Quantitative Prediction of Peptide Binding to HLA-DP1 Protein

Stefan Ivanov, Ivan Dimitrov, and Irini Doytchinova

Abstract—The exogenous proteins are processed by the host antigen-processing cells. Peptidic fragments of them are presented on the cell surface bound to the major hystocompatibility complex (MHC) molecules class II and recognized by the CD4+ T lymphocytes. The MHC binding is considered as the crucial prerequisite for T-cell recognition. Only peptides able to form stable complexes with the MHC proteins are recognized by the T-cells. These peptides are known as T-cell epitopes. All T-cell epitopes are MHC binders, but not all MHC binders are T-cell epitopes. The T-cell epitope prediction is one of the main priorities of immunoinformatics. In the present study, three chemometric techniques are combined to derive a model for *in silico* prediction of peptide binding to the human MHC class II protein HLA-DP1. The structures of a set of known peptide binders are described by amino acid z-descriptors. Data are processed by an iterative self-consisted algorithm using the method of partial least squares, and a quantitative matrix (QM) for peptide binding prediction to HLA-DP1 is derived. The QM is validated by two sets of proteins and showed an average accuracy of 86 percent.

Index Terms—MHC binding prediction, iterative self-consistent algorithm, partial least squares method, z-descriptors

1 Introduction

THE adaptive immune system provides life-long protective immunity to reinfection with the same pathogen. The immune responses are mediated by B and T lymphocytes. T cells recognize a complex between a specific major histocompatibility complex (MHC) protein and a particular antigen-derived peptidic fragment named epitope. Epitopes are generated inside the cell *via* antigen processing. It takes one of two major pathways, endogenous or exogenous, each of which has several crucial steps, imperative to the immune response [1].

The exogenous pathway of antigen processing concerns mainly peptidic antigens, originating from outside the cell. In short, it involves four main steps: uptake of the foreign molecule by endocytosis, enzymatic cleavage to oligopeptides, binding to MHC class II protein, and presenting the peptide-MHC complex on the cell surface. The cleavage is carried out in the endosomes, under the complex regulation of a multitude of factors. The ingested protein is rapidly degraded by endosomic and lysosomic proteases [2], [3]. Most proteases have interchangeable activities [4]. As the endosome is an extremely hostile environment, the peptides rapidly bind to MHC class II molecules, before being completely cleaved to amino acids and dipeptides. The peptides binding to MHC class II molecules consist of 13 to 26 amino acids [5]. The α- and β-chains of MHC class II molecule are synthesized in the endoplasmic reticulum (ER), being adjoined to a special peptide, and known as the invariant chain (Ii). It blocks the binding cleft of the MHC molecule, thereby preventing the binding of endogenous peptides. The Ii also plays a role in transporting the MHC molecule from the ER to the Golgi apparatus and the incorporation of the molecule in the endosome, which contains ingested proteins and their peptides. The Ii is removed prior to the binding of these peptides; then, the complex is externalized and recognized by CD4+ T-cells. Not all complexes are recognized, only those containing epitopes.

Human MHCs, known as human leukocyte antigens (HLA), are extremely polymorphic and polygenic. The IMGT/HLA database lists over 1,700 HLA class II proteins [6]. HLA class II contains three loci: DR, DQ, and DP. DR and DQ proteins are well studied, while DP was initially considered of lesser importance in immune responses. However, it is now clear that HLA-DP proteins have important roles in mediating the immune response to many diseases, such as graft-versus-host disease [7], sarcoidosis [8], juvenile chronic arthritis [9], Graves' disease [10], hard metal lung disease [11], and, especially, chronic beryllium disease [12]. Recently, the X-ray structure of the HLA-DP2 (DPA*0103, DPB1*0201) in complex with a self-peptide derived from the HLA-DR α-chain has been determined [13]. Although the overall structure of DP2 is similar to that of other MHC class II proteins, it contains a unique solvent-exposed acidic pocket containing three glutamic acids ($\mathrm{Glu}^{26\beta}$, $\mathrm{Glu}^{68\beta}$, and $\mathrm{Glu}^{69\beta}$). This pocket may be able to bind Beryllium and present it to T cells, providing a mechanistic explanation underlying chronic Beryllium disease [13], [14]. X-ray data also reveal that the DP2 binding site comprises four binding pockets: deep, hydrophobic pockets p1 and p6; large, shallow, negatively charged p4; and deep, narrow and polar pocket p9.

One of the key issues in T-cell epitope prediction is the prediction of MHC binding, as it is considered as a prerequisite for T-cell recognition. All T-cell epitopes are MHC binders, but not all MHC binders are T-cell epitopes. The T-cell epitope prediction is one of the main priorities of immunoinformatics. Currently, there are two approaches to this problem—sequence-based and structure-based. Sequence-based methods include binding motif recognition [15], [16], [17], [18], [19], quantitative matrices [20], [21], [22], [23], [24], [25], [26], artificial neural networks (ANN) [27], support vector machines [28], etc. Structure-based methods involve peptide docking [29], [30], [31], [32], threading algorithms [33], [34], [35], and molecular dynamics [36], [37], [38], [39]. Recent comparative study found that ANN-based predictions are the most reliable among the other methods [40].

