Forecast of cytotoxic T lymphocyte epitope using sequence weighting and artificial neural network based on EasyPred modeler

Satarudra Prakash Singh[1*] , Garima Singh[2], Bhartendu Nath Mishra[3]

[1]Department of Biotechnology, Mahatma Gandhi Central University, Motihari 845401, Bihar, India
[2]Amity Institute of Biotechnology, Lucknow Campus, Amity University, Lucknow 226028, Uttar Pradesh, India
[3]Institute of Engineering and Technology, Dr. A.P.J. Abdul Kalam Technical University, Lucknow 226021, Uttar Pradesh, India

*Correspondence: Satarudra Prakash Singh, Department of Biotechnology, Mahatma Gandhi Central University, Motihari 845401, Bihar, India. sprakashsingh@mgcub.ac.in

Abstract

Aim: Cytotoxic T lymphocytes (CTL) examine the major histocompatibility complex (MHC) class I ligands on nucleated cells to detect antigens derived from pathogens and cancer cells. Accurate prediction of T-cell epitopes is therefore crucial for the development of a wide range of biopharmaceuticals, including vaccines.

Methods: The present study involved the development of position-specific scoring matrices (PSSM) and artificial neural networks (ANN) based models for 22 MHC class I molecules, including the integrated forecast of CTL epitopes using the EasyPred modeler. Similarity-reduced peptides dataset was used to train and evaluate models with performance assessed using the area under the receiver operating characteristic curve (Aroc) as the primary metric.

Results: Comparative analysis revealed that the ANN-based predictor achieved superior performance for the HLA-A*0202 molecule by achieving the maximum Aroc value of 0.97 as compared to the PSSM predictor, having a value of 0.93. Furthermore, most natural MHC binders were identified within the top 5% with an average relative rank (%) of 2.23 and 3.13 for predictors PSSM and ANN, respectively, on the NetCTLpan dataset. Likewise, evaluation on the SARS-CoV-2 dataset of HLA-A*0201 revealed that the PSSM predictor (2.46%) performed better than the other contemporary CTL epitope forecast methods like naturally eluted ligands (EL) of NetMHCpan 4.0 (2.66%), NetCTLpan 1.1 (2.69%), and binding affinity (BA) of NetMHCpan 4.0 (3.33%), respectively.

Conclusions: The application of these predictive models offers a significant reduction of approximately 97% in the resources typically required for epitope identification, including costs related to materials, labor, and time. As such, these models represent a valuable advancement in the rational design of more efficient, cost-effective, and innovative biotherapeutics.

Keywords

epitope, MHC, machine learning, vaccine, forecast

Introduction

In the current trend, epitope-based bio-therapeutic products such as prophylactic vaccines and antibodies are critical in health care [1]. Specifically, vaccines based on the T-cell epitope are the most important method of triggering cellular immune response and clearing intracellular infections [2]. The responsibility of carrying out this task is specifically assigned to cytotoxic T lymphocytes (CTL) [3]. A crucial role of CTL is their ability to induce apoptosis of infected/altered cells, facilitated by helper T lymphocytes (HTL) that produce cytokines to modulate the activity of immune response cells [4]. The process of antigen processing and presentation, which involves the recognition of CTL epitopes, encompasses three essential stages: the proteasomal degradation of an antigen, the translocation of the antigenic fragment via the transporters associated with antigen processing (TAP) transport system, and the binding of the fragment to major histocompatibility complex (MHC) class I molecules [5]. The proteasome is responsible for producing the C-terminus of peptides [6]. A portion of these peptides is capable of being transported into the endoplasmic reticulum (ER) via the TAP, where subsequent N-terminal trimming takes place [7, 8], resulting in peptides of suitable length (approximately 8 to 10 residues) which can eventually bind to MHC class I molecules [9]. The resulting MHC class I-peptide complexes are subsequently translocated to the cell surface, which can be recognized by epitope-specific CTL receptors [10]. The immunodominance of a peptide is largely influenced by its capability to interact with an MHC class I molecule [11]. Therefore, the development of an assay to evaluate peptides binding to MHC class I molecules is essential for accurately identifying CTL epitopes [12]. However, CTL epitope mapping for the prominent proteome pathogen, for instance, *Plasmodium falciparum* (~5,300 proteins), is expensive and arduous [13]. Consequently, researchers are presently employing contemporary computational models to forecast effective epitopes that can trigger specific immune responses [14, 15].

Formerly, numerous in silico tools and methods have been developed for predicting binding peptides to MHC class I and/or CTL epitopes based upon data retrieved from several immunological databases such as IEDB [16–18] and other repositories [19, 20]. Several research groups have created systematic and quantitative benchmarks to assess the forecasting performance of binding peptides to MHC class I molecules [21, 22]. Despite many available computational methods, forecasting CTL epitopes remains challenging to date [23]. Given that the majority of existing methods rely on training data, we can anticipate remarkable models solely by acquiring exceptional and current general data [24]. Therefore, a crucial step in creating a more precise and dependable forecasting model is the collection of experimentally reliable training and validation datasets.

The present study involved an extensive collection of MHC class I binding and non-binding peptides available at the Repository for Epitope Datasets (RED) and subsequent generation of position-specific scoring matrices (PSSM) [25, 26] and artificial neural networks (ANN) [27] models for predicting MHC class I binding peptides. Subsequently, a conclusive model for integrated prediction of CTL epitopes was developed utilizing experimentally validated weight matrices for TAP binding affinity (BA) and/or forecasts of constitutive and immuno-proteasomal cleavage as a filter [28]. The CTL epitopes forecasted by the present model are likely to have a C-terminus cleavage generated by the proteasome (constitutive and immunoproteasome), a moderate BA for TAP, and a high BA for a specific MHC class I molecule, which can be used in designing a universal vaccine that ultimately regulates the efficient HLA cross-presentation [29, 30].

Materials and methods

Data retrieval of MHC class I binding and nonbinding peptides

A dataset of MHC class I binding and nonbinding peptides, with reduced similarity, was compiled for 22 molecules from the supplementary data provided by the MHCIPREDS web server, which is accessible at RED (https://web.archive.org/web/20130828192234/http://ailab.cs.iastate.edu/red/). The log-transformed MHC BA (LTMBA) values between 0 (no affinity) and 1 (very high affinity) were obtained using the relation, 1 – log (aff)/log (50,000), where aff is the experimentally measured BA in terms of half-

maximal inhibitory concentration (IC_{50}) with nM (nmol/L) unit. An LTMBA threshold value of 0.426, equivalent to an IC_{50} value of 500 nM, was used to classify binders and non-binders, which means peptides with LTMBA values ≥ 0.426 (IC_{50} ≤ 500 nM) were classified as binders (Table 1).

Table 1. MHC class I binding and nonbinding peptide dataset used in the study for training and evaluation of algorithms.

S. No.	MHC molecule	Training set		Evaluation set		
		Total no. of data	No. of binding data#	Total no. of data	No. of binding data#	No. of nonbinding data
1	HLA-A*0201	2,280	1,155	560	280	280
2	HLA-A*0203	1,080	555	283	141	142
3	HLA-A*0206	976	478	244	122	122
4	HLA-A*2902	266	134	66	33	33
5	HLA-A*6801	919	458	229	114	115
6	HLA-A*6802	678	334	168	84	84
7	HLA-A*3301	324	162	81	40	41
8	HLA-A*0101	258	128	65	32	33
9	HLA-A*0202	1,082	550	281	140	141
10	HLA-A*3002	232	115	58	29	29
11	HLA-A*3101	811	405	203	101	102
12	HLA-A*0301	1,006	503	250	125	125
13	HLA-A*1101	1,293	643	324	162	162
14	HLA-A*2402	317	157	80	40	40
15	HLA-B*0702	368	182	92	46	46
16	HLA-B*1501	296	148	72	36	36
17	HLA-B*3501	351	176	87	43	44
18	HLA-B*4002	111	56	28	14	14
19	HLA-B*4501	108	55	26	13	13
20	HLA-B*5101	161	78	40	20	20
21	HLA-B*5301	174	87	42	21	21
22	HLA-B*5401	147	73	36	18	18

MHC: major histocompatibility complex; #: no. of peptides with an IC_{50} ≤ 500 nM.

Similar peptides were removed to generate a similarity-reduced cross-validation dataset for each MHC allele, and the data were randomly partitioned. The present study used a 5-fold cross-validation approach, i.e., the dataset was split into five subsets. Initially, the dataset was scanned for equal binders and non-binders based on LTMBA value ≥ 0.426 for each 22 MHC class I allele. Then, four out of five (4/5) of each dataset were used for the training, and 1/5 data points (peptides) were used for the blind evaluation (because none of the peptides in the evaluation set were included in the training set at any stage) as described by Nielsen et al. [31] (Supplementary file 1 in https://data.mendeley.com/datasets/dxz3dk3tcm/1).

Construction of MHC class I binding motifs (PSSM)

Initially, based on a training dataset of fixed length (9-mer) MHC class I binding peptide sequences (positive) and using a statistical model of sequence weighting methods available at EasyPred modeler (https://services.healthtech.dtu.dk/services/EasyPred-1.0/), 22 PSSM (9 × 20) were constructed, which represents the frequencies of residues observed for a position in a multiple sequence alignment (MSA) assuming no correlations exist between the different peptide positions [32]. The score S of a peptide to a motif is usually calculated as the sum of the log-odds ratio.

$$S = \log_k\left(\prod_p \frac{P_{pa}}{q_a}\right) = \sum_p \log_k\left(\frac{P_{pa}}{q_a}\right) \tag{1}$$

Where p_{pa} is the probability of finding amino acid a (a can be any of the 20 amino acids) at position p (p can be 1 to 9) in the motif, and q_a is the background frequency of amino acid a, $\log_k$ is the logarithm with base k. The scores are often normalized to half-bit by multiplying all scores by $2/\log_k$ (equation 2). In half-bit units, the log-odds score S is calculated as:

$$S = 2\sum_P \log_2\left(\frac{P_{pa}}{q_a}\right) \qquad (2)$$

The three sequence weighting methods, namely Henikoff & Henikoff 1/nr [26], Hobohm [25] clustering at 62% identity, and no clustering, were used to compensate for the over-representation among MSA, along with four different weights on pseudo counts (50, 100, 150, and 200) [33–35]. In the Henikoff & Henikoff 1/nr method [26], an amino acid a on position p in sequence k contributes a weight w_{kp} = 1/nr, where n is the number of different amino acids at a given position (column) in the alignment and r is the number of occurrences of amino acid a in that column. The weight of a sequence is then assigned as the sum of the weights over all positions in the alignment. However, in the Hobohm 62% clustering method [25], each peptide k in a cluster is assigned a weight w_k = 1/n_c, where n_c is the number of sequences in the cluster containing peptide k. When the amino acid frequencies are calculated, each amino acid in sequence k is weighted by w_k. The Henikoff & Henikoff method is as fast as the computation time increases linearly with the number of sequences. In contrast, in the Hobohm clustering algorithm, computation time increases as the square of the number of sequences. Additionally, the binding potential (score) of any peptide sequence (query) to a specific MHC allele was determined by aligning the corresponding PSSM with the peptide sequence and summing the scores that correspond to the residue type and position in the PSSM. To narrow down the probable binders from the list of scored and ranked peptides, a binding cut-off score was established that encompasses 85% of the peptide sequences in the training dataset (positive data) [28, 36].

