Predicting Class I Major Histocompatibility Complex (MHC) Binders Using Multivariate Statistics: Comparison of Discriminant Analysis and Multiple Linear Regression

Irini A. Doytchinova*

Faculty of Pharmacy, Medical University of Sofia, 2 Dunav st., 1000 Sofia, Bulgaria

Darren R. Flower

The Jenner Institute, University of Oxford, Compton, Berkshire, RG20 7NN, United Kingdom

The accurate in silico identification of T-cell epitopes is a critical step in the development of peptide-based vaccines, reagents, and diagnostics. It has a direct impact on the success of subsequent experimental work. Epitopes arise as a consequence of complex proteolytic processing within the cell. Prior to being recognized by T cells, an epitope is presented on the cell surface as a complex with a major histocompatibility complex (MHC) protein. A prerequisite therefore for T-cell recognition is that an epitope is also a good MHC binder. Thus, T-cell epitope prediction overlaps strongly with the prediction of MHC binding. In the present study, we compare discriminant analysis and multiple linear regression as algorithmic engines for the definition of quantitative matrices for binding affinity prediction. We apply these methods to peptides which bind the well-studied human MHC allele HLA-A*0201. A matrix which results from combining results of the two methods proved powerfully predictive under cross-validation. The new matrix was also tested on an external set of 160 binders to HLA-A*0201; it was able to recognize 135 (84%) of them.

INTRODUCTION

Mass vaccination, which takes account of herd immunity, is now regarded as the principal prophylactic measure against endemic infectious disease. Successful vaccination involves deliberate exposure to immunogens and induces protective immunity. Subsequent exposure to pathogens will engender an enhanced recall response that clears the disease and mitigates its pathological effects. Existing vaccines are based on killed or live-attenuated microorganisms or subunits purified from microbes. A new approach to vaccine development isolates pathogen-derived peptides from host cells and uses them to induce immunity. Thus, peptide vaccines target epitopes to which a protective response can be induced. A key limitation of the peptide approach is that it is coupled tightly to the peptide specificity of particular major histocompatibily complex (MHC) alleles, so that some peptides may fail to be universally effective at inducing protective immunity.

MHC proteins hold a key position in the immune system. They bind protein fragments and present them at the cell surface for recognition by T cells. On the basis of their biological properties and chemical structure, MHC proteins are grouped into two classes: class I and class II. Both classes are polygenic (many genes) and extremely polymorphic (many alleles for each gene).[1] Peptides that bind to MHC class I proteins consist of 8−11 amino acids, although longer peptides up to 15 amino acids in length are now starting to be identified, and originate from two sources: self-proteins and antigenic proteins.[2] Their processing pathway involves the degradation of proteins by the proteasome, followed by transport mediated by the transporter associated with antigen processing (TAP) protein to the endoplasmic reticulum, where peptides are bound by MHC class I molecules. MHC−peptide complexes are then exocytosed to the cell surface, where they interact with T cells. The surface of T cells is enriched in a particular kind of receptor protein: the T-cell receptor (TCR), which acts by binding to peptide−MHC (pMHC) complexes expressed on the surface of other cells.

However, not all presented peptides are recognized by T cells. Those that are recognized are referred to as epitopes. Ligation of T-cell receptors by peptide−MHC complexes leads to many events downstream: clustering of TCRs, rapid segregation of signaling molecules into lipid rafts, and a whole swathe of complex intracellular events. Our understanding of the recognition process remains poor. The most crucial event is the phosphorylation of several residues in the CD3ζ chain, resulting in the activation of ZAP-70, which initiates further signaling. The activated T cell down-regulates surface TCR expression and up-regulates several surface markers, such as CD69 and CD25.

TCRs exhibit fine specificity for pMHC's. The immune system expresses large numbers of different TCRs, each with its own pMHC specificity. The repertoire of dominant and nondominant TCR specificities is largely shaped by the infection and vaccination history of an individual. The number of TCRs present in appreciable quantities will, however, still be many orders of magnitude greater than the three different class I MHCs. Usually, little or nothing is known about the repertoire of TCR specificities.

