Toward the Quantitative Prediction of T-Cell Epitopes: CoMFA and CoMSIA Studies of Peptides with Affinity for the Class I MHC Molecule HLA-A*0201

Irini A. Doytchinova*,† and Darren R. Flower

Edward Jenner Institute for Vaccine Research, Compton, Berkshire, RG20 7NN, UK

Received January 16, 2001

A set of 102 peptides with affinity for the class I MHC HLA-A*0201 molecule was subjected to three-dimensional quantitative structure−affinity relationship (3D QSAR) studies using comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA). A test set of 50 peptides was used to determine the predictive value of the models. The CoMFA models gave q^2 and r^2pred below 0.5. The best CoMSIA model has $q^2 = 0.542$ and r^2pred $= 0.679$, and includes hydrophobic, steric, and H-bond donor fields. The hydrophobic interactions play a dominant role in peptide−MHC molecule binding. CoMSIA coefficient contour maps were used to analyze the structural features of the peptides accounting for the affinity in terms of the three positively contributing physicochemical properties: local hydrophobicity, steric bulk and hydrogen-bond-donor ability.

Introduction

The membrane-bound proteins encoded by the major histocompatibility complex (MHC) function as antigen-presenting markers through which the immune system distinguishes body cells from invading antigens (class I MHC proteins) and immune system cells from other cells (class II MHC proteins). They form stable complexes with proteolytically digested protein fragments composed of approximately 8−10 (for class I)[1] and 13−17 (for class II)[2] amino acids. The MHC complexes are synthesized intracellularly, bind peptide in the endoplasmic reticulum, and are exported to the cell surface, where they are recognized by T-cells. Thus, T-cell-mediated immunity is governed by a multimolecular interaction involving antigenic peptides, MHC molecules, T-cell receptors (TCRs), and a number of coreceptors such as CD4 and CD8.

The short peptides presented to TCRs are termed epitopes. A subset of peptides—termed T-cell epitopes—bind well to MHC molecules, are recognized by TCRs, and elicit T-cell responses. Highly immunogenic proteins, which may prove to be useful putative vaccines, contain significant numbers of epitopes that bind MHCs well, as well as T-cell epitopes. The accurate prediction of peptide binding to MHC molecules is thus an important and necessary, but not sufficient, prerequisite to accurate epitope prediction and the design of vaccines.

The class I MHC proteins are encoded by three separate but homologous genetic loci, HLA-A, HLA-B, and HLA-C.[3] The gene map of the MHC region has been reported recently.[4] There are also three human class II MHC proteins encoded by different genes. A striking feature of MHC genes is their high level of polymorphism among individuals of the same species, and they are generally regarded as the most polymorphic genes of higher vertebrates. In humans, for example, over 86 A alleles, 185 B alleles, and 45 C alleles have been characterized.[5]

The allele HLA-A*0201 is one of the most frequent class I alleles in many different populations.[6] For example, it is expressed in approximately 50% of Caucasians[7] and it has been demonstrated to play a critical role in antigen presentation of both viral antigens[8] and tumor antigens from a variety of cancers.[9−12] Its three-dimensional structure, complexed with a limited number of different peptide ligands, has been solved by X-ray crystallography,[13,14] and its ligand specificity has been analyzed by studies describing its naturally processed ligands[15−17] as well as by quantitative in vitro binding assays.[18,19]

Peptides that bind to HLA-A*0201 have a restricted size of 9 ± 1 amino acids and require free N- and C-termini. In addition to a specific size, a combination of two main anchor residues is required. These anchors have been described as Leu at position 2 and Leu or Val at the C-terminal end.[20] The presence of anchors is necessary, but not sufficient, for high-affinity binding. Prominent roles for several other positions (1, 3, and 7), so-called secondary anchor residues, has also been demonstated.[14,16,21] Although a large number of peptides have been synthesized and tested, relatively little is known about the nature of the forces involved in the peptide−MHC molecule interaction. To explore this adequately, the methods of 3D QSAR have been used to study these interactions in a quantitative manner. They provide easily interpretable coefficient contour maps identifying those areas of the peptides that require a particular physicochemical property to increase binding.

In this paper, two 3D QSAR methods were applied, CoMFA and CoMSIA, to investigate the local physicochemical properties involved in the interaction between the peptide and the MHC molecule. The widely used CoMFA (*co*mparative *m*olecular *f*ield *a*nalysis) method calculates steric and electrostatic properties according

* Corresponding author. Tel: +44 1635 577954. Fax +44 1635 577908. E-mail:irini.doytchinova@jenner.ac.uk.
† Department of Chemistry, Faculty of Pharmacy, Medical University, 1000 Sofia, Bulgaria.

10.1021/jm010021j CCC: $20.00
Published on Web 09/19/2001

to Lennard-Jones and Coulomb potentials.[22] The more recently reported CoMSIA approach (*co*mparative *m*olecular *s*imilarity *i*ndices *a*nalysis) calculates similarity indices in the space surrounding each of the aligned molecules in the data set.[23−25] In the present study both 3D QSAR methods were used to develop quantitative models for HLA-A*0201 binding, allowing us to predict the affinity of new, untested peptides and, through the analysis of contribution maps, to suggest new ideas for further synthesis of high-affinity peptides.

Computational Methods

Peptides and Binding Affinities. The structures of 152 peptides used in the study and their binding affinities were compiled from different papers.[10−12,26−32] All the peptides chosen consisted of nine amino acids. The binding affinities (IC_{50}) were assessed by a quantitative assay based on the inhibition of binding of a radiolabeled standard peptide (FLPSDYFPSV) to detergent-solubilized MHC molecules described elsewhere.[15,16] The log values of $1/IC_{50}$ (pIC_{50}) were used in the QSAR correlations, as they are related to changes in the free energy of binding. Table 1 lists the structures and the experimental pIC_{50} of the peptides used in the study. The peptides were divided into a training and test set. One hundred and two peptides where used as the training set and 50 peptides as the test set. Two criteria were applied in the choice of test peptides. First, that the range of binding affinities in the test set should not exceed the range of affinities in the training set. Second, that each amino acid at each position in the test set should also be present at that position in the training set.

Molecular Modeling and Alignment Rule. All molecular modeling calculations were performed on a Silicon Graphics octane workstation using SYBYL 6.6 molecular modeling software.[33] The X-ray structure of the nonameric viral peptide TLTSCNTSV[14] was used as the starting conformation. Unfortunately, any quantitative data for its binding affinity was available, except for the information that this peptide binds the HLA-A*0201 molecule. Because of that, this peptide was not included in either the training or the test set. The structures of the remaining peptides were built using the BIOPOLYMER option in SYBYL and were subjected to full geometry optimization using the standard Tripos molecular mechanics force field (Powell method,[34] no electrostatics and 0.05 kcal/mol Å energy gradient convergence criterion). The peptide backbone was fixed in the X-ray conformation, using the option "Aggregates" in the Minimize Energy menu. The aggregate consists of the α-carbon atoms, the carbonyl carbon and oxygen atoms, and the amide nitrogen and hydrogen atoms. The side chain rotamers were fixed in an extended conformation before the optimization. The partial atomic charges used in CoMFA and CoMSIA were computed using the AM1 semiempirical method[35] available in the MOPAC program. MOPAC V6 was used as implemented in SYBYL. Single-point calculations were performed.

The alignment of the 152 peptides was based on the corresponding backbone atoms (the same as in the aggregate) in the conformation derived by X-ray data.

CoMFA Settings. CoMFA was performed using the QSAR option of SYBYL version 6.6. The steric and electrostatic probe-ligand interaction energies (kcal/mol) were calculated using a Lennard-Jones 6−12 potential and a Coulombic potential with a $1/r$ distance-dependent dielectric, respectively. The steric and electrostatic energies were truncated at 30 kcal/mol. The probe atom was a sp^3 carbon with a +1 charge.

CoMSIA Settings. CoMSIA was performed using the QSAR option of SYBYL version 6.6. Five physicochemical properties (steric, electrostatic, hydrophobic, and hydrogen-bond donor and acceptor) were evaluated, using a common probe atom with 1 Å radius, charge +1, hydrophobicity +1, hydrogen-bond donor and acceptor properties +1. Similarity indices were calculated using Gausian-type distance dependence between the probe and the atoms of the molecules of the data set. This functional form requires no arbitrary definition of cutoff limits, and the similarity indices can be calculated at all grid points inside and outside the molecule. The value of the attenuation factor α was set to 0.3.