Training of ANN for the forecast of MHC class I binding peptides

An ANN is a computer simulation of a system of interconnected processing units that can be trained to extract and remember a pattern present within a data set. It can subsequently recognize that pattern when presented with new data. In binding a peptide to the MHC molecule, the amino acids might compete for the space available in the binding groove. Therefore, the mutual information in the binding motif will allow the identification of higher-order sequence correlations. Neural networks with hidden layers are used to describe sequence patterns with such higher-order correlations (https://teaching.healthtech.dtu.dk/mpmbioinformatics/22801_04.pdf). Thus, the ANN simulations were performed for all 22 MHC class I molecules using the neural network options available in the EasyPred modeler [27]. Based on the other studies of MHC-binding peptides forecast, two ANN architectures, namely 180-1-1 and 180-2-1, were considered in the current study, which represent respective numbers of neurons in the input, hidden, and output layers, respectively [37, 38]. In these architectures, the input 9-mer peptide sequences were presented with 180 values (where individual amino acid was encoded as a binary string of length 20 with an exclusive position set to 0.9 and other positions set to 0.05). At the same time, the output from each neuron was transformed using the standard sigmoid function [39]. The training algorithm employed to generate the final network was the steepest descent method that learns from a training set of input-output pairs by modifying the network weight parameters such that the network generates a numerical value for each input close to the desired target output. Throughout the training process, the LTMBA value of a peptide served as a target value and was deliberated as a strong binder and a potent epitope for MHC class I alleles [40]. Nevertheless, the learning process utilized was error backpropagation [41]. In addition, the following parameters were used for ANN simulations: i) The top 80% of the training set was used to train the neural network, and the bottom 20% was used to stop the training to avoid fitting; ii) The learning rate was 0.005; iii) The maximum number of iterations was set to 300.

Evaluation parameters

The nonparametric performance measure, the area under the receiver operating characteristic curve (Aroc) and correlation coefficient (CC) values, was used to evaluate the predictive performance of the EasyPred modeler-based PSSM and ANN predictors. The receiver operator characteristic (ROC) curve is a plot of the

true positive rate [TP/(TP + FN)] on the Y-axis versus the false positive rate [FP/(TN + FP)] on the X-axis for the entire range of the decision thresholds where, true positive (TP) is an experimentally proven binding peptide forecasted as a binder, false positive (FP) is an experimentally proven nonbinding peptide forecasted as a binder, true negative (TN) is an experimentally proven nonbinding peptide forecasted as a non-binder and false negative (FN) is an experimentally proven binding peptide forecasted as nonbinder [42, 43]. An Aroc value can be interpreted as the probability of distinguishing a true positive from a false positive. For the calculation of Aroc, peptides were classified into binders and non-binders at a cutoff value of 500 nM. This affinity threshold is associated with the most well-known T-cell epitopes. The values Aroc ≥ 0.90 indicate excellent, 0.90 > Aroc ≥ 0.80 good, 0.80 > Aroc ≥ 0.70 marginal, and Aroc < 0.70 poor predictions [42]. The CC is another widely used measure of the association between pairs of values (predicted versus experimental). It is calculated as:

$$CC = \frac{\sum_i (a_i - \overline{a})(p_i - \overline{p})}{\sqrt{\sum_i (a_i - \overline{a})^2} \sqrt{\sum_i (p_i - \overline{p})^2}} \tag{3}$$

where the value p_i is found using a prediction method of choice, and the a_i is the known corresponding target value. However, the overlined letters denote average values. The value of 1 corresponds to a perfect correlation and –1 to a perfect anti-correlation, and 0 value corresponds to a random prediction [42].

Generalization test and Epstein-Barr virus case study

The generalizability test was used to assess the ability of the PSSM and ANN models, trained for the allele HLA-A*0201, to generalize to the other 21 alleles. Training was performed on the same dataset, followed by testing on evaluation datasets of all 22 alleles. In addition to the above, the performance of the PSSM and ANN predictors was validated through observations revealed in the very recent study conducted by Wohlwend et al. [44] called the Epstein-Barr virus (EBV) dataset. This research indicates that the dataset comprises 11 distinct immunogenic epitopes that are restricted by 5 HLA class I alleles derived from EBV, as identified through IFNγ ELISpot assays. The study employs PSSM and ANN models to effectively identify both confirmed and new HLA class I epitopes from EBV, which have been experimentally validated through in vitro and ex vivo studies.

Prediction and relative rank measurement of PSSM and ANN predictors on NetCTLpan and SARS-CoV-2 dataset

The additional performance evaluation of the present EasyPred modeler-based PSSM and ANN predictors, along with recently published NetMHCpan 4.0 methods [eluted ligands (EL) and BA] (https://services.healthtech.dtu.dk/services/NetMHCpan-4.0/), was estimated on NetCTLpan and SARS-CoV-2 dataset in terms of comparative ranking of the reported ligands among all nonamers included in the source protein as described by Larsen et al. [45, 46]. This NetCTLpan dataset was gathered from the supplementary material of the NetCTLpan1.0 method (https://services.healthtech.dtu.dk/suppl/immunology/NetCTLpan.php) reported as SYFPEITHI 9-mer training and evaluation dataset [47, 48]. The dataset involved 413 sequences of naturally processed T cell epitopes restricted by 7 HLA class I molecules that were common between the present study and the NetCTLpan1.0 tool (Supplementary file 2 in https://data.mendeley.com/datasets/dxz3dk3tcm/1). At the time of conducting the present study, initially, no T cell epitope data were available for SARS-CoV-2, but there was a significant amount of information available on T cell epitopes for betacoronaviruses that cause similar diseases in humans, like SARS-CoV. Thus, we have compiled the SARS-CoV-2 dataset from the study of Grifoni et al. [49] that contains 11 T cell epitopes (9-mer) derived from surface glycoprotein (NCBI ID: QHD43416) and nucleocapsid phosphoprotein (NCBI ID: QHD43423) that showed 100% identity with SARS-CoV epitopes. Additionally, evaluation of PSSM and ANN predictors on the recent SARS-CoV-2 dataset of Gfeller et al. [50] was performed for prediction and % rank analysis of CD8+ T cell epitope in their source protein using allele-specific cutoff score.

Integrated forecast of CTL epitope processing

The CTL epitope processing forecast integrated the three significant forecast steps in the filtering approach, including MHC class I BA, TAP transport efficiency, and C-terminal proteasomal cleavage (Figure 1). The details are described below:
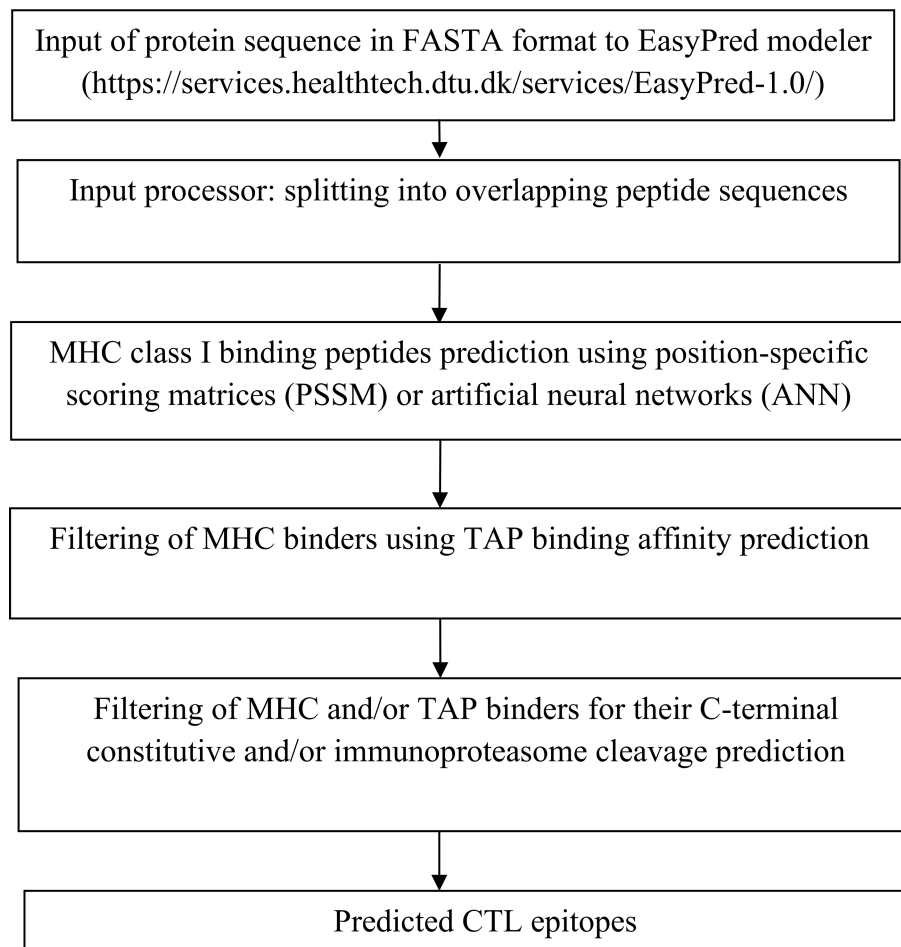
Figure 1. Flow chart indicating the integrated forecast of cytotoxic T lymphocytes (CTL) epitopes based on the EasyPred modeler. MHC: major histocompatibility complex; TAP: transporters associated with antigen processing.

1. Amino acid sequences of a given target protein were parsed into 9-mer overlapping peptides. Peptides were then scored using EasyPred-based PSSM and ANN predictors, and values above the defined threshold (Table 2) for each 22 MHC class I allele were forecasted as binders. In the case of PSSM predictors, we calculated the threshold scores in terms of predicting peptides that bind with 85% of all epitopes in the training set because an established threshold associated with immunogenicity (i.e., $IC_{50} \leq 500$ nM) covers 80–90% of all immunogenic epitopes [51]. However, in the case of ANN predictors, the threshold value of an IC_{50} less than 500 nM was considered for binder forecast.

2. Forecasted MHC class I binders were again scored by the quantitative weight matrix (9 × 20) for TAP BA in terms of $-\log IC_{50}$ (pIC_{50}) described by Peters et al. [52] (Table S1) and/or Doytchinova et al. [53] (Table S2).