* Corresponding author e-mail: idoytchinova@pharmfac.net.

CLASS I MAJOR HISTOCOMPATIBILITY COMPLEX BINDERS

J. Chem. Inf. Model., Vol. 47, No. 1, 2007 **235**

Nonetheless, the pivotal event in epitope recognition is the formation of the ternary complex between TCR, peptide, and MHC. The binding of TCRs to pMHC is relatively weak (micromolar range) compared to peptide binding to the MHC (nanomolar range). Thus, peptide binding is a necessary, but not a wholly sufficient, condition for a peptide to be an epitope, and it is now accepted that workable T-cell epitope prediction methods depend on the accurate prediction of the most discriminatory step of the antigen presentation process: peptide binding to MHCs.

The accurate identification of T-cell epitopes is a critical step in the development of subunit and peptide-based vaccines.[3] Typically, the first stage of such studies is the in silico identification of potential MHC binders from the sequence of a studied protein. This is followed by labor-, time-, and resource-consuming experiments which aim to verify the T-cell recognition of the predicted peptides. As the verity of initial in silico predictions improves, so subsequent experimental work should become more efficient and more successful.

In spite of its singular importance, the accurate prediction of T-cell epitopes has yet to be solved and remains a powerful challenge to bioinformaticians. As is well-known, a large number of computational approaches to the prediction of T-cell epitopes have been developed over the past 15 years. Recently, a new generation of methods for T-cell epitope prediction, which are based on integrated multistep approaches, have been developed.[4−7] These methods are predicated upon reproducing key aspects of the antigen presentation pathway and using them to filter out potential peptides. Typically, three steps are modeled: proteasome cleavage, TAP transport, and MHC binding. These steps may be used in parallel or in a successive manner. EpiJen[5] is such a multistep T-cell epitope prediction algorithm recently developed in our group. It is based on quantitative matrices applied successively to select epitopes and is available free online via http://www.jenner.ac.uk/EpiJen.

During the development of EpiJen models for MHC binding affinity prediction, two types of statistical methods were used depending on the available data.[5] Quantitative data (continuous values like IC_{50}'s) were available for certain alleles, while for the rest only the sequences of binders were known (discontinuous values). Binding models based on continuous values were derived by multiple linear regression (MLR) and those based on discrete values by discriminant analysis (DA). "Leave-one-out" cross-validation, and external validation tests, indicated a higher predictive rate for the DA models than for the MLR models.

These results have important potential consequences for the development of MHC peptide binding models and warrant independent investigation. In particular, we have sought to explore whether this behavior is a common observation or is an allele-specific feature. We undertook a comparative study on the predictability of DA- and MLR-derived quantitative matrices (QMs) based on a large external test set of peptides binding to the HLA-A*0201 protein. As a suitable system with which to compare these approaches, we chose the human MHC allele HLA-A*0201, which is by far the best studied class I human MHC molecule; it has, in relative abundance, both continuous and discrete peptide binding data. We found an approach combining both DA and MLR to offer significant advantages in peptide affinity prediction.

DATA SETS AND METHODS

Training Set. A total of 326 nonamers which bound to the HLA-A*0201 protein were obtained from the AntiJen database[8] (http://www.jenner.ac.uk/AntiJen). This formed the positive training set (binders) for the DA model. In order to create a similarly sized negative training set (nonbinders), source proteins corresponding to the binders were processed using EpiJen models for proteasome cleavage, TAP, and MHC binding prediction. Only peptides with predicted pIC_{50} values above 5.3 (IC_{50} < 5000 nM) were selected. Binders were excluded. Only nonbinders with well-represented amino acids at each position (more than 15 times) were selected, resulting in a negative set of 321 nonbinders. The training set is given as Supporting Information.

