

## A Combined Computer-Aided Approach to Drive the Identification of Potential Epitopes in Protein Therapeutics

Bárbara Athayde Vaz Galvão da Silva, Ana Marisa Chudzinski-Tavassi, Kerly Fernanda Mesquita Pasqualoto

Molecular Biology Laboratory – Unit of Innovation and Development, Butantan Institute and Centre of Excellence in New Target Discovery – CENTD, Butantan Institute, Sao Paulo, Brazil.

Received, March 12, 2018; Revised, July 11, 2018; Accepted July 11, 2018; Published, July 12, 2018.

**Abstract – Background:** The identification of fragment sequences, or motifs, within a therapeutic protein that may elicit an immune response when processed by T-cells can be provided by computer-aided approaches. Immunogenicity is a significant problem associated with protein therapeutics and should be investigated in the early stage of protein-based drug development to avoid treatment resistance and potentially life-threatening immune responses. **Purpose:** To provide a combined computer-aided protocol for investigating the immunogenic profile of a recombinant Kunitz-type inhibitor, which has been reported as promising antitumor agent by our research group. **Methods:** The combination of databases searching (IEDB and SYFPEITHI) and molecular docking simulations was exploited, herein. This combined protocol has allowed the identification of potential epitopes before *in vitro/in vivo* evaluation. Predictors of human proteasome cleavage transport and major histocompatibility complex (MHC) binding were considered as overall score assigning the corresponding intrinsic potential of being a T cell epitope to each fragment sequence. The peptides or motifs better classified in the two databases were docked into the three-dimensional (3D) structure of MHC (class I and II) complex to verify the calculated binding affinity. The binding interactions regarding the molecular recognition process by T-cells were also exploited through the MHC:ligand:T-cell complexes. **Results:** Regarding the Kunitz-type sequence, four motifs were identified as potentially epitopes for MHC-I and three motifs were found for MHC-II. But, those motifs were classified as moderately immunogenic. **Final remarks:** The combined computer-aided protocol has significantly reduced the number of potential epitopes to be considered for further analysis and could be useful to identify immunogenic fragments (high, moderate and low) in protein pharmaceuticals before *in vitro/in vivo* experimentation.

### INTRODUCTION

Protein therapeutics (enzymes, toxins, monoclonal antibodies) have been considered as significantly promising for improving human health offering as advantages more specificity and less toxicity when compared to small compounds. These called protein-based drugs are mostly safe and nontoxic, whether natural or recombinant, but depending on the dose and administration frequency (number of doses administered), they often induce harmful anti-drug antibodies (ADA), which may compromise the drug's efficiency and provide serious adverse events related to the cross-reactivity with autologous proteins (endogenous counterpart). ADA may change not only the drug's pharmacokinetic profile, but also its pharmacodynamic response interfering or neutralizing the biological effect (1,2). Thus, the immunogenicity of protein-based drug candidates must be considered, primarily in the early stage of drug development process, to avoid both the

treatment resistance and potentially life-threatening immune responses.

The immune response mechanisms regarding protein therapeutics can be characterized by the (i) activation of the classical immune system by “foreign” proteins, similar the immune response elicited against pathogens or vaccines (involving T-cells, B-cells, and innate immune system); and (ii) alteration of B- and/or T-cell tolerance, such as the response elicited to autologous self-proteins in certain autoimmune diseases. The two mechanisms overlap but are slightly distinct from one another. Human immune responses to autologous proteins might also involve overcoming the regulation of adaptive immune responses by regulatory T-cells (T<sub>reg</sub>), though (1,3).

---

**Corresponding Author:** Kerly Fernanda Mesquita Pasqualoto, Alchemy - Innovation, Research & Development Ltd., Prof. Lineu Prestes Ave. 2242, Room 113, CIETEC/University of São Paulo, São Paulo, SP - 05508-000, Brazil; kfmipasqualoto@alchemydrugs.com.br; kerlypasqualoto@gmail.com.

Concerning the classic immune pathway, the production of anti-therapeutic protein responses is the top of a series of events, which eventually lead to B-cell activation and antibody secretion (4,5).

The T-cell response to a protein therapeutic antigen (antigenic fragment/peptide) generally depends on (i) the binding of T-cell epitopes to the major histocompatibility complex (MHC), (ii) the presentation of the MHC:epitope complex at the cell surface, and (iii) the recognition of the MHC:epitope complex by either an effector, or T<sub>reg</sub>. The experimental binding affinity of T-cell epitopes to MHC, or human leucocyte antigen (HLA), can be evaluated through MHC or HLA binding assays (6,7), and the findings can be retrieved from databases used for immunogenicity prediction, for instance.

MHC class I molecules are expressed by all nucleated cells whereas class II molecules are primarily expressed by professional antigen-presenting cells (APC), such as dendritic cells (DC), macrophages, and B-cells (8,9). MHC class I and II are similar in function presenting peptide fragments (epitopes) at the cell surface to be recognized by CD8<sup>+</sup> and CD4<sup>+</sup> T-cells, respectively. Those peptides are from different sources, though, generally intracellular for MHC class I and exogenous for MHC class II, obtained via different pathways. Interestingly, there is a link between the two pathways (MHC-I and MHC-II), termed cross-presentation, whereby exogenous antigens (previously phagocytosed) can be also presented by MHC-I molecules. Therefore, the T-cell epitopes presentation by both, MHC-I and MHC-II, contributes to the initiation of an immune response (8,10,11).

The fragment sequences within a therapeutic protein that, when processed by T lymphocytes may elicit an immune response, can be predicted by applying computer-aided approaches. Regarding protein-based drugs, it has been reported that the primary amino acid sequence itself can be a strong determinant concerning the protein's immunogenic potential (binding affinity by MHC or HLA and presentation of T-cell epitopes) (12). Of note, there are databases available and widely used, such as IEDB (Immune Epitope Database) (13) and SYFPEITHI (14), loaded with information regarding the binding affinity by MHC or HLA and providing tools for mapping potentially T-cells epitopes (1). The databases allow the prediction of epitopes considering different application types, including tumor cell lines. However, this kind of prediction has

been reported mostly regarding vaccine candidates, and it is often used aside the experimental assays (15-18), instead of as a driven procedure before experimentation.

Herein, a recombinant Kunitz-type inhibitor (19,20), which has been reported as a promising antitumor agent by our research group (21), was investigated regarding its immunogenic profile combining the search in two databases (13,14) and molecular docking simulations to identify potentially epitopes before *in vitro/in vivo* evaluation. Kunitz-type domains are ubiquitous in nature, acting as both serine protease inhibitors and toxins in animal venoms. Structurally, they consist of between 50 and 70 amino acid residues, adopting a conserved fold generally with two antiparallel  $\beta$ -sheets and one or two helical regions, which are stabilized with three disulphide bridges. They are involved in several physiological processes, such as inflammation, blood coagulation, and fibrinolysis (22,23).

## METHODS

### *Input data*

The computer-aided approach was applied to a recombinant Kunitz-type inhibitor deposited in GenBank, code AAT68575.1, by Batista, I.F.C. and Chudzinski-Tavassi, A.M. (submission date: 02-MAR-2004) (24). The primary sequence contains 129 amino acid residues, where the residues 1-21 correspond to the signal peptide. Then, the input data was the primary (linear) amino acid sequence (FASTA format) of the recombinant protein, discarding the signal peptide.

Two free-accessed databases, SYFPEITHI (14) and IEDB (13), were chosen to perform the immunogenicity prediction regarding the number of reports and number of citations (25-27). Both databases have information on binding affinity of epitopes by the HLA complex (experimental binding affinity data) (15). In this step, we also informed the MHC source species of interest to perform the analysis.

### *Selection of alleles*

Two MHC properties make difficult to avoid the immune response: (i) MHC is polygenic, meaning it contains several different MHC-I and MHC-II genes (every individual possesses a set of MHC molecules with different ranges of peptide-binding specificities); and, (ii) MHC is highly polymorphic, meaning there are multiple variants of each gene within the population.

The MHC-I and MHC-II alleles were selected according to their frequency (28). Of note, allele frequency represents the incidence of a gene variant in a population. We selected twelve alleles for MHC-I *HLA-A\*02:01* (25.2%); *HLA-A\*24:02* (16.8%); *HLA-A\*01:01* (16.2%); *HLA-A\*03:01* (15.4%); *HLA-A\*11:01* (12.9%); *HLA-B\*08:01* (11.5%); *HLA-B\*40:01* (10.3%); *HLA-B\*44:02* (9.2%); *HLA-B\*35:01* (6.5%); *HLA-B\*51:01* (5.5%); *HLA-B\*53:01* (5.4%); *HLA-B\*15:01* (5.2%) (29,30,13) and five for MHC-II *HLA-DRB1\*15:01* (8%); *HLA-DRB1\*07:01* (7%); *HLA-DRB1\*03:01* (7%); *HLA-DRB1\*11:01* (6%); *HLA-DRB1\*09:01* (6%) (31,32) in order to cover at least 90% of individuals.

### **Size of sequence fragments (peptides/epitopes)**

Regarding the size of fragments generated, the core of residues in the T-cell epitope sequence that mainly define the binding affinity and stability to MHC pockets is limited in length to 9 or 10 amino acids (aa). Generally, epitopes presented by MHC-I have 9 aa residues in length, and those presented by MHC-II may range from 12 to 20 aa residues (15). In this study, we have considered sequence fragments containing up to 15 aa residues for MHC-II.

Due to the variability concerning the interaction points into the MHC-II binding site (less specific), the predictability findings have a confidence limit of 50%. The MHC-I prediction, on the other hand, has 80% confidence limit since the core, containing nine residues, has been considered providing more specific interactions (14,30,13).

Of note, the prediction of T-cell epitopes based on fragment sequences (peptides) is computationally feasible when sufficient information on a set of sequence fragments regarding the MHC (I or II) binding affinity is available. The IEDB database (30,13), for instance, may provide reliable background findings for developing T-cell epitope prediction tools.