In the present study, three chemometric techniques are combined to derive a sequence-based model for *in silico* prediction of peptide binding to HLA-DP1. The structures of a set of known peptide binders are described by amino acid z-descriptors. Data are processed by an iterative self-consisted algorithm using the method of partial least squares (PLS) and a quantitative matrix (QM) for peptide binding prediction to HLA-DP1 is derived. The QM is validated by two sets of proteins and showed an average accuracy of 86 percent. Additionally, the derived QM gives unambiguous interpretation of the quantitative contribution of each amino acid at each binding core position.

2 Methods

2.1 Data Sets

All data sets used in the present study were extracted from the Immune Epitope Database (IEDB) [41] in the period April-October 2012 using the following keywords: MHC binding search; MHC binding assay; MHC allele name: HLA-DPA1*02:01/DPB1*01:01; assay: purified MHC competitive binding measuring half maximal inhibitory concentration (IC_{50}) [nM] using radioactivity.

2.1.1 Peptide Training Set

The training set consisted of 406 unique 15-mer peptides with affinity to HLA-DP1 measured as IC_{50} value [42]. Each 15-mer peptide was represented as a set of overlapping nonamers acquiring the same IC_{50} value as the parent 15-mer. The IC_{50} values were in the range 0.215-10,000 nM and for the purpose of

• *The authors are with the School of Pharmacy, Medical University of Sofia, 2 Dunav St., Sofia 1000, Bulgaria. E-mail: {sivanov, idimitrov, idoytchinova}@pharmfac.net.*

Manuscript received 26 Feb. 2013; revised 16 May 2013; accepted 20 June 2013; published online 2 July 2013.
For information on obtaining reprints of this article, please send e-mail to: tcbb@computer.org, and reference IEEECS Log Number TCBB-2013-02-0061. Digital Object Identifier no. 10.1109/TCBB.2013.78.

TABLE 1
The Three z-Descriptors for the 20 Naturally Occurring Amino Acids [43]

aa	z_1	z_2	z_3	aa	z_1	z_2	z_3
Ala	0.07	-1.73	0.09	Leu	-4.19	-1.03	-0.98
Arg	2.88	2.52	-3.44	Lys	2.84	1.41	-3.14
Asn	3.22	1.45	0.84	Met	-2.49	-0.27	-0.41
Asp	3.64	1.13	2.36	Phe	-4.92	1.30	0.45
Cys	0.71	-0.97	4.13	Pro	-1.22	0.88	2.23
Gln	2.18	0.53	-1.14	Ser	1.96	-1.63	0.57
Glu	3.08	0.39	-0.07	Thr	0.92	-2.09	-1.40
Gly	2.23	-5.36	0.30	Trp	-4.75	3.65	0.85
His	2.41	1.74	1.11	Tyr	-1.39	2.32	0.01
Ile	-4.44	-1.68	-1.03	Val	-2.69	-2.53	-1.29

the study were converted into p-units $(-logIC_{50})$. The training set consisted of 2,842 nonamer peptides.

2.1.2 Internal Protein Test Set

The parent proteins of the 15-mers from the training set were used as an internal test set. Their number was 34. Each protein was represented as a set of overlapping nonamers. The binding affinity of the nonamers was predicted by the derived model. A binding threshold of $pIC_{50} = 6.301 (IC_{50} = 500\ \text{nM})$ was selected. Nonamers with affinity above or equal to 6.301 are considered as predicted binders, and below 6.301 as predicted nonbinders. If any of the predicted nonamers is included in the known binding 15-mer sequences, the predicted peptide was considered a true binder; otherwise, it was considered a false binder.

2.1.3 External Protein Test Set

Additional number of known peptide binders to HLA-DP1 with different length and their parent proteins were compiled from IEDB. The number of known peptide binders was 45 originating from 36 proteins. The validation of the model was performed by the same procedure as the internal protein test set.

2.2 *z*-Descriptors

The z-descriptors, defined by Hellberg et al. [43], represent the principal physical and chemical properties of the 20 naturally occurring amino acids (see Table 1). These descriptors have been derived by PCA on a set of physicochemical properties as molecular weight, pK_a values, ^{13}C NMR-shifts, etc. The first principal component (1PC), named z_1, is dominated by the hydrophobicity of amino acids. The second principal component (2PC), named z_2, relates best to amino acid size. The third principal component (3PC), named z_3, explains the electronic properties of amino acids. The scores of these components define the set of z-descriptors for each amino acid. In the present study, the three z-descriptors were used to describe the nonamer peptide amino acid sequences. The X-matrix of the training set contained 27 z-descriptors (9 positions $\times$ 3 z-descriptors) for 2,842 nonamers.

