PROTEOCHEMOMETRIC ANALYSIS OF PEPTIDES BINDING TO HUMAN LEUKOCYTE ANTIGEN (HLA) PROTEINS FROM LOCUS DP

V. Yordanov, I. Dimitrov, I. Doytchinova*

Faculty of Pharmacy, Medical University of Sofia,
2 Dunav Str., Sofia 1000, Bulgaria
*To whom all correspondence should be sent:
email: idoytchinova@pharmfac.mu-sofia.bg

Abstract. The human leukocyte antigen (HLA) system is an important part of the immune system involved in the presentation of antigen fragments (oligopeptides) to the T-cells. The proteins encoded by the HLA class II genes of locus DP are associated with a significant number of autoimmune diseases, as well as with the susceptibility or resistance to a number of infectious agents. The aim of the present study is to analyse the structure – affinity relationships of antigen peptides binding to 7 most common in the human population HLA-DP proteins. The analysis is performed by proteochemometrics. A set of 3,864 15-mer peptides, known binders and non-binders to HLA-DP proteins, are compiled from IEDB. The set is pre-processed and divided into training (80%) and test (20%) subsets. The training set is used to derive a proteochemometric model by iterative self-consistent partial least squares-based algorithm. The derived model has good explaining capacity (R^2 = 0.899 and Q^2 = 0.892), but moderate predictive ability validated by the test set (R_{pred} = 0.515). The proteochemometrics is a suitable method for structure-affinity analysis of peptides binding to multiple HLA proteins.

Key words: HLA-DP, proteochemometrics, iterative self-consistent algorithm, partial least squares

Introduction

The antigen is a substance enable to generate an adaptive immune response. It is a specific molecule marker and is most often a protein [1]. Basic steps in the immune response are the processing of antigens in the host cells, presentation on the cell surface and subsequent recognition by immune system cells. A key role in these processes play a specific family of proteins, which bind to fragments of antigen (oligopeptides) and present them on the cell surface where they are recognized by the relevant immune cells. This family of proteins known as a Major Histocompatibility Complex (MHC) and the coding genes are known as MHC genes. In human, MHC genes are referred to as HLA (Human Leukocyte Antigen) and are located in the sixth chromosome [2]. HLA genes encode several classes of HLA proteins, the most important of them are class I and class II.

Important features of the MHC proteins are their poligenicity (encoded by multiple genes) and extreme polymorphism (multiple alleles for each locus). In fact, they are the most polymorphic genes in nature. The database IMGT/HLA stores more than 3000 different HLA alleles of class I and II [3]. MHC proteins differ by up to 30 amino acid residues,

the majority of which are located at the peptide binding site. This variability allows the MHC proteins to bind to a wide range of antigenic peptides, which is essential for the immune recognition [2,4]. There are 6 loci encoding HLA proteins of class I and II. The first three loci, HLA-DP, HLA-DR and HLA-DQ, encoded the class II proteins. The other three loci, HLA-A, HLA-B and HLA-C, encoded the class I proteins.

MHC class II molecules consist of two transmembrane glycoprotein chains: α and β (Figure 1). Each chain consists of two domains: 1 and 2. The domains 1 are variable and form a peptide binding site, while the domains 2 are constant and bind to the membrane. The most significant difference between the MHC proteins of class I and class II is that the peptide binding site in class I is closed at both ends, while that in class II is open-ended. Peptides that can bind to MHC molecules of class I are a length of 8 to 11 residues, while those of class II reach a length of up to 25 residues.

The aim of the present study is to analyse the structure – affinity relationships of the peptides binding to proteins of locus DP. HLA-DP is the least polymorphic allele of class II. The crystallographic structure of the complex peptide - HLA-DP2 protein has been solved recently (pdb code: 3lqz) [5] and is shown in Figure 1. The proteins of the family of HLA-DP are associated with a significant number of autoimmune diseases, as well as with the susceptibility or resistance to a number of infectious agents. Susceptibility to a disease can be explained by the inability of a given MHC protein to express peptide fragments generated by the pathogen, due to low affinity between the two molecules. In contrast, peptide fragments, forming stable complexes with MHC proteins are recognized by T-lymphocytes and the infected cells are destroyed. This explains the inherent resistance to certain diseases [6]. The HLA-DP1 and HLA-DP5 alleles are associated with susceptibility to hyperthyroidism [7,8]; allele HLA-DP2 – with susceptibility to Beryllium disease [9,10], juvenile rheumatoid arthritis [11], sarcoidosis [12], atopic myelitis [13]; allele HLA-DP3 – with susceptibility to juvenile rheumatoid arthritis [14], allele HLA-DP41 – with protection from celiac disease [15], allele HLA-DP42 – with protection from Beryllium disease [16]. These five DP alleles cover more than 90% of the human population [17].