### **Sequence fragments classification and cut-off values**

When using IEDB database (30,13), the prediction method can be chosen by the user. Then, regarding the MHC-I epitopes prediction, the NetMHCpan method (33) based on artificial neural network (ANN) was chosen to classify the sequence fragments, since there was not any corresponding predictor to the Kunitz-type protein investigated. That method provides the binding affinity prediction expressed as IC<sub>50</sub> values (nM). A binding affinity (IC<sub>50</sub>) threshold of 500 nM, for instance, identifies

peptide binders recognized by T-cells and could be considered to select peptides (6). Paul *et al.* (2013) (34) have showed that absolute binding affinity threshold correlates better with immunogenicity, and for getting even better correlation, MHC-specific thresholds should be used. Then, herein, we considered the specific IC<sub>50</sub> cut-off values for the 38 most common HLA-A and HLA-B alleles, representative of the nine major subtypes found in the IEDB database (13) to classify the fragment sequences.

The stabilized matrix method, SMM-align (35), is mostly recommended to perform the MHC-II epitopes prediction and was selected to score the sequence fragments. The MHC-II binding site is open at both ends making the correct alignment of a peptide in the binding cavity a crucial part, especially regarding the identification of the core of residues as an MHC-II binding motif. The method predicts quantitative peptide:MHC binding affinity values, meaning it calculates the IC<sub>50</sub> value (nM) for each epitope, making it ideally suited to rationally discover promising epitopes.

When using the SYFPEITHI database (14), the potential epitopes are sorted by evaluating the position of each amino acid and its chemistry properties experimentally determined (lipophilicity, basicity, and so on). That information is converted to a score using a proper algorithm, implemented in the database. The algorithm also computes a maximum score value for each allele based on available information from natural binders. For example, the maximum score value that a certain peptide could present considering the allele *HLA-A\*02* would be 36 (14). Cut-off values are not reported for this kind of score calculation. Then, in order to select the peptides better classified, we have considered not only the sequence fragments presenting score values higher than 20, but also if they would appear in multiple alleles.

### **MHC class I: intracellular processing prediction**

IEDB also performs the intracellular processing prediction regarding the MHC class I pathway. Predictors of human proteasome cleavage, binding affinity by the transporter associated with antigen presentation (TAP) protein, and antigen presentation by MHC-I at cell surface, were used to produce an overall score for each peptide regarding its intrinsic potential of being a T-cell epitope.

In this step, we selected the immune proteasome type to make the prediction. The predictions are based on *in vitro* proteasomal digest of the enolase

and casein proteins as described by Tenzer et al. (2005) (36).

The TAP score estimates effective  $-\log(\text{IC}_{50})$  values of a peptide or its N-terminal prolonged precursors regarding the binding to TAP. It has been shown that high affinity of a certain peptide translates into high transport rates (30,13). The prediction of antigen presentation by MHC-I, however, is quite similar to the MHC-I binding prediction supra-mentioned.

### **Molecular Docking Simulations**

The molecular docking method provides a binding affinity estimative regarding the ligand-receptor/protein complex through a scoring or energy function. Larger ligands having many rotatable bonds (freedom degrees), for instance, demand significantly computational time-consuming. Then, a rigid docking procedure can be considered to speed up the calculation procedure. Several fragment peptides, or motifs, may have immunogenic potential and could interact with MHC molecules inducing immune responses. There is an urgent need to find faster predicting protocols that may allow to identify and classify the most immunogenic fragments before carrying out the *in vitro* and *in vivo* experiments, which are also expensive and time-consuming (37).

The findings from both databases were compared, and the fragment sequences (peptides) classified as promising T-cell epitopes for MHC-I and MHC-II were considered to perform molecular docking simulations, allowing the assessment to the three-dimensional (3D) structural information in the molecular binding recognition process by MHC and T-cell, as well.

The presence of the T-cell receptor ( $T_{CR}$ ) in the MHC-ligand complex was crucial for selecting the 3D structure to perform molecular docking simulations. It was considered 3D structures having resolution under 3.0 Å (X-ray diffraction method).

The Cartesian coordinates of the *HLA-A\*02* complex (MHC-I:ligand:Tcell), deposited in the Protein Data Bank (38) (PDB), PDB ID 1OGA (resolution at 1.40 Å) (39), were retrieved and used as reference/template to perform the molecular docking simulations for the MHC class I. The ligand was used as reference to build up the 3D molecular models of the fragment sequences previously selected as promising epitopes (database search).

Based on data reported from the crystallographic structure of the complex (39), three calculation conditions were considered regarding the number of

water molecules participating into the molecular interaction process, as follows: (i) complex MHC-I:ligand and two water molecules establishing interactions only with the ligand; (ii) complex MHC-I:ligand and five water molecules, where the additional three molecules participate in interactions into the HLA-A active site; and, (iii) complex MHC-I:ligand: $T_{CR}$  lymphocyte and ten water molecules, being five molecules corresponding to those from the second calculation condition, and five more related to interactions established between the  $T_{CR}$  lymphocyte protein chains and the rest of the complex (ligand and HLA-A active site). Of note, the last calculation condition is likely more representative in terms of T-cell recognition process.

For the molecular docking simulations of MHC class II, the Cartesian coordinates of the *HLA-DRB1* complex (MHC-II:ligand: $T_{CR}$ ) deposited in the Protein Data Bank (38) (PDB), code PDB ID 1FYT (resolution at 2.6 Å) (40), were retrieved and used as reference/template. The ligand contains 13 aa residues, and the interactions established between the complex and water molecules have been also reported (40). The ligand was used as reference to build up the 3D molecular models of the sequences previously selected as promising epitopes. Although those peptides have size of 15 aa residues, the core of residues (generally containing 9 aa) is indeed the motif responsible for establishing the interactions into the *HLA-DRB1* active site.

Molecular docking simulations were carried out using CLC Drug Discovery Workbench software (version 2.4, Qiagen Aarhus A/S, 2014) (41). Regarding the molecular modelling method used, it classifies promising ligands according to the complex binding energy values, which are related to the ligands' calculated binding affinity (complex formation). Furthermore, the intermolecular interactions and interatomic distances can be also exploited using this approach.

Re-docking procedure using the original PDB ligand was performed to establish the optimum conditions for computing the peptides binding affinity values. The simulation protocol was the following: (i) 1,000 iterations; (ii) rigid approach, because the number of rotatable bonds; and (iii) Nelder-Mead simplex method (42), implemented in the package, for function minimization. Of note, not only the root-mean square deviation (RMSD) values (43), regarding the atomic position differences between the reference ligand and each peptide, but also the score function using the  $\text{PLANTS}_{\text{PLP}}$  method (44), were considered as docking evaluation criteria.

A particularly ligand binding mode into the protein binding pocket can be connected to a score value. The score, herein, mimics the potential energy change when the target protein and ligand come together, meaning that a very negative score corresponds to a strong binding whereas a less negative, or even positive, score value corresponds to a weak or non-existing binding. The total score value comprises the following types of contribution: hydrogen bond score, metal interaction score, steric interaction score, and ligand conformation penalty score. Concerning the last contribution, it scores the complementarity between the binding site and ligand by rewarding and punishing different types of heavy atom contacts having inter-atom distance less than 5.5 Å.

Regardless the method limitations, it can be considered as a good alternative to access the 3D structural information related to the molecular recognition process for rationally driving the selection of potentially promising epitopes.

## RESULTS AND DISCUSSION

### *MHC class I ligand identification*

Initially, the databases divided the primary amino acid sequence of the Kunitz-type inhibitor in all possible fragments containing nine amino residues in size, which is the preferred size for MHC-I presentation (15), giving approximately 1,200 peptides (fragment sequences or motifs or potential epitopes).

Regarding the findings from SYFPEITHI database (14), eleven peptides were selected considering the twelve investigated alleles for MHC-I, according to the criteria previously described in the Methods section. The score values and corresponding distribution of the peptides in each allele are presented in Figure 1S (Supporting Information). Concerning Figure 1S, the fragments p5, p57, and p67, were identified as potential ligands for seven of the twelve alleles investigated. Six fragments (p5, p86, p71, p81, p69 and p57) were pointed out as MHC-I ligands for the most frequent allele in the population (HLA-A\*02). Furthermore, the fragments p5 and p86 have showed score above 20 and were selected for the next stage of the study. The same kind of evaluation was performed to the other fragments. After that, the findings obtained from SYFPEITHI database (14) were compared to those from IEDB (13).

The IEDB database (13) considers specific IC<sub>50</sub> cut-off values for each allele (29). The position of the first aa residue corresponding to each fragment better classified as well as the predicted IC<sub>50</sub> (nM) and IC<sub>50</sub> (nM) cut-off values are listed in S1 Table (Supporting Information).

Regarding S1 Table, after applying the IC<sub>50</sub> cut-off, there were not any distinct sequence fragments identified as good ligands for MHC-I considering the three most frequent alleles in population (*HLA-A\*02:01*, *HLA-A\*24:02*, and *HLA-A\*01:01*). The sequence fragments (peptides) p1, p32, p34, and p81 were found in multiples alleles suggesting they would have a significant potential to be presented at least by two different HLA complexes to T<sub>CR</sub> lymphocytes (Figure 2S; Supporting Information).

Despite the differences concerning the frequency of the epitopes in each allele, none of the fragment sequences was excluded in this step since they have presented IC<sub>50</sub> predicted values below the specified cutoff value (29). Therefore, the comparison of the motifs identified from the two databases was made and six fragment sequences were common to both databases, as showed in Table 1.

### *MHC class I intracellular processing prediction*

Predictors of human proteasome cleavage and binding affinity by TAP protein were considered in association with the prediction of fragment sequences as potential epitopes to be presented by MHC-I at cell surface. An overall score for each peptide (fragment sequence or epitope or motif) was generated concerning the intrinsic potential of each peptide being a T-cell epitope. Thus, the combined score has increased the chance of finding promising immunogenic sequences (15). The different fragment sequences found in this step were also considered to perform molecular docking simulations.

The binding affinity of certain peptides (sequence fragments or motifs) by MHC-I, to be further presented at cell surface, involves an intracellular processing where firstly an exogenous protein would be internalized by phagocytosis or endocytosis; then, it would be cleaved by the proteasome generating smaller fragments/peptides having different sizes. Those peptides may bind to the TAP protein to be transported to the endoplasmic reticulum (ER). In the ER, depending on the binding affinity, the peptides can be recognized by MHC-I, which will be responsible for presenting them at the cell surface (membrane) (10,45).