2.3 Iterative Self-Consistent (ISC) Algorithm

The initial training set of 27 z-descriptors (X variables), pIC_{50} values (Y variable) and 2,842 peptides (observations) was used to extract the first model by applying PLS-based multiple linear regression (MLR) [44]. PLS forms new X variables (PC) as linear combinations of old variables and then uses them to predict biological activity. The optimum number of principal components (PCs) was derived by cross validation in seven groups. The first model was used to predict the pIC_{50}*s* of the initial set, and the best predicted nonamers from each parent 15-mer binder was collected into a second training set. This second set was used to extract the second quantitative structure activity relationship (QSAR) model,

no.	IEDB	nonamer	pIC50	z11	z21	z31	z12	z22	z32
1	1737117	AAASVPAAD	5.09	0.07	-1.73	0.09	0.07	-1.73	0.09
2	1737117	AASVPAADK	5.09	0.07	-1.73	0.09	0.07	-1.73	0.09
3	1737117	ASVPAADKF	5.09	0.07	-1.73	0.09	1.96	-1.63	0.57
4	1737117	SVPAADKFK	5.09	1.96	-1.63	0.57	-2.69	-2.53	-1.29
5	1737117	VPAADKFKT	5.09	-2.69	-2.53	-1.29	-1.22	0.88	2.23
6	1737117	PAADKFKTF	5.09	-1.22	0.88	2.23	0.07	-1.73	0.09

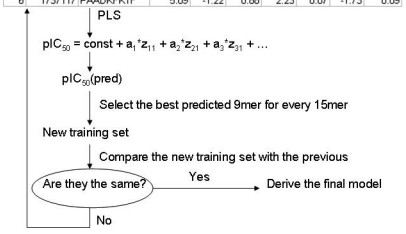

Fig. 1. Iterative self-consistent algorithm.

which predicted again the pIC_{50}*s* of the initial training set. The best predicted nonamers from each parent binder were selected and placed in a third training set. The selection procedure was repeated until maximum overlapping of the peptides in two consecutive training sets was achieved. Then, the final model was derived and considered further in the study.

The ISC algorithm was illustrated in Fig. 1. The PLS calculations were performed by SIMCA-P 13.0 (Umetrics AB, version 13.0.0.0, 2012). The algorithm was automated by tools written in C++.

2.4 Validation

The predictive ability of the derived final model was assessed by receiver operating characteristic (*ROC*) statistics [45]. Four outcomes are possible in *ROC* statistics: *true positives* (*TP*, true binders predicted as binders); *true negatives* (*TN*, true nonbinders predicted as nonbinders); *false positives* (*FP*, true nonbinders predicted as binders); and *false negatives* (*FN*, true binders predicted as nonbinders). Three classification functions were used in the present study: *sensitivity* (true positives/total positives), *specificity* (true negatives/total negatives), and *accuracy* (true positives and negatives/total). *Sensitivity*, *specificity*, and *accuracy* were calculated at different thresholds and the *area under the ROC curve* (*sensitivity/1-spesificity*) (A_{ROC}) was calculated. A_{ROC} is a quantitative measure of predictive ability and varies from 0.5 for random prediction to 1.0 for perfect prediction.

3 Results

3.1 QSAR Model for Prediction of Peptide Binding to HLA-DP1

The ISC algorithm achieved self-consistency at the 39th iteration. The maximum overlapping was 97 percent. The derived final model is shown in Table 2. The optimal number of principal components is 2. The contribution of each amino acid at each peptide position was calculated by substitution of the corresponding z-descriptors in the model. The amino acid contributions are compiled into a QM given in Table 3.

According to our model, hydrophobic amino acids as Ile, Leu, Phe, Trp, Tyr, Val, and Met are preferred at positions 1, 3, 4, 6, and 8. Polar residues like Arg, Lys, Gln, and Glu are well accepted at positions 2, 7, and 9. Position 5 tolerates both hydrophobic and positively charged residues.

TABLE 2
Final QSAR Model for Peptide Binding to HLA-DP1

	p1	*p2*	*p3*	*p4*	*p5*	*p6*	*p7*	*p8*	*p9*
z_1	-0.110	0.111	-0.145	-0.124	-0.080	-0.114	0.072	-0.203	0.049
z_2	0.049	-0.014	0.026	0.145	-0.011	-0.034	0.064	0.091	-0.082
z_3	0.039	-0.097	0.118	-0.006	-0.193	0.063	-0.110	0.050	-0.089

The constant of the model is 5.904.