Additive Method by DA−PLS. The training set consisted of 326 binders and 321 nonbinders. The additive method is described in details elsewhere.[9] Briefly, each peptide was represented as a binary string of length 180 (9 positions × 20 amino acids). A term is equal to 1 when a particular amino acid at a particular position is present and 0 when it is absent. As a dependent variable, binders were coded as a value of 1 and nonbinders took a value of 0. The matrix was solved using partial least squares (PLS) as implemented in SYBYL, version 7.0.[10] The optimum number of principal components (PCs) used to derive the model was defined as the number leading to the lowest standard error of prediction after "leave-one-out" cross-validation (LOO-CV). The prediction rate of the model was measured by LOO-CV using receiver operating characteristic (ROC) curves.[11] The variables *sensitivity* (true binders/total binders) and *1-specificity* (false binders/total nonbinders) were plotted at different thresholds, and the area under the curve (AUC_{ROC}) was used as a quantitative measure of predictive ability. It can vary from 0.5 for a random prediction to 1.0 for a perfect prediction. The *accuracy* of the prediction [(true binders + true nonbinders)/total] at a threshold of 0.5 was also calculated. The non-cross-validated model was used to predict the binding affinity of peptides from the test set.

Test Set. The peptides included in the test set were collected from the SYFPEITHI database.[12] Only nonamer peptides binding to the HLA-A*0201 allele were selected. All peptides already present in the training set were excluded. The final test set comprised 160 binders and their source proteins. This set was used to compare the predictive power of DA- and MLR-derived QMs. It is given as Supporting Information.

RESULTS

The DA-based QM for peptide binding affinity prediction to HLA-A*0201, developed in this study, is shown in Table 1. The optimum number of PCs is five. The LOO-CV gave $AUC_{ROC} = 0.889$ and an *accuracy* of 85% at a threshold of 0.5. As the negative set was selected to have all amino acids at all positions, there are no missing values in the DA model, although a few amino acids are absent at the anchor positions in the positive set. This is a definite advantage compared to the MLR QM, which is based only on binders and many terms, particularly for anchor positions, are missing. In

Table 1. DA-Derived QM for the Prediction of Affinity for Peptide Binding to the HLA-A*0201 Allele[a]

Aa	p1	p2	p3	p4	p5	p6	p7	p8	p9
Ala	0.052	0.088	0.062	0.122	−0.025	0.049	0.061	0.016	0.417
Arg	0.153	−0.142	−0.053	0.192	−0.058	−0.166	−0.087	−0.007	−0.017
Asn	0.203	−0.284	−0.039	0.048	−0.103	0.030	−0.124	−0.178	−0.141
Asp	−0.141	−0.305	0.073	0.049	0.067	0.063	0.048	−0.179	−0.159
Cys	−0.265	−0.244	−0.230	−0.197	−0.234	−0.057	0.077	−0.161	0.234
Gln	0.029	0.095	−0.010	−0.154	0.198	0.060	0.027	0.023	−0.412
Glu	−0.242	−0.138	−0.107	0.107	0.031	−0.027	0.110	0.193	−0.295
Gly	0.051	−0.034	−0.053	0.041	0.004	−0.011	0.022	−0.018	−0.195
His	−0.178	−0.086	−0.032	−0.109	0.048	−0.167	−0.051	−0.052	−0.216
Ile	−0.006	0.274	0.195	0.091	−0.004	0.168	0.180	0.205	0.419
Leu	0.076	0.378	0.006	0.154	−0.077	0.001	0.091	0.122	0.478
Lys	0.105	−0.075	−0.250	−0.003	0.074	−0.131	−0.219	0.041	−0.012
Met	−0.133	0.348	−0.007	−0.253	−0.215	−0.062	−0.087	−0.202	0.063
Phe	0.170	−0.296	0.008	−0.050	0.002	0.108	0.171	0.005	0.070
Pro	−0.110	0.054	0.066	0.087	0.048	0.137	−0.066	0.038	−0.324
Ser	0.112	−0.150	0.120	0.082	−0.013	0.072	0.090	0.021	−0.033
Thr	0.077	0.373	0.106	−0.007	0.072	0.145	0.032	0.013	0.186
Trp	−0.162	−0.079	0.087	−0.229	0.158	−0.421	−0.365	−0.039	−0.322
Tyr	0.115	−0.039	−0.069	−0.028	0.001	0.086	0.035	0.074	−0.175
Val	0.094	0.263	0.126	0.059	0.027	0.125	0.055	0.084	0.435

[a] The constant for the model is −0.180. The model has $AUC_{ROC} = 0.889$ and an accuracy of 85% at threshold 0.5.