**Table 1.** Fragment sequences (motifs) identified as good ligands for MHC-I in both databases

SYFPEITHI		IEDB	
Peptide	Sequence	Peptide	Sequence
5	KAVCNLPKL	1	MANSKAVCN
<b>57</b>	<b>RLCEEQTHF</b>	19	CSNKTEIRW
<b>59</b>	<b>CEEQTHFHF</b>	32	TACEAFIFK
<b>67</b>	<b>FESPKLICF</b>	34	CEAFIFKGC
68	ESPKLICFK	<b>57</b>	<b>RLCEEQTHF</b>
69	SPKLICFKV	<b>59</b>	<b>CEEQTHFHF</b>
71	KLICFKVQD	63	THFHFESPK
<b>81</b>	<b>WILNDIMKK</b>	<b>67</b>	<b>FESPKLICF</b>
86	IMKKNLTGI	72	LICFKVQDY
<b>89</b>	<b>KNLTGISLK</b>	<b>81</b>	<b>WILNDIMKK</b>
<b>100</b>	<b>EEDADSGEI</b>	<b>89</b>	<b>KNLTGISLK</b>
		<b>100</b>	<b>EEDADSGEI</b>

Common sequences are in bold letters.

**Table 2.** Proteasome cleavage and TAP transport prediction ranking using the IEDB database (13)

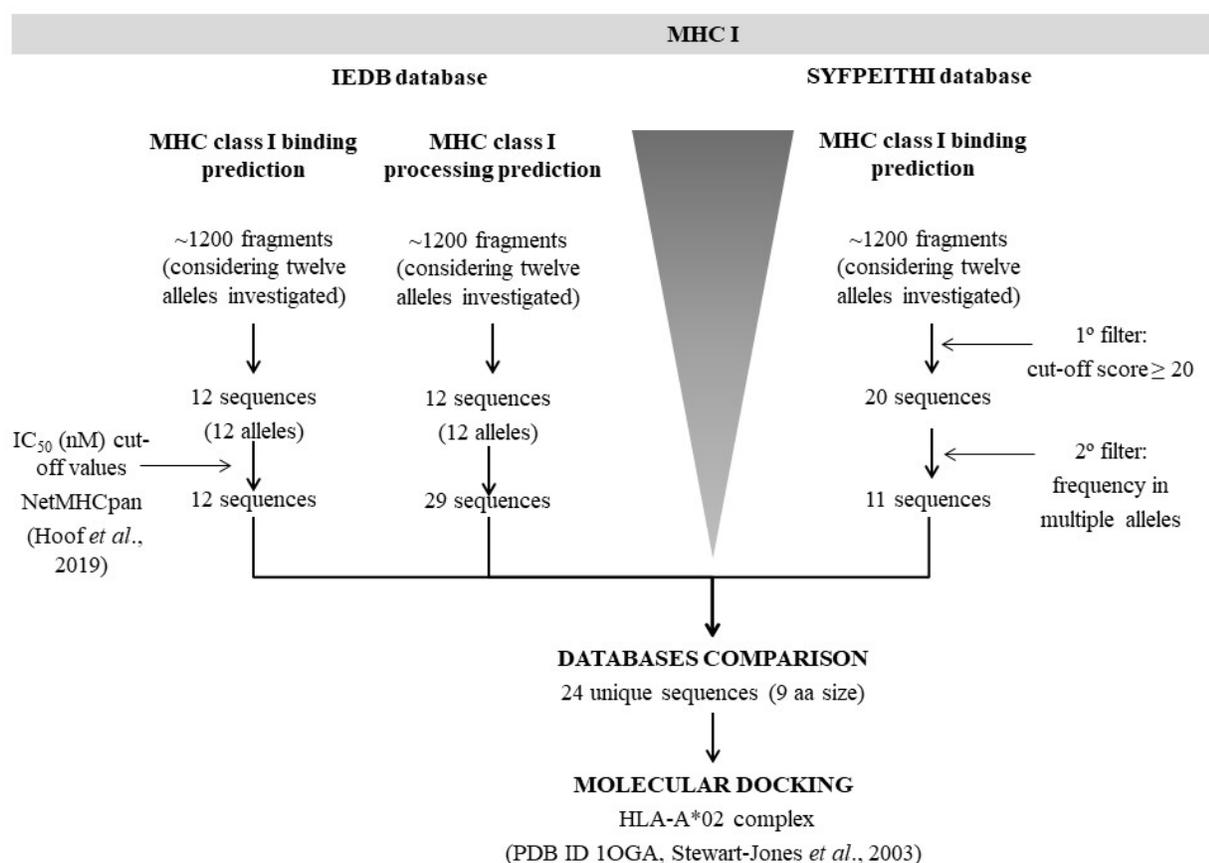
Peptide	Sequence	Score
72*	LICFKVQDY	<b>2.38</b>
21	NKTEIRWYY	<b>2.30</b>
20	SNKTEIRWY	<b>2.28</b>
57**	RLCEEQTHF	<b>2.21</b>
31	GTACEAFIF	<b>2.04</b>
67**	FESPKLICF	<b>2.02</b>
59**	CEEQTHFHF	<b>1.87</b>
29	YNGTACEAF	<b>1.86</b>
41	GCGGNDNNF	<b>1.82</b>
5**	KAVCNLPKL	<b>1.74</b>
75	FKVQDYWIL	<b>1.65</b>
50	DRVDDCQRL	<b>1.52</b>
18	TCSNKTEIR	1.49
88	KKNLTGISL	1.49
64	HFHFESPKL	1.48
43	GGNDNNFDR	1.44
83	LNDIMKKNL	1.43
49	FDRVDDCQR	1.42
73	ICFKVQDYW	1.38
19	CSNKTEIRW	1.36
2	ANSKAVCNL	1.36
86	IMKKNLTGI	1.25
65	FHFESPKLI	1.22
26	RWYYNGTAC	1.15
74	CFKVQDYWI	1.13
6	AVCNLPKLA	1.07
81	WILNDIMKK	1.05
17	ETCSNKTEI	1.03
69	SPKLICFKV	1.02

Alleles *HLA-A\*02:01*, *HLA-A\*24:02*, *HLA-A\*01:01*, and *HLA-A\*03:03*. Sequences also found in the MHC-I binding procedure using IEDB\* and SYFPEITHI\*\* database.

The analysis was made considering the alleles *HLA-A\*02:01*, *HLA-A\*24:02*, *HLA-A\*01:01*, and *HLA-A\*03:03*, which cover 70% of the population frequency (13,30). Higher score values suggest that the fragments generated in the intracellular processing would be likely presented via MHC-I pathway (29). The intracellular processing prediction score combines the proteasome cleavage, TAP transport, and MHC binding predictions (13,15). The score values ranged from -0.30 to 2.38 regarding 390 fragment sequences generated. A cut-off score value of 1.0 was set to reduce the number of fragments to be considered in the subsequent steps. Therefore, twenty-nine fragments were selected regarding all alleles investigated, and they are listed in Table 2. The fragment sequences also found in the MHC-I binding procedure (previous step), using both databases, are pointed out with asterisk: one (\*) for those peptides found in IEDB (13) and two (\*\*) for those found in SYFPEITHI (14).

Twelve fragment sequences presented score values higher than 1.50 (in bold; Table 2) and among them the peptides p5, p57, p59, and p67, were also found in the MHC-I binding step carried out in SYFPEITHI (\*\*) database (14), and the peptide p72 was also found in the previous step performed in the IEDB (\*) database (13). The sequences identified as good MHC-I ligands were compared to the fragments found in this step (MHC-I intracellular processing prediction) before being considered to follow to the next step, molecular docking simulations.

A summary regarding the steps considered for searching potentially promising MHC-I ligands using the tools available in the two databases is presented in Figure 1. Concerning the large number of fragments retrieved from the databases it is quite important adopting selection filters to rationally reduce the number of fragment sequences to be considered in molecular docking simulations.



**Figure 1. Searching steps for potential MHC-I ligands.** Summary containing the steps and filters considered to select potentially promising fragment sequences as MHC-I (*HLA-A\*02*) ligands using the both databases, IEDB (13) and SYFPEITHI (14).

### MHC class I molecular docking simulations

After the findings comparison from both databases, twenty-four fragment sequences (peptides or epitopes or motifs) (see Figure 1) classified as promising T-cell epitopes for MHC-I were considered to perform molecular docking simulations (41).

Then, in this step, it was possible to assess the 3D structural information regarding the molecular binding recognition process. The selection of potentially promising epitopes has become more rational and reliable, since the interactions involved in whole molecular system (*HLA-A\*02:ligand:T<sub>CR</sub>*), PDB ID 1OGA (39), were taken into account. The calculation conditions were previously described in the Methods section, and the findings for the first and second conditions are shown in Supporting Information (Table 2S and 3S). Of note, the molecular docking conditions can be directly compared to the epitopes prediction findings from the two databases since the interactions involved in the complex *HLA-A\*02:ligand* can also be assessed (calculation condition 1 and 2). Therefore, eighteen of twenty-four peptides selected in both databases have presented favourable score energy values (negative values) as well as lower RMSD values, regarding the *HLA-A\*02:ligand* complex formation. Concerning the calculation conditions, the findings from molecular docking were mostly in agreement with those obtained from the databases (MHC-I binding prediction), but reducing the number of promising sequences to be considered to further evaluation.

The third calculation condition includes the water molecules establishing molecular interaction with the whole system: the protein chains of the T<sub>CR</sub> lymphocyte, the *HLA-A\*02* binding site and the ligand (39). This condition is the most representative of the molecular system investigated regarding the stimulation of an immune response. Not only the ligand binding affinity by MHC-I is important in the molecular recognition process, but also the ligand presentation by MHC-I to the T<sub>CR</sub> lymphocyte. Concerning that, the databases have not enough information to predict the interactions involved in the entire complex *HLA-A\*02:ligand:T<sub>CR</sub>* lymphocyte. The findings regarding the third calculation condition are shown in Table 3. Among the twenty-four peptides investigated only four fragment sequences (p1, p100, p69, p86) presented favourable score energy values. Of note, their RMSD values were lower than 1 Å. Based on these findings,

they could be considered as potentially immunogenic.

