Modeling Ligand−Receptor Interaction for Some MHC Class II HLA-DR4 Peptide Mimetic Inhibitors Using Several Molecular Docking and 3D QSAR Techniques

Hsin-Yuan Wei, Keng-Chang Tsai, and Thy-Hou Lin*

Institute of Molecular Medicine & Department of Life Science, National Tsing Hua University, Hsinchu 30013, Taiwan

Received April 20, 2005

The ligand−receptor interaction between some peptidomimetic inhibitors and a class II MHC peptide presenting molecule, the HLA-DR4 receptor, was modeled using some three-dimensional (3D) quantitative structure−activity relationship (QSAR) methods such as the Comparative Molecular Field Analysis (CoMFA), Comparative Molecular Similarity Indices Analysis (CoMSIA), and a pharmacophore building method, the Catalyst program. The structures of these peptidomimetic inhibitors were generated theoretically, and the conformations used in the 3D QSAR studies were defined by docking them into the known structure of HLA-DR4 receptor through the GOLD, GLIDE Rigidly, GLIDE Flexible, and Xscore programs. Some of the parameters used in these docking programs were selected by docking an X-ray ligand into the receptor and comparing the root-means-square difference (RMSD) computed between the coordinates of the X-ray and docked structure. However, the goodness of a docking result for docking a series of peptidomimetic inhibitors into the HLA-DR4 receptor was judged by comparing the Spearman's rank correlation coefficient computed between each docking result and the activity data taken from the literature. The best CoMFA and CoMSIA models were constructed using the aligned structures of the best docking result. The CoMSIA was conducted in a stepwise manner to identify some important molecular features that were further employed in a pharmacophore building process by the Catalyst program. It was found that most inhibitors of the training set were accurately predicted by the best pharmacophore model, the Hypo1 hypothesis constructed. The deviation or conflict found between the actual and predicted activities of some inhibitors of both the training and the test sets were also investigated by mapping the Hypo1 hypothesis onto the corresponding structures of the inhibitors.

INTRODUCTION

The MHC class II molecules are cell surface proteins which perform an essential function in immunological detection using T-helper cells. They are encoded by the genes HLA-DR, -DQ, and -DP. Each MHC molecule consists of a α- and β-chain. In the case of the DR molecule, the two chains are encoded by the genes HLA-DRA and HLA-DRB1, and only DRB1 is polymorph i.e., only the gene has a number of different alleles existing in the population.[1] In addition, each individual possesses two DRB1 alleles, one from each parent.

The serological typification of the DR alleles leads to the differentiation between 10 different classes, HLA-DR1-DR10.[2] Molecular genetic typification shows that these classes can be further split e.g., DR2 has been divided into DR15 and DR16.[3] Within these classes it is possible to distinguish between a number of subtypes. Up until now, there are 33 subtypes of DR4 which have been described and are termed as HLA-DRB1*0401-*0433.[3−5] Rheumatoid arthritis (RA) or chronic polyarthritis is an intermittent systemic autoimmune disease which occurs in approximate 1% of the population.[6−9] The aetiology of the disease is unknown. It has been shown that there is a genetic disposition for RA caused by several alleles of the HLA-DRB1 region.[7,8] RA is associated with the HLA-DRB1*04 subtypes DRB1*0401, *0404, *0405, and *0408 and also in some different ethnic groups with the subtypes DRB1*0101, *0102, and DRB1 × 1001.[8,9]

Recently, the general features of the molecular recognition between antigenic peptide and the binding site on several MHC class II molecules have been elucidated through crystallization of several MHC molecular complexes.[10−14] Both the MHC α and β chains contribute to the peptide binding site, which is made up of a β sheet floor topped by two roughly parallel α helical regions.[15−17] The peptide binding motifs for some heptapeptides binding to the DR alleles have also been determined through phage display libraries and synthetic peptides.[18−21] Peptides bind in an extended conformation in the groove between the two helices, with about 10 residues able to interact with the MHC protein, while the peptide termini extend from the binding site.[20,21] The conformation places 4−6 of the peptide side chains into pockets within the overall groove. The residues lining these pockets vary between allelic variants, providing different peptide sequence binding specificity. The interaction buries about 70% of the peptide surface area in the central region of a bound peptide, leaving the remainder available for interaction with the antigen receptors on T-cells.[21]

The binding of peptides to human and mouse MHC class II molecules is characterized by several conserved side-chain binding pockets namely p1−p9 within the overall peptide-

* Corresponding author fax: 886-3-571-5934; e-mail: thlin@life.nthu.edu.tw.

10.1021/ci050140y CCC: $30.25
Published on Web 08/24/2005

binding groove.[18,19,22] The pockets are numbered along the peptide relative to a large usually hydrophobic pocket near the peptide-binding site. The importance of residues at p1, p2, p4, p6, and p7 on binding has been addressed by panning M13 phage expressed random peptide libraries.[22] An immunodominant peptide epitope of hemagglutinin (HA) (HA306-318) from influenza A virus H3N2 has been found to bind with different DR alleles of the MHC class II molecules.[23] The α/β T-cell receptor (TCR) HA1.7 specific for the HA antigen peptide is HLA-DR1 restricted but cross-reactive with the HA peptide presented by the MHC class II molecule HLA-DR4.[23] The overall structures of the HA1.7/DR4/HA and HA1.7/DR1/HA complexes are found to be very similar though there is a difference in the amino acid sequence of DR1 and DR4 located deeply inside the peptide binding groove and out of reach by direct contact by the TCR.[23]

The autoimmune diseases such as RA, multiple sclerosis, and diabetes mellitus may be treated by selectively inhibiting antigen presentation by competitive blockade of the peptide binding site of a disease-associated MHC molecule using some nonantigenic ligands.[22,24] A series of such compounds has been synthesized by Bolin et al.[22] to bind with HLA-DR4 (DRB1*0401) and inhibit the T-cell proliferation in response to the corresponding protein antigens. These peptide analogues are taken up by the antigen-presenting cells (APCs) and are loaded onto MHC class II molecules in competition with the antigenic peptide fragments of ovalbumin and hen egg white lysozyme.[25] The binding of peptides to HLA-DR1 has been strengthened by incorporating an *N*-methyl substitution at p7 of the peptide.[26] The *N*-methyl group oriented in the p6/p7 pocket is shown to displace one of the waters usually bound in this pocket, and the corresponding MHC-peptide complexes generated are able to activate the antigen-specific T-cells.[26] In this report, we have employed several 3D QSAR techniques on the aligned structures of 30 peptidomimetic inhibitors of HLA-DR4 designed by Bolin et al.[22] for constructing some 3D QSAR models for these compounds. The peptidomimetic inhibitors were divided into the training and the test sets, and the structures in each set were aligned and analyzed by the CoMFA[27] and CoMSIA[28] methods to derive the best 3D QSAR model for the peptidomimetic inhibitors. Further, the pharmacophore features obtained from the best CoMSIA model were used to construct some pharmacophore hypotheses using the Catalyst 4.9 program.[29] The top hypothesis thus generated was mapped onto the structures of several highly active peptidomimetic inhibitors selected from both the training and the test sets. The predicted activities for both the training and the test sets by the top hypothesis were found to be good in statistics as those predicted by the best CoMSIA model from which the hypothesis was derived. The feasibility of using the constructed 3D QSAR models to design the peptidomimetic inhibitors was discussed.