Table 2. Statistics for Predictions on the Test Set

QM	model	true binders	false binders	true nonbinders	false nonbinders	sensitivity %	specificity %	accuracy %
single	DA	126	3850	74 531	34	79	95	95
	MLR	126	3850	74 531	34	79	95	95
combined	parallel	135	3841	74 540	25	84	95	95
	successive DA+MLR	132	3844	74 537	28	83	95	95
	successive MLR+DA	133	3846	74 535	30	81	95	95

general, missing terms reduce the generality and prediction accuracy of the QM.[9]

The DA-based model was used to predict binders present in the test set. Results were compared with predictions made by a QM for HLA-A*0201 binding affinity prediction previously derived by MLR[13] (Table 2). The EpiJen server contains this MLR QM and was used to make MLR predictions. For this purpose, source proteins corresponding to test set binders were processed using EpiJen. The MLR QM in EpiJen was then substituted by DA QM, and EpiJen was used offline to predict test set binders. In both predictions, the thresholds for the proteasome cleavage and TAP binding were 0.1 and 5.0, respectively. Only the top 5% of the predicted peptides were considered. Equal numbers for true and false binders and nonbinders were achieved by both models, with a *sensitivity* of 79%. Because the number of nonbinders in a protein is significantly higher than the number of binders, the *specificity* and the *accuracy* of the predictions were 95%. The conclusion drawn from these results was that the two QMs are equally predictive and the observed superiority of the DA models seen previously[5] is allele-specific.

Analysis of the false nonbinders identified by the two methods indicated that one-third was comprised of different peptides. In order to reduce this number, and improve the sensitivity of predictions, we tried different combinations of the DA and MLR QMs: one parallel and two successive. In the parallel combination, the two matrices were merged into one, but in the successive combination, the matrices were used in series: first DA followed by MLR, then MLR followed by DA. Results are shown in Table 2. All combinations gave better predictions than the single QMs

with the parallel combination being the best. The parallel model predicted 135 out of 160 binders (*sensitivity* of 84%), while the two successive models recognized 132 (*sensitivity* of 83%) and 130 (*sensitivity* of 81%) binders, respectively, for the DA+MLR and MLR+DA models. The parallel matrix replaced the old MLR QM formerly implemented in EpiJen.

DISCUSSION

One of the prime tasks for computational vaccinology is the improvement of methods for epitope prediction and identification. The accuracy of such prediction will impact greatly on the success of subsequent experimental work. This imperative has fueled considerable efforts to perfect epitope prediction, yet problems remain. Early and inaccurate methods addressed direct T-cell epitope prediction using methods based on amphipathicity[14] or sequence patterns or "motifs".[15] Subsequently, a diverse range of methods have been developed which predict MHC binders instead of epitopes.[16] First among such approaches was the use of peptide binding motifs,[17,18] which sought to characterize the peptide specificity of different MHC alleles in terms of dominant anchor positions which express strong preferences for certain amino acids: that is, the well-known human class I MHC HLA-A*0201 has anchor residues at peptide positions p2 and p9. At p2, acceptable amino acids are Leu and Met, while at p9, acceptable amino acids are Val and Leu. Secondary anchors (residues favorable, but not essential, for binding) may also be present. Many papers have extended the list of motifs to include the specificity of many alleles from humans and other animals. However, despite their success, there are still many fundamental problems associated

CLASS I MAJOR HISTOCOMPATIBILITY COMPLEX BINDERS

J. Chem. Inf. Model., Vol. 47, No. 1, 2007 **237**

with the motif approach. The most significant such problem is that motifs produce many false positives, and probably also many false negatives, although peptides predicted to be inactive are seldom tested. Being motif-positive is neither necessary nor sufficient for a peptide to possess MHC affinity.