LaFuente and Reche (2009) (11) have reported that the residues at terminal positions in the ligands' sequences seem to directly establish interactions with MHC. Otherwise, the residues at fourth, eighth, and sixth positions are related to the recognition process by the T<sub>CR</sub> lymphocyte. The peptides which presented negative energy score values had the same interaction mode previously reported by LaFuente and Reche (2009) (11). The peptide p1 [MANSKAVCN], for instance, was the best ranked by the docking score function (Table 3; -60.56 kcal/mol) and has showed molecular interactions with the following residues in the *HLA-A\*02* binding site: Tyr7, Tyr59, Tyr84, Thr143, Lys146, Tyr159, and Tyr171. Those residues are involved in antigen presentation (46). Furthermore, interactions were also observed with the amino acid residues Gln52, Gln96, Ser95, and Ser99 of the T<sub>CR</sub> lymphocyte protein chain (Figure 2).

The peptides p100 [EEDADSGEI], p69 [SPKLCIFKV], and p86 [IMKKNLTGI] presented additional molecular interactions with four, three, and five amino acid residues, respectively, conserved in the *HLA-A\*02* binding site (46). The main difference seems to be related to the interactions established with the amino acid residues of the T<sub>CR</sub> lymphocyte chain, which could explain the different score energy values as well as the RMSD variation (accommodation into the binding site) in comparison to the best ranked peptide (p1). Regarding the molecular interactions with T<sub>CR</sub>, the fourth residue of peptide p1 interacts with the residues Ser95, Gln96, Gln52, and Ser99, while its sixth residue interacts with the Ser99 residue of T<sub>CR</sub>. For peptide p100 (-21.19 kcal/mol), the fourth residue only interacts with the Gln52 residue of T<sub>CR</sub>, while the sixth residue interacts with the residues Gln52 and Ser99. Concerning the peptides p69 (-18.27 kcal/mol) and p86 (-5.95 kcal/mol), the interactions with T<sub>CR</sub> are established only by the amino acid residues at the fourth position (Gly97 and Gln52 to p69; Gln96 to p86).

The 3D representation regarding the molecular interactions established by peptide p1 in the complex *HLA-A\*02:p1:T<sub>CR</sub>* lymphocyte is shown in Figure 3 (A-C), using Discovery Studio Visualizer software.<sup>44</sup> The molecular docking of p1 in the *HLA-A\*02* binding site is presented in Figure 3A, and the solvent accessibility surface around the ligand p1 in Figure 3B. The interactions involved in the entire complex (antigen recognition process), including the

T<sub>CR</sub> lymphocyte, are shown in Figure 3C. The MHC-I binding site exposes p1 to be recognized by the lymphocyte, and it can be observed that the terminal portions of ligand p1 establish interactions with aa residues from MHC-I whereas the central portion of p1 interacts with the T lymphocyte chain.

### MHC class II ligand identification

The two databases cut the primary sequence of the protein in all possible fragments containing fifteen amino acid residues in size (preferred size for MHC-II; ranging from 12 to 20 residues) resulting in approximately 450 fragments. The alleles chosen in both databases to perform the MHC-II ligand prediction were *HLA-DRB1\*15:01*, *HLA-DRB1\*07:01*, *HLA-DRB1\*03:01*, and *HLA-DRB1\*11:01*. Of note, the allele *DRB1\*09:01* was not available in the SYFPEITHI database (14).

Regarding the findings from SYFPEITHI database (14), thirteen peptides (fragment sequences or motifs) presented score above 20. The peptide p80

has appeared in three alleles and the peptides p4 and p69 have appeared in two alleles. The score values and corresponding distribution of the peptides in each allele are presented in Figure 3S (Supporting Information section). Concerning the IEDB database (13), the predicted binding affinity of each peptide via MHC-II pathway is provided through IC<sub>50</sub> values (nM). According to IEDB 2005-2015® (30), when peptides present IC<sub>50</sub> values below 50 nM they are classified as having high binding affinity; when they have IC<sub>50</sub> values between 50 and 500 nM they are considered as having moderate (intermediate) binding affinity; and when they present IC<sub>50</sub> values higher than 500 nM they are discriminated as having low binding affinity by MHC-II (48). Twelve peptides were identified as having moderate binding affinity by MHC-II (IC<sub>50</sub> values between 50 and 500 nM) and seven from twelve appeared in the allele *HLA-DRB\*11:01* (Figure 4S; Supporting Information section).

**Table 3.** Findings from molecular docking simulations considering the third calculation condition (complex *HLA-A\*02:ligand:T<sub>CR</sub>* lymphocyte and water molecules) (39)

Peptide	Sequence	Total score* (kcal.mol <sup>-1</sup> )	RMSD (Å)	Hydrogen bond score (kcal.mol <sup>-1</sup> )	Steric interaction score (kcal.mol <sup>-1</sup> )	Ligand conformation penalty (kcal.mol <sup>-1</sup> )
L10GA	GILGFVFTL	-173.15	0.16	-42.36	-145.9	15.11
p1	MANSKAVCN	<b>-60.56</b>	0.14	-51.53	-33.46	24.43
p100	EEDADSGEI	<b>-21.19</b>	0.18	-39.44	-27.61	45.85
p69	SPKLICFKV	<b>-18.27</b>	0.31	-26.70	-34.35	42.78
p86	IMKKNLTGI	<b>-5.95</b>	0.35	-27.95	1.84	20.17
p59	CEEQTHFHF	8.59	0.76	-23.94	7.6	24.93
p71	KLICFKVQD	20.96	0.29	-37.59	16.29	42.27
p31	GTACEAFIF	22.73	0.35	-30.96	30.74	22.96
p89	KNLTGISLK	38.57	0.91	-30.7	9.19	60.08
p41	GCGGNDNNF	42.11	0.88	-34.05	60.15	16.01
p57	RLCEEQTHF	48.73	0.57	-28.15	44.73	32.14
p50	DRVDDCQRL	49.28	0.52	-39.2	30.88	57.59
p34	CEAFIFKGC	61.04	0.58	-19.42	61.67	18.8
p29	YNGTACEAF	68.66	1	-15.65	32.34	51.96
p19	CSNKTEIRW	86.28	0.27	-27.34	80.12	33.5
p20	SNKTEIRWY	108.26	0.5	-24.82	72.74	60.34
p81	WILNDIMKK	124.9	0.39	-32.93	107.79	50.04
p5	KAVCNLPKL	191.3	6.98	-20.23	193.74	17.79
p32	TACEAFIFK	192.23	16.32	-4.11	176.27	20.07
p21	NKTEIRWYY	198.67	0.7	-18.34	188.75	28.26
p63	THFHFESPK	202.07	1.12	-15.65	179.4	38.32
p67	FESPKLICF	221.53	18.03	-26.37	218.4	29.51
p75	FKVQDYWIL	254.72	1.02	-23.33	220.73	57.32
p68	ESPKLICFK	257.98	2.33	-18.4	170.69	105.7
p72	LICFKVQDY	392.15	17.07	-12.03	372.62	31.56

\* Total score corresponds to the sum of contributions: hydrogen bond, steric interaction, and ligand conformation penalty.

Five peptides from the same sequence region (76 to 93) of the Kunitz-type inhibitor protein have identified as good MHC-II ligands sharing a common sequence around the moiety DIMKK (see Table 4). Even though the sequences were not the same, the DIMKK region could be considered as a promising epitope. Table 4 shows the fragment sequences better classified as ligands for MHC-II using the two databases.

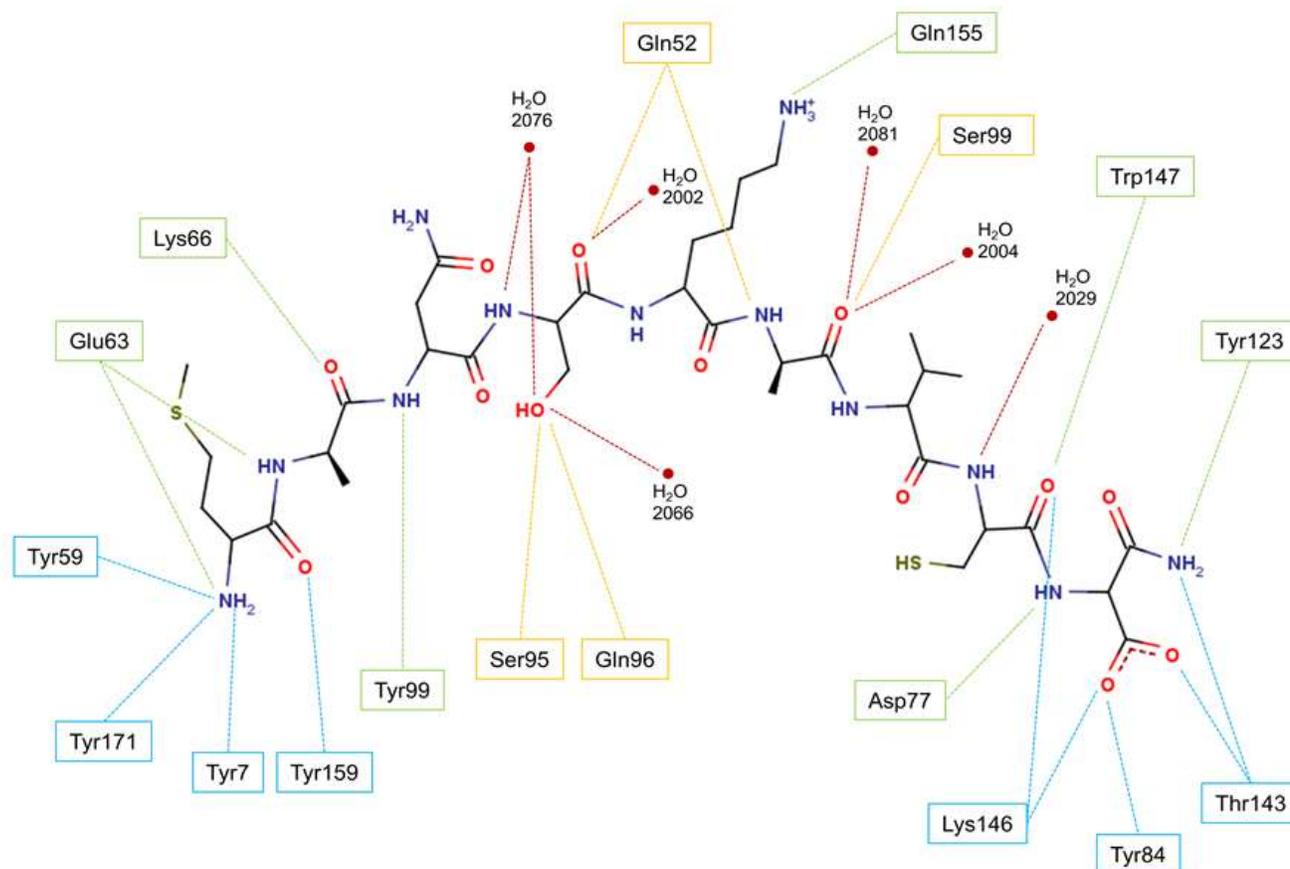
A summary regarding the steps considered for searching potentially promising MHC-II ligands using the tools available in the two databases is presented in Figure 4, pointing out the filters adopted to rationally reduce the number of fragment sequences to be considered in each next step of analysis.