PLS Calculations and Validations. PLS methodology was used for all 3D QSAR analyses. The grid had a resolution of 2.0 Å and extended beyond the molecular dimensions by 4.0 Å in all directions. Column filtering was set to 2.0 kcal/mol. CoMFA and CoMSIA models were developed using the conventional stepwise procedure. The optimum number of components used do derive the nonvalidated model was defined as the number of components leading to the highest cross-validated r^2 (q^2) and the lowest standard error of prediction (SEP). The q^2 values were derived after "leave-one-out" cross-validation (LOO-CV). A more robust CV test using "leave-half-out" (LHO-CV) was performed to estimate the extent of chance correlation in the best model. The mean of q^2 values from 100 runs is given as q^2LHO. The non-cross-validated models were assessed by the explained variance r^2, standard error of estimate (S), and F ratio. A bootstrap analysis[36] was performed in 10 runs for the best model, and the mean r^2 is given as r^2bootstrap. The non-cross-validated analyses were used to make predictions of the binding affinities of the peptides from the test set and to display the coefficient contour maps. The actual versus predicted binding affinities of the test peptides were fitted by linear regression, and the "predictive" r^2, S, and F ratio were determined.

QSAR Coefficient Contour Maps. The visualization of the results of the best CoMSIA model (hydrophobic + steric + H-bond donor fields) has been performed using the "StDev*Coeff" mapping option contoured by contribution. Favored and disfavored levels fixed at 80% and 20%, respectively, were used for the steric and hydrophobic fields. The hydrogen-bond-donor field is presented in a map with favored and disfavored levels fixed at 80% and 0%, respectively. The contours of the CoMSIA steric maps are shown in green (more bulk is favored) and yellow (less bulk is favored). The hydrophobic fields are colored yellow (hydrophobic amino acids enhance affinity) and white (hydrophilic groups enhance affinity). The hydrogen-bond field contours show regions where hydrogen-bond acceptors (cyan) on the receptor enhance the binding.

Results

Peptides associated with class I MHC proteins assume nearly extended but twisted conformations.[14] As a result, consecutive peptide side chains protrude in largely opposite directions, analogous to the alternating side chains of a β strand. After the full geometry optimization, a rotation at almost 90° from the initial 180° of the aromatic rings of Phe and Tyr at P1, P4, P6, and P7 was the most significant change in the side chains conformations. This resulted from the residues adopting a minimum-energy conformation as defined by the force field we have used. This distortion was less expressed in the aromatic side chains at P3, P5 (because of the steric hindrance between them), and P8.

The training set was checked initially for outliers. Peptides with residuals between experimental and predicted pIC_{50} values above 1 log unit were considered as outliers. For this purpose models including only one field (monomodels) were developed—two CoMFA and five CoMSIA models. Two peptides (**2**, VCMTVDSLV; **3**, HLESLFTAV) were outliers in all models, one (**10**, ALPYWNFAT) was an outlier in six models, and one (**8**, AMFQDPQER) in five. There are three possible reasons why these peptides are outliers: an incorrectly measured experimental value, a different binding conformation, or a significant difference in the physicochemical properties. The last reason seems to be im-

Table 1. Peptides Used in the Study, Their Experimental pIC_{50} Values, and the Differences between Experimental and Predicted pIC_{50} for the Hydrophobic + Steric + H-Bond-Donor CoMSIA Fields Model

no.	peptide	expl pIC_{50}	refs	differences according to hydrophobic + steric + H-bond-donor fields model	no.	peptide	expl pIC_{50}	refs	differences according to hydrophobic + steric + H-bond-donor fields model
				Training Set					
				Low Affinity: $pIC_{50} < 6.301$ ($IC_{50} > 500$ nM)					
1	VALVGLFVL	5.146	12	−1.294	**10**	ALPYWNFAT	5.869	12	outlier (−0.862)[b]
2	VCMTVDSLV	5.146	12	outlier (−1.550)[b]	**11**	SLNFMGYVI	5.881	29	0.116
3	HLESLFTAV	5.301	29	outlier (−2.149)[b]	**12**	NLQSLTNLL	6.000	29	−0.494
4	GTLVALVGL	5.342	12	−0.097	**13**	FVTWHRYHL	6.025	12	−0.370
5	LLSCLGCKI	5.447	12	−1.356	**14**	WLEPGPVTA	6.082	28	−1.141
6	LQTTIHDII	5.501	26	−0.156	**15**	QVMSLHNLV	6.170	12	−0.691
7	TLLVVMGTL	5.580	12	−1.236	**16**	DPKVKQWPL	6.176	30	0.239
8	AMFQDPQER	5.740	26	outlier (−1.408)[b]	**17**	ITSQVPFSV	6.196	28	−0.323
9	SLHVGTQCA	5.842	11	−0.291	**18**	ALAKAAAAI	6.211	30	−0.507
				Intermediate Affinity: $7.301 > pIC_{50} > 6.301$ (50 nM $< IC_{50} <$ 500 nM)					
19	GLGQVPLIV	6.301	31	−1.230	**39**	TLGIVCPIC	6.815	26	0.427
20	MLDLQPETT	6.335	26	−0.152	**40**	CLTSTVQLV	6.832	10	0.139
21	LLSSNLSWL	6.342	29	−0.050	**41**	ILLLCLIFL	6.845	27	0.329
22	GLACHQLCA	6.380	10	−0.734	**42**	FLCKQYLNL	6.875[a]	27, 29	−0.674
23	LIGNESFAL	6.415	12	−1.126	**43**	FAFRDLCIV	6.886	26	−0.350
24	ALAKAAAAV	6.419	30	−0.592	**44**	FLEPGPVTA	6.898	28	−0.349
25	LLAVGATKV	6.477	31	1.088	**45**	ALAKAAAAA	6.947	30	0.148
26	KLPQLCTEL	6.484	26	−0.831	**46**	LMAVVLASL	6.954	31	0.094
27	ALAKAAAAL	6.511	30	0.306	**47**	YVITTQHWL	6.983	12	−0.098
28	WILRGTSFV	6.556	27	−0.030	**48**	LLCLIFLLV	6.996	27	−0.355
29	IISCTCPTV	6.580	27	0.025	**49**	HLAVIGALL	7.000	31	0.763
30	FLGGTPVCL	6.623	27	−0.265	**50**	ITAQVPFSV	7.020	28	0.272
31	ALIHHNTHL	6.623	10	0.294	**51**	YLEPGPVTL	7.058	28	0.326
32	NLSWLSLDV	6.639[a]	27, 29	−0.131	**52**	YTDQVPFSV	7.066	28	−0.305
33	YMIMVKCWM	6.663	10	−0.327	**53**	NLYVSLLLL	7.114	29	0.090
34	VLQAGFFLL	6.682	27	0.551	**54**	ILHNGAYSL	7.127	10	0.041
35	GTLGIVCPI	6.714	26	0.483	**55**	SIISAVVGI	7.159	10	−0.029
36	VILGVLLLI	6.785	11	−1.259	**56**	VVMGTLVAL	7.174	12	0.528
37	VTWHRYHLL	6.793	12	−0.366	**57**	YLEPGPVTI	7.187	28	0.060
38	PLLPIFFCL	6.796	27	−0.597	**58**	GLSRYVARL	7.248[a]	27, 29	0.705
				High Affinity: $pIC_{50} > 7.301$ ($IC_{50} <$ 50 nM)					
59	LLAQFTSAI	7.301	29	0.053	**81**	FVWLHYYSV	7.824	12	−0.100
60	VLLDYQGML	7.328	27	−0.459	**82**	MLGTHTMEV	7.845	31	0.647
61	YLEPGPVTV	7.342	28	−0.068	**83**	LLFGYPVYV	7.886	30	−0.659
62	ILSPFMPLL	7.347	27	0.221	**84**	ILKEPVHGV	7.921	30	0.414
63	YLSPGPVTA	7.383	28	0.063	**85**	YLMPGPVTV	7.932	28	−0.138
64	ALAKAAAAM	7.398	30	1.184	**86**	WLDQVPFSV	7.939	28	0.303
65	IIDQVPFSV	7.398	28	−0.366	**87**	KTWGQYWQV	7.955	31	1.030
66	SVYDFFVWL	7.444	12	0.495	**88**	ALMPLYACI	8.000	29	0.320
67	ITWQVPFSV	7.463	28	0.204	**89**	YLAPGPVTA	8.032	28	0.349
68	ITYQVPFSV	7.480	28	0.034	**90**	YLYPGPVTV	8.051	28	0.005
69	GLYSSTVPV	7.481	27	−0.504	**91**	LLMGTLGIV	8.097	26	0.916
70	VMGTLVALV	7.553	12	−0.232	**92**	YLWPGPVTV	8.125	28	−0.122
71	LLLCLIFLL	7.585	27	0.274	**93**	FLLTRILTI	8.149	27	0.104
72	SLDDYNHLV	7.585	12	0.893	**94**	GLLGWSPQA	8.237	27	0.313
73	VLIQRNPQL	7.644	10	−0.196	**95**	ILYQVPFSV	8.310	28	−0.398
74	SLYADSPSV	7.658[a]	27,29	−0.103	**96**	GILTVILGV	8.347	11	0.578
75	RLLQETELV	7.682	10	0.291	**97**	NMVPFFPPV	8.398	12	−0.189
76	ILSQVPFSV	7.699	28	0.217	**98**	ILDQVPFSV	8.481	28	0.658
77	IMDQVPFSV	7.719	28	0.075	**99**	YLFPGPVTA	8.495	28	0.503
78	QLFEDNYAL	7.764	10	1.079	**100**	YLDQVPFSV	8.638	28	0.125
79	ALMDKSLHV	7.770	11	0.880	**101**	ILFQVPFSV	8.699	28	0.187
80	YAIDLPVSV	7.796	12	0.317	**102**	ILWQVPFSV	8.770	28	0.448
						MEAN	7.068		−0.012
						ST_DEV	0.856		0.551
						HIGH	8.770		1.184
						LOW	5.146		−1.356
				Test Set					
				Low Affinity: $pIC_{50} < 6.301$ ($IC_{50} > 500$ nM)					
103	LLGCAANWI	5.301	29	−0.991	**108**	LLVVMGTLV	5.869	12	−1.053
104	SAANDPIFV	5.342	12	−0.841	**109**	GIGILTVIL	6.000	11	−0.137
105	TTAEEAAGI	5.380	11	−0.329	**110**	TVILGVLLL	6.072	11	−0.103
106	LTVILGVLL	5.580	11	−0.084	**111**	WTDQVPFSV	6.145	28	−0.576
107	HLLVGSSGL	5.792	29	−0.582	**112**	AIAKAAAAV	6.176	30	−0.716