3. Peptides with scores less than the selected threshold value (default, $pIC_{50} < 4$) were considered as TAP binders. The TAP transports peptides into the ER that are potentially N-terminally extended from the ligand that ends up in MHC. This means that the peptide that binds to MHC does not necessarily need to be a suitable substrate for TAP [52].

4. Subsequently, the proteasomal (constitutive or immuno) cleavage score of 12-mer overlapping peptide fragments generated from the target protein sequence was calculated along with recording

of C-terminal position at the cleavage site (i.e., position of the sixth amino acid in the target protein) by using a quantitative cleavage weight matrix (12 × 20) published by Toes et al. [54] (Table S3 and S4).

5. The query 12-mer peptide with a score above a chosen threshold value (corresponding to 1–10%) was forecasted to be cleaved between the sixth and seventh amino acids [55].

6. Finally, MHC class I and TAP binding peptides identified through the above steps involving proteasomal (constitutive or immuno) cleavage site position at their C-terminal (steps 1–5) were considered CTL epitopes.

Table 2. MHC class I allele-specific binding affinity cutoff of PSSM predictor, along with worldwide frequency in the human population sourced from IEDB and Paul et al. [64].

S. No.	MHC class I allele	Worldwide population frequency of allele (%)	PSSM predictor binding cutoff score
1	HLA-A*0201	25.2	−0.007
2	HLA-A*0203	3.3	1.379
3	HLA-A*0206	4.9	1.625
4	HLA-A*2902	2.9	3.839
5	HLA-A*6801	4.6	3.362
6	HLA-A*6802	3.3	2.305
7	HLA-A*3301	3.2	4.660
8	HLA-A*0101	16.2	3.933
9	HLA-A*0202	0.28	3.984
10	HLA-A*3002	5.0	4.77
11	HLA-A*3101	4.7	2.651
12	HLA-A*0301	15.4	3.800
13	HLA-A*1101	12.9	2.832
14	HLA-A*2402	16.8	2.465
15	HLA-B*0702	13.3	2.835
16	HLA-B*1501	5.2	1.829
17	HLA-B*3501	6.5	2.363
18	HLA-B*4002	3.5	5.415
19	HLA-B*4501	0.63	5.736
20	HLA-B*5101	5.5	8.074
21	HLA-B*5301	5.4	7.923
22	HLA-B*5401	0.56	7.050

MHC: major histocompatibility complex; PSSM: position-specific scoring matrices.

Results

## Construction and evaluation of PSSM and ANN predictors

Sequence weighting methods have been employed in creating a PSSM from frequencies of residues surveyed for a position in an MSA of MHC binders [32]. The present study utilizes three sequence weighting methods (Henikoff & Henikoff 1/nr, clustering at 62% identity, and no clustering) available at the EasyPred modeler (https://services.healthtech.dtu.dk/services/EasyPred-1.0/) for the construction of MHC allele-specific PSSM, including four corrections for low counts (weight on pseudo counts) 50, 100, 150, and 200 [34, 35]. As peptides in the evaluation dataset are not included in the training dataset, it is equivalent to a blind test [27]. For each MHC class I allele, only those PSSM were selected as predictors that gave maximum predictive performance evaluated in terms of Aroc and CC values (Table S5). The Aroc value operates independently of the anticipated scale, as it assesses the ranking of predictors and remains unaffected by the dataset's composition, including varying ratios of binders and non-binders. The Aroc value provides a crucial metric for evaluating forecast quality, with a score of 0.5 indicating random forecasts and 1.0 representing perfect forecasts. Conversely, the CC value of one corresponds to perfect correlation, a value of zero indicates a random estimate, and a value of minus one signifies perfect anti-correlation. The Aroc value

principally captures the probability that, given two peptides, one a binder and the other a non-binder, the forecasted score will be higher for the binder than the non-binder [43]. Furthermore, the MSA of each HLA class I allele binding motif was visualized by using the graphical depiction (sequence logo) method [56], where the height of a column of letters represents the information content (I) at that position. The height of each letter within a column is proportional to the frequency of the corresponding amino acid at that position and colored according to their physicochemical properties, such as acidic (DE)-red, basic (HKR)-blue, hydrophobic (ACFILMPVW)-black, and neutral (GNQSTY)-green (Figure S1). The similarity in peptide binding preferences is observed when comparing the MHC class I binding motif logo developed in the presented study with the logo stored in the MHC motif viewer. In the MHC motif viewer database, pairwise assessments of MHC binding motifs facilitate immediate analysis of epitope selection data in patient cohorts with HLA diversity [57]. For example, when comparing the binding motif of human (HLA-A*2402) to the chimpanzee (Patr-A*0701) allele, an unmistakable resemblance between the binding motifs of the two alleles was spotted, as noted by Sidney et al. [58]. A similar conserved motif for MHC class I peptide binding has also been shown between humans and rhesus macaques, as demonstrated by Dzuris et al. [59]. Consequently, the ability to differentiate various MHC binding specificities (motifs) has applications that range from the design of experiments for peptide binding assays to personalized medicine for a significant population, which includes the selection of peptides that are immunogenic in both humans and model organisms [60, 61].

Moreover, in the case of ANN simulations, two architectures (180-2-1 and 180-1-1) were initially used, and the input peptide sequences (binders with $IC_{50} \leq 500$ nM and rest non-binders) were offered in a conventional encoding, as described by Nielsen et al. [27]. Further, the network weights were renewed using gradient descent backpropagation algorithms. For a given peptide sequence of 9-mer, the ANN weights were renewed to lower the sum of squared errors between the forecasted and experimentally measured BA (target value). The training of the neural networks was performed using a five-fold cross-validation (Supplementary file 1 in https://data.mendeley.com/datasets/dxz3dk3tcm/1). For each of the five training and test subsets, a series of network training sessions is conducted, each utilizing two distinct hidden neurons (1 and 2) along with a single bin for balanced training. For every series, a single network exhibiting the highest test performance (measured by the highest CC and the lowest square error) was ultimately chosen as an ANN predictor (refer to Table S6). Subsequently, the predictive performance of the corresponding ANN network for each MHC allele was assessed based on Aroc (see Table S6). From these findings, it can also be deduced that increasing the number of hidden neurons within the ANN architecture does not have a significant impact on performance. Consequently, the Aroc value for the ANN predictor featuring a single hidden neuron was utilized for comparison with the PSSM predictor. It is crucial to note that, initially, a BA (IC_{50}) threshold for the MHC class I allele was established, where IC_{50} values exceeding 500 nM were identified in competitive assays (involving an MHC and isolated peptides) and were compared with immunogenicity across various marker peptides and nonimmunogenic peptides [62]. This affinity threshold has been identified as being associated with the majority of known T-cell epitopes [63]. Nevertheless, this threshold value has also been demonstrated to be specific to MHC class I alleles [64]. An alternative way to assess the efficacy of peptide binding to MHC class I involves measuring the stability of peptide-MHC complexes over time, but affinity and stability do not rank the peptides in the same order within their source proteins and remain debatable [65, 66]. To compare the performance of ANN and PSSM predictors, the identical evaluation dataset from MHCIPREDS-IEDB was utilized. Consequently, upon comparison, the predictive performance of the ANN models was determined to be superior to that of the PSSM models for most of the MHC class I molecules (Figure 2). The ANN performance in terms of Aroc is maximum (0.97) for the alleles HLA-A*0202 and HLA-A*0203, as well as the minimum (0.56) for allele HLA-A*2902, whereas the maximum PSSM predictor performance in terms of Aroc is (0.93) for the allele HLA-A*0203 and the minimum (0.49) for allele HLA-B*4501 (Figure 2).

From the above results, it is clearly stated that PSSM predictors, to a high degree, describe the binding motif of the corresponding HLA class I alleles, though they demonstrate lower Aroc than the ANN predictors. This is to be expected since the ANN can take higher-order sequence associations into account
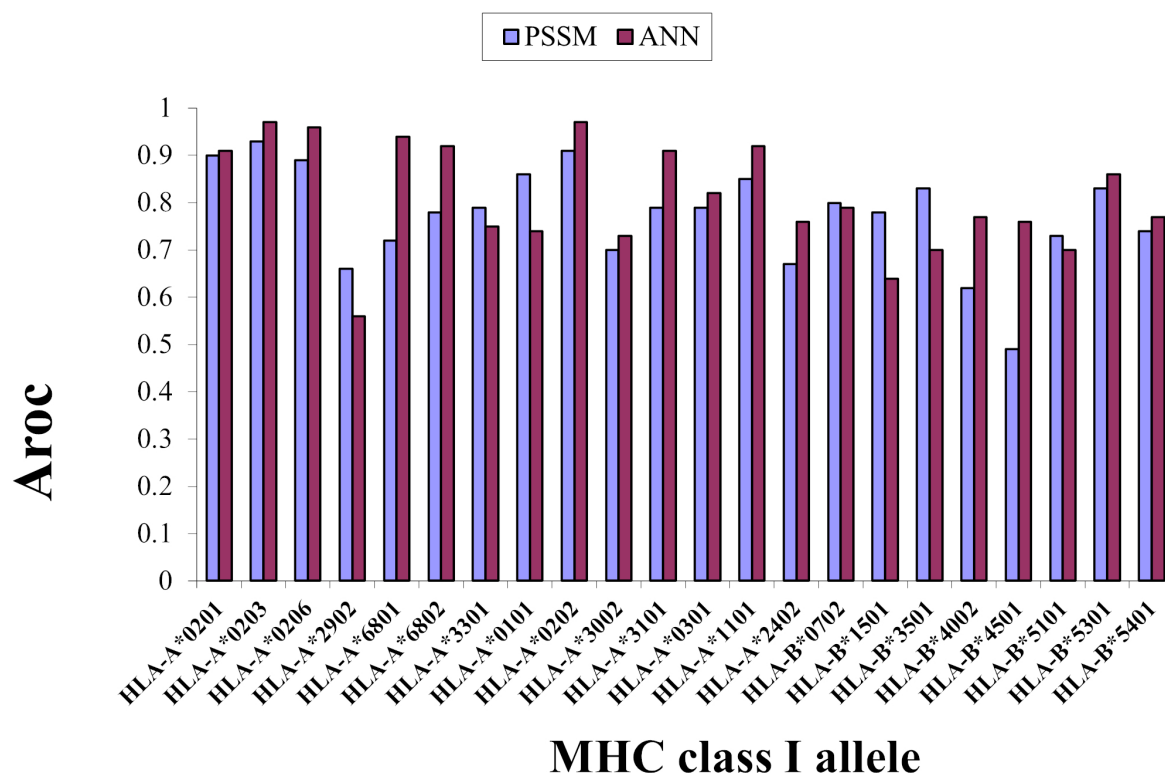
Figure 2. Histogram of the predictive performance measured in terms of the Aroc value of PSSM and ANN predictors for 22 MHC class I alleles trained and evaluated on the MHCIPREDS-IEDB dataset. Aroc: area under the receiver operating characteristic curve; PSSM: position-specific scoring matrices; ANN: artificial neural networks; MHC: major histocompatibility complex.

for a fixed length of MHC binding peptides [27]. Moreover, not only the size of the training and validation dataset but also the choice of specific algorithms can influence the efficiency of the forecast. To integrate PSSM-based MHC class I predictor in CTL epitope processing identification, we calculated the binding thresholds in terms of predicting peptides that bind with an IC_{50} value less than 500 nM, an established threshold associated with immunogenicity for 85% of all epitopes in the training set [51]. These binding threshold scores for each of the 22 MHC class I molecules were determined for PSSM predictors to demarcate the range of putative binders among the top-scoring peptides (Table 2). Paul et al. [64] also revealed similar observations that different MHC molecules bind ligands at different (forecasted) binding thresholds (scores). If a single criterion for MHC class I alleles has to be deliberated, then it is preferable to select absolute BA, as in the case of the ANN predictor, where $IC_{50} \leq 500$ nM could be regarded as a reasonably good "universal" binding threshold. In a similar study, Bonsack et al. [67] confirmed the decisive threshold of $IC_{50} \leq 500$ nM for MHC class I binding peptides through in vitro validation. However, predictive efficacy is increased using allele-specific affinity thresholds.