### MHC class II molecular docking simulations

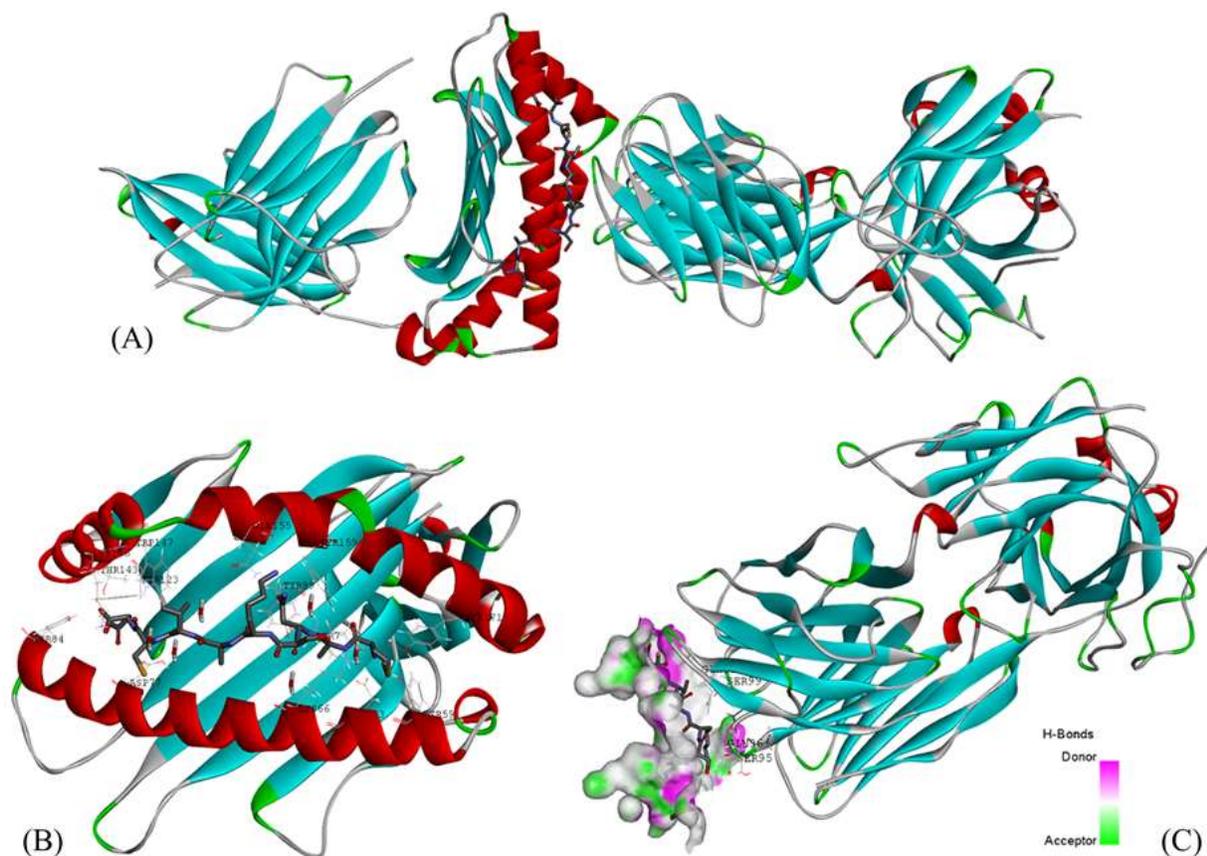
MHC-II molecules are highly polymorphic, and that feature may contribute to degenerate their binding site, which is less specific (more open groove)

allowing the binding of multiple fragment sequences (peptides) to multiple MHC-II subtypes (49). Therefore, concerning the promiscuity related to the interaction regions into the MHC-II binding site, the confidence limit regarding the prediction of potential epitopes by the database used to carry out the analysis is only 50% (13,14,30). In this regard, the 3D structural information concerning the molecular system MHC-II:ligan:T<sub>CR</sub> lymphocyte can play a role to rationally select the epitopes more promising to be good ligands of MHC-II and, then, recognized by T<sub>CR</sub> lymphocyte.

Molecular docking simulations were performed constraining flexible bonds in the primary chain of each peptide. Of note, the size of peptides and related conformational flexibility have impact directly on the time-consumed in calculation as well as in findings reliability (50,51). Molecular docking approach, however, is indeed a relatively fast procedure and can provide new insights regarding the molecular system investigated.



**Figure 2. Molecular interactions established by peptide p1 in the complex HLA-A\*02:p1:T<sub>CR</sub> lymphocyte.** Amino acid residues conserved in HLA-A\*02 binding site are highlighted in blue; residues in green correspond to additional interactions encountered between HLA-A\*02:p1 residues highlighted in yellow belong to the T<sub>CR</sub> binding site; water molecules are presents in red dots. The distances are measured in Å (CLC Drug Discovery Workbench software, version 2.4, 2014) (41).

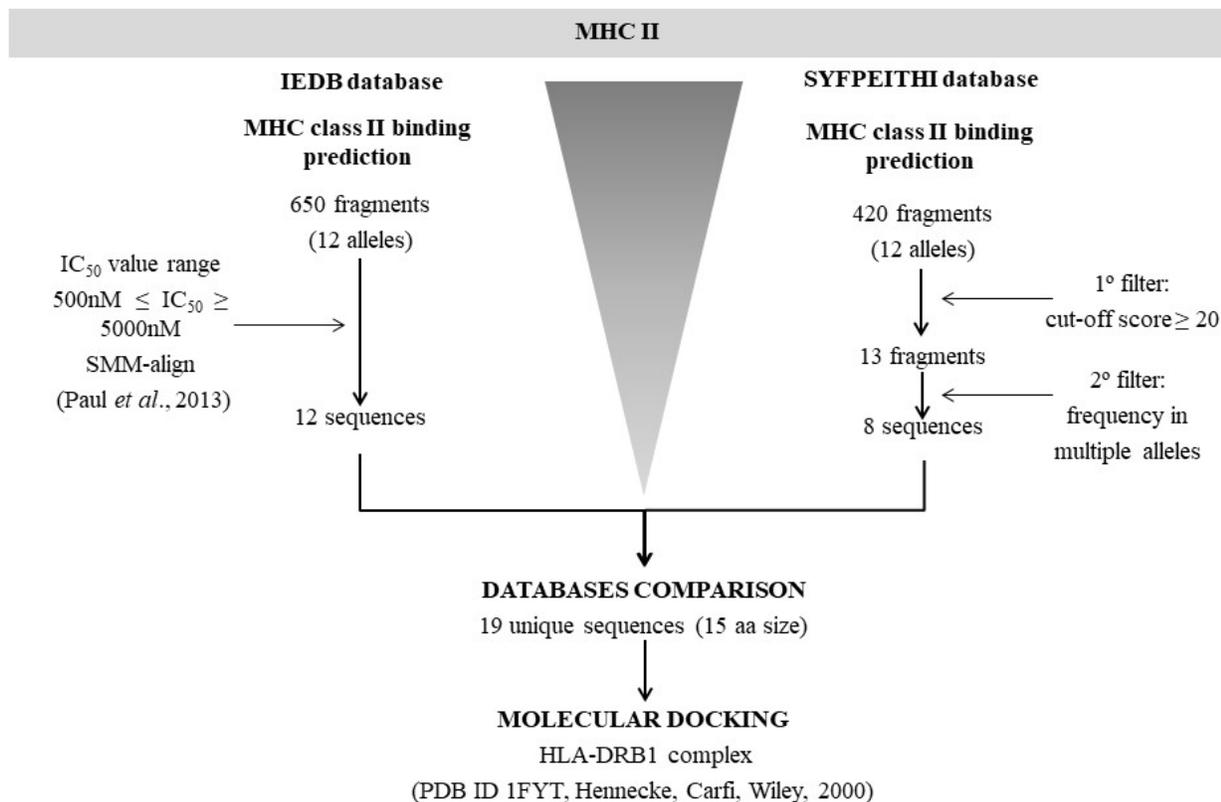


**Figure 3.** Three-dimensional representation of the molecular interactions established by the peptide p1 (stick model) in the complex *HLA-A\*02*:p1:TCR (A); molecular docking of p1 in the *HLA-A\*02*:TCR binding site (B); solvent accessibility surface around the peptide p1 (hydrogen bond acceptor regions in green, and hydrogen bond donor moieties in purple); antigen recognition process showing interactions established between the central portion of p1 and the TCR lymphocyte (C) (Discovery Studio Visualizer, version 4.5, 2015) (47).

**Table 4.** Fragment sequences identified as good ligands for MHC-II in both databases

SYFPEITHI		IEDB	
Peptide	Sequence	Peptide	Sequence
4	SKAVCNLPKLAGDET	1	MANSKAVCNLPKLAG
10	LPKLAGDETCNKTE	2	ANSKAVCNLPKLAGD
22	KTEIRWYNGTACEA	19	CSNKTEIRWYNGTA
26	RWYNGTACEAFIFK	20	SNKTEIRWYNGTAC
69	SPKLICFKVQDYWIL	23	TEIRWYNGTACEAF
74	CFKVQDYWILNDIMK	59	CEEQTHFHFPKLI
79	DYWILND <b>DIMKKN</b> LTI	61	EQTHFHFPKLI
80	YWILND <b>DIMKKN</b> LTI	62	QTHFHFPKLI
		76	KVQDYWILND <b>DIMKKN</b>
		77	VQDYWILND <b>DIMKKN</b> L
		78	QDYWILND <b>DIMKKN</b> L
		79	DYWILND <b>DIMKKN</b> LTI

DIMKK region is shown in bold letters.