MATERIALS AND METHODS

The structures and activities of all the peptidomimetic inhibitors of HLA-DR4 studied were listed in Table 1. The ligand ACE-HAC-ARG-MPQ-MET-ALA-SER-BUG of X-ray structure 1D6E[22] was used as a template for constructing all the structures of peptidomimetic inhibitors of HLA-DR4 studied. The structures were constructed within the active site of 1D6E by replacing side chains of the template with other groups as has been described by others. The hydrogen atoms were added for each structure. The total number of structures constructed was 30, and some different substitution analogues of the heptapeptide Inh. #1 (Ac-(Cha)RAMASL-NH2) were included (Table 1). The activity of each of these peptidomimetic inhibitors was recorded as a ratio of IC_{50} value to that of Ac-(Cha)RAMASL-NH2.[22] The range of activity ratio of these inhibitors was 0.07−2.32 where potent compounds were expressed with larger ratio values (Table 1). These inhibitors were divided into the training and the test sets with each containing 15 inhibitors and covering roughly the same range of activity ratio.

Table 1. Sequences and Activities of the Peptidomimetic Inhibitors against the HLA-DR4 Receptor Studied

Inh #	structure	mean relative potency[a]
1	Ac-(Cha)RAMASL-NH2	1.00
2	Ac-(L-Nba)RAMASL-NH2	1.71
11	Ac-(L-m-F-Phe)RAMASL-NH2	1.19
12	Ac-(Cha)VAMASL-NH2	0.91
13	Ac-(Cha)OAMASL-NH2	0.67
14	Ac-(Cha)O(Me)2AMASL-NH2	0.38
15	Ac-(Cha)KAMASL-NH2	0.45
16	Ac-(Cha)K(Me)2AMASL-NH2	0.62
17	Ac-(Cha)(alle)AMASL-NH2	1.27
20	Ac-(Cha)R(Phg)MASL-NH2	1.10
21	Ac-(Cha)R(MePhg)MASL-NH2	2.32
25	Ac-(Cha)RA(Nle)ASL-NH2	0.40
26	Ac-(Cha)RAIASL-NH2	1.01
27	Ac-(Cha)RALASL-NH2	0.54
28	Ac-(Cha)RAM(Cacm)SL-NH2	1.73
29	Ac-(Cha)RAMPSL-NH2	0.70
30	Ac-(Cha)RAM(Pip)SL-NH2	1.08
34	Ac-(Cha)RAMASL-N(CH3)2	1.60
36	Ac-(Cha)RAMA-NH2	0.07
40	Ac-(Cha)RMMASL-NH2	0.88
58	Ac-(Cha)R(MeA)(Haic)S(MeL)-NH2	0.18
60	Ac-(Cha)R(MePhg)MAS(tLeu)-NH2	2.02
62	Ac-(Cha)R(MeA)M(β-PhPro)S(MeA)-NH2	0.20
63	Ac-(Cha)R(MePhg)(Haic)S(MeA)-NH2	0.37
64	Ac-(Cha)R(MePhg)(Haic)S(tLeu)-NH2	0.23
65	Ac-(Cha)R(Tic)(Haic)S(MeA)-NH2	0.27
68	Ac-(Cha)R(MePhg)(Haic)S(MeL)-NH2	0.11
69	Ac-(Cha)R(MeA)M(β-PhPro)S(tLeu)-NH2	0.54
70	Ac-(Cha)R(Tic)M(β-PhPro)S(MeL)-NH2	0.22
71	Ac-(Cha R(Tic)M(β-PhPro)S(tLeu)-NH2	0.64

[a] The activity of each peptidomimetic inhibitor is expressed as a ratio of IC_{50} against that of Inh. #1 as taken from the literature.[22]

Each structure constructed was subjected to a brief energy minimization with the receptor together using the SYBYL 6.9.1 program.[30] Then, the crystal ligand was docked into the active site of HLA-DR4 using the GOLD 2.1[31] and GLIDE[32] programs, and the corresponding RMSD between the crystal and docked conformation was computed. This docking was used to set the parameters Number of Operations and Population Size respectively as 1 600 000 and 1000, and the others were chosen as the default settings. The constructed and energy minimized inhibitor structures were subsequently docked into the active site of HLA-DR4 using these parameter settings. The MMFF94[33] charges were deployed for each inhibitor. The docking results were compared using the docking scoring functions by GOLD, GLIDE Rigidly, and Xscore[34] after the conformations were generated by GOLD. Most of the parameters used in the GLIDE Rigidly docking were also default settings except

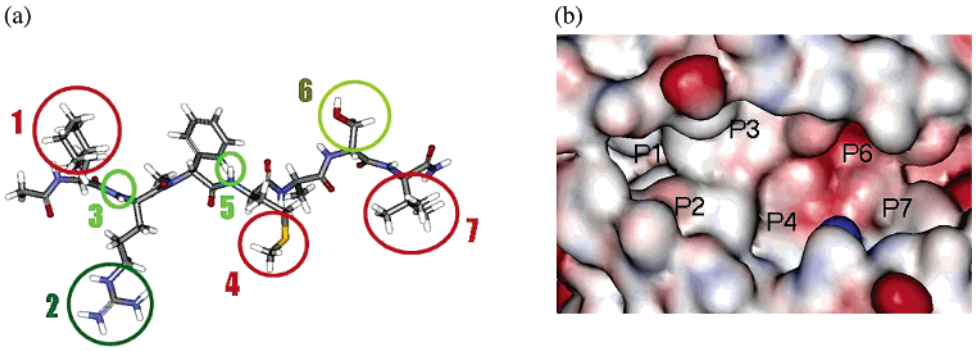

Figure 1. (a) The binding positions reported by Bolin et al.[22] are highlighted on the structure of Inh. #1. Positions p1, p4, and p7 are hydrophobic, positions 3 and 5 are donor, and position 2 is a positive ionizable region. (b) The peptide binding site of HLA-DR4 receptor where the corresponding binding pockets/positions highlighted in (a) are marked.

the followings: Skip Ligands > 200 Atoms and 35 Rotatable Bonds and Scaling Of vdW Radii For Nonpolar Ligand Atoms Scale By 0.65. The rank of each docking score was compared with that of the activity ratio by computing the Spearman's rank correlation coefficient rs between them defined as follows

$$\mathrm{rs} = 1 - \frac{6\sum di^2}{n(n^2 - 1)} \qquad (1)$$

where *di* was the rank difference for the *i*th observation under two different criteria e.g., the rank difference between the activity ratio and a scoring function, and *n* was the total number of inhibitors compared.