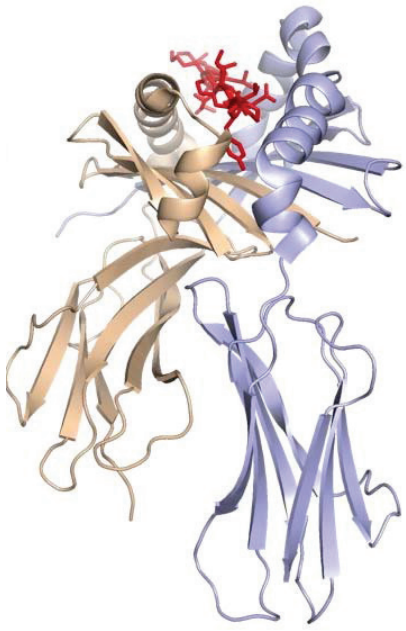

Figure 1. X-ray structure of the complex peptide – HLA-DP1 protein (pdb: 3lqz). The complex consists of two transmembrane chains (blue and beige) and one peptide (red).

The structure – affinity relationship analysis in the present study was performed by proteochemometrics. The proteochemometrics (PCM) is a quantitative structure – activity relationship (QSAR) method developed by Lapinsh and colleagues for simultaneous modelling of the bioactivity of multiple ligands against multiple protein targets [18,19]. PCM can be regarded as an extension of QSAR, which combines the information from the ligands and target molecules in a single X matrix, allowing the extrapolation of the results and a prediction of biological activity of new compounds for new target molecules [20]. The chemical structure of the ligands and proteins in the PCM is described by three descriptor blocks: ligand (L), protein (P) and ligand-protein (LP) (Figure 2.6) [19].

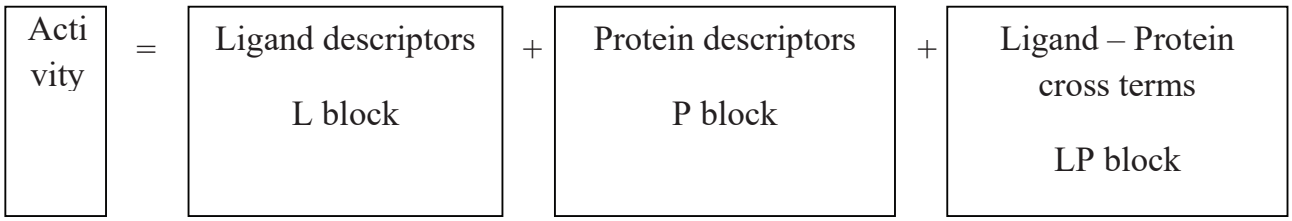

Figure 2. Descriptor blocks in proteochemometrics.

Materials and methods

HLA-DP proteins object of the study

The most common HLA-DP proteins were included in the present study: DP1 (DPA * 02:01/DPB1 * 01:01), DP2 (DPA * 01:03/DPB1 * 02:01), DP3 (DPA * 01:03/DPB1 * 03:01), DP41 (DPA * 01:03/DPB1 * 04:01), DP42 (DPA * 01:03/DPB1 * 04:02, assigned as DP42a and DPA * 03:01/DPB1 * 04:02 assigned as DP42b) and DP5 (DPA * 02:01/DPB1 * 05:01).

The protein sequences were derived from the database IMGT/HLA (http://www.ebi.ac.uk /imgt/ hla) (Figure 3). The peptide binding site on HLA protein is formed by the first 80 residues of the α-chain and of the first 90 residues of the β-chain [21,22]. Among them are four polymorphic residue in α-chain and 15 polymorphic in β-chain. These are Ala/Met$^{11\alpha}$, Met/Gln$^{31\alpha}$, Gln/Arg$^{50\alpha}$, Leu/Ser$^{66\alpha}$, Val/Leu$^{8\beta}$, Tyr/Phe$^{9\beta}$, Gly/Leu$^{11\beta}$, Tyr/Phe/Leu$^{35\beta}$, Ala/Val$^{36\beta}$, Ala/Asp/Glu$^{55\beta}$, Ala/Glu$^{56\beta}$, Glu/Asp$^{57\beta}$, Ile/Leu$^{65\beta}$, Lys/Glu$^{69\beta}$, Val/Met$^{76\beta}$, Asp/Gly$^{84\beta}$, Glu/Gly$^{85\beta}$, Ala/Pro$^{86\beta}$ and Val/Met$^{87\beta}$.