Table 1 (Continued)

no.	peptide	expl pIC_{50}	refs	differences according to hydrophobic + steric + H-bond-donor fields model	no.	peptide	expl pIC_{50}	refs	differences according to hydrophobic + steric + H-bond-donor fields model
				Intermediate Affinity: 7.301 > pIC_{50} > 6.301 (50 nM < IC_{50} < 500 nM)					
113	VLHSFTDAI	6.380	12	−1.129	**123**	TLDSQVMSL	6.793	12	−0.200
114	AAAKAAAAV	6.398	30	0.195	**124**	HLYQGCQVV	6.832	10	−0.209
115	ILTVILGVL	6.419	11	0.249	**125**	QLFHLCLII	6.886	27	−0.502
116	MLLAVLYCL	6.477	32	−0.727	**126**	ITDQVPFSV	6.947[a]	28, 31, 32	0.184
117	AVAKAAAAV	6.495	30	−0.155	**127**	ALCRWGLLL	7.000	10	0.285
118	AAGIGILTV	6.581[a]	11, 32	0.456	**128**	NLGNLNVSI	7.119	29	−0.467
119	ILDEAYVMA	6.623	10	−0.535	**129**	HLYSHPIIL	7.131	29	−0.276
120	YLEPGPVTA	6.668[a]	28, 31, 32	−0.633	**130**	ITFQVPFSV	7.179	28	−0.199
121	LLWFHISCL	6.682	27	−0.222	**131**	FTDQVPFSV	7.212	28	0.117
122	TLHEYMLDL	6.726	26	−0.518					
				High Affinity: pIC_{50} > 7.301 (IC_{50} < 50 nM)					
132	YMLDLQPET	7.310	26	−1.007	**143**	VVLGVVFGI	7.845	10	0.058
133	RLMKQDFSV	7.342	31	−0.520	**144**	MMWYWGPSL	7.921	27	0.669
134	KLHLYSHPI	7.352[a]	27, 29	−0.079	**145**	ILAQVPFSV	7.939	28	−0.041
135	ITMQVPFSV	7.398	28	0.762	**146**	WLSLLVPFV	8.048[a]	27, 30	0.632
136	YMNGTMSQV	7.398	32	0.640	**147**	FLLSLGIHL	8.053[a]	27, 29	0.086
137	KIFGSLAFL	7.478	10	0.211	**148**	ILMQVPFSV	8.125	28	0.315
138	ALVGLFVLL	7.585	12	0.359	**149**	YLFPGPVTV	8.237	28	0.205
139	YLSPGPVTV	7.642	28	0.353	**150**	YLMPGPVTA	8.367	28	0.513
140	GLYSSTVPV	7.699	29	−0.414	**151**	YLWPGPVTA	8.495	28	0.429
141	YLYPGPVTA	7.772	28	−0.124	**152**	FLDQVPFSV	8.658	28	0.405
142	YLAPGPVTV	7.818	28	0.146					
						MEAN	7.014		−0.122
						ST_DEV	0.862		0.489
						HIGH	8.658		0.762
						LOW	5.301		−1.129

[a] The IC_{50} value is an average value of all the cited IC_{50} values. [b] Predicted pIC_{50} for the initial outliers.

portant for peptides **2**, **8**, and **10** because they include a single presentation of a particular amino acid in the training set amino acids (Cys at the 2nd position in **2**, Arg at the 9th position in **8**, and Thr at the 9th position in **10**). His at the 1st position in **3** is also rarely presented; there are only two peptides in the training set, and the other one, peptide **49**, behaves normally. The four outliers were dropped from the training set, and the further development of the models was performed on the training set consisting of 98 peptides.

CoMFA Study. The steric and electrostatic CoMFA monomodels have $q^2 = 0.477$, NC = 4, $r^2 = 0.827$ and $q^2 = 0.477$, NC = 4, $r^2 = 0.768$, respectively. Because of the high percentage of explained variance, it is difficult to conclude what is the major property influencing the binding affinity.

As was expected, the combination between the monomodels did not give any improvement ($q^2 = 0.480$, NC = 6, $r^2 = 0.911$).

The test set (peptides **103**−**152)** was used to evaluate the predictive power of the non-cross-validated CoMFA monomodels. Although the monomodels had high r^2 values, the predictions made by them gave "predictive" r^2 below 0.5. Therefore, the CoMFA monomodels are probably overestimated due to introduced "noise".

CoMSIA Study. The CoMSIA study was performed using the same PLS protocol and stepwise procedure as in the CoMFA analysis. The same test set of 50 peptides was used to evaluate the predictive power of the non-cross-validated CoMSIA models.

The stepwise development of the CoMSIA models is presented in Table 2. The best model has a q^2 value of 0.542 and includes hydrophobic, steric, and electrostatic fields. The contribution of the steric and H-bond-donor fields in the improvement of q^2 and r^2 is small. Interestingly, the all-fields model has lower q^2 although it explains 89% of the variance. This means that the electrostatic and H-bond-acceptor fields do not only contribute essentially nothing but they may even reduce the correlation.

We used the best model (hydrophobic + steric + H-bond-donor fields) to predict the binding affinities of the test set. The values are given in Table 1. Forty seven peptides (94%) have differences below 1 log unit. The graph of predicted-versus-measured pIC_{50} values is presented as Figure 1. The "predictive" $r^2 = 0.679$, $s = 0.390$, $F = 101.535$. The predicted pIC_{50} values for the training set are also shown in Table 1.