The steric and electrostatic potential fields of CoMFA were calculated by the SYBYL 6.9 program using a regularly spaced lattice of 2.0 Å. The lattice was extended to 4 Å units beyond the van der Waals volume of each molecule in the X, Y, and Z directions. A C.3 carbon atom of radius 1.52 Å and charge +1.0 was used as a probe to calculate both the steric and electrostatic fields. The truncation for both the steric and electrostatic contributions computed was set at ±30 kcal/mol. The electrostatic contribution at the lattice intersections where maximum steric interactions were computed was ignored. Both the CoMFA steric and electrostatic fields computed were scaled by the standard option given in the program. The same lattice where each molecule was submerged for CoMFA was used for CoMSIA. A C.3 atom of radius 1.0 Å and charge +1 was used as the probe to compute the CoMSIA similarity indices defined by Klebe et al.[28] The similarity indices were calculated using the Gausian-type distance dependence between the probe and atoms of the molecules of a data set. This functional form requires no arbitrary definition of cutoff limits, and the similarity indices can be calculated at all lattice points inside and outside the molecule.[28] The attenuation factor α was set as 0.3. In the SYBYL CoMSIA module,[30] the third power of the atomic radii was computed as the steric indices, the atomic partial charges were treated as the electrostatic indices, the atom-based parameters developed by Viswanadhan et al.[35] were used as the hydrophobic indices, and a rule-based method derived experimentally[36] was used as the hydrogen bond donor and acceptor indices.

In proceeding CoMFA and CoMSIA, the structures were aligned based on the docking conformations generated by the GOLD program. The coordinates of backbone atoms C1, C2, N5, C6, C7, N14, C15, C16, N37, C38, C39, N43, C44, C45, N60, C61, and C62 of each peptidomimetic inhibitor were fitted to those of the template structure (Inh. #1 of Table 1) using the SYBYL Fit module.[30] The CoMFA and CoMSIA results were cross-validated using the SYBYL PLS module.[30] The CoMFA and CoMSIA descriptors were treated as the independent variables, while the activity ratios (Table 1) were treated as the dependent ones in all the PLS regression analyses for deriving the 3D QSAR models. The optimum number of components used to derive a 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. The nonvalidated models were assessed by the conventional correlation coefficient r^2, standard error of estimate, and *F*-values. The results of nonvalidated analyses were used to make prediction of the binding affinities for the test set inhibitors and to display the coefficient contour maps.

The same training set used in CoMFA and CoMSIA was used for constructing some pharmacophore models by the Catalyst 4.9 program. All the parameters used were default settings except that Unc was set at 1.4. There were three pharmacophore features namely H (hydrophobic), D (hydrogen-bond donor), and I (positive ionizable group) selected for the hypothesis generation process. The pharmacophores were automatically generated by the HypoGen module[29] of Catalyst 4.9 program for the training set. The top 10 scored hypotheses generated for each inhibitor of the set were exported.

RESULTS AND DISCUSSION

The important binding pockets or binding positions p1, p2, p4, p6, and p7 of a HLA-DR4 (DRB1*0401) receptor described by Bolin et al.[22] where the corresponding residues from a peptidomimetic inhibitor can bind with are depicted in Figure 1a,b. While pockets p2 and p6 are characterized as being favored by binding with the positive ionizable and small polar or neutral residues, pockets p1, p4, and p7 are found to be hydrophobic (Figure 1b). In addition, the peptide backbone of p2 and p4 positions can make some hydrogen-bond contacts with some side chain residues of the binding pocket (Asn82, Asn62, and Gln9) (Figure 1a). The activity ratios of the 30 peptidomimetic inhibitors studied vary with a single substitution made at a particular binding position for a residue on the template inhibitor Inh. #1 (Table 1). Docking the inhibitor into the active site of HLA-DR4 receptor 1D6E using GOLD 2.1 and GLIDE programs gives

Table 2. Comparison for the Docking Results Using the GOLD, GLIDE Flexible, GLIDE Rigidly, and Xscore on the Structures of Peptidomimetic Inhibitors Generated

	GOLD	GLIDE Flexible	GLIDE Rigidly	XSCORE						
				HP	HM	HS	HP/HM	HP/HS	HS/HM	HP/HM/HS
RMSd	1.24	3.069	0.4801	NA						
rs[a]	0.61	NA	0.84	0.67	0.65	0.6	0.66	0.61	0.62	0.61

[a] The goodness of docking results is judged by comparing both the RMSD and rs, the Spearman's rank correlation coefficient computed (see Materials and Methods section of text for definition).

Table 3. Summary of CoMFA and Stepwise CoMSIA Statistics for the Training Set Peptidomimetic Inhibitors

			HLA-DR4 leave one out		HLADR4 cross- validation		HLA-DR4 no validation		
CoMFA HLA-DR4 60		CoMSIA	NC[a]	q^2_{loo}	NC	q^2	SEP[a]	r^2	F
leave one out q^2_{loo}	0.513	S[a]	6	0.777	6	0.673	0.019	0.999	1465.239
cross-validational (q^2)	0.502	E	5	−0.147	2	−0.097	0.115	0.968	40.460
conventional r^2	0.998	H	6	0.714	6	0.686	0.011	1.000	4588.885
standard error	0.030	D	6	0.218	4	0.172	0.172	0.929	17.450
principal components	6	A	5	0.107	5	0.065	0.186	0.907	17.482
F-values	597.919	H + S	6	0.720	6	0.709	0.019	0.999	1525.574
		H + E	6	0.538	6	0.484	0.033	0.987	506.708
		H + D	6	0.607	6	0.594	0.052	0.994	204.017
		H + A	6	0.504	6	0.492	0.046	0.995	263.545
		S + E	6	0.333	6	0.302	0.059	0.992	157.397
		S + D	6	0.461	4	0.409	0.075	0.986	97.284
		S + A	6	0.460	6	0.404	0.063	0.990	136.634
		H + S + E	6	0.589	6	0.585	0.032	0.998	551.854
		H + S + D	6	0.621	6	0.604	0.043	0.996	303.508
		H + S + A	6	0.543	6	0.467	0.036	0.997	415.950
		H + S + A + E	6	0.442	6	0.394	0.046	0.995	259.037
		H + S + A + D	6	0.498	5	0.298	0.056	0.993	177.283
		All fields	6	0.415	6	0.325	0.061	0.991	147.852

[a] Abbreviations used are SEP: standard error of prediction, S: steric, E: electrostatic, H: hydrophobic, A: H-bond acceptor, and D: H-bond donor index.

a RMSD of 1.24 and 3.07 between the crystal and docked ligand, respectively. The goodness of a docking result is judged by both the RMSD and Spearman's rank correlation coefficient rs (eq 1) computed between some docking scoring functions and the activity ratio. The docking scoring functions of GOLD, GLIDE Rigidly, and Xscore for the docked conformations generated by GOLD are employed. The rank of each docking score for an inhibitor set is compared with that of the activity ratio for computing an rs value between them. The rs values computed for the docking scoring functions GOLD, GLIDE Rigidly, Xscores (HP score, HM score, HS score, HP/HM score, HP/HS score, HS/HM score, and HP/HS/HM score) used are 0.61, 0.84, 0.67, 0.65, 0.60, 0.66, 0.61, 0.62, and 0.61, respectively (Table 2). The rs values computed using Xscores appears to be very similar to that computed for the GOLD docking, indicating that the docking parameters chosen for using the program is adequate. The docked ligand−receptor complex of Inh #2 Ac-(L-Nba)-RAMASL-NH2 generated by GOLD docking is analyzed by the Ligplot 4.22 program[37] to reveal that there are hydrophobic contacts between ligand and receptor at the binding pockets p1, p4, and p7 (data not shown here). There are also hydrogen-bond contacts found between the backbone atoms NH and CO of positions p2 or p4 of the docked ligand with those of the receptor residues Asn282 or Gln9 and Asn62 (data not shown here).