TABLE 3
Amino Acid Contributions for Peptide Binding to HLA-DP1

	p1	*p2*	*p3*	*p4*	*p5*	*p6*	*p7*	*p8*	*p9*
Ala	-0.088	0.023	-0.044	-0.259	-0.003	0.057	-0.115	-0.168	0.136
Arg	-0.327	**0.619**	-0.760	0.029	**0.406**	-0.630	**0.746**	-0.526	0.241
Asn	-0.250	0.257	-0.330	-0.195	-0.435	-0.363	0.233	-0.480	-0.037
Asp	-0.252	0.161	-0.220	-0.304	-0.758	-0.304	0.077	-0.518	-0.126
Cys	0.036	-0.308	**0.360**	-0.255	-0.842	0.213	-0.463	-0.027	-0.255
Gln	-0.258	**0.346**	-0.437	-0.186	0.040	-0.338	**0.316**	-0.451	0.164
Glu	-0.322	**0.344**	-0.445	-0.325	-0.236	-0.368	0.255	-0.593	0.124
Gly	-0.494	0.292	-0.427	-1.053	-0.174	-0.050	-0.212	-0.927	**0.519**
His	-0.136	0.137	-0.173	-0.055	-0.426	-0.264	0.163	-0.275	-0.124
Ile	**0.365**	-0.371	**0.479**	**0.315**	**0.571**	**0.497**	-0.315	**0.697**	0.013
Leu	**0.371**	-0.357	**0.465**	**0.377**	**0.534**	**0.450**	-0.261	**0.708**	-0.032
Lys	-0.365	**0.601**	-0.747	-0.128	**0.364**	-0.569	**0.639**	-0.605	**0.303**
Met	0.244	-0.234	**0.306**	0.273	0.280	0.266	-0.152	**0.461**	-0.062
Phe	**0.620**	-0.609	**0.801**	**0.795**	0.290	**0.542**	-0.323	**1.140**	-0.385
Pro	0.264	-0.364	**0.463**	0.264	-0.343	0.249	-0.277	**0.439**	-0.330
Ser	-0.272	0.185	-0.259	-0.482	-0.247	-0.131	-0.024	-0.518	0.177
Thr	-0.257	0.266	-0.353	-0.407	0.221	-0.121	0.087	-0.448	**0.340**
Trp	**0.732**	-0.661	**0.884**	**1.111**	0.172	**0.467**	-0.205	**1.340**	-0.604
Tyr	0.266	-0.187	0.263	**0.508**	0.082	0.079	0.046	**0.495**	-0.258
Val	0.122	-0.140	0.172	-0.024	**0.492**	**0.311**	-0.214	0.251	0.191

Contributions above 0.300 are given in bold.

3.2 Internal Validation

The 34 parent proteins of the 15-mers from the training set were used as an internal test set. Each protein was represented as a set of overlapping nonamers. The binding affinity of the nonamers was predicted by the derived model. Peptides with affinity above or equal to 6.301 are considered as predicted binders, and below 6.301 as predicted nonbinders. If the predicted binder is included in the known binding 15-mer sequences, the predicted peptide is a true binder; otherwise, it is a false binder. The average values for *sensitivity*, *specificity*, *accuracy* and A_{ROC} are given in Table 4.

3.3 External Validation

The external validation was performed according to the procedure of internal validation. Additional 45 known binders to HLA-DP1 with different length and 36 parent proteins were compiled from IEDB. The average values for *sensitivity*, *specificity*, *accuracy*, and A_{ROC} are given in Table 4.

3.4 Comparison to NetMHCII Server

Using the same test set for external validation, we compared our model to the "state-of-the-art" server for MHC class II binding prediction: NetMHCII [46]. NetMHCII is sequence-based and uses ANN; it has been shown previously to be the-best-in-class [40]. It predicts 15-mer binding peptides, while our model predicts 9-mers. Because of the different length of the predicted peptides, the performance of the two models was compared only in terms of *sensitivity* (identified binders). The binding threshold was set to $pIC_{50} = 5.301 (IC_{50} = 5{,}000$ nM), as there were several low binders in the external test set. NetMHCII identified 39 of the 45 binders ($sensitivity = 0.867$), while our model identified all of them ($sensitivity = 1$).

TABLE 4
Internal and External Validation of the Model for Peptide Binding to HLA-DP1

	Internal validation	*External validation*
known binders	406	45
parent proteins	34	36
average sensitivity	1.000	0.944
average specificity	0.802	0.706
average accuracy	0.901	0.825
average A_{ROC}	0.903	0.783

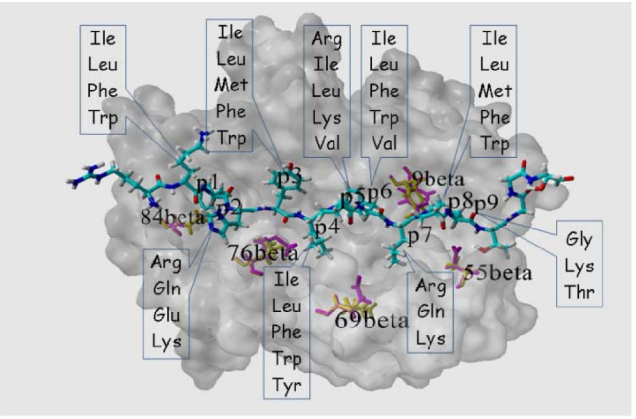

Fig. 2. The complex peptide-HLA-DP1 modeled by homology to complex peptide-HLA-DP2 (pdb code: 3LQZ). The protein is given as molecular surface, and the peptide RK**FHYLPFLPS**TGGS as sticks. Polymorphic protein residues and peptide binding core positions are labeled.