The peptide binding site on DP proteins consists of 54 residues: 25 of them belong to chain α, while 29 belong to chain β. The contacts between the bound peptide and the DP protein were identified by program Chimera (UCSF) [23] and confirmed by data in the literature [22]. Only three residues from chain α and nine residues from chain β are polymorphic among the 7 most common DP alleles (Figure 3). These are: Ala/Met$^{11\alpha}$, Met/Gln$^{31\alpha}$, Leu/Ser$^{66\alpha}$, Tyr/Phe$^{9\beta}$, Gly/Leu$^{11\beta}$, Tyr/Phe/Leu$^{35\beta}$, Ala/Val$^{36\beta}$, Ala/Asp/Glu$^{55\beta}$, Ile/Leu$^{65\beta}$, Lys/-Glu$^{69\beta}$, Val/Met$^{76\beta}$ и Asp/Gly$^{84\beta}$.

```
AA Pos.              10         20         30         40         50         60         70         80
 DPA1*01:03   IKADHVSTYA AFVQTHRPTG EFMFEFDEDE MFYVDLDKKE TVWHLEEFGQ AFSFEAQGGL ANIAILNNNL NTLIQRSNHT
 DPA1*02:01   ---------- ---------- ---------- Q--------- ---------R ---------- ---------- ----------
 DPA1*03:01   ***------- M--------- ---------- ---------- ---------- ---------- -----S---- ----------

AA Pos.              10         20         30         40         50         60         70         80         90
 DPB1*01:01    RATPENYVYQ GRQECYAFNG TQRFLERYIY NREEYARFDS DVGEFRAVTE LGRPAAEYWN SQKDILEEKR AVPDRVCRHN YELDEAVTLQ
 DPB1*02:01    -------LF- ---------- ---------- ----FV---- ---------- ----DE---- --------E- -----M---- ---GGPM---
 DPB1*03:01    ---------- L--------- ---------- ----FV---- ---------- ----DED--- ----L----- ---------- ----------
 DPB1*04:01    -------LF- ---------- ---------- ----F----- ---------- ---------- ---------- -----M---- ---GGPM---
 DPB1*04:02    -------LF- ---------- ---------- ----FV---- ---------- ----DE---- ---------- -----M---- ---GGPM---
 DPB1*05:01    -------LF- ---------- ---------- ----LV---- ---------- ----E----- ---------- -----M---- ----------
```

Figure 3. Protein sequences forming the peptide binding site of the 7 most common HLA-DP proteins.

Peptide binding set

A dataset 4,304 15-mer peptides binding to the 7 studied HLA-DP proteins were collected from the database IEDB (http://www.immuneepitope.org) of the National Institutes of Health of the United States. In the case of several overlapping peptides of various lengths, only the longest peptide was included. The binding affinities of the peptides were given in IC_{50} values and converted to $pIC_{50} = \log(1/IC_{50})$. The duplicate peptides were removed and the number of peptides was reduced to 3,864 15-mers. As the peptide binding site on HLA could be occupied by 9 amino acids only (the rest are floating), each 15-mer peptide was represented as a set of 7 overlapping nonamers with the same pIC_{50} value (3,864 x 7 = 27,048 nonamers). A binding affinity threshold of 5.3 for pIC_{50} was considered for classification of the peptides [24], which formed the below two subsets:

a) binders, $pIC_{50} > 5.3$, n = 1,847 (12,929 nonamers), and

b) non-binders, $pIC_{50} \leq 5.3$, n = 2,017 (14,119 nonamers).

Routine check for duplicates between the two groups identified multiple data conflicts: many nonamers derived from the same allele were found to be present both in binders and non-binders subsets. Such common nonamer sequences were considered to be non-binding ones and, therefore, excluded from the binders subset in order to decrease the noise and remained only in the non-binding set. The latter step involved exclusion of 2,358 duplicate nonamers, but the overall reduction of 15-mers was relatively low (from 3,864 to 3,852).

As a next step, the peptides were grouped by alleles and arranged in descending order. Every fifth peptide (20%) of each group was relocated to a test set for external validation. The remaining peptides formed the training set. Thus, the training set consisted of 3,082 15-mer peptides and the test set – of 770 15-mer peptides. The training set was used to derive the proteochemometrics model, while the test set was used to validate it.

The steps of initial data pre-processing are illustrated in the below flowchart:

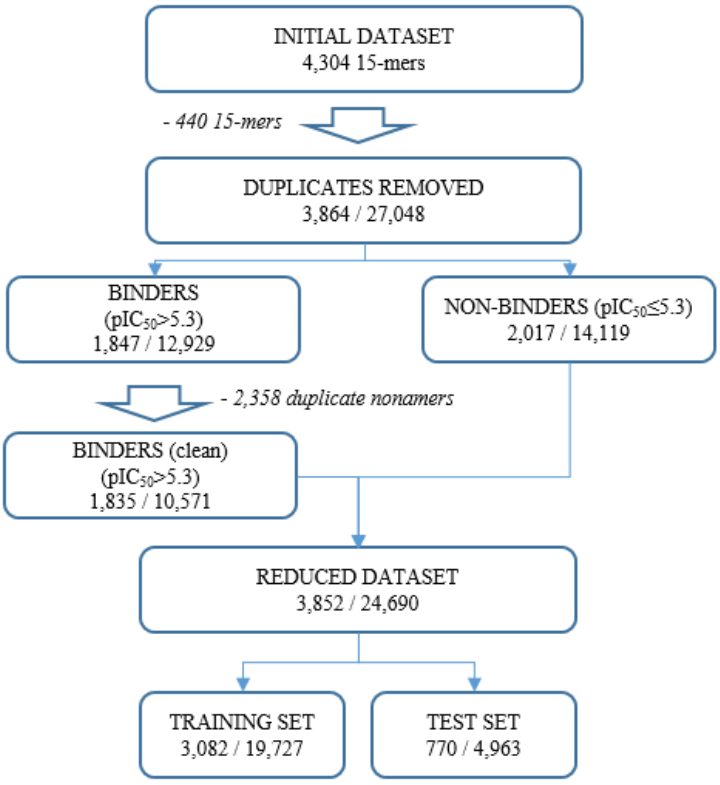

Figure 4. Data pre-processing.