To proceed with CoMFA, the docked conformation generated for each inhibitor of each set by GOLD docking is aligned against that of the template structure (Inh. #1 of Table 1) by treating the coordinates of some backbone atoms selected as the correspondence points. The aligned structural sets are then analyzed by the SYBYL CoMFA, CoMSIA, and PLS programs, and the results with significant statistics obtained are kept for further analyses. The statistics of the best CoMFA and CoMSIA results is presented in Table 3. The best CoMFA result yields a leave-one-out (loo) validated r^2 (q^2_{loo}) of 0.513, a cross-validated r^2 (q^2) of 0.502, and a conventional r^2 (r^2) of 0.998 (Table 3). The CoMSIA is conducted in a stepwise manner namely a single field index, a combination of any two field indexes, a combination of any three field indexes, a combination of any four field indexes, and then a combination of all field indexes is employed step-by-step in the analyses. There are five different field indexes [steric, denoted as S; electrostatic, denoted as E; hydrophobic, denoted as H; H-bond acceptor, denoted as A; and H-bond donor, denoted as D] being chosen for the stepwise CoMSIA, and the results are presented in Table 3. As judged by q^2 values computed, no statistically significant CoMSIA result is obtained for each single field index selected except the H and S one (Table 3). A combination of H and S field indexes with each of the other three ones is then conducted. Apparently, the statistics of these results are better than those of the single field ones (Table 3). The third step is conducted by adding each of the rest field indexes to the combined field indexes of H and S because a combination of the two gives the best CoMSIA statistics obtained (Table 3) up to the step. However, no apparent improvement in CoMSIA statistics is obtained for further steps using a combination of either D, A, or E field index with the combined H+S field indexes or even a

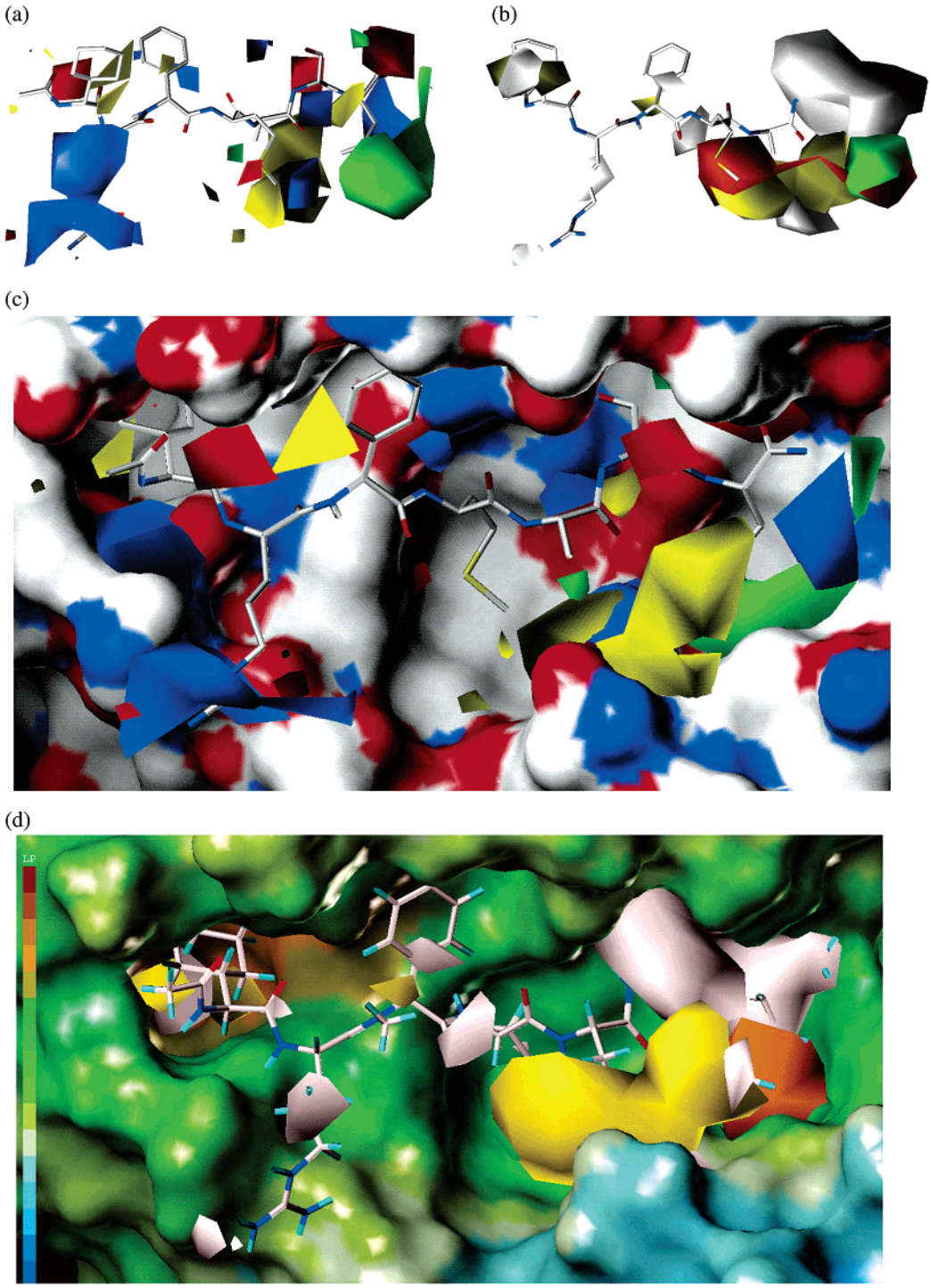

Figure 2. (a) The CoMFA contours of the best CoMFA result of the training set (Table 3). (b) The CoMSIA contours of the best CoMSIA result of the training set (Table 3). (c) Projection of the CoMFA contours depicted in (a) over the electrostatic potential surface map (blue surfaces are favored positive charge; red surfaces are favored negative charge) of the HLA-DR4 receptor active site. (d) Projection of the CoMSIA contours depicted in (b) over the lipophilicity surface map (brown surfaces are hydrophobic favored regions; blue surfaces are hydrophilic favored regions) of the HLA-DR4 receptor active site.

combination of all the five field indexes (Table 3). Therefore, as revealed by the stepwise CoMSIA results, the interaction of the HLA-DR4 inhibitors of the training set with their common receptor is best described by a combination of H and S field indexes. In fact, with a q^2 value of 0.709, the CoMSIA statistics of the combined H+S fields is better than the CoMFA one (Table 3).