Thus, based on these observations, the PSSM-based MHC class I binding forecasts were applied without rescaling, thereby preserving prospective fundamental biological differences between MHC class molecules. However, ANN-based MHC class I binding peptide forecasts were performed with $IC_{50} \leq 500$ nM (LTMBA value ≥ 0.426).

Evaluation of the EBV epitope dataset

The PSSM and ANN predictors successfully identified the EBV epitopes as revealed in the study conducted by Wohlwend et al. [44], above the chosen threshold (Table 2), except the epitope IACPIVMRY restricted by the HLA-B*1501 allele with a comparable BA score of 0.411 compared to the ANN predictor threshold (LTMBA value) of 0.426 (Table 3). This clearly indicates the real-life application of models in the identification of both established and novel HLA class I epitopes from EBV.

Table 3. List of EBV epitopes used in the present study derived from Wohlwend et al. [44].

S. No.	CD8+ T cell epitope	HLA binding allele	PSSM predictor score	ANN predictor score (binding affinity)
1	AFDQATRVY	HLA-A*0101	8.505	0.438
2	HLSQAAFGL	HLA-A*0201	6.252	0.633
3	SIIPRTPDV		6.581	0.46
4	YVLDHLIVV		9.036	0.769
5	RYSIFFDYM	HLA-A*2402	8.583	0.45
6	TYPVLEEMF		7.784	0.455
7	IPQCRLTPL	HLA-B*0702	12.551	0.454
8	RPPIFIRRL		10.057	0.495
9	IACPIVMRY	HLA-B*1501	6.259	0.411
10	SQISNTEMY		10.109	0.486
11	VQTAAAVVF		6.888	0.428

EBV: Epstein-Barr virus; PSSM: position-specific scoring matrices; ANN: artificial neural networks.

Generalizability evaluation

The generalization ability of the MHC class I binding prediction model is essential for epitope prediction, as there are many HLA alleles with inadequate data for training an allele-specific model [68]. Therefore, we have performed a detailed analysis of the performance of PSSM/ANN predictors trained on one allele and their ability to accurately predict other alleles in their evaluation datasets. Pan-specific algorithms can predict peptide binding to HLA alleles for which limited or even no experimental data are available [69]. In Figure 3, the Aroc values for the models trained on the HLA-A*0201 dataset are given for 22 different alleles. The ANN model excellently performs over the PSSM model for alleles of the HLA-A*02 type, but for the other alleles, the performance of both models is poor, except for HLA-B*4002, 4501, 5101. The prediction capabilities are good to marginal for some alleles, suggesting that cross-allele prediction is feasible in some cases. This may be due to MHC supertype classification systems that make clustered sets of HLA molecules with largely overlapping peptide repertoires. These classification systems normally depend on descriptions such as published motifs and/or analysed shared repertoires of binding peptides, etc. [70–73]. Generally, HLA-A and -B alleles are not clustered in the same supertype, but our PSSM/ANN predictor trained on HLA-A*0201 allele was able to make reasonable predictions for HLA-B*4002, 4501, and 5101 alleles.

Rank measure analysis of PSSM and ANN predictors with NetMHCpan 4.0 and NetCTLpan 1.1

NetCTLpan dataset

Although the existence of homologous peptides between training and testing datasets has been avoided to provide real-world estimates of forecast performance metrics, the relative ranking of diverse predictors is principally unaffected by the existence of homologous peptides [63]. Thus, the additional performance evaluation of the present PSSM/ANN predictors was estimated in terms of the relative ranking of the reported ligands between all nonamers included in the source protein as described by Larsen et al. [45, 46]. This evaluation indicates how large a portion of the peptides for a provided protein needs to be verified to identify the new epitopes [74]. For each of the stated ligands in the NetCTLpan dataset (Supplementary file 2 in https://data.mendeley.com/datasets/dxz3dk3tcm/1), the source protein was identified, and the affinity of all nonamers contained in the source protein was forecasted (assuming that all nonamers, except for the reported ligand, are non-binders). Here, we define the term reliability of a forecast method as the probability of identifying an epitope in each protein within a certain top percentage of the peptides [45]. The rank measure performance of the NetMHCpan 4.0 method based on EL (1.11%) and BA (3.19%) showed similar results (comparatively higher average values) as compared to PSSM (2.23%) and ANN (3.13%) predictors. The results from these evaluations are encouraging because the most natural MHC binders compiled in the NetCTLpan dataset were identified within the top 5% (Table S7; Table 4). Therefore, in terms of wet laboratory work, nearly ~97% less expenses are consumed on materials, labor,
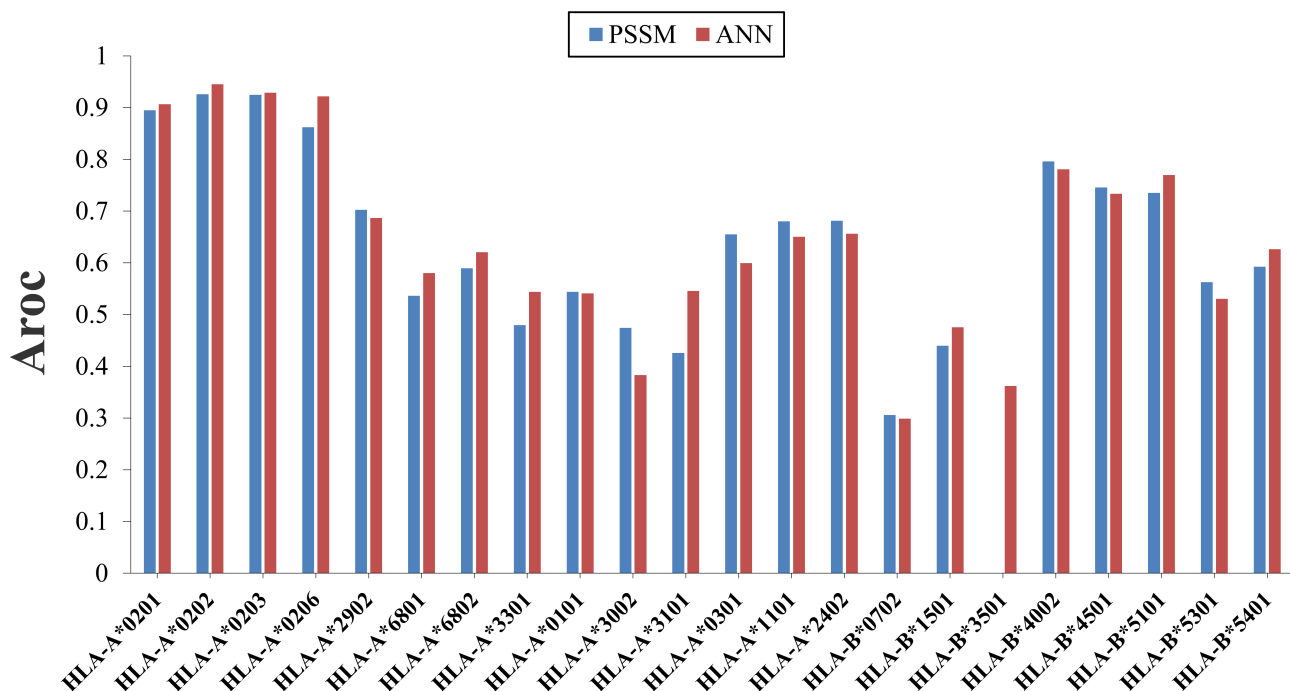
Figure 3. The cross-allele performance of the PSSM and ANN prediction models, trained on the HLA-A*0201 dataset and tested on the evaluation dataset of all 22 alleles. Aroc: area under the receiver operating characteristic curve; PSSM: position-specific scoring matrices; ANN: artificial neural networks.

and time for the peptides that require experimental verification to detect new epitopes in an antigen. However, the NetCTLpan method reports a rank measure of 3.7% for the peptides that need to be experimentally verified to detect new epitopes with 90% likelihood. Thus, for a hypothetical protein of 300 peptides, this means that, on average, 7 and 9 peptides need to be tested to identify the epitope using PSSM and ANN predictors, respectively. The corresponding numbers reported for NetCTL, NetMHCpan, and NetCTLpan were 17, 13, and 11 peptides, respectively [46, 47]. Using the NetCTLpan tool, the experimental effort to discover 90% of new epitopes can be minimized by 15% and 40%, respectively, compared to the NetMHCpan and NetCTL tools [47]. However, the corresponding peptide numbers are 3 and 10 to identify new CTL epitopes using NetMHCpan 4.0 (EL) and NetMHCpan 4.0 (BA), respectively. Thus, the overall performance of EasyPred modeler-based predictors (PSSM and ANN) was found to be similar to NetMHCpan 4.0 (BA) but lower than NetMHCpan 4.0 (EL).

Table 4. Rank measure analysis of MHC class I predictors (PSSM and ANN) based on the NetCTLpan dataset.

S. No.	MHC class I allele	No. of antigen	Average relative rank of epitopes in their source antigen (%)			
			PSSM	ANN	NetMHCpan 4.0 (EL)	NetMHCpan 4.0 (BA)
1	HLA-A*0101	29	0.36	1.84	0.3246	0.3159
2	HLA-A*0201	254	3.30	4.17	3.1785	17.405
3	HLA-A*1101	14	1.72	0.80	0.399	0.4765
4	HLA-A*6801	12	3.05	3.11	0.9291	0.907
5	HLA-A*0301	65	2.77	3.68	1.5704	1.4966
6	HLA-B*0702	25	1.59	4.71	0.6781	0.9225
7	HLA-B*4501	14	2.83	3.66	0.7079	0.8363
Average			2.23	3.13	1.11	3.19

MHC: major histocompatibility complex; PSSM: position-specific scoring matrices; ANN: artificial neural networks; EL: eluted ligands; BA: binding affinity.

