

JM9910775