4 Discussion

In the present study, the ISC algorithm was applied on a set of peptides binding to HLA-DP1 protein. The peptide structures were described by the amino acid z-descriptors. The ISC algorithm was used to identify the peptide binding core. Self-consistency was achieved at the 39th iteration. The derived model was validated by two test sets: internal, containing the 34 parent proteins of the binders from the training set, and external, containing 36 proteins and 45 known binders. There was no overlapping between the test sets.

For MHC class II proteins, the anchor binding positions are 1, 4, 6, and 9 [47], [48]. The corresponding peptide residues bind into well-defined pockets. The X-ray data for HLA-DP2 (DPA1*01:03, DPB1*02:01) show that pockets 1 and 6 are deep and hydrophobic, while pockets 4 and 9 are shallow and polar [13]. Five polymorphic residues exist between HLA-DP2 and HLA-DP1 (see Fig. 2). All of them are located in the binding site and belong to β-chain. The mutations in HLA-DP1 comparing to HLA-DP2 are as follows: $\mathrm{Phe}^{9\beta} \rightarrow \mathrm{Tyr}^{9\beta}$; $\mathrm{Asp}^{55\beta} \rightarrow \mathrm{Ala}^{55\beta}$; $\mathrm{Glu}^{69\beta} \rightarrow \mathrm{Lys}^{69\beta}$; $\mathrm{Met}^{76\beta} \rightarrow \mathrm{Val}^{76\beta}$ and $\mathrm{Gly}^{84\beta} \rightarrow \mathrm{Asp}^{84\beta}$.

The protein residue 9β is part of the peptide binding pocket 9 (see Fig. 2). The mutation $\mathrm{Phe}^{9\beta} \leftarrow \mathrm{Tyr}^{9\beta}$ reduces the pocket space and brings some more negative charge. Pocket 9 accepts a wide range of residues from short Ser, Thr, Val, Ala, Gly [13] to larger aliphatic, polar, or even charged residues as Lys and Arg [49]. The protein residue 55β also is part of pocket 9 (see Fig. 2). The mutation $\mathrm{Asp}^{55\beta} \leftarrow \mathrm{Ala}^{55\beta}$ leads to a reduction of the negative charge inside the pocket [50] and to the loss of the salt-bridge hydrogen bond between $\mathrm{Asp}^{55\beta}$ and $\mathrm{Arg}^{76\alpha}$ [51]. This salt-bridge bond exists in most DR, DQ, and DP molecules [52]. According to our model, DP1 pocket 9 accepts short amino acids as Gly, Thr, Ser, Val, and Ala, as well as polar and charged ones as Lys, Arg, Gln, and Glu.

The protein residue 69β contacts peptide positions 4 and 7 (see Fig. 2). The mutation $\mathrm{Glu}^{69\beta} \leftarrow \mathrm{Lys}^{69\beta}$ affects the binding preferences at these two positions. As the side chain of Lys is oriented toward p7, the preferences here are for polar and charged amino acids. The binding pocket 4 is more affected by the mutation $\mathrm{Met}^{76\beta} \leftarrow \mathrm{Val}^{76\beta}$ and makes it less negatively charged with preferences for hydrophobic amino acids.

The protein residue 84β is located at the bottom of peptide binding pocket 1 (see Fig. 2). The mutation $\mathrm{Gly}^{84\beta} \leftarrow \mathrm{Asp}^{84\beta}$ is supposed to make the pocket polar and attractive for positively charged amino acids as it was observed for HLA-DP5 (DPA1*02:01/DPB1*05:01) [51] where the same mutation exists.

Here, the preferences for pocket 1 are dominated by hydrophobic residues which are the universal anchors at this position for all MHC class II proteins [54].

In conclusion, the ISC algorithm applied in this study in combination with the z-descriptors used for peptide structure description and the PLS-MLR used for developing of the model gave quantitative and easy-to-interpret amino acid contributions at each position of the peptide binding core. These contributions are in good agreement with experimentally derived data and could be used for *in silico* prediction of peptides binding to HLA-DP1 protein.

ACKNOWLEDGMENTS

This work was supported by the National Research Fund of the Bulgarian Ministry of Education and Science under Grant 02-1/2009.

REFERENCES

[1] C.A. Janeway Jr., P. Travers, M. Walport, and M.J. Shlomchik, *Immunobiology: The Immune System in Health and Disease,* fifth ed. Garland Science, 2001.

[2] P. Guermonprez, L. Saveanu, M. Kleijmeer, J. Davoust, P. van Endert, and S. Amigorena, "ER-Phagosome Fusion Defines an MHC Class I Cross-Presentation Compartment in Dendritic Cells," *Nature,* vol. 425, pp. 397-402, 2003.

[3] S. Burgdorf, A. Kautz, V. Böhnert, P.A. Knolle, and C. Kurts, "Distinct Antigen Uptake and Intracellular Routing Mechanisms in CD4 and CD8 T Cell Activation," *Science,* vol. 316, pp. 612-616, 2007.

[4] M.H. Glickman and A. Ciechanover, "The Ubiquitin-Proteasome Proteolytic Pathway: Destruction for the Sake of Construction," *Physiological Rev.,* vol. 82, pp. 373-428, 2002.