SARS-CoV-2 dataset

In a comparative evaluation (average rank measure analysis) of EasyPred modeler-based predictors (PSSM and ANN) on the SARS-CoV-2 dataset of Grifoni et al. [49], the PSSM predictor (2.46%) performed better than other T-cell epitope forecast methods, NetMHCpan 4.0 (EL) (2.66%), NetCTLpan 1.1 (2.69%), NetMHCpan 4.0 (BA) (3.33%), as well as its own ANN predictor (4.67%), respectively (Table 5). Similar results were also observed for evaluation on the SARS-CoV-2 dataset of Gfeller et al. [50]. In which both the predictors (PSSM and ANN) identified all the CD8+ T cell epitopes above their cutoff score, except for a few: LYLYALVYF (A*2402: 0.418), LWLLWPVTL (A*2402: 0.402), FTSDYYQLY (A*2402: 0.326), and YFPLQSYGF (A*2402: 0.38) in the case of ANN. Moreover, % rank measure analysis in their source protein revealed that the PSSM predictor identified the epitopes within 3% (average 2.4%) (Table 6). However, Nosrati et al. [75] recently established that ANN was the most accurate algorithm for distinguishing epitopes and non-epitopes of the Crimean-Congo hemorrhagic fever virus with an accuracy of 90%. Such evaluation knowledge is of urgent significance that would assist COVID-19 vaccine developers in facilitating the evaluation of vaccine candidate immunogenicity against human populations [76].

Table 5. Comparative using rank measure evaluation of PSSM and ANN predictors with up-to-date tools NetMHCpan 4.0 (EL/BA) and NetCTLpan 1.1 on SARS-CoV-2 dataset from the study of Grifoni et al. [49].

S. No.	T cell epitope	Source protein	Average relative rank of epitopes in their source antigen (%)				
			PSSM predictor	ANN predictor	NetMHCpan 4.0 (EL)	NetMHCpan 4.0 (BA)	NetCTLpan 1.1
1	ALNTLVKQL	Spike	2.77	6.80	1.42	3.87	2.06
2	VLNDILSRL	Spike	0.79	0.95	0.16	0.47	0.32
3	LITGRLQSL	Spike	2.29	5.06	8.06	9.64	5.06
4	RLNEVAKNL	Spike	4.98	9.33	1.11	7.06	1.42
5	NLNESLIDL	Spike	1.58	4.35	0.79	1.58	1.11
6	FIAGLIAIV	Spike	0.40	0.28	0.40	0.16	0.79
7	ALNTPKDHI	Nucleocapsid	3.16	8.03	2.43	4.62	6.08
8	LQLPQGTTL	Nucleocapsid	2.92	5.12	1.70	2.91	1.12
9	LALLLLDRL	Nucleocapsid	7.06	10.71	12.41	5.56	10.94
10	LLLDRLNQL	Nucleocapsid	0.24	0.24	0.24	0.24	0.24
11	GMSRIGMEV	Nucleocapsid	0.71	0.49	0.49	0.49	0.49
Average			2.46	4.67	2.66	3.33	2.69

PSSM: position-specific scoring matrices; ANN: artificial neural networks; EL: eluted ligands; BA: binding affinity.

Table 6. Evaluation of PSSM and ANN predictors on the SARS-CoV-2 dataset of Gfeller et al. [50] using threshold (cutoff score) and % rank measure analysis in the source protein.

S. No.	CD8+ T cell epitope	Source protein (SWISS-PROT accession no.)	HLA binding allele	PSSM predictor		ANN predictor	
				Score	% rank	Score	% rank
1	LYLYALVYF	AP3A (P0DTC3)	A*2402	9.218	0.75	0.418	13.48
2	LWLLWPVTL	VME1 (P0DTC5)		8.366	1.4	0.402	22.43
3	LPPAYTNSF	SPIKE (P0DTC2)	B*0702, B*3501, B*5301	9.718, 10.505, 9.574	0.32, 0.24, 0.87	0.457, 0.447, 0.495	2.21, 19.37, 6.56
4	FTSDYYQLY	AP3A (P0DTC3)	A*0101, A*2402, A*2902	14.473, 2.999, 11.658	0.37, 16.48, 0.37	0.491, 0.326, 0.540	34.08, 46.81, 22.47
5	YFPLQSYGF	SPIKE (P0DTC2)	A*2402	7.166	1.42	0.38	34.78
6	SASKIITLK	AP3A (P0DTC3)	A*0301, A*1101	6.574, 5.973	2.25, 1.9	0.672, 0.790	0.37, 0.37

PSSM: position-specific scoring matrices; ANN: artificial neural networks.

CTL epitope processing forecast

The PSSM and ANN predictors of the EasyPred modeler described in the present study have been shown to perform best when high-sensitivity forecasts for CTL epitope identification are focused (Figure 4). Most of the MHC molecules achieve a sensitivity of 80% at the threshold determined by the score of the top 85% in the training set, while some alleles (HLA-A*0101, A*1101, and A*6801) approached the sensitivity of 100%. If focusing on optimal sensitivity, it was shown that the forecast algorithm should exclude both proteasomal cleavage and TAP forecasts, reducing the method to the MHC binding forecast alone (Figure 4). Whether this observation reflects actual biological aspects of the specificity overlap between the three pathway players or it simply occurs because the forecast of MHC class I affinity has gained accuracy during the recent years, the predictors for TAP transport efficiency and proteasomal cleavage have not much changed or been renewed [77].
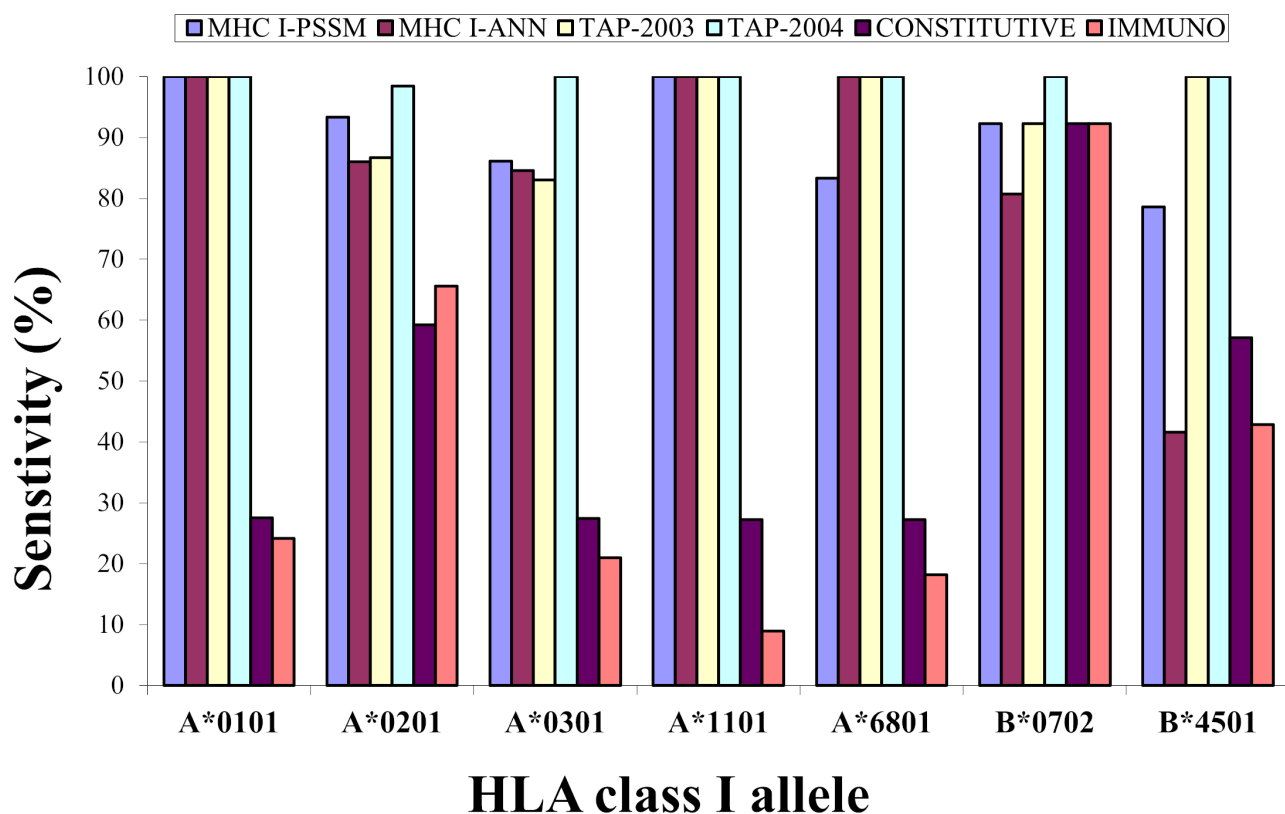
Figure 4. Sensitivity analysis of PSSM and ANN predictors, TAP-2003 [52] and TAP-2004 [53] matrices, as well as constitutive- and immuno-proteasomal cleavage matrices in identifying CTL epitopes. MHC: major histocompatibility complex; PSSM: position-specific scoring matrices; ANN: artificial neural networks; TAP: transporters associated with antigen processing; CTL: cytotoxic T lymphocytes.