Obvious limitations in the performance of motifs have led many to seek other, better ways of predicting peptide−MHC affinity. Operating ostensibly at the sequence level, these methods utilize a rich variety of underlying methodologies: sequence profiles,[19,20] QMs,[21−23] and artificial intelligence techniques, such as artificial neural networks (ANNs)[24−27] or hidden Markov models. Methods operating on the three-dimensional structure of peptides are also prevalent and include free energy scoring,[28] threading,[29] and 3D quantitative structure−activity relationship studies.[30−32] The most recent development in epitope prediction has been the application of another artificial intelligence technique: support vector machines (SVM).[33−35]

In this paper, we have compared two fundamentally different approaches to generating predictive methods: DA and MLR. As before, we have used the well-studied human MHC allele HLA-A*0201 to exemplify our methodology. We derived QMs using DA and MLR methods and found them to be equally predictive. However, combining both matrices gave significantly better predictions. The new boosted matrix has been included in our recently launched server for T-cell epitope prediction: EpiJen. EpiJen is a reliable four-step algorithm and belongs to the next generation of in silico T-cell epitope prediction methods. These methods model the whole antigen processing pathway, including proteasome cleavage, TAP transport, and MHC binding. The advantages of such integrated methods are higher accuracy and a lower rate of false positive predictions; that is, the number of peptides that must be tested in order to identify true epitopes is reduced greatly.

All prediction methods share a common problem: the initial data set used for model development. The ideal set of data should be created by experimental design; it should cover all available chemical (or peptide) space, and it should be derived by the reproducible, automated screening of ligands. In general, such data is simply not available. Instead, one has to deal with quite diverse and miscellaneous data generated by different labs using different experimental methodologies. This complicates calculations, introduces noise into the models, and can greatly reduce the accuracy of prediction. Even when powerful robust statistical techniques like PLS, ANN, and SVM are applied to such data, the predictions can still be unsatisfactory.

Comparing the DA QM, developed in this study, with a previously derived MLR QM for HLA-A*0201 binding affinity prediction,[13] several differences are apparent. Although they are built from almost identical sets of binders, they are very different. Agreement exists only for the anchor positions (p2 and p9). For other positions, no correlations exist. Indeed, certain amino acids make opposite contributions to affinity. For example, Gln, Arg, Ser, and Thr at position 1 are favored according to the DA QM and disfavored according to the MLR QM. The same is true for Ser, Thr, and Val at position 3; Ala, Ile, and Leu at position 4; and so forth. Practically speaking, these differences are due mainly to the set of nonbinders used by the DA. Despite

these differences, however, both QMs perform equally well. Two-thirds of false nonbinders, as predicted by DA and MLR QMs, are common. Combining the matrices aims to reduce the number of different false nonbinders found in the other third. This new QM was implemented in the EpiJen server and is freely accessible via the URL http://www.jenner.ac.uk/EpiJen.

Conceptually, these differences arise, in part, because they are attempting to predict different quantities. MLR is trying to model a continuous real value while DA is undertaking classification. The MLR model will weight contributions from high-affinity peptides more than lower-affinity peptides. DA will weight contributions equally according to class. Thus, DA may produce models that better represent mean class membership and thus have different residue-by-residue contributions within its QM.

DA outperforms MLR under certain conditions, but not, as we have shown, in all circumstances. This may reflect problems with the data, which, in general, have not been designed to cover all available chemical (peptide) space. It may also be related to the differences in the type of data we are exploring. Since DA is classifying by class, it can use alternate data sources equally. This is another practical advantage to using DA. For example, any affinity measure can be used to identify binders, be that an IC_{50} value or a half-life or a BL_{50} value. Likewise, a peptide eluted from a cell-surface MHC must obviously be bound, with reasonable affinity, before being eluted. Any peptide found, from, say, overlapping peptide scanning, to be an epitope must, again, have been bound by an MHC prior to being recognized. Thus, all of these conceptually distinct means of classification yield equivalent definitions of binder versus nonbinder. This allows a DA-based approach to rationalize a much wider array of data. This is a potentially useful feature of such an approach, though not one exploited in our controlled analysis, as presented here. By combining multiple data sources for binders with the EpiJen-based definition of nonbinders, we can envisage the development of much more robust, much more general, much more complete—and thus much more successful—future methods for the prediction of MHC binders.