Descriptors of the chemical structure

The chemical structure of peptides and proteins used in present study was described by three *z*-scales. The z-scales are amino acid descriptors derived by principal component analysis of 29 amino acid descriptors [24]. The three z-scales describe hydrophobicity, size and electronic properties of amino acids [25] (Table 1).

Table 1. The three *z*-scales for the 20 naturally occurring amino acids [24,25].

Amino acid	z_1 hydrophobicity	z_2 molecular size	z_3 polarity	Amino acid	z_1 hydrophobicity	z_2 molecular size	z_3 polarity
A Ala	0.07	-1.73	0.09	M Met	-2.49	-0.27	-0.41
C Cys	0.71	-0.97	4.13	N Asn	3.22	1.45	0.84
D Asp	3.64	1.13	2.36	P Pro	-1.22	0.88	2.23
E Glu	3.08	0.39	-0.07	Q Gln	2.18	0.53	-1.14
F Phe	-4.92	1.30	0.45	R Arg	2.88	2.52	-3.44
G Gly	2.23	-5.36	0.30	S Ser	1.96	-1.63	0.57
H His	2.41	1.74	1.11	T Thr	0.92	-2.09	-1.40
I Ile	-4.44	-1.68	-1.03	W Trp	-4.75	3.65	0.85
K Lys	2.84	1.41	-3.14	V Val	-2.69	-2.53	-1.29
L Leu	-4.19	-1.03	-0.98	Y Tyr	-1.39	2.32	0.01

The peptides used in the present study were 15-mers. However, only 9 residues are able to fill the binding site forming the binding core. The rest of the residues are flanking outside the binding site and do not affect the peptide affinity. Each 15-mer was presented as a set of overlapping 9-mer peptides as is illustrated in Figure 4. Each 9-mer peptide acquired the pIC_{50} value of the parent 15-mer peptide.

15-mer peptide	pIC_{50}	Overlapping 9-mer peptides
GGSILKISNKYHTKG	7.369	G G S I L K I S N
	7.369	G S I L K I S N K
	7.369	S I L K I S N K Y
	7.369	I L K I S N K Y H
	7.369	L K I S N K Y H T
	7.369	K I S N K Y H T K
	7.369	I S N K Y H T K G

Figure 4. Presentation of a 15-mer peptide as a set of overlapping 9-mer peptides.

Each residue in the nonamer is described by the three *z*-scales (z_1, z_2 and z_3) and the set of 27 (9 × 3) descriptors formed the ligand block L. The 19 polymorphic residues of HLA-DP proteins were also encoded by *z*-scales. The set of 57 (19 x 3) descriptors formed the protein block P. The contact between peptides and DP proteins are listed in Table 2. These contacts were described by 57 (19 x 3) cross terms and form the ligand-protein block LP.

Table 2. Peptide – DP protein contacts.

Peptide position	*Protein position*
p1	31α, 76β, 84β
p2	76β
p3	76β
p4	69β, 76β
p5	-
p6	11α, 66α, 11β
p7	66α, 55β, 65β, 69β
p8	55β
p9	9β, 35β, 36β, 55β

The final X matrix consists of three blocks of descriptors L, P and LP, or 141 descriptors in total. The pIC_{50} values were the Y variables. The system of equations was solved by partial least squares multiple linear regression (PLS-MLR) using SIMCA v13 [26]. The derived model had the following view:

$$pIC_{50} = b + \Sigma(a_1*L) + \Sigma(a_2*P) + \Sigma(a_3*LP),$$

where a_n are the coefficients describing the contribution of each descriptor to binding affinity. The positive a_n corresponds to positive impact on the binding affinity, while the negative a_n decrease the affinity. The model was evaluated by R^2 (explained variance), Q^2 (correlation coefficient after cross-validation in 7 groups) and R_{pred} (test set correlation coefficient). The number of the principal components (PC) was determined at maximum value of Q^2.

Results and discussion

The proteochemometric model was derived by iterative self-consisted PLS-based (ISC-PLS) algorithm [27]. The main steps of ISC-PLS are given at Figure 5:

1. The initial training (working) set WS_0 was consisted of 19,727 9-mer peptides generated from the 3,082 15-mer peptides. It was used to derive the initial model M_0.

2. The model M_0 was used to calculate the pIC_{50} values of the 9-mers from the initial training set WS_0. The 9-mers with pIC_{50calc} closest to the pIC_{50exp} of the corresponding 15-mer were compiled into a new training set WS_1. WS_1 consisted of 3,082 9-mer peptides.

3. WS_1 was used to derive a new proteochemometric model M_1. M_1 predicted the pIC_{50} values of the 9-mers from the initial training set WS_0. The 9-mers with pIC_{50calc} closest to the pIC_{50exp} of the corresponding 15-mer were compiled into a new training set WS_2.