Discussion

Antigen processing happens before the MHC binding, revealing the pool of peptides that can become T-cell epitopes. MHC class I-restricted CTL epitopes, in general, are obtained from protein pieces produced by the protease activity of the constitutive and/or immunoproteasome. The activity of several amino peptidases with the resulting loss of information shapes the N-terminus of every MHC class I-restricted peptide. In contrast, the C-terminus results from the proteasomal cleavage [78]. Moreover, it is believed that the immuno-proteasome is more responsible for producing CTL epitopes [54, 79, 80]. As a first approximation, proposed by Assarsson et al. [81], about 15% of all peptides that can be made from a protein are TAP transported into the ER, and about 2.5% of peptides that are made will bind to an MHC molecule. Further, about 50% of MHC binding peptides presented on the cell surface will be recognized by a CD8+ T cell receptor (TCR) and considered CTL epitopes. Using TAP binding as a filter, Doytchinova et al. [53] have also shown that the forecast of the peptide unable to bind with TAP decreases the number of peptides binding to

MHC by 10–30% (depending on MHC allele). Although there are several bioinformatics programs available for the forecast of proteasomal cleavage sites in antigens [82, 83], substantial success has been realized in blending these into an integrated CTL epitope forecast system. RankPep [36], ProPred1 [29], MAP [84], and PEPVAC [85] programs provide a platform for a concurrent forecast of proteasomal cleavage and MHC binding. Other virtual models of the endogenous antigen processing pathway have also been developed, incorporating proteasomal cleavage, TAP transport, and MHC class I binding forecasts [86, 87] such as MAPPP [84], WAPP [88], EpiJen [28], MHC-pathway [89], and NetCTL [45]. Many of these methods have also proved their efficiency in screening new epitopes and designing poly-epitope vaccines for cancer [90, 91], tuberculosis [92], Ebola [93], dengue [94], novel coronavirus [95], etc. In a large-scale benchmark evaluation of a publicly available MHC class I pathway presentation forecast server, Larsen et al. [46] showed that the NetCTL method substantially outperformed all these methods, closely followed by the MHC pathway. Further, the NetCTL method has also proven successful in the identification of CTL epitopes from Influenza [96], HIV [97], and Orthopoxvirus [98]. NetCTLpan [47] is also available for integrating many MHC class I allele binding, TAP transport efficiency, and proteasomal cleavage forecasts to an overall forecast of CTL epitopes. However, all these methods are limited because they allow for the forecast of peptide binding to pre-calculated parameters of MHC molecules and have one or more limitations. However, the current algorithms, PSSM and ANN predictors of the EasyPred modeler, are independent of using pre-calculated MHC binding parameters, and the user can make their own PSSM and ANN parameters for the forecast. Therefore, an integrated model was developed in the present study to predict CTL epitope primarily based on forecasting MHC class I binding and constitutive, as well as immuno-proteasomal cleavage and TAP binding as filters. This type of filtering algorithm has been shown to improve the forecast of CTL epitopes by decreasing the false positives and improving the process of drug and vaccine design strategies [28, 99]. It is worth noting that proteasomal cleavage, TAP transport, and MHC binding have largely undergone co-evolution so that MHC molecules have evolved to bind peptides in the ER. As a result, their combination does not provide vastly improved forecasts [89]. Thus, we recommend using the MHC class I binding forecasts as a primary tool to select candidate peptides for wet lab validations. We also recommend CTL processing forecasts to reduce further candidates to test in vitro/in vivo [51].

An increasing body of literature substantiates the prediction of direct T cell immunogenicity, which refers to the relative capacity of a specific set of peptides that are bound within an MHC complex and recognized by the TCR. In the MHC-peptide-TCR complex, the residues P1, P2, and P9 of the peptide are most likely to interact directly with MHC binding, whereas the other residues, P3 to P8, are more likely to engage with the TCR. These investigations have shown that certain amino acids, including tryptophan, phenylalanine, and isoleucine, are prevalent in immunogenic peptides, while other residues, such as serine, methionine, and lysine, are less common. This phenomenon may be attributed to the longer side chains of tryptophan and phenylalanine, which have a higher likelihood of interacting with the TCR [100]. Consequently, it is possible to generate more immunogenic peptides through amino acid substitutions in naturally occurring MHC class I-restricted epitopes, thereby enhancing the stability and affinity of the MHC-peptide-TCR complex. This approach enhances the immunogenicity of MHC class I transitional affinity binders [101]. Furthermore, Rasmussen et al. [102] indicated that the stability of the peptide-MHC class I complex is crucial for eliciting T-cell responses. Therefore, parameters related to BA and stability are vital for characterizing and predicting peptide immunogenicity [103]. Given these more dependable structural models, peptide-based vaccines are likely the most favored and extensively researched due to their ease of synthesis and biocompatibility [104].

However, the efficacy of peptide vaccines tends to be compromised by various factors such as rapid clearance, extracellular and enzymatic degradation, poor solubility (due to the presence of hydrophobic peptides), reduced immunogenicity, and reduced uptake of the peptide by antigen-presenting cells (APC) [105]. With current advancements in nanobiotechnology-based delivery systems, such as self-assembled peptide nanoparticles (SANP) using spacers for separation of epitopes for proteasomal cleavage, which can efficiently deliver peptide vaccines to APC, could help to surmount the limitations and elicit T cell responses in humans [106–109]. Successful production and efficiency of such vaccine delivery systems require

deliberation of other variables, e.g., material toxicity, particle shape and size, surface charge, and stiffness or rigidity, which are crucial in various biological processes, such as bioavailability, biodistribution, and cytotoxicity, including activation of adaptive immune responses [110, 111] and specific other biomedical applications [112].

In conclusion, this study emphasizes the creation of an innovative forecasting model aimed at identifying peptide interactions with MHC class I molecules through the application of sequence weighting and ANN methodologies. As a result of this research, we have successfully established new forecasting parameters, specifically PSSM and ANN weights, for predicting MHC binding peptides. The models developed can be utilized alongside the EasyPred modeler to forecast MHC class I binding peptides, which includes the processing of CTL epitopes. Overall, the findings discussed highlight the importance of predicting MHC peptide binding for epitope identification, while also acknowledging the challenges that persist, indicating significant opportunities for enhancement and integration with other structure-based approaches. These forecasting techniques will relieve vaccinologists and immunologists from the burdens of uninformed experimentation, enabling them to devise improved, quicker, and more innovative methods for discovering new reagents, diagnostics, and vaccines.

Abbreviations

ANN: artificial neural networks

APC: antigen-presenting cells

Aroc: area under the receiver operating characteristic curve

BA: binding affinity

CC: correlation coefficient

CTL: cytotoxic T lymphocytes

EBV: Epstein-Barr virus

EL: eluted ligands

ER: endoplasmic reticulum

IC_{50}: half-maximal inhibitory concentration

LTMBA: log-transformed major histocompatibility complex binding affinity

MHC: major histocompatibility complex

MSA: multiple sequence alignment

PSSM: position-specific scoring matrices

RED: Repository for Epitope Datasets

TAP: transporters associated with antigen processing

TCR: T cell receptor

Supplementary materials

The supplementary materials for this article are available at: https://www.explorationpub.com/uploads/Article/file/1003215_sup_1.pdf.

Declarations

Acknowledgments

Author contributions

SPS: Conceptualization, Methodology, Investigation, Writing—review & editing. GS: Data curation, Investigation. BNM: Supervision, Writing—review & editing. All authors read and approved the submitted version.

Conflicts of interest

The authors declare that they have no conflicts of interest.

Ethical approval

Not applicable.

Consent to participate

Not applicable.

Consent to publication

Not applicable.

Availability of data and materials

Supplemental information can be found online at https://data.mendeley.com/datasets/dxz3dk3tcm/1 at Mendeley data (DOI: 10.17632/dxz3dk3tcm.1). Data are licensed under an Attribution-NonCommercial 3.0 Unported licence. All figures in the study were created using the MS-Excel program.

Funding

Not applicable.

Copyright

Publisher's note

Open Exploration maintains a neutral stance on jurisdictional claims in published institutional affiliations and maps. All opinions expressed in this article are the personal views of the author(s) and do not represent the stance of the editorial team or the publisher.

References

1. Kazi A, Chuah C, Majeed ABA, Leow CH, Lim BH, Leow CY. Current progress of immunoinformatics approach harnessed for cellular- and antibody-dependent vaccine design. Pathog Glob Health. 2018; 112:123–31. [DOI] [PubMed] [PMC]
2. Gilbert SC. T-cell-inducing vaccines - what's the future. Immunology. 2012;135:19–26. [DOI] [PubMed] [PMC]
3. Rudolph MG, Stanfield RL, Wilson IA. How TCRs bind MHCs, peptides, and coreceptors. Annu Rev Immunol. 2006;24:419–66. [DOI] [PubMed]
4. Kägi D, Ledermann B, Bürki K, Seiler P, Odermatt B, Olsen KJ, et al. Cytotoxicity mediated by T cells and natural killer cells is greatly impaired in perforin-deficient mice. Nature. 1994;369:31–7. [DOI] [PubMed]
5. Lankat-Buttgereit B, Tampé R. The transporter associated with antigen processing: function and implications in human diseases. Physiol Rev. 2002;82:187–204. [DOI] [PubMed]
6. Craiu A, Akopian T, Goldberg A, Rock KL. Two distinct proteolytic processes in the generation of a major histocompatibility complex class I-presented peptide. Proc Natl Acad Sci U S A. 1997;94: 10850–5. [DOI] [PubMed] [PMC]

7. van Endert PM, Riganelli D, Greco G, Fleischhauer K, Sidney J, Sette A, et al. The peptide-binding motif for the human transporter associated with antigen processing. J Exp Med. 1995;182:1883–95. [DOI] [PubMed] [PMC]

8. Schatz MM, Peters B, Akkad N, Ullrich N, Martinez AN, Carroll O, et al. Characterizing the N-terminal processing motif of MHC class I ligands. J Immunol. 2008;180:3210–7. [DOI] [PubMed]

9. Matsumura M, Fremont DH, Peterson PA, Wilson IA. Emerging principles for the recognition of peptide antigens by MHC class I molecules. Science. 1992;257:927–34. [DOI] [PubMed]

10. Pamer E, Cresswell P. Mechanisms of MHC class I--restricted antigen processing. Annu Rev Immunol. 1998;16:323–58. [DOI] [PubMed]

11. Yewdell JW, Bennink JR. Immunodominance in major histocompatibility complex class I-restricted T lymphocyte responses. Annu Rev Immunol. 1999;17:51–88. [DOI] [PubMed]

12. Zhong W, Reche PA, Lai CC, Reinhold B, Reinherz EL. Genome-wide characterization of a viral cytotoxic T lymphocyte epitope repertoire. J Biol Chem. 2003;278:45135–44. [DOI] [PubMed]

13. Tang Y, Wang H. Construction and immunogenicity prediction of Plasmodium falciparum CTL epitope minigene vaccine. Sci China C Life Sci. 2001;44:207–15. [DOI] [PubMed]

14. Zhang GL, Khan AM, Srinivasan KN, August JT, Brusic V. MULTIPRED: a computational system for prediction of promiscuous HLA binding peptides. Nucleic Acids Res. 2005;33:W172–9. [DOI] [PubMed] [PMC]

15. Kar P, Ruiz-Perez L, Arooj M, Mancera RL. Current methods for the prediction of T-cell epitopes. Pept Sci. 2018;110:e24046. [DOI]

16. Gowthaman U, Agrewala JN. In silico methods for predicting T-cell epitopes: Dr Jekyll or Mr Hyde? Expert Rev Proteomics. 2009;6:527–37. [DOI] [PubMed]

17. Yang X, Yu X. An introduction to epitope prediction methods and software. Rev Med Virol. 2009;19: 77–96. [DOI] [PubMed]

18. Peters B, Sidney J, Bourne P, Bui H, Buus S, Doh G, et al. The immune epitope database and analysis resource: from vision to blueprint. PLoS Biol. 2005;3:e91. [DOI] [PubMed] [PMC]

19. Zhang L, Udaka K, Mamitsuka H, Zhu S. Toward more accurate pan-specific MHC-peptide binding prediction: a review of current methods and tools. Brief Bioinform. 2012;13:350–64. [DOI] [PubMed]

20. Bhasin M, Raghava GPS. Prediction of CTL epitopes using QM, SVM and ANN techniques. Vaccine. 2004;22:3195–204. [DOI] [PubMed]