3D QSAR Coefficient Contour Maps. The three physicochemical properties having positive contributions to the affinity are presented in coefficient contour maps, generated by the hydrophobic + steric + H-bond-donor fields model. Peptide **100** (YLDQVPVTA) is shown inside the different fields. The peptide is positioned with the N-terminus to the left and the C-terminus to the right as it is oriented in the binding cleft on the HLA-A*0201 molecule.[37] The fields are considered in the order of their significance for the affinity.

Local Hydrophobicity. This property is of the greatest importance for the binding affinity (Figure 2, upper). The well-tolerated hydrophobic groups are shown as a yellow contour. Such areas exist at positions 2 (P2), 3 (P3), 7 (P7), and 9 (P9). The favored hydrophilic groups are shown as white contours. They are at positions 4 (P4), 5 (P5), 8 (P8), and far from P9.

Steric Bulk. Sterically favorable regions are depicted in green, whereas disfavorable regions are in yellow (Figure 2, middle). Green areas exist at P1, P2, and

Table 2. Stepwise Development of the CoMSIA Models

fields	CV-LOO			no validation			fractions				
	NC	q^2	SEP	r^2	S	F	steric	electrostatic	hydrophobic	H-bond donor	H-bond acceptor
steric	6	0.374	0.662	0.797	0.376	59.680	1.000				
electrostatic	2	0.207	0.729	0.508	0.574	493.040		1.000			
hydrophobic	4	0.529	0.568	0.795	0.374	90.306			1.000		
H-bond donor	3	0.155	0.756	0.557	0.548	39.328				1.000	
H-bond acceptor	2	0.198	0.733	0.436	0.615	36.680					1.000
hydrophobic + steric	4	0.530	0.567	0.820	0.351	106.133	0.282		0.718		
hydrophobic + electrostatic	3	0.430	0.621	0.763	0.400	101.132		0.455	0.545		
hydrophobic + H-bond donor	4	0.504	0.583	0.829	0.342	113.071			0.598	0.402	
hydrophobic + H-bond acceptor	4	0.462	0.607	0.806	0.364	96.568			0.696		0.304
hydrophobic + steric + electrostatic	3	0.466	0.601	0.792	0.376	119.065	0.162	0.369	0.469		
hydrophobic + steric + H-bond donor	5	0.542	0.563	0.870	0.294	129.080	0.190		0.472	0.338	
		0.423[a]		0.923[b]							
hydrophobic + steric + H-bond acceptor	4	0.507	0.581	0.831	0.340	114.426	0.219		0.529		0.252
hydrophobic + steric + H-bond donor + electrostatic	5	0.510	0.582	0.883	0.284	139.068	0.129	0.273	0.341	0.257	
hydrophobic + steric + H-bond donor + H-bond acceptor	3	0.505	0.579	0.798	0.370	123.916	0.150		0.385	0.282	0.184
all fields	5	0.486	0.596	0.885	0.282	141.648	0.116	0.240	0.290	0.224	0.129

[a] q^2_{LHO}: the mean q^2 value from 100 runs of "leave-half-out" cross-validation. [b] $r^2_{bootstrap}$: the mean r^2 value from 10 runs of bootstrap analysis.

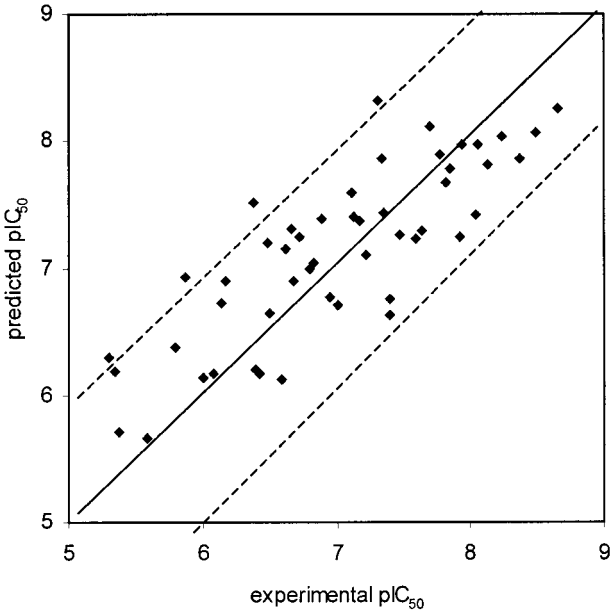

Figure 1. Predicted versus experimental pIC_{50} values for the test set of 50 peptides. Predictions are made according to the hydrophobic + steric + H-bond-donor fields CoMSIA model, predictive $r^2 = 0.679$. The upper and lower lines mark deviations of 1 log unit from the ideal prediction.

between P3, P5, and P7. Three large yellow disfavored areas exist at P4, P8, and P9 and two smaller at P3 and P7. The disfavored area at P7 is in the plane of these aromatic rings that are perpendicular to the extended side chain.

Hydrogen-Bond-Donor Ability. In CoMSIA the hydrogen-bond-donor field describes areas where hydrogen-bond-acceptor groups should be located on the receptor[38] (Figure 2, lower). These areas are colored cyan. Such groups are favored near the N terminal, in the vicinity of the OH-group of Tyr at P1, between P3, P5, and P7, near P6, and beyond P8.

Discussion

Predictability of the 3D QSAR Models. As 3D QSAR analyses, CoMFA and CoMSIA have many common features. They both use huge matrixes of data generated at regularly spaced grid points using some distance dependent function. For the statistical calculation of the structure−activity relationship, they both use a PLS protocol able to cope with the large volume of data. To obtain high predictive power they both need a perfect preliminary alignment of the structures included in the investigated set. In general, their technical performance is identical.

The main differences between CoMFA and CoMSIA are in the form of the defined probe and in the type of the distance dependent function used. In CoMFA, the probe atom is an sp^3 carbon with a +1 charge. For each molecule belonging to the set under study, two values of the interaction energy are calculated at each grid point—one a van der Waals Lennard-Jones interaction and one an electrostatic Coulombic interaction. Because of the hyperbolic functional form, both potentials obtain very large, essentially nonsensical values at or beyond the van der Waals surface. To avoid these values, arbitrarily fixed cutoffs are defined (here 30 kcal/mol for the both functions). Due to the different slopes of the potentials, these cutoffs are exceeded for the different terms at different distances from the molecules.[39]

In CoMSIA, similarity indices are calculated instead of interaction energies. Each molecule from the training set is compared to a common probe with a radius of 1 Å and charge, hydrophobicity, and hydrogen-bond property equal to +1. The functional form here is selected to be Gaussian with an attenuation factor $\alpha = 0.3$.[23] Compared to the Lennard-Jones and Coulomb potentials, the Gaussian-type function has the advantage of using all grid points inside and outside the molecules, and no arbitrary cutoffs are required. Five different similarity fields are calculated: steric, electrostatic, hydrophobic, hydrogen-bond donor, and hydrogen-bond acceptor. These fields cover the major contributions to ligand−protein binding.[24]

The predictability of the models derived by CoMFA and CoMSIA is the most important criterion for assessment of both methods. The present study indicates that the steric and electrostatic CoMFA fields are not enough