**Figure 4.** Searching steps for potential MHC-II ligands. Summary containing the steps and filters considered to select potentially promising fragment sequences as MHC-II (*HLA-DRB1*) ligands using the two databases.

The nineteen fragment sequences selected from the two databases were, firstly, docked into the MHC-II binding site (PDB ID 1FYT; resolution at 2.6 Å) (40) without considering T<sub>CR</sub> lymphocyte. The peptides were sorted according to their calculated binding affinity (energy score value related to the MHC-II:ligand complex formation) and accommodation/orientation in the binding site (RMSD values). The energy scoring function and RMSD values are listed in Table 5.

Ten peptides (p79, p78, p19, p74, p80, p77, p10, p22, p20, p76) have showed favourable (negative) energy score and RMSD values lower than 1 Å. The negative values obtained for total energy score indicate the formation of the *HLA-DRB1*:ligand complexes. Those peptides, consequently, would be more likely presented to the T<sub>CR</sub> lymphocyte. However, in order to be more certain, molecular docking simulations were also run considering the whole complex, *HLA-DRB1*:ligand:T<sub>CR</sub> lymphocyte, and the findings are shown in Table 6.

The peptides p2, p4, p61, and p62 presented RMSD values higher than 20 Å (43), indicating they did not fit well into the complex binding site. The

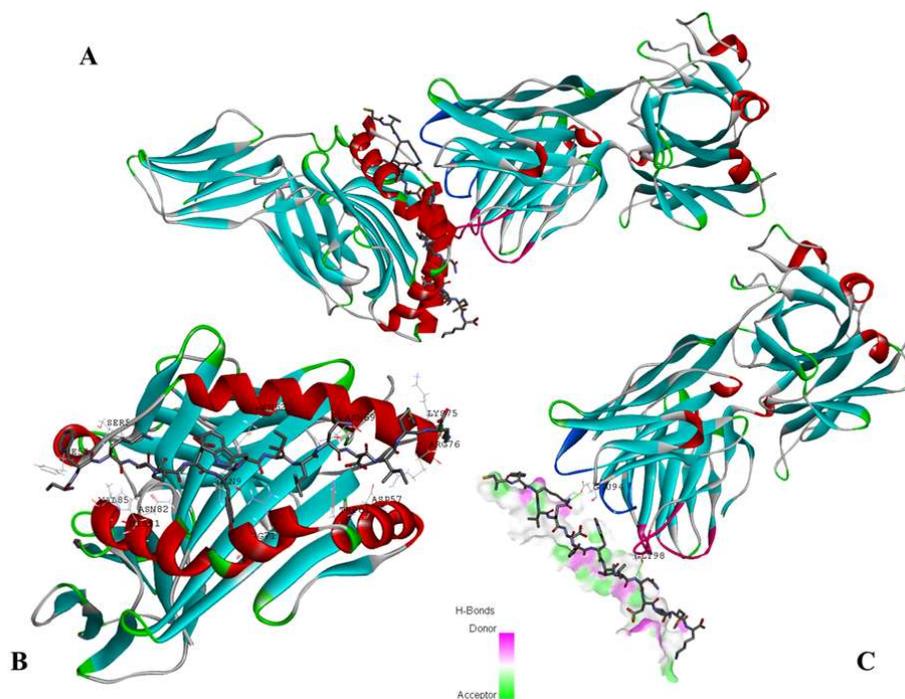
steric hindrance regarding the residues' side chains in the ligands and in the binding site was responsible for that lack of accommodation. Interestingly, when the entire complex (MHC-II:ligand:T<sub>CR</sub> lymphocyte) was considered in docking simulations, only three peptides have showed favorable total energy score (negative values): p74 (-147.54 kcal/mol), p79 (-87.39 kcal/mol), and p78 (-21.21 kcal/mol), reinforcing these peptides as potentially immunogenic epitopes. The 3D representation regarding the molecular interactions established by the peptide p74 (more negative energy value; RMSD = 0.05 Å) in the complex *HLA-DRB1*:p74:T<sub>CR</sub> is shown in Figure 5, using Discovery Studio Visualizer software (47).

The peptide p74 [CFKVQDYWILNDIMK] establishes hydrogen bond interactions with the following residues into the *HLA-DRB1* binding site: Gln9, Phe51, Ser53, Asp57, Trp61, Asn62, Asn69, Arg71, Lys75, Arg76, His81, Asn82, and Val85 (the interatomic distances are lower than 3 Å). Furthermore, hydrogen bond interactions are also established with the Glu94 and Gly98 residues in the T<sub>CR</sub> lymphocyte binding site.

**Table 5.** Findings from molecular docking simulations regarding the complexes formed by the selected sequence fragments (15 aa residues) and *HLA-DRB1*

Peptide	Sequence	Total score* (kcal.mol <sup>-1</sup> )	RMSD (Å)	Hydrogen bond score (kcal.mol <sup>-1</sup> )	Steric interaction score (kcal.mol <sup>-1</sup> )	Ligand conformation penalty (kcal.mol <sup>-1</sup> )
L1FYT	PKYVKQNTLKLAT	-140.84	0.11	-37.42	-142.23	38.81
p79	DYWILNDIMKKNLTG	<b>-108.24</b>	0.11	-42.23	-121.37	55.36
p78	QDYWILNDIMKKNLT	<b>-102.78</b>	0.07	-36.09	-124.03	57.34
p19	CSNKTEIRWYYNGTA	<b>-91.45</b>	0.05	-30.91	-115.83	55.29
p74	CFKVQDYWILNDIMK	<b>-71.72</b>	0.03	-36.7	-125.37	90.36
p80	YWILNDIMKKNLTGI	<b>-55.42</b>	0.18	-30.94	-106.11	81.63
p77	VQDYWILNDIMKKNL	<b>-54.69</b>	0.16	-39.23	-96.39	80.93
p10	LPKLAGDETCSENKTE	<b>-45.62</b>	0.04	-37.48	-111.1	102.96
p22	KTEIRWYYNGTACEA	<b>-37.24</b>	0.93	-35.45	-56.89	55.09
p20	SNKTEIRWYYNGTAC	<b>-21.18</b>	0.03	-57.12	-9.03	44.97
p76	KVQDYWILNDIMKKN	<b>-19.04</b>	0.28	-31.12	-75.02	87.11
p23	TEIRWYYNGTACEAF	3.06	26.65	-7.88	-40.63	51.57
p1	MANSKAVCNLPKLAG	5.82	0.14	-31.32	-90.8	127.95
p2	ANSKAVCNLPKLAGD	11.89	0.23	-29.87	-57.23	98.99
p61	EQTHHFESPCLICF	85.7	0.18	-19.29	-36.49	141.47
p59	CEEQTHHFESPCLI	99.29	1.59	-23.99	-56.93	180.21
p4	SKAVCNLPKLAGDET	128.11	0.84	-8.83	-46.63	183.57
p62	QTHHFESPCLICFK	171.57	0.16	-13.95	-33.31	218.83
p69	SPKLCFKVQDYWIL	315.28	5.66	-7.97	-42.79	366.05
p26	RWYYNGTACEAFIFK	757.73	5.05	-9.67	-36.1	803.5

\* Total score corresponds to the sum of contributions: hydrogen bond, steric interaction, and ligand conformation penalty.



**Figure 5.** Three-dimensional representation of the molecular interactions established by the peptide p74 (stick model) in the complex *HLA-DR1*:p74:T<sub>CR</sub> (A). Molecular docking of p74 in the *HLA-DR1* binding site (B); solvent accessibility surface around the peptide p74 and hydrogen bond interactions established with the residues of T<sub>CR</sub> chains (C). The loop regions Ser25 to Pro30 and Glu94 to Lys103 of T<sub>CR</sub> chain-D are highlighted in blue and loop regions of chain-E (Gln25 to Glu30; Tyr50 to Met54; and, Ser94 to Pro100) are shown in pink (Discovery Studio Visualizer, version 4.5, 2015) (47).

**Table 6.** Findings from molecular docking simulations regarding the nineteen sequence fragments (15 aa residues) as ligands in the whole complex *HLA-DRB1*:ligand: $T_{CR}$  lymphocyte

Peptide	Sequence	Total score* (kcal.mol <sup>-1</sup> )	RMSD (Å)	Hydrogen bond score (kcal.mol <sup>-1</sup> )	Steric interaction score (kcal.mol <sup>-1</sup> )	Ligand conformation penalty (kcal.mol <sup>-1</sup> )
L1FYT	PKYVKQNTLKLAT	-171.78	0.09	-53.85	-156.74	38.81
p74	CFKVQDYWILNDIMK	<b>-147.54</b>	0.05	-40.89	-162.63	55.98
p79	DYWILNDIMKKNLTG	<b>-87.39</b>	0.23	-39.76	-102.99	55.36
p78	QDYWILNDIMKKNLT	<b>-21.21</b>	0.36	-34.25	-44.3	57.34
p76	KVQDYWILNDIMKKN	10.84	0.42	-36.60	-39.66	87.11
p10	LPKLAGDETC SNKTE	21.91	0.21	-43.79	-37.27	102.96
p77	VQDYWILNDIMKKNL	55.01	0.48	-44.1	18.18	80.93
p80	YWILNDIMKKNLTGI	71.3	0.62	-37.74	27.40	81.63
p2	ANSKAVCNLPKLAGD	78.75	<b>20.44</b>	-10.27	-9.97	98.99
p19	CSNKTEIRWYYNGTA	149.97	0.91	-22.03	116.71	55.29
p4	SKAVCNLPKLAGDET	184.35	<b>28.63</b>	-13.24	14.01	183.57
p1	MANSKAVCNLPKLAG	191.1	0.6	-38.07	101.22	127.95
p23	TEIRWYYNGTACEAF	203.91	0.38	-34.07	186.42	51.57
p22	KTEIRWYYNGTACEA	243.94	1.12	-29.21	218.06	55.09
p62	QTHHFHESPCLICFK	246.27	<b>20.66</b>	-10.00	37.45	218.83
p26	RWYYNGTACEAFIFK	246.46	1.76	-24.56	160.65	110.36
p20	SNKTEIRWYYNGTAC	251.74	0.43	-52.41	259.18	44.97
p59	CEEQTHHFHESPCLI	292.73	0.79	-35.03	147.54	180.21
p61	EQTHHFHESPCLICF	489.42	<b>20.61</b>	-19.58	367.52	141.47
p69	SPKLICFKVQDYWIL	600.33	1.99	-10.24	244.52	366.05
p4	SKAVCNLPKLAGDET	184.35	<b>28.63</b>	-13.24	14.01	183.57

\* Total score corresponds to the sum of contributions: hydrogen bond, steric interaction, and ligand conformation penalty.