21. Trolle T, Metushi IG, Greenbaum JA, Kim Y, Sidney J, Lund O, et al. Automated benchmarking of peptide-MHC class I binding predictions. Bioinformatics. 2015;31:2174–81. [DOI] [PubMed] [PMC]

22. Lin HH, Ray S, Tongchusak S, Reinherz EL, Brusic V. Evaluation of MHC class I peptide binding prediction servers: applications for vaccine research. BMC Immunol. 2008;9:8. [DOI] [PubMed] [PMC]

23. Paul S, Sidney J, Sette A, Peters B. TepiTool: A Pipeline for Computational Prediction of T Cell Epitope Candidates. Curr Protoc Immunol. 2016;114:18.19.1–24. [DOI] [PubMed] [PMC]

24. Dimitrov I, Garnev P, Flower DR, Doytchinova I. MHC Class II Binding Prediction-A Little Help from a Friend. J Biomed Biotechnol. 2010;2010:705821. [DOI] [PubMed] [PMC]

25. Hobohm U, Scharf M, Schneider R, Sander C. Selection of representative protein data sets. Protein Sci. 1992;1:409–17. [DOI] [PubMed] [PMC]

26. Henikoff S, Henikoff JG. Position-based sequence weights. J Mol Biol. 1994;243:574–8. [DOI] [PubMed]

27. Nielsen M, Lundegaard C, Worning P, Lauemøller SL, Lamberth K, Buus S, et al. Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. Protein Sci. 2003;12: 1007–17. [DOI] [PubMed] [PMC]

28. Doytchinova IA, Guan P, Flower DR. EpiJen: a server for multistep T cell epitope prediction. BMC Bioinformatics. 2006;7:131. [DOI] [PubMed] [PMC]

29. Singh H, Raghava GPS. ProPred1: prediction of promiscuous MHC Class-I binding sites. Bioinformatics. 2003;19:1009–14. [DOI] [PubMed]

30. Jaishwal P, Jha K, Singh SP. Revisiting the dimensions of universal vaccine with special focus on COVID-19: Efficacy versus methods of designing. Int J Biol Macromol. 2024;277:134012. [DOI] [PubMed]

31. Nielsen M, Lundegaard C, Blicher T, Lamberth K, Harndahl M, Justesen S, et al. NetMHCpan, a method for quantitative predictions of peptide binding to any HLA-A and -B locus protein of known sequence. PLoS One. 2007;2:e796. [DOI] [PubMed] [PMC]

32. Gribskov M, McLachlan AD, Eisenberg D. Profile analysis: detection of distantly related proteins. Proc Natl Acad Sci U S A. 1987;84:4355–8. [DOI] [PubMed] [PMC]

33. Lüthy R, Xenarios I, Bucher P. Improving the sensitivity of the sequence profile method. Protein Sci. 1994;3:139–46. [DOI] [PubMed] [PMC]

34. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997;25:3389–402. [DOI] [PubMed] [PMC]

35. Nielsen M, Lundegaard C, Worning P, Hvid CS, Lamberth K, Buus S, et al. Improved prediction of MHC class I and class II epitopes using a novel Gibbs sampling approach. Bioinformatics. 2004;20: 1388–97. [DOI] [PubMed]

36. Reche PA, Reinherz EL. Definition of MHC supertypes through clustering of MHC peptide-binding repertoires. Methods Mol Biol. 2007;409:163–73. [DOI] [PubMed]

37. Brusic V, van Endert P, Zeleznikow J, Daniel S, Hammer J, Petrovsky N. A neural network model approach to the study of human TAP transporter. In Silico Biol. 1999;1:109–21. [PubMed]

38. Brusic V, Zeleznikow J, Sturniolo T, Bono E, Hammer J. Data cleaning for computer models: a case study from immunology. Proceedings of the 6th International Conference on Neural Information Processing (ICONIP'99/ANZIIS'99/ANNES'99/ACNN'99); 1999 Nov 22–24; Perth, Australia. Piscataway (NJ): IEEE; 1999. pp. 603–9. [DOI]

39. Lund O, Nielsen M, Lundegaard C, Kesmir C, Brunak S. Immunological Bioinformatics. Cambridge: The MIT Press; 2005.

40. Zhao W, Sher X. Systematically benchmarking peptide-MHC binding predictors: From synthetic to naturally processed epitopes. PLoS Comput Biol. 2018;14:e1006457. [DOI] [PubMed] [PMC]

41. Rumelhart DE, Hinton GE, Williams RJ. Parallel distributed processing: explorations in the microstructure of cognition. Cambridge: MIT Press; 1986. pp. 318–62.

42. Swets JA. Measuring the accuracy of diagnostic systems. Science. 1988;240:1285–93. [DOI] [PubMed]

43. Lasko TA, Bhagwat JG, Zou KH, Ohno-Machado L. The use of receiver operating characteristic curves in biomedical informatics. J Biomed Inform. 2005;38:404–15. [DOI] [PubMed]

44. Wohlwend J, Nathan A, Shalon N, Crain CR, Tano-Menka R, Goldberg B, et al. Deep learning enhances the prediction of HLA class I-presented CD8+ T cell epitopes in foreign pathogens. Nat Mach Intell. 2025;7:232–43. [DOI] [PubMed] [PMC]

45. Larsen MV, Lundegaard C, Lamberth K, Buus S, Brunak S, Lund O, et al. An integrative approach to CTL epitope prediction: a combined algorithm integrating MHC class I binding, TAP transport efficiency, and proteasomal cleavage predictions. Eur J Immunol. 2005;35:2295–303. [DOI] [PubMed]

46. Larsen MV, Lundegaard C, Lamberth K, Buus S, Lund O, Nielsen M. Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. BMC Bioinformatics. 2007;8:424. [DOI] [PubMed] [PMC]

47. Stranzl T, Larsen MV, Lundegaard C, Nielsen M. NetCTLpan: pan-specific MHC class I pathway epitope predictions. Immunogenetics. 2010;62:357–68. [DOI] [PubMed] [PMC]

48. Rammensee H, Bachmann J, Emmerich NP, Bachor OA, Stevanović S. SYFPEITHI: database for MHC ligands and peptide motifs. Immunogenetics. 1999;50:213–9. [DOI] [PubMed]

49. Grifoni A, Sidney J, Zhang Y, Scheuermann RH, Peters B, Sette A. A Sequence Homology and Bioinformatic Approach Can Predict Candidate Targets for Immune Responses to SARS-CoV-2. Cell Host Microbe. 2020;27:671–80.e2. [DOI] [PubMed] [PMC]

50. Gfeller D, Schmidt J, Croce G, Guillaume P, Bobisse S, Genolet R, et al. Improved predictions of antigen presentation and TCR recognition with MixMHCpred2.2 and PRIME2.0 reveal potent SARS-CoV-2 CD8$^+$ T-cell epitopes. Cell Syst. 2023;14:72–83.e5. [DOI] [PubMed] [PMC]

51. Fleri W, Paul S, Dhanda SK, Mahajan S, Xu X, Peters B, et al. The Immune Epitope Database and Analysis Resource in Epitope Discovery and Synthetic Vaccine Design. Front Immunol. 2017;8:278. [DOI] [PubMed] [PMC]

52. Peters B, Bulik S, Tampe R, Endert PMV, Holzhütter H. Identifying MHC class I epitopes by predicting the TAP transport efficiency of epitope precursors. J Immunol. 2003;171:1741–9. [DOI] [PubMed]

53. Doytchinova I, Hemsley S, Flower DR. Transporter associated with antigen processing preselection of peptides binding to the MHC: a bioinformatic evaluation. J Immunol. 2004;173:6813–9. [DOI] [PubMed]

54. Toes RE, Nussbaum AK, Degermann S, Schirle M, Emmerich NP, Kraft M, et al. Discrete cleavage motifs of constitutive and immunoproteasomes revealed by quantitative analysis of cleavage products. J Exp Med. 2001;194:1–12. [DOI] [PubMed] [PMC]

55. Kuttler C, Nussbaum AK, Dick TP, Rammensee HG, Schild H, Hadeler KP. An algorithm for the prediction of proteasomal cleavages. J Mol Biol. 2000;298:417–29. [DOI] [PubMed]

56. Schneider TD, Stephens RM. Sequence logos: a new way to display consensus sequences. Nucleic Acids Res. 1990;18:6097–100. [DOI] [PubMed] [PMC]

57. Rapin N, Hoof I, Lund O, Nielsen M. MHC motif viewer. Immunogenetics. 2008;60:759–65. [DOI] [PubMed] [PMC]

58. Sidney J, Peters B, Frahm N, Brander C, Sette A. HLA class I supertypes: a revised and updated classification. BMC Immunol. 2008;9:1. [DOI] [PubMed] [PMC]

59. Dzuris JL, Sidney J, Appella E, Chesnut RW, Watkins DI, Sette A. Conserved MHC class I peptide binding motif between humans and rhesus macaques. J Immunol. 2000;164:283–91. [DOI] [PubMed]

60. Frahm N, Yusim K, Suscovich TJ, Adams S, Sidney J, Hraber P, et al. Extensive HLA class I allele promiscuity among viral CTL epitopes. Eur J Immunol. 2007;37:2419–33. [DOI] [PubMed] [PMC]

61. Lundegaard C, Lund O, Buus S, Nielsen M. Major histocompatibility complex class I binding predictions as a tool in epitope discovery. Immunology. 2010;130:309–18. [DOI] [PubMed] [PMC]

62. Sette A, Vitiello A, Reherman B, Fowler P, Nayersina R, Kast WM, et al. The relationship between class I binding affinity and immunogenicity of potential cytotoxic T cell epitopes. J Immunol. 1994; 153:5586–92. [PubMed]

63. Wang P, Sidney J, Kim Y, Sette A, Lund O, Nielsen M, et al. Peptide binding predictions for HLA DR, DP and DQ molecules. BMC Bioinformatics. 2010;11:568. [DOI] [PubMed] [PMC]

64. 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: 5831–9. [DOI] [PubMed] [PMC]

65. van der Burg SH, Visseren MJ, Brandt RM, Kast WM, Melief CJ. Immunogenicity of peptides bound to MHC class I molecules depends on the MHC-peptide complex stability. J Immunol. 1996;156: 3308–14. [PubMed]

66. Harndahl M, Rasmussen M, Roder G, Pedersen ID, Sørensen M, Nielsen M, et al. Peptide-MHC class I stability is a better predictor than peptide affinity of CTL immunogenicity. Eur J Immunol. 2012;42: 1405–16. [DOI] [PubMed]

67. Bonsack M, Hoppe S, Winter J, Tichy D, Zeller C, Küpper MD, et al. Performance Evaluation of MHC Class-I Binding Prediction Tools Based on an Experimentally Validated MHC-Peptide Binding Data Set. Cancer Immunol Res. 2019;7:719–36. [DOI] [PubMed]