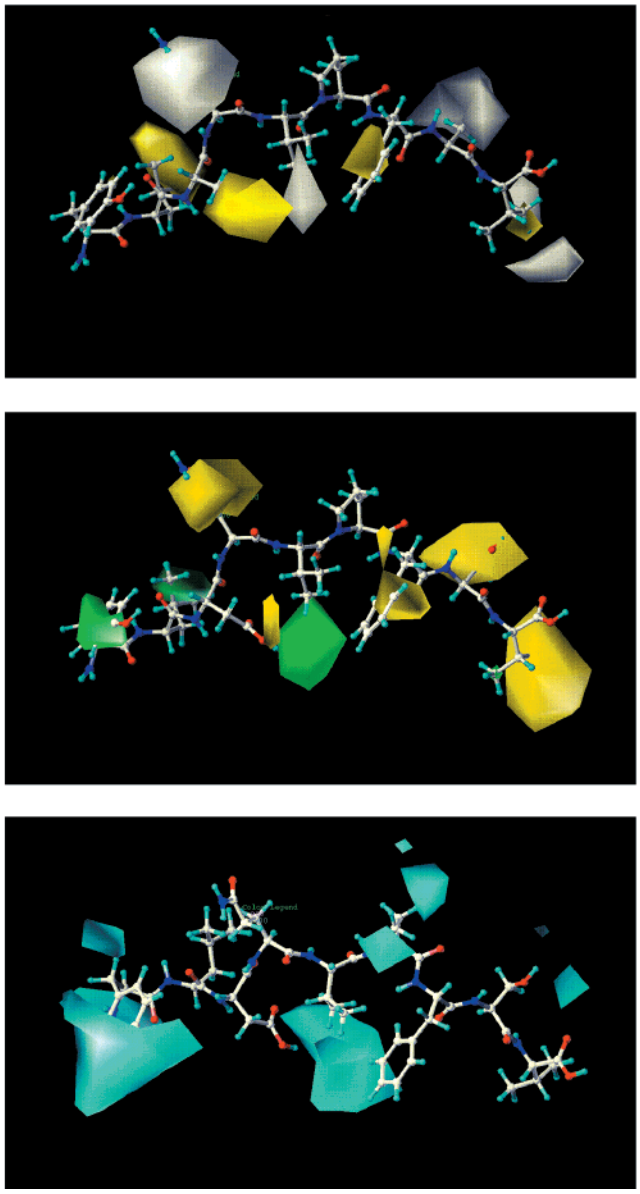

Figure 2. CoMSIA stdev*coeff contour maps. Peptide **100** is shown inside the field. Upper: hydrophobic map. Yellow and white polyhedra indicate regions where hydrophobic or hydrophilic groups, respectively, will enhance the affinity. Middle: steric map. Green and yellow polyhedra indicate regions where more steric bulk or less steric bulk, respectively, will enhance the affinity. Lower: H-bond donor ability map. Cyan polyhedra indicate regions where hydrogen bond acceptor groups on the receptor will enhance the binding.

to describe fully the peptide−protein interaction. The "predictive" r^2 for the test set is below 0.5. This result is not surprising, taking into account the dominant role of hydrophobic interactions, as is evident from the CoMSIA models. In the CoMSIA study, the model with the highest q^2 value was used to predict the binding affinities of the test set. The "predictive" r^2 is 0.679. Surprisingly, the steric and H-bond donor fields have small contributions to the affinity, but the electrostatic and H-bond acceptor fields contribute negatively.

Peptide Structure. All nine side chains of the bound peptides are found to contact HLA-A*0201.[14] The antigen-binding groove has a 30 Å long surface accessible to a solvent probe. Six pockets have been described in the surface denoted A through F (Figure 3).[21] Some of the pockets are nonpolar and are expected to form hydrophobic contacts. Others contain polar atoms and could make hydrogen bonds with the side chains. The preferences for a particular amino acid at any one position are considered here in terms of the three physicochemical properties that contribute positively to the affinity: local hydrophobicity, steric bulk, and hydrogen-bond-donor ability (Table 3).

Position 1. The steric map and the map of hydrogen-bond-donor ability (Figure 3) indicate that bulk side chains bearing a hydrogen-bond-forming group is preferred at position 1. The most suitable amino acid for this position seems to be Tyr. There should be a hydrogen-bond acceptor in the receptor near the N-terminal. Topologically this position corresponds to pocket A of the cleft of the peptide-binding site on HLA-A*0201, as described by Sapper et al.[13] (Figure 3). The pocket surface is predominantly polar: five Tyr hydroxyl groups (Tyr7, Tyr59, Tyr99, Tyr159, and Tyr171), a carboxyl group (Glu63), and an ϵ-amino group (Lys66). Tyr7, Tyr59, and Tyr171 form a network of hydrogen bonds that directly involve two hydrogen bonds to the peptide N-terminus. Tyr159 hydrogen bonds to the carbonyl oxygen of the first peptide amino acid residue (P1).[21] Recently it has been reported that the substitution of Ile at P1 with Phe or Tyr in the HIV reverse transcriptase (RT) peptide(309−317) (ILKEPVHGV) increased by 3-fold the cell surface half-life of complexes.[40] A $\pi-\pi$ stacking interactions between Trp167 and the aromatic P1 residues was proposed to account for this change.[40] Tourdot et al.[41] report that the P1Y substitution in 10 nonimmunogenic low-affinity peptides exhibited a 2.3−55-fold higher binding affinity and/or stabilized the HLA-A2.1 for at least 2 h more than the corresponding native peptides.

Position 2. Large hydrophobic side chains are preferred for position 2. They fall into pocket B of the peptide-binding site on HLA-A*0201 (Figure 3). This pocket has a polar rim and hydrophobic inner walls made up of Val67, Phe9, and Met45. Leu, Ile, and Met are well-accommodated in this pocket, as indicated by many experimental data.[16−20]

Position 3. Amino acids with hydrophobic aromatic rings at position 3, which fall into pocket D, will enhance the binding affinity of the peptides. This pocket is predominantly hydrophobic,[14] accommodating Phe and Trp[42] well. The hydrogen-bond-donor field map indicates that a Tyr also will be suitable, because of its ability to form hydrogen bonds.

Position 4. Hydrophilic nonbranched amino acids making hydrogen bonds are well-tolerated at position 4. Lys66 might form hydrogen bonds with Asp and Asn, but Arg65 could interact with the longer Glu,[14] Gln, and Arg. Kirksey et al.[40] suggest hydrogen-bond formation between Tyr at P1 and Glu at P4 bridged by a water molecule, which should make the bound peptide more rigid and more easily recognized by T-cells.

Position 5. The steric maps indicated that amino acids with branched or aromatic side chains are well-tolerated at position 5. According to the hydrophobic and hydrogen-bond property maps, there should be a hydrophilic group making a hydrogen bond with the

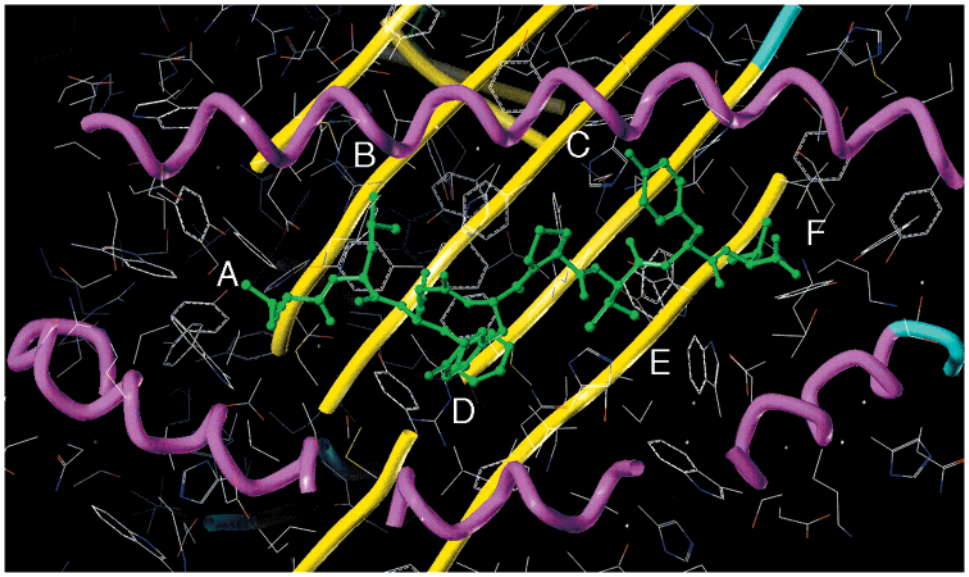

Figure 3. Peptide-binding site on the HLA-A2.1 molecule (X-ray data taken from ref 14). The α-helix chains are given in magenta and the β-sheets in yellow tubes. The binding site contains six pockets (labeled A through F). The peptide **83** (green ball-and-stick style) is shown inside the cleft.