In conclusion, DA- and MLR-derived QMs for affinity prediction of peptide binding to HLA-A*0201 are, in terms of *sensitivity*, *specificity*, and *accuracy*, equipotent in their predictive power. The parallel combination of both matrices gives better prediction results, achieving a *sensitivity* of 84%. The combination of methods outlined in this paper points the way to significant future advances in the accurate prediction of peptide−MHC binding affinity.

Supporting Information Available: The training and test sets of peptides used in the present study are given. This information is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Janeway, C. A., Jr.; Travers, P.; Walport, M.; Capra, J. D. *Immunobiology. The Immune System in Health and Disease*; Current Biology Publications: London, 1999; pp 115−159.

(2) Shastri, N.; Schwab, S.; Serwold, T. Producing Nature's Gene-Chips: The Generation of Peptides for Display by MHC Class I Molecules. *Annu. Rev. Immunol.* **2002**, *20*, 463−493.

(3) Sette, A.; Livingston, B.; McKinney, D.; Appella, E.; Fikes, J.; Sidney, J.; Newman, M.; Chesnut, R. The Development of Multi-Epitope Vaccines: Epitope Identification, Vaccine Design and Clinical Evaluation. *Biologicals* **2001**, *29*, 271−276.

(4) Peters, B.; Sette, A. Generating Quantitative Models Describing the Sequence Specificity of Biological Process with the Stabilized Matrix Method. *BMC Bioinf.* **2005**, *6*, 132.

(5) Doytchinova, I. A.; Guan, P.; Flower, D. R. EpiJen: A Server for Miltistep T Cell Epitope Prediction. *BMC Bioinf.* **2006**, *7*, 131.

(6) Larsen, M. V.; Lundegaard, C.; Lamberth, K.; Buus, S.; Brunak, S.; Lund, O.; Nielsen, M. An Integrative Approach to CTL Epitope Prediction: A Combined Algorithm Integrating MHC Class I Binding, TAP Transport Efficiency, and Proteasomal Cleavage Predictions. *Eur. J. Immunol.* **2005**, *35*, 2295−2303.

(7) Dönnes, P.; Kohlbacher, O. Integrated Modeling of the Major Events in the MHC Class I Antigen Processing Pathway. *Protein Sci.* **2005**, *14*, 2132−2140.

(8) Toseland, C. P.; Taylor, D. J.; McSparron, H.; Hemsley, S. L.; Blythe, M. J.; Paine, K.; Doytchinova, I. A.; Guan, P.; Hattotuwagama, C. K.; Flower, D. R. AntiJen: A Quantitative Immunology Database Integrating Functional, Thermodynamic, Kinetic, Biophysical, and Cellular Data. *Immun. Res.* **2005**, *1*, 4.

(9) Doytchinova, I. A.; Blythe, M. J.; Flower, D. R. Additive Method for the Prediction of Protein−Peptide Binding Affinity. Application to the MHC Class I Molecule HLA-A*0201. *J. Proteome Res.* **2002**, *1*, 263−272.

(10) *SYBYL 7.0*; Tripos Inc.: St. Louis, MO.

(11) Bradley, A. P. The Use of the Area Under the ROC Curve in the Evaluation of Machine Learning Algorithms. *Pattern Recognit.* **1997**, *30*, 1145−1159.

(12) 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.

(13) Doytchinova, I. A.; Flower, D. R. The HLA-A2-Supermotif: A QSAR Definition. *Org. Biomol. Chem.* **2003**, *1*, 2648−2654.

(14) DeLisi, C.; Berzofski, J. A. T-Cell Antigenic Sites Tend To Be Amphipathic Structures. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 7048−7052.

(15) Rothbard, J. B.; Taylor, W. R. A Sequence Pattern Common to T Cell Epitopes. *EMBO J.* **1988**, *7*, 93−100.

(16) Flower, D. R.; Doytchinova, I. A.; Paine, K.; Taylor, P.; Blythe, M. J.; Lamponi, D.; Zygouri, C.; Guan, P.; McSparron, H.; Kirkbride, H. *Computational Vaccine Design. Drug Design: Cutting Edge Approaches*; Royal Society of Chemistry: Cambridge, U. K., 2002; pp 136−180.