68. Roomp K, Antes I, Lengauer T. Predicting MHC class I epitopes in large datasets. BMC Bioinformatics. 2010;11:90. [DOI] [PubMed] [PMC]

69. Zhang H, Lundegaard C, Nielsen M. Pan-specific MHC class I predictors: a benchmark of HLA class I pan-specific prediction methods. Bioinformatics. 2009;25:83–9. [DOI] [PubMed] [PMC]

70. Singh SP, Mishra BN. Prediction of MHC binding peptide using Gibbs motif sampler, weight matrix and artificial neural network. Bioinformation. 2008;3:150–5. [DOI] [PubMed] [PMC]

71. Wang M, Claesson MH. Classification of human leukocyte antigen (HLA) supertypes. Methods Mol Biol. 2014;1184:309–17. [DOI] [PubMed] [PMC]

72. Shen Y, Parks JM, Smith JC. HLA Class I Supertype Classification Based on Structural Similarity. J Immunol. 2023;210:103–14. [DOI] [PubMed]

73. Singh SP, Mishra BN. Major histocompatibility complex linked databases and prediction tools for designing vaccines. Hum Immunol. 2016;77:295–306. [DOI] [PubMed]

74. Tscharke DC, Croft NP, Doherty PC, Gruta NLL. Sizing up the key determinants of the CD8(+) T cell response. Nat Rev Immunol. 2015;15:705–16. [DOI] [PubMed]

75. Nosrati M, Mohabatkar H, Behbahani M. Introducing of an integrated artificial neural network and Chou's pseudo amino acid composition approach for computational epitope-mapping of Crimean-Congo haemorrhagic fever virus antigens. Int Immunopharmacol. 2020;78:106020. [DOI] [PubMed]

76. Grifoni A, Weiskopf D, Ramirez SI, Mateus J, Dan JM, Moderbacher CR, et al. Targets of T Cell Responses to SARS-CoV-2 Coronavirus in Humans with COVID-19 Disease and Unexposed Individuals. Cell. 2020;181:1489–501.e15. [DOI] [PubMed] [PMC]

77. Nielsen M, Lundegaard C, Lund O, Keşmir C. The role of the proteasome in generating cytotoxic T-cell epitopes: insights obtained from improved predictions of proteasomal cleavage. Immunogenetics. 2005;57:33–41. [DOI] [PubMed]

78. Brouwenstijn N, Serwold T, Shastri N. MHC class I molecules can direct proteolytic cleavage of antigenic precursors in the endoplasmic reticulum. Immunity. 2001;15:95–104. [DOI] [PubMed]

79. van Hall T, van Bergen J, van Veelen PA, Kraakman M, Heukamp LC, Koning F, et al. Identification of a novel tumor-specific CTL epitope presented by RMA, EL-4, and MBL-2 lymphomas reveals their common origin. J Immunol. 2000;165:869–77. [DOI] [PubMed]

80. Chen W, Norbury CC, Cho Y, Yewdell JW, Bennink JR. Immunoproteasomes shape immunodominance hierarchies of antiviral CD8(+) T cells at the levels of T cell repertoire and presentation of viral antigens. J Exp Med. 2001;193:1319–26. [DOI] [PubMed] [PMC]

81. Assarsson E, Sidney J, Oseroff C, Pasquetto V, Bui H, Frahm N, et al. A quantitative analysis of the variables affecting the repertoire of T cell specificities recognized after vaccinia virus infection. J Immunol. 2007;178:7890–901. [DOI] [PubMed]

82. Holzhütter HG, Frömmel C, Kloetzel PM. A theoretical approach towards the identification of cleavage-determining amino acid motifs of the 20 S proteasome. J Mol Biol. 1999;286:1251–65. [DOI] [PubMed]

83. Nussbaum AK, Kuttler C, Hadeler KP, Rammensee HG, Schild H. PAProC: a prediction algorithm for proteasomal cleavages available on the WWW. Immunogenetics. 2001;53:87–94. [DOI] [PubMed]

84. Hakenberg J, Nussbaum AK, Schild H, Rammensee H, Kuttler C, Holzhütter H, et al. MAPPP: MHC class I antigenic peptide processing prediction. Appl Bioinformatics. 2003;2:155–8. [PubMed]

85. Reche PA, Reinherz EL. PEPVAC: a web server for multi-epitope vaccine development based on the prediction of supertypic MHC ligands. Nucleic Acids Res. 2005;33:W138–42. [DOI] [PubMed] [PMC]

86. Brusic V, Bajic VB, Petrovsky N. Computational methods for prediction of T-cell epitopes--a framework for modelling, testing, and applications. Methods. 2004;34:436–43. [DOI] [PubMed]

87. Heemels MT, Ploegh H. Generation, translocation, and presentation of MHC class I-restricted peptides. Annu Rev Biochem. 1995;64:463–91. [DOI] [PubMed]

88. Dönnes P, Kohlbacher O. Integrated modeling of the major events in the MHC class I antigen processing pathway. Protein Sci. 2005;14:2132–40. [DOI] [PubMed] [PMC]

89. Tenzer S, Peters B, Bulik S, Schoor O, Lemmel C, Schatz MM, et al. 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] [PubMed] [PMC]

90. Adotévi O, Mollier K, Neuveut C, Cardinaud S, Boulanger E, Mignen B, et al. Immunogenic HLA-B*0702-restricted epitopes derived from human telomerase reverse transcriptase that elicit antitumor cytotoxic T-cell responses. Clin Cancer Res. 2006;12:3158–67. [DOI] [PubMed]

91. Hundemer M, Schmidt S, Condomines M, Lupu A, Hose D, Moos M, et al. Identification of a new HLA-A2-restricted T-cell epitope within HM1.24 as immunotherapy target for multiple myeloma. Exp Hematol. 2006;34:486–96. [DOI] [PubMed] [PMC]

92. Mustafa AS, Shaban FA. ProPred analysis and experimental evaluation of promiscuous T-cell epitopes of three major secreted antigens of Mycobacterium tuberculosis. Tuberculosis (Edinb). 2006;86:115–24. [DOI] [PubMed]

93. Sundar K, Boesen A, Coico R. Computational prediction and identification of HLA-A2.1-specific Ebola virus CTL epitopes. Virology. 2007;360:257–63. [DOI] [PubMed]

94. Wen J, Jiang L, Zhou J, Yan H, Fang D. Computational prediction and identification of dengue virus-specific CD4(+) T-cell epitopes. Virus Res. 2008;132:42–8. [DOI] [PubMed] [PMC]

95. Baruah V, Bose S. Immunoinformatics-aided identification of T cell and B cell epitopes in the surface glycoprotein of 2019-nCoV. J Med Virol. 2020;92:495–500. [DOI] [PubMed] [PMC]

96. Wang M, Lamberth K, Harndahl M, Røder G, Stryhn A, Larsen MV, et al. CTL epitopes for influenza A including the H5N1 bird flu; genome-, pathogen-, and HLA-wide screening. Vaccine. 2007;25: 2823–31. [DOI] [PubMed]

97. Perez EE, Wang J, Miller JC, Jouvenot Y, Kim KA, Liu O, et al. Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. Nat Biotechnol. 2008;26:808–16. [DOI] [PubMed] [PMC]

98. Tang ST, Wang M, Lamberth K, Harndahl M, Dziegiel MH, Claesson MH, et al. MHC-I-restricted epitopes conserved among variola and other related orthopoxviruses are recognized by T cells 30 years after vaccination. Arch Virol. 2008;153:1833–44. [DOI] [PubMed] [PMC]

99. Lehnert E, Tampé R. Structure and Dynamics of Antigenic Peptides in Complex with TAP. Front Immunol. 2017;8:10. [DOI] [PubMed] [PMC]

100. Calis JJA, Maybeno M, Greenbaum JA, Weiskopf D, Silva ADD, Sette A, et al. Properties of MHC class I presented peptides that enhance immunogenicity. PLoS Comput Biol. 2013;9:e1003266. [DOI] [PubMed] [PMC]

101. Lasso P, Cárdenas C, Guzmán F, Rosas F, Thomas MC, López MC, et al. Effect of secondary anchor amino acid substitutions on the immunogenic properties of an HLA-A*0201-restricted T cell epitope derived from the Trypanosoma cruzi KMP-11 protein. Peptides. 2016;78:68–76. [DOI] [PubMed]

102. Rasmussen M, Fenoy E, Harndahl M, Kristensen AB, Nielsen IK, Nielsen M, et al. Pan-Specific Prediction of Peptide-MHC Class I Complex Stability, a Correlate of T Cell Immunogenicity. J Immunol. 2016;197:1517–24. [DOI] [PubMed] [PMC]

103. Antunes DA, Devaurs D, Moll M, Lizée G, Kavraki LE. General Prediction of Peptide-MHC Binding Modes Using Incremental Docking: A Proof of Concept. Sci Rep. 2018;8:4327. [DOI] [PubMed] [PMC]

104. Kuai R, Ochyl LJ, Bahjat KS, Schwendeman A, Moon JJ. Designer vaccine nanodiscs for personalized cancer immunotherapy. Nat Mater. 2017;16:489–96. [DOI] [PubMed] [PMC]

105. Kapadia CH, Perry JL, Tian S, Luft JC, DeSimone JM. Nanoparticulate immunotherapy for cancer. J Control Release. 2015;219:167–80. [DOI] [PubMed]

106. Bissati KE, Zhou Y, Paulillo SM, Raman SK, Karch CP, Roberts CW, et al. Protein nanovaccine confers robust immunity against *Toxoplasma*. NPJ Vaccines. 2017;2:24. [DOI] [PubMed] [PMC]

107. Zhu X, Vo C, Taylor M, Smith BR. Non-spherical micro-and nanoparticles in nanomedicine. Mater Horiz. 2019;6:1094–121. [DOI]

108. Liu J, Miao L, Sui J, Hao Y, Huang G. Nanoparticle cancer vaccines: Design considerations and recent advances. Asian J Pharm Sci. 2020;15:576–90. [DOI] [PubMed] [PMC]

109. Jha K, Jaishwal P, Yadav TP, Singh SP. Self-assembling of coiled-coil peptides into virus-like particles: Basic principles, properties, design, and applications with special focus on vaccine design and delivery. Biophys Chem. 2025;318:107375. [DOI] [PubMed]

110. Saleh T, Shojaosadati SA. Multifunctional nanoparticles for cancer immunotherapy. Hum Vaccin Immunother. 2016;12:1863–75. [DOI] [PubMed] [PMC]

111. Schudel A, Francis DM, Thomas SN. Material design for lymph node drug delivery. Nat Rev Mater. 2019;4:415–28. [DOI] [PubMed] [PMC]

112. Mu R, Zhu D, Abdulmalik S, Wijekoon S, Wei G, Kumbar SG. Stimuli-responsive peptide assemblies: Design, self-assembly, modulation, and biomedical applications. Bioact Mater. 2024;35:181–207. [DOI] [PubMed] [PMC]