The major  $T_{CR}$  lymphocyte amino acid residues involved in the molecular recognition process (Glu30, Gly98, Ser96) interact with ligands at position 11 (40). Regarding position 11, the peptide p74 has established interaction only with the residue Gly98. Besides that, hydrogen bonding interaction at position 3 was observed involving the Glu94 residue from the  $T_{CR}$  binding site. The interactions established between p74 and *HLA-DRI*: $T_{CR}$  lymphocyte are shown in Figure 6.

The peptide p69, which presented unfavorable total energy score (600.33 kcal/mol, more positive value; RMSD value = 1.99 Å), interacts with the Ser96 residue at position 11 and Gly98 at position 10. It also established interactions at position 8 with the Ser95, Glu102, and Asn99 residues into the  $T_{CR}$  lymphocyte binding site. The intermolecular interactions established by the peptides p74 and p69 into the *HLA-DRI* and  $T_{CR}$  lymphocyte binding sites, concerning the aa residues and water molecules, are summarized in Table 7. Of note, the peptide p69 establishes intermolecular interactions with only three residues into the *HLA-DRI* protein (Ser53, Arg71 and Tyr60), in contrast with the thirteen

residues involved in the binding of p74 (more favorable total energy score).

## FINAL REMARKS

The combined computer-aided protocol described herein has significantly reduced the number of fragments or motifs to be considered as potentially promising epitopes concerning the Kunitz-type inhibitor used as example. Also, the fragment sequences were classified as having moderate immunogenic activity. Four potentially promising epitopes (p1, p100, p69 and p86) were found for MHC-I and three (p74, p79 and p78) for MHC-II. It is noteworthy that the three epitopes predicted for MHC-II using the combined computer-aided approach were among the sequences experimentally assayed and classified as potential epitopes (preliminary data; not published). However, more experimental assays should be conducted to fully validate the combined computer-aided protocol. Since the protocol comprises the 3D structural information related not only to the MCH:ligand complex, but also to the MHC:ligand: $T_{CR}$

lymphocyte complex, the prediction power from this combined approach seems to be improved. The assessment of ligand-lymphocyte interaction may allow to correlate the MHC ligand's binding affinity as well as the ligand's chance to be presented to TCR (immunogenic activity).

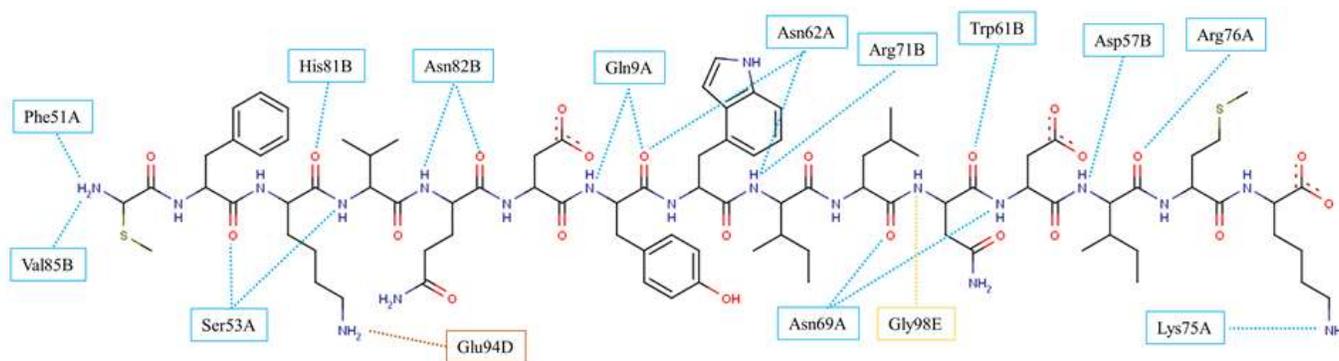
Then, regarding the early stage of protein-based drugs development, for instance, the combined computer-approach could be applied to drive the selection of more promising epitopes to be synthesized and experimentally evaluated, as well as to find better constructions (through molecular modifications) derived to the original protein sequence to avoid or reduce the appearance of ADA.

## ACKNOWLEDGEMENTS

B.A.V.G. da Silva and K.F.M. Pasqualoto contributed equally to this work. Silva, B. A. V. G. thanks BNDES - Banco Nacional de Desenvolvimento [# 13.2.0711.1, 2013] for her scholarship.

## CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.



**Figure 6.** Interactions established by the peptide p74 into the *HLA-DR1*:p74:TCR lymphocyte complex. Amino acid residues from the *HLA-DR1* binding site are highlighted in blue and from the TCR lymphocyte binding site in yellow. Additional interactions with TCR are highlighted in orange (CLC Drug Discovery Workbench software, version 4.5, 2014) (41).

**Table 7.** Summary of the aa residues which participate in intermolecular interactions established by the peptides p74 and p69 into the *HLA-DR1* and TCR lymphocyte binding sites

aa position in the peptides' sequence	<i>HLA-DR1</i> :p74:TCR complex	<i>HLA-DR1</i> :p69:TCR complex	aa position in the peptides' sequence	<i>HLA-DR1</i> :p74:TCR complex	<i>HLA-DR1</i> :p69:TCR complex
1	Phe51	—	9	Asn62	—
2	Val85	—	10	Arg71	—
3	Ser53	Ser53	11	Asn69	Gly98*
4	His81	H <sub>2</sub> O 228	12	Trp61	Ser96*
5	Glu94*	—	13	Gly98*	Tyr60
6	Ser53	—	14	Asn69	Tyr60
7	Asn82	H <sub>2</sub> O 228	15	Asp57	—
8	—	—	—	Arg76	—
		Ser95*		—	—
		Glu102*		Lys75	—
		Asn99*		—	—
		Arg71			

\* Amino acid residue in the TCR binding site.

## REFERENCES

1. De Groot AS, Scott DW. Immunogenicity of Protein Therapeutics. *Trends in Immunol*, 2007; 28: 482–90. Doi:10.1016/j.it.2007.07.011.
2. Jawa V, Cousens LP, Awwad M, Wakshull E, Kropshofer H, De Groot AS. T-Cell Dependent Immunogenicity of Protein Therapeutics: Preclinical Assessment and Mitigation. *Clin Immunol*, 2013; 149: 534–55. Doi:10.1016/j.clim.2013.09.006.
3. Barbosa MDF, Celis E. Immunogenicity of Protein Therapeutics and the Interplay between Tolerance and Antibody Responses. *Drug Discov Today*, 2007; 12: 674-81. Doi:10.1016/j.drudis.2007.06.005.
4. Krishna M, Nadler SG. Immunogenicity to Biotherapeutics - The Role of Anti-Drug Immune Complexes. *Front Immunol*, 2016; 7: 21. Doi:10.3389/fimmu.2016.00021.
5. Koren E, Zuckerman LA, Mire-Sluis AR. Immune Responses to Therapeutic Proteins in Humans - Clinical Significance, Assessment and Prediction. *Curr Pharm Biotechnol*, 2002; 3: 349-60. Doi: 10.2174/1389201023378175.
6. Sette A, Vitiello A, Rehman B, Fowler P, Nayarsina R, Kast WM, Melief CJ, Oseroff C, Yuan L, Ruppert J, Sidney J, del Guercio MF, Southwood S, Kubo RT, Chesnut RW, Grey HM, Chisari FV. The Relationship between Class I Binding Affinity and Immunogenicity of Potential Cytotoxic T Cell Epitopes. *J Immunol*, 1994; 153 (12): 5586–92. Doi:153:5586-92.
7. Lundegaard C, Lund O, Nielsen M (2012) Predictions versus High-Throughput Experiments in T-Cell Epitope Discovery: Competition or Synergy? *Expert Rev Vaccines* 11 (1): 43–54. Doi:10.1586/erv.11.160.
8. Neefjes J, Jongstra ML, Paul P, Bakke, O. Towards a Systems Understanding of MHC Class I and MHC Class II Antigen Presentation. *Nat Rev Immunol*, 2001; 11 (12): 823–36. Doi:10.1038/nri3084.
9. Roche PA, Furuta K. The Ins and Outs of MHC Class II-Mediated Antigen Processing and Presentation. *Nat Rev Immunol*, 2015; 15 (4): 203–16. Doi:10.1038/nri3818.
10. Grommé M, Neefjes J. Antigen Degradation or Presentation by MHC Class I Molecules via Classical and Non-Classical Pathways. *Mol Immunol*, 2002; 39: 181–202. Doi:10.1016/S0161-5890(02)00101-3.
11. Lafuente E, Reche P. Prediction of MHC-Peptide Binding: A Systematic and Comprehensive Overview. *Curr Pharm Des*, 2009; 15 (28): 3209–20. Doi:10.2174/138161209789105162.
12. De Groot AS, Moise L. Prediction of Immunogenicity for Therapeutic Proteins: State of the Art. *Curr Opin Drug Discov Devel*, 2007; 10 (3): 332–40.
13. Vita R, Overton JA, Greenbaum JA, Ponomarenko J, Clark JD, Cantrell JR, et al. The Immune Epitope Database (IEDB) 3.0. *Nucleic Acids Res*, 2015; 43 (D1): D405–12. Doi:10.1093/nar/gku938.
14. Rammensee HG, Bachmann J, Emmerich NPN, Bachor OA, Stevanović S. SYFPEITHI: Database for MHC Ligands and Peptide Motifs. *Immunogenetics*, 1999; 50 (3–4): 213–19. Doi:10.1007/s002510050595.
15. Kim Y, Sette A, Peters B. Applications for T-Cell Epitope Queries and Tools in the Immune Epitope Database and Analysis Resource. *J Immunol Methods*, 2011; 374 (1–2): 62–9. Doi:10.1016/j.jim.2010.10.010.
16. Hansson L, Rabbani H, Fagerberg J, Österborg A, Mellstedt H. T-Cell Epitopes within the Complementarity-Determining and Framework Regions of the Tumor-Derived Immunoglobulin Heavy Chain in Multiple Myeloma. *Blood*, 2003; 101 (12): 4930–36. Doi:10.1182/blood-2002-04-1250.
17. Wang RF, Rosenberg SA. Human Tumor Antigens for Cancer Vaccine Development. *Immunol Rev*, 1999; 170: 85–100. Doi:10.1111/j.1600-065X.1999.tb01331.x.
18. Berzofsky JA, Ahlers JD, Derby MA. Approaches to improve engineered vaccines for human immunodeficiency virus and other viruses that cause chronic infections. *Immunol Rev*, 1999; 170: 151–72. Doi:10.1111/j.1600-065X.1999.tb01336.x.
19. Pasqualoto KFM, Balan A, Barreto SA, Simons SM, Chudzinski-Tavassi AM. Structural Findings and Molecular Modeling Approach of a TFPI-like Inhibitor. *Protein Pept Lett*, 2014; 21 (5): 452–57. Doi:10.2174/0929866520666131210115334.
20. Batista IFC, Ramos OHP, Ventura JS, Junqueira-de-Azevedo ILM, Ho PL, Chudzinski-Tavassi AM. A New Factor Xa Inhibitor from *Amblyomma cajennense* with a Unique Domain Composition. *Arch Biochem Biophys*, 2010; 493 (2): 151–56. Doi:10.1016/j.abb.2009.10.009.
21. Chudzinski-Tavassi AM, De-Sá-Júnior PL, Simons SM, Maria DA, de Souza Ventura J, Batista IFC, et al. A New Tick Kunitz Type Inhibitor, *Amblyomin-X*, Induces Tumor Cell Death by Modulating Genes Related to the Cell Cycle and Targeting the Ubiquitin-Proteasome System. *Toxicol*, 2010; 56 (7): 1145–54. Doi:10.1016/j.toxicol.2010.04.019.
22. Buchanan A, Revell JD. Novel therapeutic proteins and peptides in novel approaches and strategies for biologics, vaccines and cancer therapies. In: Singh M, Salnikova M, editors. *Novel therapeutics proteins and peptides*. 1st ed. Academic Press, Elsevier Inc., 2015, pp. 171-197.
23. Ranasinghe SL, Fischer K, Gobert GN, McManus DP. Functional expression of a novel Kunitz type protease inhibitor from human blood fluke *Schistosoma mansoni*. *Parasit Vectors*, 2015; 8:408. Doi: 10.1186/S13071-015-1022-z.
24. Batista IFC, Chudzinski-Tavassi AM, Faria F, Simons SM, Barros-Batestti DM, Labruna MB, et al.