The CoMFA and CoMSIA contour maps agree with each other on the identification of favor regions for steric interaction (displayed with green contours by both CoMFA and CoMSIA) which are around position p7 (Figure 1a and 2a,b). Both the CoMFA and CoMSIA contour maps also show that there are disfavor regions for steric interaction (displayed as yellow contours by both results) somewhere around the p1, p3, p4, and p5 and near position p6 of the HLA-DR4 inhibitors (Figure 2a,b). The favor region for positive charge identified by CoMFA (displayed by blue contours) is around position p2 (Figure 2a) which agrees with that displayed in Figure 1a. Position p6 is identified by CoMSIA as a disfavor region for hydrophobic interaction

Table 4. Validation of the Hypo1 Hypothesis Using the CatScramble Program Implemented in the Catalyst Program

hypothesis no.	total cost	Δcost	rms deviation	correlation (*r*)
1	58.51	48.30	0.90	0.96
2	83.86	22.95	1.80	0.82
3	84.57	22.24	1.88	0.80
4	87.51	19.30	1.89	0.80
5	87.52	19.29	1.91	0.79
6	93.05	13.76	2.14	0.73
7	97.22	9.59	2.26	0.70
8	97.44	9.37	2.25	0.70
9	101.54	5.27	2.34	0.67
10	102.88	3.93	2.33	0.67

(displayed by white contours, Figure 2b) which agrees with that displayed in Figure 1a where the region is recognized as a favorable binding site for small polar groups. The favor regions for hydrophobic interaction around positions p4 and p7 are also identified by CoMSIA (displayed by orange contours) (Figure 2b) and matched with those reported previously by others[22] or displayed in Figure 1a. A projection of the CoMFA contours over the electrostatic potential surface map of the active site of HLA-DR4 receptor (Figure 1b) is presented in Figure 2c where favored regions for positive charge are expressed as blue, while favored negative charge ones are expressed as red surfaces. This shows that the blue contours of CoMFA at position p2 representing favor regions for positive charge (Figure 2a) are correctly mapped onto the receptor regions of pro blue surfaces where positive charge is favored (Figure 2c). The orange contours of CoMSIA at position p4 and p7 representing favor regions for hydrophobic interaction (Figure 2b) are also correctly mapped onto the receptor regions of brown surfaces where hydrophobic interaction is presumably favored (Figure 2d).

Based on the CoMSIA results obtained (Table 3), the structural features selected for constructing a pharmacophore for the peptidomimetic inhibitors studied using the Catalyst 4.9 program are H, D, and I. A comparison for the statistical significance of the top 10 hypotheses generated is given in Table 4. The cost difference between null and total cost should be greater than 40, and the configuration cost should be smaller than 17 bits for a good hypothesis generated. While the configuration cost computed for each hypothesis type is always smaller than 17 bits, the cost difference between null and total cost is 58.51 and that between null and fixed cost is 56.37. Therefore, the top hypothesis of the hypotheses (designated as Hypo1 hereafter) generated meets the criteria of being a good hypothesis.

The actual and predicted biological activities by the best CoMFA (Table 3 and Figure 2a) and CoMSIA (Table 3 and Figure 2b) models and the Hypo1 hypothesis for each training set inhibitor are listed and compared in Table 5, while those for each test set inhibitor are given in Table 6, respectively. In theses tables, all the predicted activity ratios are listed in an enhancing order from top to bottom as those of the actual ones. Apparently, most of the predicted activity ratios of the training set by the best CoMFA and CoMSIA models are in accord with the actual ones (Table 5). Regression of the predicted versus actual activity ratios onto a linear line yields a coefficient of 0.99 for both the CoMFA and CoMSIA results. However, more discrepancies between the predicted and actual activity ratios are found for the test set by the best CoMFA than CoMSIA model (Table 6). A

Table 5. Predicted Activity Ratios (Pred RA) by the Best CoMFA and CoMSIA Models and by the Hypo1 Hypothesis Are Compared with the Actual Activity Ratio (Act RA) for the Training Set Peptidomimetic Inhibitors

				Catalyst Pharmacophore Hypothesis Hypo1				
HLA-DR4 Inh #	Act RA	CoMFA Pred RA	CoMSIA S+H Pred RA	Act RA	Pred RA	error	act activity scale[a]	pred activity scale[a]
36	0.07	0.07	0.07	0.07	0.08	1.2	+	+
68	0.11	0.09	0.09	0.11	0.16	1.5	++	++
62	0.2	0.21	0.20	0.20	0.16	1.3	++	++
64	0.23	0.23	0.23	0.23	0.19	1.2	++	++
69	0.54	0.48	0.51	0.54	0.70	1.3	++	++
13	0.67	0.73	0.72	0.67	1.00	1.5	++	+++
1	1.00	1.02	0.96	1.00	1.00	1.0	+++	+++
30	1.08	1.08	1.10	1.08	0.87	1.3	+++	++
20	1.10	1.16	1.07	1.10	1.05	1.1	+++	+++
11	1.19	1.20	1.25	1.19	0.87	1.4	+++	++
17	1.27	1.18	1.22	1.27	1.30	1.0	+++	+++
34	1.60	1.65	1.56	1.60	1.00	1.6	+++	+++
2	1.71	1.62	1.69	1.71	1.05	1.6	+++	+++
60	2.02	2.09	2.04	2.02	2.60	1.3	+++	+++
21	2.32	2.17	2.31	2.32	3.55	1.5	+++	+++

[a] Definition of the activity scale is given by the Catalyst program and described in the text.

Table 6. Predicted Activity Ratios (Pred RA) by the Best CoMFA and CoMSIA Models and by the Hypo1 Hypothesis Are Compared with the Actual Activity Ratio (Act RA) for the Test Set Peptidomimetic Inhibitors

				Catalyst Pharmacophore Hypothesis Hypo1				
HLA-DR4 Inh #	Act RA	CoMFA Pred RA	CoMSIA S+H Pred RA	Act RA	Pred RA	error	act activity scale[a]	pred activity scale[a]
58	0.18	0.45	0.29	0.18	0.39	2.2	++	++
70	0.22	0.48	0.63	0.22	0.47	2.1	++	++
65	0.27	0.14	0.07	0.27	0.16	1.7	++	++
63	0.37	0.17	0.09	0.37	0.23	1.6	++	++
14	0.38	0.82	0.90	0.38	0.89	2.3	++	++
25	0.40	0.64	0.92	0.40	0.21	1.9	++	++
15	0.45	0.73	0.55	0.45	0.33	1.4	++	++
27	0.54	0.58	0.88	0.54	0.65	1.2	++	++
16	0.62	1.75	0.78	0.62	0.33	1.9	++	++
71	0.64	0.65	0.62	0.64	0.35	1.8	++	++
29	0.70	0.80	1.23	0.70	0.63	1.1	++	++
40	0.88	1.05	1.32	0.88	0.41	2.1	++	++
12	0.91	0.70	1.19	0.91	0.59	1.5	++	++
26	1.01	1.43	1.72	1.01	0.76	1.3	+++	++
28	1.73	1.20	2.07	1.73	1.08	1.6	+++	+++

[a] Definition of the activity scale is given by the Catalyst program and described in the text.