4. Likewise, WS_2 was used to derive a new proteochemometric model M_2. M_2 predicted the pIC_{50} values of the 9-mers from the initial training set WS_0. The 9-mers with pIC_{50calc} closest to the pIC_{50exp} of the corresponding 15-mer were compiled into a new training set WS_3.

5. Step 4 was repeated until full self-consistency between two subsequent training sets WS_{n-1} and WS_n was achieved. The self-consistency was monitored by % identity between two subsequent training sets.

6. When 100% (or close to it) consistency was achieved, the final model M_n was derived.

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

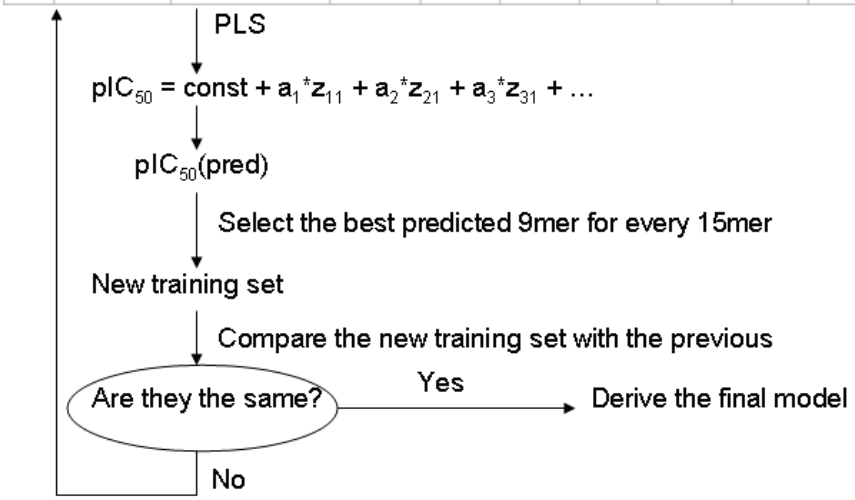

Figure 5. Iterative self-consisted PLS-based (ISC-PLS) algorithm.

The parameter Q^2 rapidly reached 0.88 within the first 10 iterations, and then it grew gradually with a very slow pace (Figure 6). Self-consistency of 100% similarity in two consecutive peptide sets was achieved at the 150th iteration. The final model had $R^2 = 0.899$, $Q^2 = 0.892$ at $PC = 6$.

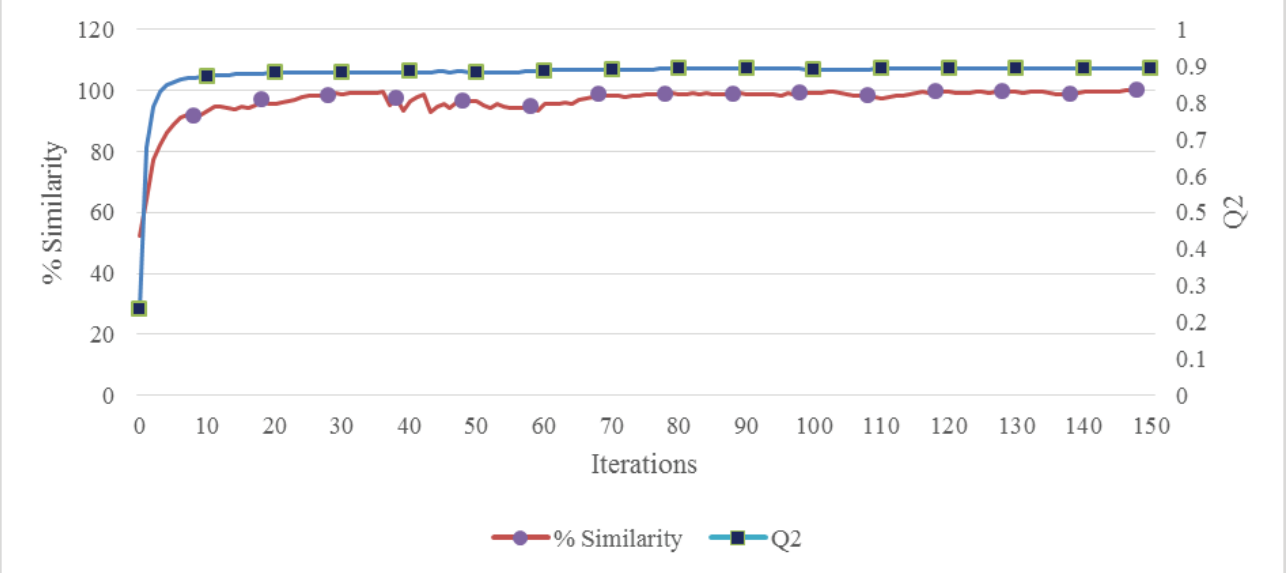

Figure 6. PCM model development by PLS-ISC algorithm. % Similarity (●), Q^2 (■).