(17) Sette, A.; Buus, E.; Appella, J. A.; Smith, R.; Chesnut, R.; Miles, C.; Colon, S. M.; Grey, H. M. Prediction of Major Histocompatibility Complex Binding Regions of Protein Antigens by Sequence Pattern Analysis. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 3296−3300.

(18) 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.

(19) Reche, P. A.; Glutting, J. P; Reinherz, E. L. Prediction of MHC Class I Binding Peptides Using Profile Motifs. *Hum. Immunol.* **2002**, *63*, 701−709.

(20) Reche, P. A.; Glutting, J. P.; Reinherz, E. L. Enhancement to the RANKPEP Resource for the Prediction of Peptide Binding to MHC Molecules Using Profiles. *Immunogenetics* **2004**, *56*, 405−419.

(21) Marshall, K. W.; Wilson, K. J.; Liang, J.; Woods, A.; Zaller, D.; Rothbard, J. B. Prediction of Peptide Affinity to HLA-DRB1*0401. *J. Immunol.* **1995**, *154*, 5927−5933.

(22) Doytchinova, I. A.; Blythe, M. J.; Flower, D. R. Additive Method for the Prediction of Protein−Peptide Binding Affinity. Application to the MHC Class I Molecule HLA-A*0201. *J. Proteome Res.* **2002**, *1*, 263−272.

(23) Doytchinova, I. A.; Flower, D. R. Towards the in Silico Identification of Class II Restricted T Cell Epitopes: A Partial Least Squares Iterative Self-Consistent Algorithm for Affinity Prediction. *Bioinformatics* **2003**, *19*, 2263−2270.

(24) Bisset, L. R.; Fierz, W. Using a Neutral Network to Identify Potential HLA-DR1 Binding Sites within Proteins. *J. Mol. Recognit.* **1993**, *6*, 41−48.

(25) Gulukota, K.; DeLisi, C. Neural Network Method for Predicting Peptides that Bind Major Histocompatibility Complex Molecules. *Methods Mol. Biol.* **2001**, *156*, 201−209.

(26) Honeyman, M. C.; Brusic, V.; Stone, N. L.; Harrison, L. C. Neural Network-Based Prediction of Candidate T-Cell Epitopes. *Nat. Biotechnol.* **1998**, *16*, 966−969.

(27) 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.

(28) 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.

(29) Altuvia, Y.; Schueler, O.; Margalit, H. Ranking Potential Binding Peptides to MHC Molecules by a Computational Threading Approach. *J. Mol. Biol.* **1995**, *249*, 244−250.

(30) Doytchinova, I.; Flower, D. R. Towards the Quantitative Prediction of T-Cell Epitopes: CoMFA and CoMSIA Studies of Peptides with Affinity to Class I MHC Molecule HLA-A*0201. *J. Med. Chem.* **2001**, *44*, 3572−3581.

(31) Doytchinova, I. A.; Flower, D. R. A Comparative Molecular Similarity Index Analysis (CoMSIA) Study Identifies an HLA-A2 Binding Supermotif. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 535−544.

(32) Guan, P.; Doytchinova, I. A.; Flower, D. R. HLA-A3-Supermotif Defined by Quantitative Structure−Activity Relationship Analysis. *Protein Eng.* **2003**, *16*, 11−18.

(33) Dönnes, P.; Elofsson, A. Prediction of MHC Class I Binding Peptides Using SVMHC. *BMC Bioinf.* **2002**, *3*, 25.

(34) Zhao, Y.; Pinilla, C.; Valmori, D.; Martin, R.; Simon, R. Application of Support Vector Machines for T-Cell Epitope Predictions. *Bioinformatics* **2003**, *19*, 1978−1984.

(35) Liu, W.; Meng, X.; Xu, Q.; Flower, D. R.; Li, T. Quantitative Prediction of Mouse Class I MHC Peptide Binding Affinity Using Support Vector Machine Regression (SVR) Models. *BMC Bioinf.* **2006**, *7*, 182.