Table 3. Physicochemical Properties Preferred at Each Position in the Peptides Binding to HLA-A*0201 Molecule

position	steric bulk	local hydrophobicity	H-bond acceptor group on the receptor
P1			
Side chain falls into pocket A.	Aromatic side chains are favored.		One in the vicinity of N-terminal and another one in the vicinity of OH-group of Tyr.
P2			
Side chain falls into pocket B.	Long side chains are favored.	Hydrophobic side chains are preferred.	
P3			
Side chain falls into pocket D.	Aromatic side chains are favored.	Hydrophobic side chains are preferred.	One in the vicinity of OH-group of Tyr.
P4			
	Branched side chains are disfavored.	Hydrophilic side chains are preferred.	
P5			
	Branched and aromatic side chains are favored.	Hydrophilic side chains are preferred.	One in the vicinity of OH-group of Tyr.
P6			
Side chain falls into pocket C.			One small in the vicinity of OH-group of Tyr and another one near the backbone NH.
P7			
Side chain falls into pocket E.	Short side chains are favored.	Hydrophobic side chains are preferred.	One in the vicinity of OH-group of Tyr.
P8			
	Short side chain is required.	Hydrophilic side chains are preferred.	One in the vicinity of $CONH_2$ of Gln and COOH of Glu.
P9			
C-terminal falls into pocket F.	Short side chain is required.	Hydrophobic side chains are preferred.	

$CONH_2$ of Gln155. Therefore, Tyr, Asn, and Arg are suitable residues for this position, but Ser, Thr, Val, and Phe will also be tolerated.

Position 6. The side chain of the amino acid at position 6 falls into pocket C (Figure 3).[13] This pocket is predominantly polar, made up of Thr73, His70, His74, and Arg97. Amino acids carrying H-bond donor groups, like Thr and Tyr, will be favored here.

Position 7. Pocket E corresponds topologically to position 7 (Figure 3). Two-thirds of the surface area in the pocket is hydrophobic, but Arg97 provides a large polar patch on one side of the pocket.[13] Short hydrophobic side chains are favored here. Ala, Val, or Pro seem to be well-accepted at this position.

Position 8. Hydrophilic amino acids with short side chains making hydrogen bonds are required for position 8. Ser and Thr will be favored here. Trp147 hydrogen bonds to the P8 carbonyl oxygen.[21]

Position 9. The C-terminal of the peptide falls into pocket F of the binding site. Tyr84, Thr143, and positively charged Lys146 bind to the carboxyl group (Figure 3).[21] Amino acids with hydrophobic short side chains such as Ala and Val are required for position 9. According to our study, Leu and Ile, previously considered to be anchor residues at P9,[16,17,20] do not seem to be well-tolerated.

The contribution maps obtained by CoMSIA shows how 3D QSAR methods can identify features important

for peptide−MHC molecule interactions. They allow identification of those regions within the area occupied by the ligand that requires a particular physicochemical property to improve the affinity.

The computational prediction of MHC binding epitopes has taken several forms. Perhaps the oldest and most widespread approach is to develop deterministic motifs for particular alleles (for reviews of this approach, see refs 43 and 44). Often published motifs neglect the contribution made by subsidiary residues.[16] Our approach shows that binding affinity arises as a composite interaction of the whole peptide with the MHC binding site.

A number of studies have developed computational methods based on extensions to experimental studies, for example, using data from pooled sequencing of epitopes to derive scoring matrixes.[45,46] The approach taken by others[47−49] tries to extend, or extrapolate, such analyses to alleles for which little or no ligand binding data is available through analysis of spatially conserved positions.

Of purely computational approaches, molecular dynamics (MD) simulations are, perhaps, the most rigorous method applied to the problem.[50−53] The success of such methods has been limited by the large computational burden required for such calculations. To obtain dynamic trajectories of sufficient duration to allow free energies of binding to be calculated requires thousands of hours of computer time on serial machines. Massively parallel implementations may speed up such calculations, allowing much longer time scales to become accessible. A number of empirical methods, based on artificial intelligence techniques, such as neural networks,[54,55] hidden Markov models,[56] and stepwise discriminant analysis,[57] have been used. Based, as is the approach we have presented here, on measured ligand binding data, these methods are commonly held to be among the most accurate of those available. Between these is a set of semiempirical approaches that combine knowledge from the structure of the MHC protein with ligand binding data. These include virtual screening[58] and threading.[59] Of these published methods, only the BIMAS[19] and SYTHEPII[44] systems make their predictions, for a large number of alleles, generally available through the Internet.

As far as we can tell, only one paper detailing a traditional, nonartificial intelligence approach from QSAR has been applied to the MHC binding problem. This is the work of Rovero et al.[60] However, theirs is a very limited study using peptide QSAR, and we are unaware of any known application of CoMFA or CoMSIA to problems arising from immunology. In addition, the amino acids principal property scales (Z-score scales), a methodology developed by Svante Wold's group,[61] though applicable to problems such as these, has yet to be explored in the field of MHC binding.

The approach we have taken is highly complimentary to the many excellent methods described above and also to the growing volume of crystal structure data. Our approach uses the full three-dimensional structure of the peptide, while making use of large numbers of experimentally measured ligand binding affinities, indeed, an order of magnitude greater than is available through crystallography. For this test study, the techniques appear highly predictive, albeit within the context of the dubious quality of some of the data used, the size of the compounds studied, and the wide variation in their physical properties.

Conclusions

The application of the 3D QSAR method CoMSIA to the set of 102 peptides with affinity for the HLA-A*0201 molecule delineated those areas within the binding site that favor or disfavor the presence of a group with a particular physicochemical property. This led to suggestions concerning the occupation of each of the nine positions in the nonamer peptide by favorable amino acid residues. The CoMSIA method is equally applicable to any MHC allele for which sufficient ligand binding data is available.

The prediction of epitopes is vital to our goal of developing computational vaccinology. Computational estimation of immunogenicity can be a useful tool for the assessment of epitope, multiepitope, or subunit vaccines, whether delivered as peptide or DNA. Although the prediction of MHC binding is only one component of the process leading to T-cell activation, it probably forms the most selective filter. Peptides first need to be presented, a process containing several bottlenecks. For class I, there is proteolytic cleavage by the proteasome followed by uptake by the TAP transporter. For class II, peptide cleavage is undertaken by one or more cathepsin proteases. Once processed peptides have bound to the MHC, a subset is recognized as T-cell epitopes. An accurate prediction of immunogenicity will encompass all these aspects, but nonetheless, the estimation of MHC binding probably eliminates the greatest number of potential epitopes. The ability to predict MHC binding will enable us to analyze microbial genomes, identifying the most immunogenic proteins and thus selecting a set of favored putative vaccines. We would expect computational vaccinology to have a similar effect on the search for new vaccines as molecular modeling and QSAR have had on search for new drugs.

Acknowledgment. We should like to thank Dr. David Livingstone, Dr. Vladimir Brusic, and Prof. Peter Beverley for the helpful discussions. We should also like to thank the referees for their insightful comments.

References

(1) Jardetzky, T. S.; Lane, W. S.; Robinson, R. A.; Madden, D. R.; Wiley: D. C. Identification of Self-Peptides Bound to Purified HLA-B27. *Nature* (*London*) **1991**, *353*, 326−329.
(2) Rudensky, A. Y.; Preston-Hulbert, P.; Soon-Cheol, H.; Barlow, A.; Janeway, C. A. Jr. Sequence Analysis of Peptides Bound to MHC Class II Molecules. *Nature* (*London*) **1991**, *353*, 622−627.
(3) Garcia, K. C. Molecular Interactions Between Extracellular Components of the T-Cell Receptor Signaling Complex. *Immunol. Rev.* **1999**, *172*, 73−85.
(4) Beck, S.; Geraghty, D.; Inoko, H.; Rowen, L. Complete sequence and gene map of a human major histocompatibility complex. The MHC sequencing consortium. *Nature* **1999**, *401*, 921−923.
(5) Mason, P. M.; Parham, P. HLA Class I Region Sequences. *Tissue Antigens* **1998**, *51*, 417−466.
(6) Imanishi, T.; Akaza, T.; Kimura, A.; Tokunaga, K.; Gojobori, T. Allele and Haplotype Frequencies for HLA and Complement Loci in Various Ethnic Groups. In *HLA 1991, Proceedings of the Eleventh International Histocompatibility Workshop and Conference*; Tsuji, K.; Aizawa, M.; Sasazuki, T., Eds.; Oxford University Press: Tokyo, 1992; pp 1066−1077. Allsopp, C. E.; Harding, R. M.; Taylor, C.; Bunce, M.; Kwiatkowski, D.; Anstey, N.; Brewster, D.; McMichael, A. J.; Greenwood, B. M.; Hill, A. V.