The predictive ability of the derived model was tested by the external test set. The 15-mers in the test set were presented as overlapping 9-mers. The binding affinities of all 9-mers were predicted by the proteochemometric model. Among the rest predicted as binders, the 9-mer binding cores of the corresponding 15-mers were selected by two parallel procedures and the derived R_{pred} values were compared.

- Maximum pIC_{50}. The binding core is the 9-mer peptide with the highest predicted pIC_{50} value among the peptides generated from the same 15-mer peptide. The R_{pred} value derived by this procedure was 0.410.

- Average pIC_{50}. The binding affinity of a 15-mer peptide is calculated as an average pIC_{50} of the binding 9-mer peptides. The R_{pred} value derived by this procedure was 0.515.

The best prediction was achieved when the average predicted pIC_{50} values were considered. Even in this case, the predictive ability of the model was moderate. The reason for this moderate predictive ability could be noise in the training set due to the mixing of binders and non-binders in one set.

The proteochemometric model derived in the present study consists of three descriptor blocks: L, P and LP. The descriptors with coefficients above ± 0.1 are given in Figure 7. Among them are descriptors from blocks L and LP.

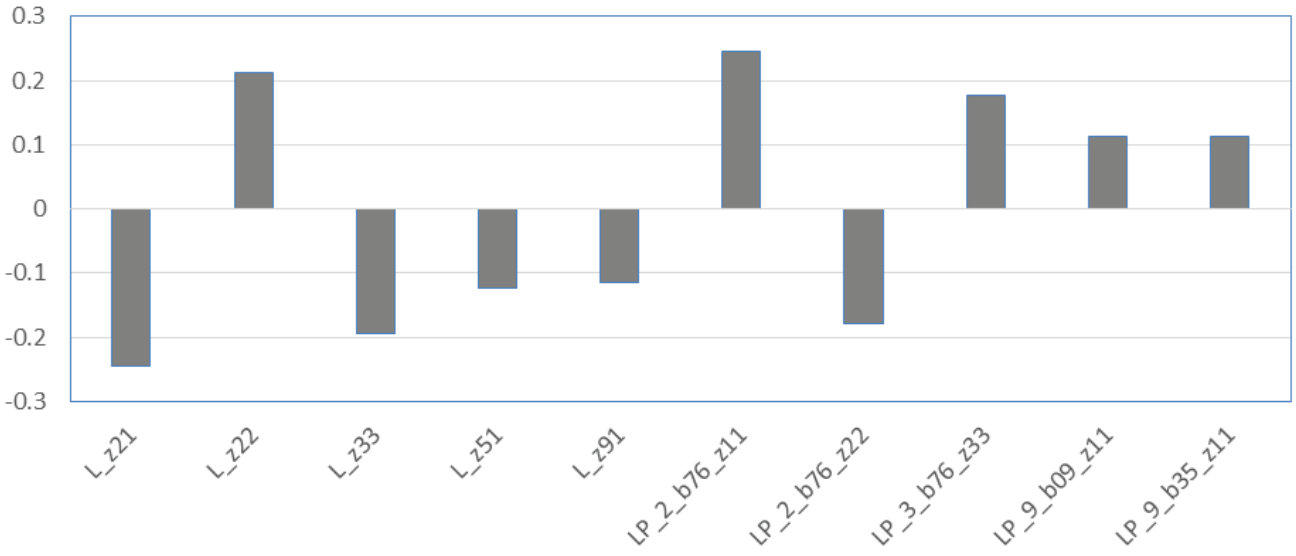

Figure 7. Descriptors with coefficients (scaled and centered) above ± 0.1.

The coefficients in the PCM model are indicative for the preferred residues in the peptide. Positive coefficients mean that amino acids with positive values of the corresponding *z* descriptor will increase pIC_{50} and will be preferred in the binding site. Negative coefficients point that residues with negative *z-scales* will increase the affinity.

According to the model, large and hydrophobic amino acids are preferred at peptide position 2 (L2) like Phe, Trp, Tyr and Pro (Table 3). At position 3 (L3) are preferred residues with negative z_3 scores. Hydrophobic amino acids are preferred also at positions 5 (L5) and 9 (L9) (Table 3).

Table 3. Coefficients of the *z-scales* from block L and preferred peptide residues. The known preferred residues[4] are given in bold.

Peptide position	z_1	z_2	z_3	preferred residues
p2 (L2)	-0.246	0.213		**Phe**, **Trp**, **Tyr**, Pro
p3 (L3)			-0.195	**Lys**, **Arg**, Thr, Gln, **Leu**, Ile, Val
p5 (L5)	-0.124			**Phe**, **Ile**, **Leu**, Met, Pro, Trp, Val, Tyr
p9 (L9)	-0.116			**Phe**, **Ile**, **Leu**, **Met**, Pro,Trp, **Val**, Tyr

The most important cross terms in the PCM model are summarized in Table 4. The peptide position 2 is solvent exposed and has no direct contact with protein position β76. Dimorphism exists at β76: Val and Met; both are hydrophobic. The positive coefficient of the cross term LP_2_b76_z11 means that hydrophobic residues are preferred here, while the negative coefficient for LP_2_b76_z22 selects the large sized hydrophobic residues, like Phe, Trp and Tyr.