linear regression coefficient of 0.35 and 0.78 is obtained for the test set predicted by the best CoMFA and CoMSIA model, respectively. The predicted activity ratios for either the training or the test set by the Hypo1 hypothesis are also given in Tables 5 and 6, respectively. The goodness of prediction can be judged through the activity scale the '+' sign labeled for each actual and predicted activity ratio as listed in both Tables 5 and 6. While highly active inhibitors predicted are labeled with the '+++' sign, the less and least active ones predicted are labeled with the '++' and '+' signs, respectively. Apparently, there are three differences in the activity scale labeled between the actual and predicted activity ratios for the training set by the Hypo1 hypothesis

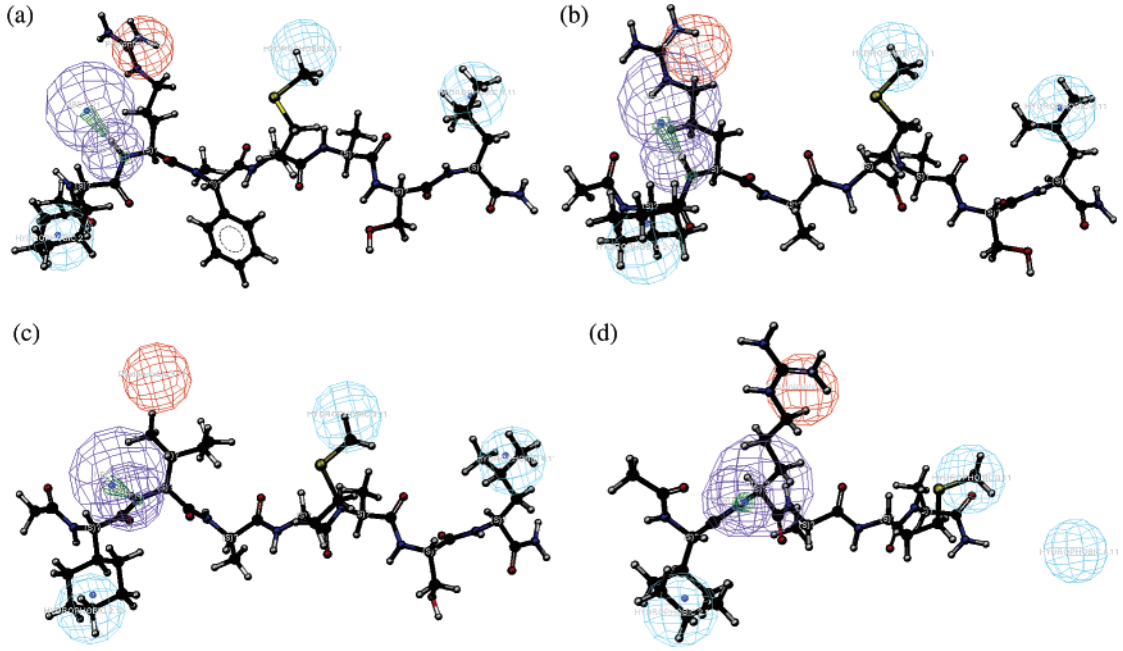

Figure 3. Mapping of the Hypo1 hypothesis onto the structures of four peptidomimetic inhibitors Inh. #21 (a), #01 (b), #17 (c), and #36 (d) (Table 5) selected from the training set. The pharmacophore features are color coded as follows: blue spheres for hydrophobic (H), violet spheres for hydrogen-bond donor (D), and red spheres for positive ionizable (I).

(Table 5). The differences are due to inhibitors Inh. #13, #30, and #11 where the actual activity ratios are labeled either higher or lower than the predicted one (Table 5). Therefore, the prediction accuracy by the Hypo1 hypothesis for the training set estimated is 80%. However, there is only one conflict in the activity scale labeled between the actual and predicted activity ratios for the test set by the Hypo1 hypothesis as shown in Table 6. This prediction difference is due to Inh. #26 (Table 6). In other words, the prediction accuracy by the Hypo1 hypothesis for the test set estimated is 93%. A linear regression of the predicted against actual activity ratios of the training set yields a coefficient of 0.96, revealing the feasibility of using the structural features selected by the stepwise CoMSIA to construct a pharmacophore hypothesis. The same regression procedure on the test set yields a correlation coefficient of 0.55. However, a better correlation coefficient of 0.72 is obtained if both Inh. #58 and #70 of the test set are excluded in the regression process.

The Hypo1 hypothesis features are mapped onto the structures of several peptidomimetic inhibitors (Inh. #21, #01, #17, and #36) as presented in Figure 3. The activity ratios of these peptidomimetic inhibitors mapped range from the most to the least potent ones (Table 1). A hydrogen-bond donor feature (displayed by violet spheres) is identified by the Hypo1 hypothesis to be around position p2 or the backbone NH group of both Inh. #21 and #01 (Figure 3a,b). The same position is also identified by the Hypo1 hypothesis as a positive ionizable feature (displayed by red spheres) (Figure 3a,b) which agrees with the CoMFA result (displayed as the blue contours in Figure 2a). The favor regions for hydrophobic interaction identified by Hypo1 hypothesis are displayed with blue spheres and are at positions p1, p4, and p7 which agree with the orange contours of the CoMSIA result though only positions p4 and p7 of the latter are identified by the contours (Figure 2b). The p1 position is identified by both CoMFA and CoMSIA as a disfavor region for steric interaction and is displayed with yellow contours by both results (Figure 2a,b). An ionizable feature displayed by red spheres by the Hypo1 hypothesis is presumably mapped at position p2 of the inhibitor #17 (Figure 3c). However, a rather bad mapping result is observed at the position (Figure 3c) because the aIle atom originally belongs to the position is removed on the inhibitor. The mutation at position p7 of Inh. #36 also causes a bad mapping of the hydrophobic feature displayed by blue spheres by the Hypo1 hypothesis at the position which results in a great reduction of activity for the inhibitor (Figure 3d). A mapping of Hypo1 hypothesis onto the structures of Inh. #30 and #13 of the training and #70 and #58 of the test sets where an apparent deviation is found between the actual and predicted activity ratio is presented in Figure 4a−d, respectively. The prediction errors for both Inh. #70 and #58 of the test set are larger than that for #26 of the same set though a conflict in the activity scale is labeled only for the latter by the program (Table 6). With a value of 1.00 and 0.67, the predicted activity ratio of Inh. #13 of the training set is only slightly deviated from the actual one (Table 5). The slight deviation could be caused by bad mapping of the positive ionizable feature represented by red spheres by the Hypo1 hypothesis onto position p2 (Figure 4b). However, the prediction conflict for Inh. #70 may be simply caused by a steric hindrance introduced by the mutated group β-PhPro at position p5 which is unrecognized by the Hypo1 hypothesis in predicting the activity ratio (Table 1 and Figure 4c). The prediction error for Inh. #58 may be also ascribed to the fact that the bulky Haic group is incorrectly placed at positions p4 and p5 which are recognized as disfavored regions for steric (Figure 2b) and unrecognized by the Hypo1 hypothesis in predicting the activity ratio (Table 1 and Figure 4d). Similarly, a mutation at position p5 by the Pip group of Inh. #30 (Table 1) results in bad mapping of the Hypo1 hypothesis features onto positions p1 and p2 though the mutated group appears to strengthen the hydrophobic inter-