Interethnic Genetic Differentiation in Africa: HLA Class I Antigens in The Gambia. *Am. J. Hum. Genet*. **1992**, *50*, 411−421. Bodmer, J. World Distribution of HLA Alleles and Implications for Disease. *Ciba Found Symp*. **1996**, *197*, 233−253.

(7) Peoples, G. E.; Goedegebuure, P. S.; Smith, R.; Linehan, D. C.; Yoshino, I.; Eberlein, T. Y. Breast and Ovarian Cancer-Specific Cytotoxic T Lymphocytes Recognize the Same HER2/neu-Derived Peptide. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 432−436.

(8) McMichael, A. J.; Parham, P.; Brodsky, F. M.; Pilch, J. R. Influenza Virus-Specific Cytotoxic T Lymphocytes Recognize HLA-Molecules. Blocking by Monoclonal Anti-HLA Antibodies. *J. Exp. Med*. **1980**, *152*, *Suppl. 2*, 195−203.

(9) Schendel, D. J.; Gansbacher, B.; Oberneder, R.; Kriegmair, M.; Hofstetter, A.; Riethmuller, G.; Segurado, O. G. Tumor-Specific Lysis of Human Renal Cell Carcinomas by Tumor-Infiltrating Lymphocytes. I. HLA-A2-Restricted Recognition of Autologous and Allogeneic Tumor Lines. *J. Immunol*. **1993**, *151*, 4209−4220.

(10) Rongcun, Y.; Salazar-Onfray, F.; Charo, J.; Malmberg, K.-J.; Evrin, K.; Maes, H.; Hising, C.; Petersson, M.; Larsson, O.; Lan, L.; Appella, E.; Sette, A.; Celis, E.; Kiessling, R. Identification of New HER2/neu-Derived Peptide Epitopes That Can Elicit Specific CTL Against Autologous and Allogeneic Carcinomas and Melanomas. *J. Immunol*. **1999**, *163*, 1037−1044.

(11) Rivoltini, L.; Kawakami, Y.; Sakaguchi, K.; Southwood, S.; Sette, A.; Robbins, P. F.; Marincola, F. M.; Salgaller, M. L.; Yannelli, J. R.; Appella, E.; Rosenberg, S. A. Induction of Tumor-Reactive CTL from Peripheral Blood and Tumor-Infiltrating Lymphocytes of Melanoma Patients by In Vitro Stimulation with an Immunodominant Peptide of the Human Melanoma Antigen MART-1. *J. Immunol*. **1995**, *154*, 2257−2265.

(12) Parkhurst, M. R.; Fitzgerald, E. B.; Southwood, S.; Sette, A.; Rosenberg, S. A.; Kawakami, Y. Identification of a Shared HLA-A*0201-Restricted T-Cell Epitope from the Melanoma Antigen Tyrosinase-Related Protein 2 (TRP2). *Cancer Res*. **1998**, *58*, 4895−4901.

(13) Saper, M. A.; Bjorkman, P. J.; Wiley: D. C. Refined Structure of the Human Histocompatibility Antigen HLA-A2 at 26 Å Resolution. *J. Mol. Biol*. **1991**, *219*, 277−319.

(14) 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.

(15) Sette, A.; Sidney, J.; del Guercio, M.-F.; Southwood, S.; Ruppert, J.; Dalberg, C.; 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.

(16) 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-A*0201 Molecules. *Cell* **1993**, *74*, 929−937.

(17) Kubo, R. T.; Sette, A.; Grey, H. M.; Appella E.; Sakaguchi, K.; Zhu, N.-Z.; Arnott, D.; Sherman, N.; Shabanowitz, J.; Michel, H.; Bodnar, W. M.; Davis, T. A.; Hunt, D. F. Definition of Specific Peptide Motifs for Four Major HLA-A Alleles. *J. Immunol*. **1994**, *152*, 3913−3924.

(18) Parker, K. C.; Bednarek, M. A.; Hull, L. K.; Utz, U.; Cunningham, B.; Zweerink, H. J.; Biddison, W. E.; Coligan, J. E. Sequence Motifs Important for Peptide Binding to the Human MHC Class I Molecule, HLA-A2. *J. Immunol*. **1992**, *149*, 3580−3587.

(19) 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.

(20) Falk, K.; Rötzschke, O.; Stefanovic, S.; Jung, G.; Rammensee, H.-G. Allele Specific Motifs Revealed by Sequencing of Self-Peptides Eluted from MHC Molecules. *Nature* **1991**, *351*, 290−296.

(21) Madden, D. R. The Three−Dimensional Structure of Peptide−MHC Complexes. *Annu. Rev. Immunol*. **1995**, *13*, 587−622.

(22) Crammer, R. D., III; Patterson, D. E.; Bunce, J. D. Comparative Molecular Field Analysis (CoMFA). 1. Effect of Shape on Binding of Steroids to Carrier Proteins. *J. Am. Chem. Soc*. **1988**, *110*, 5959−5967.

(23) Klebe, G.; Abraham, U.; Mietzner, T. Molecular Similarity Indices in a Comparative Analysis (CoMSIA) of Drug Molecules To Correlate and Predict Their Biological Activity. *J. Med. Chem*. **1994**, *37*, 4130−4146.

(24) Klebe, G.; Abraham, U. Comparative Molecular Similarity Index Analysis (CoMSIA) To Study Hydrogen-Bonding Properties and To Score Combinatorial Libraries. *J. Comput.-Aided Mol. Design* **1999**, *13*, 1−10.

(25) Böhm, M.; Stürzebecher, J.; Klebe, G. Three−Dimensional Quantitative Structure−Activity Relationship Analyses Using Comparative Molecular Field Analysis and Comparative Molecular Similarity Indices Analysis To Elucidate Selectivity Differences of Inhibitors Binding to Trypsin, Thrombin, and Factor Xa. *J. Med. Chem*. **1999**, *42*, 458−477.

(26) Kast, W. M.; Brandt, R. M. P.; Sidney, J.; Drijfhout, J.-W.; Kubo, R. T.; Grey, H. M.; Melief, C. J. M.; Sette, A. Pole of HLA-A Motifs in Identification of Potential CTL Epitopes in Human Papillomavirus Type 16 E6 and E7 Proteins. *J. Immunol*. **1994**, *152*, 3904−3911.

(27) Sette, A.; Vitiello, A.; Reherman, B.; Fowler, P.; Nayersina, R.; Kast, W. M.; Melief, C. J. M.; Oseroff, C.; Yuan, L.; Ruppert, J.; Sidney, J.; del Guercio, M.-F.; Southwood, S.; Kubo, R. T.; Chesnut, R. W.; Grey, H. M.; Chisari, F. V. The Relationship Between Class I Binding Affinity and Immunogenicity of Potential Cytotoxic T Cell Epitopes. *J. Immunol*. **1994**, *153*, 5586−5592.

(28) Parkhurst, M. R.; Salgaller, M. L.; Southwood, S.; Robbins, P. F.; Sette, A.; Rosenberg, S. A.; Kawakami, Y. Improved Induction of Melanoma-Reactive CTL with Peptides from the Melanoma Antigen gp100 Modified at HLA-A*0201-Binding Residues. *J. Immunol*. **1996**, *157*, 2539−2548.

(29) Vitiello, A.; Sette, A.; Yuan, L.; Farrness, P.; Southwood, S.; Sidney, J.; Chesnut, R. W.; Grey, H. M.; Liningston, B. Comparison of Cytotoxic T Lymphocyte Responses Induced by Peptide or DNA Immunization: Implications on Immunogenicity and Immunodominance. *Eur. J. Immunol*. **1997**, *27*, 671−678.

(30) Del Guercio, M.-F.; Sidney, J.; Hermanson, G.; Perez, C.; Grey, H. M.; Kubo, R. T.; Sette, A. Binding of a Peptide Antigen to Multiple HLA Alleles Allows Definition of an A2-Like Supertype. *J. Immunol*. **1995**, *154*, 685−693.