The preferences at peptide position 3 for residues with negative z_3 scores are confirmed by the positive coefficient of LP_3_b76_z33. Both Val and Met have negative z_3 scores and the preferred amino acids at p3 also need negative z_3 scores.

Dimorphism Tyr/Phe exists at β09 and polymorphism Tyr/Phe/Leu – at β35 in pocket 9. Both positions require hydrophobic amino acids at the corresponding peptide position 9.

A good agreement on the preferred peptide amino acids exists between the terms from the L and LP blocks. The model in the present study is well corresponding to our previous model for DP binding prediction, derived recently [28]. Our findings are in a good agreement with the binding motifs defined by Andreatta and Nielsen for five common DP allelic variants (DP1, DP2, DP41, DP42b and DP5) [29].

Conclusions

The proteochemometrics is a suitable method for deriving of quantitative structure – affinity relationships for peptides binding to 7 most common HLA-DP proteins. The iterative self-consistent PLS-based algorithm selects correctly the binding core when the binding peptide is longer than the binding site. The proteochemometric model derived in the present study had moderate predictive ability. The preferred amino acids at the bound peptide are confirmed by the X-ray structure of the complex peptide – HLA-DP2 protein.

Table 4. Coefficients of the *z-scales* from block LP and preferred peptide residues. The known preferred residues[29] are given in bold.

Cross term	DP1	DP2	DP3	DP41	DP42a	DP42b	DP5
+0.245*LP_2_b76_z11 -0.179*LP_2_b76_z22							
β76	Val	Met	Val	Met	Met	Met	Met
β76 z_1	-2.69	-2.49	-2.69	-2.49	-2.49	-2.49	-2.49
β76 z_2	-2.53	-0.27	-2.53	-0.27	-0.27	-0.27	-0.27
p2 z_1	Negative	Negative	Negative	Negative	Negative	Negative	Negative
p2 z_2	Positive	Positive	Positive	Positive	Positive	Positive	Positive
Preferred aa				**Phe**, **Trp**, **Tyr**, Pro			
+0.177*LP_3_b76_z33							
β76	Val	Met	Val	Met	Met	Met	Met
β76 z_3	-1.29	-0.41	-1.29	-0.41	-0.41	-0.41	-0.41
p3 z_3	Negative	Negative	Negative	Negative	Negative	Negative	Negative
Preferred aa				**Lys**, **Arg**, Thr, Gln, **Leu**, Ile, Val			
+0.112*LP_9_b09_z11+0.114*LP_9_b35_z11							
β09	Tyr	Phe	Tyr	Phe	Phe	Phe	Phe
β09 z_1	-1.39	-4.92	-1.39	-4.92	-4.92	-4.92	-4.92
p9 z_1	Negative	Negative	Negative	Negative	Negative	Negative	Negative
β35	Tyr	Phe	Phe	Phe	Phe	Phe	Leu
β35 *z1*	-1.39	-4.92	-4.92	-4.92	-4.92	-4.92	-4.19
p9 *z1*	Negative	Negative	Negative	Negative	Negative	Negative	Negative
Preferred aa				**Phe**, **Ile**, **Leu**, **Met**, Pro,Trp, **Val**, Tyr			

References

1. Alberts B, Johnson A, Lewis J. 2002. The Adaptive Immune System. In: Molecular Biology of the Cell. Garland Science, New York 2002.

2. Janeway Jr CA, Travers P, Walport M, Shlomchik MJ. 2001. The major histocompatibility complex and its functions. In: Immunobiology: The Immune System in Health and Disease. Garland Science, New York 2001.

3. Robinson J, Halliwell JA, Hayhurst JH, Flicek P, Parham P, Marsh SGE. 2015. The IPD and IMGT/HLA database: allele variant databases. Nucl Acids Res 2015; 43: D423-431.

4. Doytchinova IA, Flower DR. QSAR and the Prediction of T-Cell Epitopes. Curr Proteomics 2008; 5: 73–95.

5. Dai A, Murphy GA, Crawford F, Mack DG, Falta MT, Marrack P, Kappler J W, Fontenot AP. Crystal structure of HLA-DP2 and implications for chronic beryllium disease. Proc Natl Acad Sci USA 2010; 107: 7425-7430.

6. Scherer A, Frater J, Oxenius A, Agudelo J, Price DA, Günthard HF, Barnardo M, Perrin L, Hirschel B, Phillips RE, McLean AR. Swiss HIV Cohort Study, 2004. Quantifiable cytotoxic T lymphocyte responses and HLA-related risk of progression to AIDS. Proc Natl Acad Sci USA 2004; 101: 12266–12270.

7. Ratanachaiyavong S, McGregor AM. HLA-DPB1 polymorphisms on the MHC-extended haplotypes of families of patioents with Graves' disease: two distinct HLA-DR17 haplotypes. Eur J Clin Invest 1994; 24: 309-315.