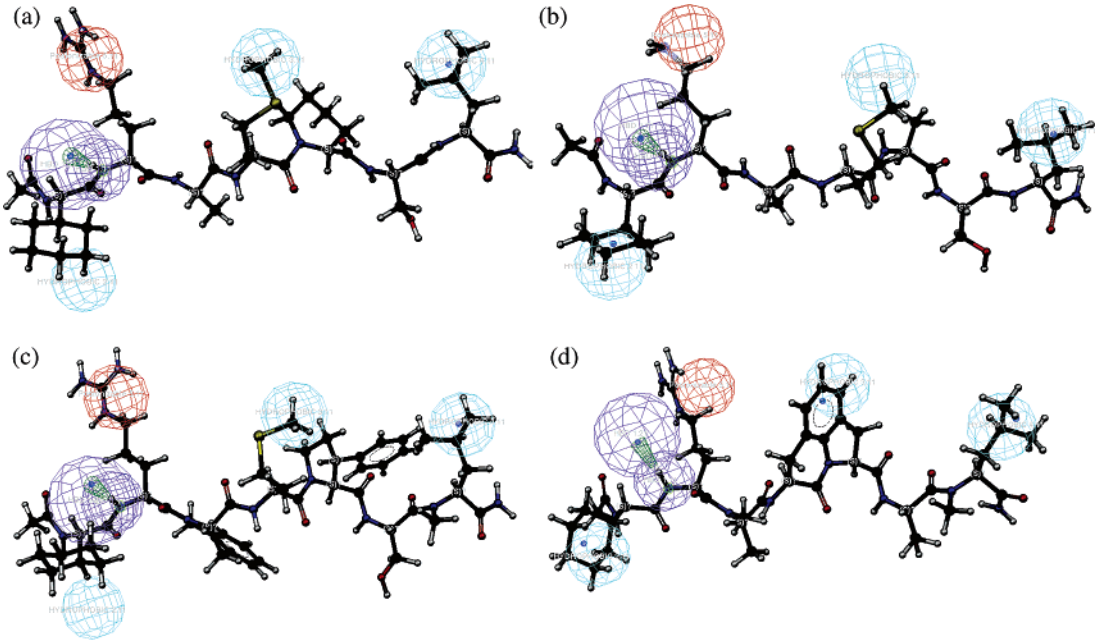

Figure 4. Mapping of the Hypo1 hypothesis onto the structures of four peptidomimetic inhibitors Inh. #30 (a), #13 (b), #70 (c), and #58 (d) (Tables 5 and 6) selected from both the training and the test sets. The pharmacophore features are color coded as follows: blue spheres for hydrophobic (H), violet spheres for hydrogen-bond donor (D), and red spheres for positive ionizable (I).

action and cause little steric hindrance for the region (Figure 4a).

CONCLUSION

In this work, we have shown that structural features identified by some 3D QSAR studies can be used to assist the construction of a pharmacophore model for a series of peptidomimetic inhibitors of HLA-DR4, a class II MHC peptide presenting molecule, by the Catalyst program. We have used some docking techniques to define the conformations of these peptidomimetic inhibitors for aligning them up for the 3D QSAR studies because their structures are rather flexible and diversified. We have also shown that the structural features selected by a stepwise CoMSIA can be plugged into the Catalyst program for an automatic generation of conformations by the latter for constructing some pharmacophore models for the peptidomimetic inhibitors studied. An accurate prediction result for the training set has been obtained by mapping the best pharmacophore model of the Hypo1 hypothesis onto the previously defined binding pockets for the receptor by others.[22] The underlying causes for some prediction conflicts for the test set where predicted activities are greater than the actual ones are also elucidated through the pharmacophore mapping process. The stepwise CoMSIA sweeps all the single or possible combinations of structural features involved in the binding process and identifies only two of them to be statistically important in the pharmacophore construction process. A q^2 value of 0.71 is obtained for structures aligned by the GOLD docking while that obtained for structures aligned by some correspondence points is only 0.57 for the same H+S field indexes employed in the CoMSIA studies. This shows that structures aligned by a field alignment method such as the GOLD docking are superior to the point alignment one in constructing the CoMSIA model. The only binding pocket that is unaccounted by the best CoMSIA model is p6 which prefers binding with some hydroxyalkyl groups as defined by the M13 phage display libraries by others.[21,22] This feature is included as D or a hydrogen-bond donor in the pharmacophore construction process by the Catalyst program. However, no significant declining in prediction accuracy of the training set is obtained if the feature is neglected in the pharmacophore construction process.

ACKNOWLEDGMENT

This work is supported in part by a grant from the National Science Council, (NSC93-2313-B007-002). The GOLD and GLIDE docking, SYBYL CoMFA, CoMSIA, and PLS plus the Catalyst 4.9 studies were conducted at the National Center for High Performance Computing, Taiwan.

REFERENCES AND NOTES

(1) Watts, C. Capture and processing of exogenous antigens for presentation on MHC molecules. *Annu. Rev. Immunol.* **1997**, *15*, 821−850.

(2) Rudensky, A.; Prestoa-Hurlburt, P.; Hong, S. C.; Barlow, A.; Janeway, C. A., Jr. Sequence analysis of peptides bound to MHC class II molecules. *Nature* **1991**, *353*, 622−627.

(3) Chicz, R. M.; Urban, R. G.; Lone, W. S.; Gorga, J. C.; Stern, L. J.; Vignali, D. A.; Strominger, J. L. Predominant naturally processed peptides bound to HLA-DR1 are derived from MHC-related molecules and are heterogeneous in size. *Nature* **1992**, *358*, 764−768.

(4) Tiwari, J.; Terasaki, P. *HLA and disease association;* Springer-Verlag: New York, 1985.

(5) Rowley, M. J.; Stockman, A.; Bond, C. A.; Tait, B. D.; Rowley, G. L.; Sherritt, M. A.; Mackay, I. R.; Muirden, K. D.; Bernard, C. C. The effect of HLA-DRB1 disease susceptibility markers on the expression of RA. *Scand. J. Rheumatol.* **1997**, *26*, 448−455.

(6) Weyand, C. M.; Goronzy, J. J. Inherited and noninherited risk factors in rheumatoid arthritis. *Curr. Opin. Rheumatol.* **1995**, *7*, 206−213.

(7) Nepom, G. T.; Gersuk, V.; Nepom, B. S. Prognostic implications of HLA genotyping in the early assessment of patients with rheumatoid arthritis. *J. Rheumatol. Suppl.* **1996**, *44*, 5−9.

(8) Wagner, U.; Kaltenhauser, S.; Sauer, H.; Arnold, S.; Seidel, W.; Hantzschel, H.; Kalden, J. R.; Wassmuth, R. HLA markers and prediction of clinical course and outcome in rheumatoid arthritis. *Arthritis Rheum.* **1997**, *40*, 341−351.