(31) Tsai, V.; Southwood, S.; Sidney, J.; Sakaguchi, K.; Kawakami, Y.; Appella, E.; Sette, A.; Celis, E. Identification of Subdominant CTL Epitopes of the GP100 Melanoma-Associated Tumor Antigen by Primary In Vitro Immunization with Peptide-Pulsed Dendritic Cells. *J. Immunol*. **1997**, *158*, 1796−1802.

(32) Kawakami, Y.; Eliyahu, S.; Jennings, C.; Sakaguchi, K.; Kang, X.; Southwood, S.; Robbins, P. F.; Sette, A.; Appella, E.; Rosenberg, S. A. Recognition of Multiple Epitopes in the Human Melanoma Antigen gp100 by Tumor-Infiltrating T Lymphocytes Associated with In Vivo Tumor Regression. *J. Immunol*. **1995**, *154*, 3961−3968.

(33) SYBYL 6.6. Tripos Inc., 1699 Hanley Road, St. Louis, MO 63144.

(34) Powell, M. J. D. Restart Procedures for the Conjugate Gradient Method. *Math. Prog*. **1977**, *12*, 241−254.

(35) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. AM1: A New General Purpose Quantum Mechanical Molecular Model. *J. Am. Chem. Soc*. **1985**, *107*, 3902−3909.

(36) Cramer, R. D., III; Bunce, J. D.; Patterson, D. E. Crossvalidation, Bootstrapping, and Partial Least Squares Compared with Multiple Regression in Conventional QSAR Studies. *Quant. Struct.-Act. Relat*. **1988**, *7*, 18−25.

(37) Latron, F.; Moots, R.; Rothbard, J. B.; Garrett, T. P. J.; Strominger, J. L.; McMichael, A. Positioning of a Peptide in the Cleft of HLA-A2 by Complementing Amino Acid Changes. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 11325−11329.

(38) The Tripos implementation of CoMSIA uses a nomenclature opposite to that used in ref 24, in accordance with modifications made by the original authors.

(39) Folkers, G.; Merz, A.; Rognan, D. CoMFA: Scope and Limitations. In *3D QSAR in Drug Design;* Kubinyi, H., Ed.; ESCOM: Leiden, 1993; pp 583−618.

(40) Kirksey, T. J.; Pogue-Caley, R. R.; Frelinger, J. A.; Collins, E. J. The Structural Basis for the Increased Immunogenicity of Two HIV−Reverse Transcriptase Peptide Variant/Class I Major Histocompatibility Complexes. *J. Biol. Chem*. **1999**, *274*, 37259−37264.

(41) Tourdot, S.; Scardino, A.; Saloustrou, E.; Gross, D. A.; Pascolo, S.; Cordopatis, P.; Lemonnier, F. A.; Kosmatopoulos, K. A General Strategy to Enhance Immunogenicity of low-affinity HLA-A2.1-Associated Peptides: Implication in the Identification of Criptic Tumor Epitopes. *Eur. J. Immunol*. **2000**, *30*, 3411−3421.

(42) Sarobe, P.; Pendleton, C. D.; Akatsuka, T.; D.; Engelhard, V. H.; Feinstone, S. M.; Berzofsky, J. A. Enhanced In Vitro Potency and In Vivo Immunogenicity of a CTL Epitope from Hepatitis C Virus Core Protein Following Amino Acid Replacement at Secondary HLA-A2.1 Binding Positions. *J. Clin. Invest*. **1998**, *102*, 1239−1248.

(43) Rammensee, H. G.; Friede, T.; Stevanoviic, S. MHC Ligands and Peptide Motifs: First Listing. *Immunogenetics* **1995**, *41*, 178−228.

(44) Rammensee, H.; Bachmann, J.; Emmerich, N. P.; Bachor, O. A.; Stevanovic, S. SYFPEITHI: Database for MHC Ligands and Peptide Motifs. *Immunogenetics* **1999**, *50*, 213−219.

(45) Udaka, K.; Wiesmuller, K. H.; Kienle, S.; Jung, G.; Tamamura, H.; Yamagishi, H.; Okumura, K.; Walden, P.; Suto, T.; Kawasaki, T. An Automated Prediction of MHC Class I−Binding Peptides Based on Positional Scanning with Peptide Libraries. *Immunogenetics* **2000**, *51*, 816−828.

(46) Reay, P. A.; Kantor, R. M.; Davis, M. M. Use of Global Amino Acid Replacements to Define the Requirements for MHC Binding and T Cell Recognition of Moth Cytochrome C (93−103). *J. Immunol.* **1994**, *152*, 3946−3957.
(47) Chelvanayagam, G. A Road map for HLA-A, HLA-B, and HLA-C Peptide Binding Specificities. *Immunogenetics* **1996**, *45*, 15−26.
(48) Kangueane, P.; Sakharkar, M. K.; Lim, K. S.; Hao, H.; Lin, K.; Chee, R. E.; Kolatkar, P. R. Knowledge-Based Grouping of Modeled HLA Peptide Complexes. *Hum. Immunol.* **2000**, *61*, 460−466.
(49) Cano, P.; Fan, B.; Stass, S. A Geometric Study of the Amino Acid Sequence of Class I HLA Molecules. *Immunogenetics* **1998**, *48*, 324−334.
(50) 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.
(51) Sezerman U, Vajda S, DeLisi C. Free energy mapping of class I MHC molecules and structural determination of bound peptides. *Protein Sci.* **1996**, *5*(7), 1272−81.
(52) Toh, H.; Kamikawaji, N.; Tana, T.; Muta, S.; Sasazuki, T.; Kuhara, S. Magnitude of Structural Changes of the T-Cell Receptor Binding Regions Determine the Strength of T-Cell Antagonism: Molecular Dynamics Simulations of HLA-DR4 (DRB1*0405) Complexed with Analogue Peptide. *Protein Eng.* **2000**, *13*, 423−429.
(53) Froloff, N.; Windemuth, A.; Honig, B. On the Calculation of Binding Free Energies Using Continuum Methods: Application to MHC Class I Protein-Peptide Interactions. *Protein Sci.* **1997**, *6*, 1293−1301.
(54) 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.
(55) Brusic, V.; Rudy, G.; Honeyman, G.; Hammer, J.; Harrison, L. Prediction of MHC Class II−Binding Peptides Using an Evolutionary Algorithm and Artificial Neural Network. *Bioinformatics* **1998**, *14*, 121−130.
(56) Mamitsuka, H. Predicting Peptides that Bind to MHC Molecules Using Supervised Learning of Hidden Markov Models. *Proteins* **1998**, *33*, 460−474.
(57) Mallios, R. R. Class II MHC Quantitative Binding Motifs Derived from a Large Molecular Database with a Versatile Iterative Stepwise Discriminant Analysis Meta-Algorithm. *Bioinformatics* **1999**, *15*, 432−439.
(58) Rognan, D.; Lauemoller, S. L.; Holm, A.; Buus, S.; Tschinke, V. Predicting Binding Affinities of Protein Ligands from Three-Dimensional Models: Application to Peptide Binding to Class I Major Histocompatibility Proteins. *J. Med. Chem.* **1999**, *42*, 4650−4658.
(59) Schueler-Furman, O.; Altuvia, Y.; Sette, A.; Margalit, H. Structure-Based Prediction of Binding Peptides to MHC Class I Molecules: Application to a Broad Range of MHC Alleles. *Protein Sci.* **2000**, *9*, 1838−1846.
(60) Rovero, P.; Riganelli, D.; Fruci, D.; Vigano, S.; Pegoraro, S.; Revoltella, R.; Greco, G.; Butler, R.; Clementi, S.; Tanigaki, N. The Importance of Secondary Anchor Residue Motifs of HLA Class I Proteins: a Chemometric Approach. *Mol. Immunol.* **1994**, *31*, 549−554.
(61) Sandberg, M.; Eriksson, L.; Jonsson, J.; Sjöström, M.; Wold, S. New Chemical Descriptors Relevant for the Design of Biologically Active Peptides. A Multivariate Characterization of 87 Amino Acids. *J. Med. Chem.* **1998**, *41*, 2481−2491.

JM010021J