[5] T.S. Jardetzky, J.H. Brown, J.C. Gorga, L.J. Stern, R.G. Urban, J.L. Strominger, and D.C. Wiley, "Crystallographic Analysis of Endogenous Peptides Associated with HLA-DR1 Suggests a Common, Polyproline II-Like Conformation for Bound Peptides," *Proc. Nat'l Academy of Sciences USA,* vol. 93, pp. 734-738, 1996.

[6] J. Robinson, K. Mistry, H. McWilliam, R. Lopez, P. Parham, and S.G.E. Marsh, "The IMGT/HLA Database," *Nucleic Acids Research,* vol. 39, no. Suppl. 1, pp. D1171-D1176, 2011.

[7] E.W. Petersdorf, A.G. Smith, E.M. Mickelson, G.M. Longton, C. Anasetti, S.Y. Choo, P.J. Martin, and J.A. Hansen, "The Role of HLA-DPB1 Disparity in the Development of Acute Graft-versus-Host Disease following Unrelated Donor Marrow Transplantation," *Blood,* vol. 81, pp. 1923-1932, 1993.

[8] P.A. Lympany, M. Petrek, A.M. Southcott, A.J.N. Taylor, K.I. Welsh, and R.M. du Bois, "HLA-DPB Polymorphism: Glu 69 Association with Sarcoidosis," *European J. Immunogenetics,* vol. 23, pp. 353-359, 1996.

[9] A.B. Begovich, T.L. Bugawan, B.S. Nepom, W. Klitz, G.T. Nepom, and H.A. Erlich, "A Specific HLA-DPβ Allele Is Associated with Pauciarticular Juvenile Rheumatoid Arthritis but Not Adult Rheumatoid Arthritis," *Proc. Nat'l Academy of Sciences USA,* vol. 86, pp. 9489-9493, 1989.

[10] R.P. Dong, A. Kimura, R. Okubo, H. Shinagawa, H. Tamai, Y. Nishimura, and T. Sasazuki, "HLA-A and DPB1 Loci Confer Susceptibility to Graves' Disease," *Human Immunology,* vol. 35, pp. 165-172, 1992.

[11] I. Potolicchio, G. Mosconi, A. Forni, B. Nemery, P. Seghizzi, and R. Sorrentino, "Susceptibility to Hard Metal Lung Disease is Strongly Associated with the Presence of Glutamate 69 in HLA-DP Beta Chain," *European J. Immunology,* vol. 27, pp. 2741-2743, 1997.

[12] L. Richeldi, R. Sorrentino, and C. Saltini, "HLA-DPB1 Glutamate 69: A Genetic Marker of Beryllium Disease," *Science,* vol. 262, pp. 242-244, 1993.

[13] S. Dai, G.A. Murphy, F. Crawford, D.G. Mack, M.T. Falta, P. Marrack, J.W. Kappler, and A.P. Fontenot, "Crystal Structure of HLA-DP2 and Implications for Chronic Beryllium Disease," *Proc. Nat'l Academy of Sciences USA,* vol. 107, pp. 7425-7430, 2010.

[14] M. Amicosante, F. Berretta, R. Dweik ,and C. Saltini, "Role of High-Affinity HLA-DP Specific CLIP-Derived Peptides in Beryllium Binding to the HLA-DPGlu69 Berylliosis-Associated Molecules and Presentation to Beryllium-Sensitized T Cells," *Immunology,* vol. 128, pp. e462-e470, 2008.

[15] A. Sette, S. Buus, E. Appella, J.A. Smith, R. Chesnut, C. Miles, S.M. Colon, and H.M. Grey, "Prediction of Major Histocompatibility Complex Binding Regions of Protein Antigens by Sequence Pattern Analysis," *Proc. Nat'l Academy of Sciences USA,* vol. 86, pp. 3296-3300, 1989.

[16] J. D'Amaro, J.G. Houbiers, J.W. Drijfhout, R.M. Brandt, R. Schipper, and J.N. Bavinck, "A Computer Program for Predicting Possible Cytotoxic T Lymphocyte Epitopes Based on HLA Class I Peptide-Binding Motifs," *Human Immunology,* vol. 43, pp. 13-18, 1995.

[17] H. Rammensee, J. Bachmann, N.P.N. Emmerich, O.A. Bachor, and S. Stevanovic, "SYFPEITHI: Database for MHC Ligands and Peptide Motifs," *Immunogenetics,* vol. 50, pp. 213-219, 1999.

[18] A.S. De Groot, A. Bosma, N. Chinai, J. Frost, B.M. Jesdale, M.A. Gonzalez, and W. Martin, "From Genome to Vaccine: In Silico Predictions, Ex Vivo Verification," *Vaccine,* vol. 19, pp. 4385-4395, 2001.

[19] M. Amicosante, C. Gioia, C. Montesano, R. Casetti, S. Topino, and G. D'Offizi, "Computer-Based Design of An HLA-Haplotype and HIV-Clade Independent Cytotoxic T-Lymphocyte Assay for Monitoring HIV-Specific Immunity," *Molecular Medicine,* vol. 8, pp. 798-807, 2002.