(9) Perdriger, A.; Chales, G.; Semana, G.; Guggenbuhl, P.; Meyer, O.; Quillivic, F.; Pawlotsky, Y. Role of HLA-DR-DR and DR-DQ

association in the expression of extraarticular manifestations and rheumatoid factor in rheumatoid arthritis. *J. Rheumatol.* **1997**, *24*, 1272−1276.
(10) Stern, L. J.; Brown, J. H.; Jardetzky, T. S.; Gorga, J. C.; Urban, R. G.; Strominger, J. L.; Wiley, D. C. Crystal structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide. *Nature* **1994**, *368*, 215−221.
(11) Jardetzky, T. S.; Brown, J. H.; Stern, L. J.; Urban, R. G.; Chi, Y. I.; Stauffacher C.; Strominger, J. L.; Wiley, D. C. Three-dimensional structure of a human class II histocompatability molecule complexed with superantigen. *Nature* **1994**, *368*, 711−718.
(12) Ghosh, P.; Amaya, M.; Mellins, E.; Wiley, D. C. The structure of an intermediate in class II MHC maturation: CLIP bound to HLA-DR3. *Nature* **1995**, *378*, 457−462.
(13) Brown, J. H.; Jardetzky, T. S.; Gorga, J. C.; Stern, L. J.; Urban, R. G.; Strominger, J. L.; Wiley, D. C. Three-dimensional structure of the human class II histocompatability antigen HLA-DR1. *Nature* **1993**, *364*, 33−39.
(14) Dessen, A.; Lawrence, C. M.; Cupo, S.; Zaller, D. M.; Wiley, D. C. X-ray crystal structure of HLA-DR4 (DRA0101, DRB10401) complexed with a peptide from human collagen II. *Immunity* **1997**, *7*, 473−481.
(15) Garboczi, D. N.; Ghosh, P.; Utz, F.; Oing, 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.
(16) 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.
(17) Reinherz, E. L.; Tan, K.; Tang, L.; Kern, P.; Liu, J.; Xiong, Y.; Hussey, E.; Smolyar, A.; Hare, B.; Zhong, R.; Joachimiak, A.; Chang, H.; Wagner, G.; Wang, J. The crystal structure of a T-cell receptor in complex with peptide and MHC class II. *Science* **1999**, *286*, 1913−1921.
(18) Hammer, J.; Takacs, B.; Sinigaglia, F.; Identification of a motif for HLA-DR1 binding peptides using M13 display libraries. *J. Exp. Med.* **1993**, *176*, 1007−1013.
(19) Hammer, J.; Valsasnini, P.; Tolba, K.; Bolin, D.; Higelin, J.; Takacs, B.; Sinigaglia, F. Promicsuous and allele-specific anchors in HLA-DR binding peptides. *Cell* **1994**, *74*, 197−203.
(20) Hammer, J.; Bono, E.; Gallazzi, F.; Belunis, C.; Nagy, Z. A.; Sinigaglia, F. Precise prediction of major histocompatability complex class II-peptide interaction based on peptide side chain scanning. *J. Exp. Med.* **1995**, *180*, 2353−2358.
(21) Hammer, J.; Callazzi, F.; Bono, E.; Karr, R. W.; Guenot, J.; Valsasin, P.; Nagy, Z. A.; Sinigaglia, F. Peptide binding specificity of HLA-DR4 molecules: correlation with rheumatoid arthritis association. *J. Exp. Med.* **1995**, *181*, 1847−1855.
(22) Bolin, D. R.; Swain, A. L.; Ramakanth, S.; Berthel, S. J.; Gillespie, P.; Huby, N. J. S.; Makofske, R.; Orzechowski, L.; Perrotta, A.; Toth, K.; Cooper, J. P.; Jiang, N.; Falcion, F.; Campbell, R.; Cox, D.; Gaizband, D.; Belunis, C. J.; Vidovic, D.; Ito, K.; Crowther, R.; Kammlott, U.; Zhang, X.; Palermo, R.; Weber, D.; Guenot, J.; Nagy, Z.; Olson, G. L. Peptide and peptide mimetic inhibitors of antigen presentation by HLA-DR class II MHC molecules. Design, structure−activity relationship, and X-ray crystal structure. *J. Med. Chem.* **2000**, *43*, 2135−2148.
(23) Hennecke, J.; Wiley, D. C. Structure of a complex of the human α/β T cell receptor (TCR) HA 1.7, influenza hemmagglutinin peptide and major histocompatability complex class II molecule, HLA-DR4 (DRA*0101 and DRB1*0401): insight into TCR cross-restriction and alloreactivity. *J. Exp. Med.* **2002**, *195*, 571−581.
(24) Adorini, L.; Muller, S.; Cardinaux, F.; Lehmann, P. V.; Falcioni, F.; Nagy, Z. A. In vivo competition between self-peptides and foreign antigens in T-cell activation. *Nature* **1988**, *334*, 623−625.
(25) Falcioni, F.; Ito, K.; Vidovic, D.; Belunis, C.; Campbell, R.; Berthel, S. J.; Bolin, D. R.; Gillespie, P. B.; Nuby, N.; Olson, G. L.; Sarabu, R.; Guenot, J.; Madison, V.; Hammer, J.; Sinigaglia, F.; Steinmetz, M.; Nagy, Z. A. Peptidomimetic compounds that inhibit antigen presentation by autoimmune disease-associated class II major histocompatability molecules. *Nature Biotechnol.* **1999**, *17*, 562−567.
(26) Zavala-Ruize, Z.; Sundberg, E. J.; Stone, J. D.; DeOliveira, D. B.; Chan, I. C.; Svendsent, J.; Mariuzza, R. A.; Stern, L. J. Exploration of the p6/p7 region of the peptide-binding site of the human class II major histocompatability complex protein HLA-DR1. *J. Biol. Chem.* **2003**, *278*, 44904−44912.
(27) Cramer, 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.
(28) 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.
(29) Catalyst, version 4.9 (software package); Accelrys, Inc. (previously known as Molecular Simulations, Inc.): San Diego, CA, 2003. http://www.accelrys.com.
(30) SYBYL 6.9.1; The Tripos Associates, 1699 S. Hanley Rd., St. Louis, MO.
(31) GOLD is distributed by the Cambridge Crystallographic Data Center, Cambridge, U.K. www.ccdc.ac.uk/prods/gold.html.
(32) GLIDE, Schrödinger LLC: Portland, 2003. http://www.schrodinger-.com/Products/glide.html.
(33) Halgren, T. A. Merck molecular force field I. Bassis, form, scope, parametrization, and performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490−519.
(34) Wang, R.; Lai, L.; Wang, S. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J. Comput.-Aided. Mol. Des.* **2002**, *16*, 11−26.
(35) Viswanadhan, V. N.; Ghose, A. K.; Revankar, G. R.; Robins, R. K. Atomic physicochemical parameters for three-dimensional structure directed quantitative structure−activity relationships. 4. Additional parameters for hydrophobic and dispersive interactions and their application for an automated superposition of certain naturally occurring nucleoside antibiotics. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 163−172.
(36) Klebe, G. The use of composite crystal-field environments in molecular recognition and the de novo design of protein ligands. *J. Mol. Biol.* **1994**, *237*, 212−235.
(37) Wallace, A. C.; Laskowski, R. A.; Thornton, J. M. LIGPLOT: a program to generate schematic diagrams of protein−ligand interactions. *Protein Eng.* **1995**, *8*, 127−134.

CI050140Y