- Expressed sequence tags (ESTs) from the salivary glands of the tick *Amblyomma cajennense* (Acari: Ixodidae). *Toxicon*, 2008; 51: 823-34. <https://doi.org/10.1016/j.toxicon.2007.12.011>.
25. Zhang GL, Ansari HR, Bradley P, Cawley GC, Hertz T, Hu X, et al. (Editorial) Machine learning competition in immunology – Prediction of HLA class I binding peptides. *J Immunol Methods*, 2011; 374: 1–4. Doi: 10.1016/j.jim.2011.09.010.
  26. Salimi N, Fleri W, Peters B, Sette A. Design and utilization of epitope-based databases and predictive tools. *Immunogenetics*, 2010; 62: 185–96. Doi: 10.1007/s00251-010-0435-2.
  27. Lin H, Ray S, Tongchusak S, Reinherz EL, Brusica V. Evaluation of MHC Class I Peptide Binding Prediction Servers: Applications for Vaccine Research. *BMC Immunol*, 2008; 9 (1): 8. Doi:10.1186/1471-2172-9-8.
  28. Gonzalez-Galarza FF, Christmas S, Middleton D, Jones AR. Allele Frequency Net: A Database and Online Repository for Immune Gene Frequencies in Worldwide Populations. *Nucleic Acids Res*, 2011; 39 suppl 1. Doi:10.1093/nar/gkq1128.
  29. Immune Epitope Database and Analysis Resource. IEDB Solutions Center Website. Selecting thresholds (cut-offs) for MHC class I and II binding predictions. 2013. <http://help.iedb.org/entries/23854373-Selecting-thresholds-cut-offs-for-MHC-class-I-and-II-binding-predictions>, 2016 (accessed 06.08.16).
  30. Vita R, Zarebski L, Greenbaum JA, Emami H, Hoof I, Salimi N, et al. The immune epitope database 2.0. *Nucleic Acids Res*, 2010; 38: 854–862. Doi: 10.1093/nar/gkp1004.
  31. Savage DA, Middleton D, Trainor F, Taylor A, McKenna PG, Darke C. Frequency of HLA-DPB1 alleles, including a novel DPB1 sequence, in the Northern Ireland population. *Hum Immunol*, 1992; 33: 235–242. [https://doi.org/10.1016/0198-8859\(92\)90330-P](https://doi.org/10.1016/0198-8859(92)90330-P).
  32. Solberg OD, Mack SJ, Lancaster AK, Single RM, Tsai Y, Sanchez-Mazas A, Thomson G. Balancing selection and heterogeneity across the classical human leukocyte antigen loci: a meta-analytic review of 497 population studies. *Hum Immunol*, 2008; 69: 443–64. Doi:10.1016/j.humimm.2008.05.001.
  33. Hoof I, Peters B, Sidney J, Pedersen LE, Sette A, Lund O, Buus S, Nielsen M. NetMHCpan, a Method for MHC Class I Binding Prediction beyond Humans. *Immunogenetics*, 2009; 61 (1): 1–13. Doi:10.1007/s00251-008-0341-z.
  34. Paul S, Weiskopf D, Angelo MA, Sidney J, Peters B, Sette A. HLA Class I Alleles Are Associated with Peptide-Binding Repertoires of Different Size, Affinity, and Immunogenicity. *J Immunol*, 2013; 191 (12): 5831–39. Doi:10.4049/jimmunol.1302101.
  35. Nielsen M, Lundegaard C, Lund O. Prediction of MHC Class II Binding Affinity Using SMM-Align, a Novel Stabilization Matrix Alignment Method. *BMC Bioinformatics*, 2007; 8 (1): 238. Doi:10.1186/1471-2105-8-238.
  36. Tenzer S, Peters B, Bulik S, Schoor O, Lemmel C, Schatz MM, Kloetzel PM, Rammensee HG, Schild H, Holzhütter HG. Modeling the MHC class I pathway by combining predictions of proteasomal cleavage, TAP transport and MHC class I binding. *Cell Mol Life Sci*, 2005; 62: 1025–37. Doi:10.1007/s00018-005-4528-2.
  37. Pagadala NS, Syed K, Tuszynski J. Software for molecular docking: a review. *Biophys Rev*, 2017; 9(2): 91-102. Doi:10.1007/s12551-016-0247-1.
  38. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov I N, Bourne PE. The Protein Data Bank. *Nucleic Acids Res*, 2000; 28 (1): 235–42. Doi:10.1093/nar/28.1.235.
  39. Stewart-Jones GB, McMichael AJ, Bell JI, Stuart DI, Jones EY. A Structural Basis for Immunodominant Human T Cell Receptor Recognition. *Nat Immunol*, 2003; 4 (7): 657–63. Doi:10.1038/ni942.
  40. Hennecke J. Structure of a Covalently Stabilized Complex of a Human Alphabeta T-Cell Receptor, Influenza HA Peptide and MHC Class II Molecule, HLA-DR1. *EMBO J*, 2000; 19 (21): 5611–24. Doi:10.1093/emboj/19.21.5611.
  41. QIAGEN Bioinformatics, CLC Drug Discovery Workbench Software®, version 2.4. QIAGEN Aarhus A/S. 2014.
  42. Nelder JA, Mead R. A Simplex Method for Function Minimization. *Comput J*, 1965; 7 (4): 308–13. Doi:10.1093/comjnl/7.4.308.
  43. Dadrass OG, Sobhani AM, Shafiee A, Mahmoudian M. Flexible Ligand Docking Studies of Matrix Metalloproteinase Inhibitors Using Lamarckian Genetic Algorithm. *DARU*, 2004; 12: 1–10. <http://daru.tums.ac.ir/index.php/daru/article/view/191>.
  44. Korb O, Stütze T, Exner TE. Empirical scoring functions for advanced protein-ligand docking with plants. *J Chem Inf Model*, 2009; 49: 84–96. Doi:10.1021/ci800298z.
  45. Serwold T, Gonzalez F, Kim J, Jacob R, Shastri N. ERAAP customizes peptides for MHC class I molecules in the endoplasmic reticulum. *Nature*, 2002; 419: 480–83. Doi:10.1038/nature01074.
  46. Bjorkman PJ, Saper MA, Samraoui B, Bennett WS, Strominger JL, Wiley DC. The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. *Nature*, 1987; 329: 512–18. Doi:10.1038/329512a0.
  47. Dassault Systèmes BIOVIA, Discovery Studio Visualizer, version 4.5. San Diego: Dassault Systèmes, 2015.
  48. Peters B, Sette A. Generating Quantitative Models Describing the Sequence Specificity of Biological Processes with the Stabilized Matrix Method. *BMC Bioinformatics*, 2005; 6 (1): 132. Doi:10.1186/1471-2105-6-132.

49. Wang P, Sidney J, Dow C, Mothé B, Sette A, Peters B. A Systematic Assessment of MHC Class II Peptide Binding Predictions and Evaluation of a Consensus Approach. *PLoS Comput Biol*, 2008; 4(4): e1000048. Doi:10.1371/journal.pcbi.1000048.
50. Dhanik A, McMurray JS, Kaviraki LE. DINC: A New AutoDock-Based Protocol for Docking Large Ligands. *BMC Struct Biol*, 2013; 13 (1): S11. Doi:10.1186/1472-6807-13-S1-S11.
51. Elokely KM, Doerksen RJ. Docking Challenge: Protein Sampling and Molecular Docking Performance. *J Chem Inf Model*, 2013; 53 (8): 1934–45. Doi:10.1021/ci400040d.