8. Takahashi M, Kimura A. HLA and CTLA4 polymorphisms may confer a synergistic risk in the susceptibility to Graves' disease. J Hum Genet 2010; 55: 323-326.

9. Fontenot AP, Kotzin BL. Chronic beryllium disease: immune-mediated destruction with implications for ordan-specific autoimmunity. Tissue Antigens 2003; 62: 449-458.

10. Petukh M, Wu B, Stefl S, Smith N, Hyde-Volpe D, Wang L, Alexov E. Chronic beryllium disease: revealing the role of beryllium ion and small peptides binding to HLA-DP2. PLoS 2014, 9, e111604.

11. Begovich AB, Bugawan TL, Nepom BS, Klitz W, Nepom GT, Erlich HA. A specific HLA-DRβ allele is associated with pauciarticular juvenile rheumatoid arthritis but not adult rheumatoid arthritis. Proc Natl Acad Sci USA 1989; 86: 9489-9493.

12. Lympany PA, Petrek M, Southcott AM, Taylor AJN, Welsh KI, du Bois RM. HLA-DPB polymorphism: Glu69 association with sarcoidosis. Eur J Immunogenet 1996; 23: 353-359.

13. Sato S, Isobe N, Yoshimura S, Kanamori Y, Masaki K, Matsushira T, Kira J. HLA-DPB1*0201 is associated with susceptibility to atopic myelitis in Japanese. J Neuroimmunol 2012; 251: 110-113.

14. Donn R. Etiology and pathogenesis of juvenile idiopathic arthritis. In: Hochberg, M.C., Silman, A.J., Smolen, J.S., Weinblatt, M.E., Weisman, M.H. (Eds.) Rheumatology, Sixth Ed., Elsevier, Philaderphia, 2015.

15. Hadley D, Hagopian W, Liu E, She JX, Simell O, Akolkar B, Ziegler AG, Rewers M, Krischer JP, Chen WM, Onenegut-Gumuscu S, Bugawan TL, Rich SS, Erlich H, Agardh D, TEDDY Study Group. HLA-DPB1*04:01 protects genetically susceptible children from celiac disease autoimmunity in the TEDDY study. Am J Gastroenterol 2015; 110: 915-920.

16. Arroyo J, Alvarez AM, Nombela C, Sanchez-Perez M. The role of HLA-DP beta residue 69 in the definition of antibody-binding epitopes. Hum Immunol 1995; 43: 219-226.

17. Sidney J, Steen A, Moore C, Ngo S, Chung J, Peters B, Sette A. Five HLA-DP molecules frequently expressed in the worldwide human population share a common HLA supertypic binding specificity. J Immunol 2010; 184: 2492–2503.

18. Lapinsh M, Prusis P, Gutcaits A, Lundstedt T, Wikberg JE. 2001. Development of proteo-chemometrics: a novel technology for the analysis of drug-receptor interactions. Biochim Biophys Acta 2001; 1525: 180–190.

19. Prusis P, Uhlén S, Petrovska R, Lapinsh M, Wikberg JE. Prediction of indirect interactions in proteins. BMC Bioinformatics 2006;

7: 167.

20. Cortés-Ciriano I, Ain QU, Subramanian V, Lenselink EB, Méndez-Lucio O, IJzerman AP, Wohlfahrt G, Prusis P, Malliavin TE, van Westen GJP, Bender A. Polypharmacology modelling using proteochemometrics (PCM): recent methodological developments, applications to target families, and future prospects. Med Chem Commun 2015; 6: 24–50.

21. Patronov A, Dimitrov I, Flower DR, Doytchinova I. Peptide binding prediction for the human class II MHC allele HLA-DP2: a molecular docking approach. BMC Str Biol 2011; 11: 32.

22. Patronov A, Dimitrov I, Flower DR, Doytchinova I. Peptide binding to HLA-DP proteins at pH 5.0 and pH 7.0: a quantitative molecular docking study. BMC Str Biol 2012; 12: 20.

23. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem 2004; 25: 1605-1612.

24. Hellberg S, Sjostrom M, Skagerberg B, Wold S. Peptide quantitative structure-activity relationships, a multivariate approach. J Med Chem 1987; 30: 1126-1135.

25. Eriksson L, Jonsson J, Sjostrom M, Wold S. Multivariate parametrization of coded and non-coded amino acids by thin layer chromatography. Prog Clin Biol Res 1989; 291: 131-134.

26. SIMCA 13.0. Umetrics, Sweden, 2012.

27. Doytchinova IA, Flower DR. 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.

28. Yordanov V, Dimitrov I, Doytchinova I. Proteochemometrics-based prediction of peptide binding to HLA-DP proteins. J Chem Inf Model 2017, in press.

29. Andreatta M, Nielsen M. Characterizing the Binding Motifs of 11 Common Human HLA-DP and HLA-DQ Molecules using NNAlign. Immunology 2012; 136: 306-311.

Corresponding author:
Irini Doytchinova
Department of Chemistry
Faculty of Pharmacy
Medical University of Sofia
2 Dunav st, Sofia 1000, Bulgaria
email: idoytchinova@pharmfac.mu-sofia.bg