[20] J. Hammer, E. Bono, F. Gallazzi, C. Belunis, Z.A. Nagy, and F. Sinigaglia, "Precise Prediction of Major Histocompatibility Complex Class II-Peptide Interaction Based on Peptide Side Chain Scanning," *J. Experimental Medicine,* vol. 180, pp. 2353-2358, 1994.

[21] K.C. Parker, M.A. Bednarek, and J.E. Coligan, "Scheme for Ranking Potential HLA-A2 Binding Peptides Based on Independent Binding of Individual Peptide Side-Chains," *J. Immunology,* vol. 152, pp. 163-175, 1994.

[22] K. Gulukota, J. Sidney, A. Sette, and C. DeLisi, "Two Complementary Methods for Predicting Peptides Binding Major Histocompatibility Complex Molecules," *J. Molecular Biology,* vol. 267, pp. 1258-1267, 1997.

[23] J.R. Schafer, B.M. Jesdale, J.A. George, N.M. Kouttab, and A.S. De Groot, "Prediction of Well-Conserved HIV-1 Ligands Using a Matrix-Based Algorithm, EpiMatrix," *Vaccine,* vol. 16, pp. 1880-1884, 1998.

[24] T. Sturniolo, E. Bono, J. Ding, L. Raddrizzani, O. Tuereci, U. Sahin, M. Braxenthaler, F. Gallazzi, M.P. Protti, F. Sinigaglia, and J. Hammer, "Generation of Tissue-Specific and Promiscuous HLA Ligand Databases Using DNA Microarrays and Virtual HLA Class II Matrices," *Nature Biotechnology,* vol. 17, pp. 555-561, 1999.

[25] P.A. Reche, J.P. Glutting, and E.L. Reinherz, "Prediction of MHC Class I Binding Peptides Using Profile Motifs," *Human Immunology,* vol. 63, pp. 701-709, 2002.

[26] I.A. Doytchinova, V.A. Walshe, N.A. Jones, S.E. Gloster, P. Borrow, and D.R. Flower, "Coupling In Silico and In Vitro Analysis of Peptide-MHC Binding: A Bioinformatic Approach Enabling Prediction of Superbinding Peptides and Anchorless Epitopes," *J. Immunology,* vol. 172, pp. 7495-7502, 2004.

[27] C. Lundegaard, K. Lamberth, M. Harndahl, S. Buus, O. Lund, and M. Nielsen, "NetMHC-3.0: Accurate Web Accessible Predictions of Human, Mouse and Monkey MHC Class I Affinities for Peptides of Length 8-11," *Nucleic Acids Research,* vol. 36, pp. W509-512, 2008.

[28] L. Nanni, "Machine Learning Algorithms for T-Cell Epitopes Prediction," *Neurocomputing,* vol. 69, pp. 866-868, 2006.

[29] R. Rosenfeld, Q. Zheng, S. Vajda, and C. DeLisi, "Flexible Docking of Peptides to Class I Major-Histocompatibility-Complex Receptors," *Genetic Analysis: Biomolecular Eng.,* vol. 12, pp. 1-21, 1995.

[30] J. Zeng, H.R. Treutlein, and G.B. Rudy, "Predicting Sequences and Structures of MHC-Binding Peptides: A Computational Combinatorial Approach," *J. Computer-Aided Molecular Design,* vol. 15, pp. 573-586, 2001.

[31] Z. Liu, B.N. Dominy, and E.I. Shakhnovich, "Structural Mining: Self-Consistent Design on Flexible Protein-Peptide Docking and Transferable Binding Affinity Potential," *J. Am. Chemical Soc.,* vol. 126, pp. 8515-8528, 2004.

[32] M. Atanasova, I. Dimitrov, D.R. Flower, and I. Doytchinova, "MHC Class II Binding Prediction by Molecular Docking," *Molecular Informatics,* vol. 30, pp. 368-375, 2011.

[33] P.E. Adrian, G. Rajaseger, V.S. Mathura, M. Sakharkar, and P. Kangueane, "Types of Inter-Atomic Interactions at the MHC-Peptide Interface: Identifying Commonality from Accumulated Data," *BMC Structural Biology,* vol. 2, article 2, 2002.

[34] O. Schueler-Furman, R. Elber, and H. Margalit, "Knowledge-Based Structure Prediction of MHC Class I Bound Peptides: A Study of 23 Complexes," *Folding and Design,* vol. 3, pp. 549-564, 1998.

[35] O. Schueler-Furman, Y. Altuvia, A. Sette, and H. Margalit, "Structure-Based Prediction of Binding Peptides to MHC Class I Molecules: Application to a Broad Range of MHC Alleles," *Protein Science,* vol. 9, pp. 1838-1846, 2000.

[36] U. Sezerman, S. Vajda, and C. DeLisi, "Free Energy Mapping of Class I MHC Molecules and Structural Determination of Bound Peptides," *Protein Science,* vol. 5, pp. 1272-1281, 1996.

[37] A. Logean and D. Rognan, "Recovery of Known T-Cell Epitopes by Computational Scanning of a Viral Genome," *J. Computer-Aided Molecular Design,* vol. 16, pp. 229-243, 2002.

[38] P.M. Petrone and A.E. Garcia, "MHC-Peptide Binding is Assisted by Bound Water Molecules," *J. Molecular Biology,* vol. 338, pp. 419-435, 2004.

[39] I. D. oytchinova, P. Petkov, I. Dimitrov, M. Atanasova, and D.R. Flower, "HLA-DP2 Binding Prediction by Molecular Dynamics Simulations," *Protein Science,* vol. 20, pp. 1918-1928, 2011.

[40] H.H. Lin, G.L. Zhang, S. Tongchusak, E.L. Reinherz, and V. Brusic, "Evaluation of MHC-II Peptide Binding Prediction Servers: Applications for Vaccine Research," *BMC Bioinformatics,* vol. 9, article S22, 2008.

[41] R. Vita, L. Zarebski, J.A. Greenbaum, H. Emami, I. Hoof, N. Salimi, R. Damle, and A. Sette, "The Immune Epitope Database 2.0," *Nucleic Acids Research,* vol. 38, pp. D854-862, 2010.

[42] A. Sette, J. Sidney, M.-F. del Guercio, S. Southwood, J. Ruppert, C. Dahlberg, and H. Grey, "Peptide Binding to the Most Frequent HLA-A Class I Alleles Measured by Quantitative Molecular Binding Assays," *Molecular Immunology,* vol. 31, pp. 813-822, 1994.

[43] S. Hellberg, M. Sjostrom, B. Skagerberg, and S. Wold, "Peptide Quantitative Structure-Activity Relationships, A Multivariate Approach," *J. Medicinal Chemistry,* vol. 30, pp. 1126-1135, 1987.

[44] L. Eriksson, E. Johnansson, N. Kettaneh-Wold, and S. Wold, "PLS," *Multi- and Megavariate Data Analysis. Basic Principles and Applications,* Umetrics AB, ed., Umetrics AB, pp. 63-101, 2006.
[45] A.P. Bradley, "The Use of the Area under the ROC Curve in the Evaluation of Machine Learning Algorithms," *Pattern Recognition,* vol. 30, pp. 1145-1159, 1997.
[46] M. Andreatta and M. Nielsen, "Characterizing the Binding Motifs of 11 Common Human HLA-DP and HLA-DQ Molecules Using NNAlign," *Immunology,* vol. 136, pp. 306-311, 2012.
[47] L.J. Stern, J.H. Brown, T.S. Jardetzky, J.C. Gorga, R.G. Urban, J.L. Strominger, and D.C. Wiley, "Crystal Structure of the Human Class II MHC Protein HLA-DR1 Complexed with An Influenza Virus Peptide," *Nature,* vol. 368, pp. 215-221, 1994.
[48] Y. Zhu, A.Y. Rudensky, A.L. Corper, L. Teyton, and I.A. Wilson, "Crystal Structure of MHC Class II I-Ab in Complex with a Human CLIP Peptide: Prediction of An I-Ab Peptide-Binding Motif," *J. Molecular Biology,* vol. 326, pp. 1157-1174, 2003.
[49] J. Sidney, A. Steen, C. Moore, S. Ngo, J. Chung, B. Peters, and A. Sette, "Five HLA-DP Molecules Frequently Expressed in the Worldwide Human Population Share a Common HLA Supertypic Binding Specificity," *J. Immunology,* vol. 184, pp. 2492-2503, 2010.
[50] G. Diaz, M. Amicosante, D. Jaraquemada, R.H. Butler, M.V. Guillen, M. Sanchez, C. Nombela, and J. Arroyo, "Functional Analysis of HLA-DP Polymorphism: A Crucial Role for DPβ Residues 9, 11, 35, 55, 56, 69 and 84-87 in T Cell Allorecognition and Peptide Binding," *Int'l Immunology,* vol. 15, pp. 565-576, 2003.
[51] K.W. Wucherpfennig, "Insights into Autoimmunity Gained from Structural Analysis of MHC-Peptide Complexes," *Current Opinion in Immunology,* vol. 13, pp. 650-656, 2001.
[52] K.H. Lee, K.W. Wucherpfennig, and D.C. Wiley, "Structure of a Human Insulin Peptide-HLA-DQ8 Complex and Susceptibility to Type I Diabetes," *Nature Immunology,* vol. 2, pp. 501-507, 2001.
[53] A. Patronov, I. Dimitrov, D.R. Flower, and I.A. Doytchinova, "Peptide Binding to HLA-DP2 Proteins at pH 5.0 and pH 7.0: A Quantitative Molecular Docking Study," *BMC Structural Biology,* vol. 12, article 20, 2012.
[54] R.M. Chicz, D.F. Graziano, M. Trucco, J.L. Strominger, and J.C. Gorga, "HLA-DP2. Self Peptide Sequences and Binding Properties," *J. Immunology,* vol. 159, pp. 4935-4942, 1997.