Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Immunogenic epitope prediction to create a universal influenza vaccine

R.R. Mintaev [a,b,*], D.V. Glazkova [a], E.V. Bogoslovskaya [a], G.A. Shipulin [a]

[a] Federal State Budgetary Institution «Center for Strategic Planning and Management of Medical and Biological Health Risks», 119833, Federal Medical-Biological Agency, Moscow, Russia
[b] I.Mechnikov Research Institute of Vaccines and Sera, 105064, Moscow, Russia

ARTICLE INFO

ABSTRACT

Influenza virus is one of the most rapidly evolving human pathogens and causes significant morbidity and mortality worldwide. This feature enables the virus to avoid natural or vaccine-induced immunity. For this reason, there is an intensive search for new approaches to create a universal influenza vaccine. Here, we propose pipelines based on modern prediction algorithms that allowed us to select 10 B-cell epitopes, 10 CD8+ T-cell epitopes and 6 CD4+ T-cell epitopes from influenza viruses that were characterized by high conservation and antigenicity. These epitopes could be used to create universal vaccines against influenza viruses. In addition, the scripts used in these pipelines are universal and can be used to select epitopes from other pathogens.

1. Introduction

One of the most socially and economically significant diseases in the world is the disease caused by influenza viruses. According to the World Health Organization (WHO) estimates (https://www.who.int/influenza/surveillance_monitoring/bod/en/), annual influenza epidemics lead to approximately 3–5 million cases of serious illnesses and 290 000 to 650 000 deaths worldwide. The morbidity rate varies among different age groups and reaches its maximum in children under 5 years old (Nair et al., 2011), in old people, and in persons with chronic diseases of the respiratory and cardiovascular systems.

Currently WHO recommends vaccination as the primary means of influenza control. However, the high variability in influenza viruses leads to the need for an annual renewal of the strain composition in vaccines. In fact, the manufactured vaccines do not always correspond to the circulating virus of the next season, which leads to their low efficiency. This is the reason for an intensive search for new approaches to create a universal influenza vaccine designed to provide protection against a wide range of influenza A and B viruses, as well as against potentially pandemic strains in animals that have not yet been transmitted to humans. There are different approaches for creating universal vaccines, which have been discussed in detail in several reviews (Vemula et al., 2017; Nachbagauer et al., 2017; Trucchi et al., 2019). The main immunological targets in the given approaches are the most highly conserved epitopes that can be delivered by introducing whole virus proteins or a set of short epitope peptides.

An interesting method with great potential for the creation of a universal vaccine is the use of a combination of separate epitopes from different influenza antigens. This enables the activation of both humoral and T-cell immune responses, in addition to controlling side effects by reducing the general antigen load, ultimately achieving cross-reactive protection against strains of different subtypes and types using conserved epitopes. For example, the use of a mixture of epitope-based peptides in the FLU-v (Pleguezuelos et al., 2020) vaccine or of a peptide sequence that consists of a set of epitopes in the M001 vaccine (Lowell et al., 2017) enabled the production of a cross-reactive immune response and protection against different groups of viruses both in model animals and in humans. However, while some of the problems with these vaccines (i.e., low immunogenicity and the lack of broad-spectrum cross-protection) can be addressed through a more intensive selection of immunogenic epitopes using a comprehensive range of bioinformatics tools, these issues remain unsolved.

The aim of the given work was to develop pipelines for selecting optimal conserved linear epitopes of B- and T-cells based on modern algorithms of prediction that could be used to create a universal vaccine against influenza virus. The given approach may also be useful for selecting epitopes from antigens of any other pathogen.

2. Material and methods

2.1. Sequences and databases

Protein sequence records were obtained from the influenza virus database (Influenza Research Database, www.fludb.org). Epitopes were selected from conserved regions of the HA, NP, and M1 viral proteins. These proteins are major viral proteins, as well as the most important influenza virus antigens that can cause a cross-reactive B- and T-cell immune response. The number of sequence records were as follows (as of April 2020): hemagglutinin type A subtype H1, 15653; H3, 13360; type B, 3448; NP type A, 12721; type B, 1601, M1 type A, 3951.

The studied epitopes were mapped on the surface of HA protein structures using the molecular visualization program (PyMOL, http://pymol.org). PDB records of hemagglutinin were obtained from the PDB database (http://www.rcsb.org). The record IDs were as follows: type A subtype H1, PDB: 5K9O, H3, PDB: 4KVN; and type B, PDB: 4FQK.

The Immune Epitope Database (IEDB, http://www.iedb.org) was used to evaluate the literature available for the epitopes.

The distribution frequencies of MHC (major histocompatibility complex) molecule allelic variants were taken from the Allele Frequency Net Database (AFND) (Takeshita et al., 2014).

To predict and mark the positions of the epitopes, the following strain sequences were used as references for position numbering: influenza A subtype H1, A/California/07/2009; H3, A/Victoria/208/2009; and influenza B, B/Brisbane/60/2008.

2.2. B-cell and T-cell epitope prediction

B-cell epitopes with a score higher than 0.5 were selected from HA using the BepiPred-2.0 program (Jespersen et al., 2017). The antigenicity of the conserved epitope regions was confirmed using the VaxiJen 2.0 server with a score higher than 0.5 (Doytchinova and Flower, 2001). A Python script was written to automatically select the epitope (https://github.com/lioj/bioinformatics/blob/master/py/B_Cell_epitopes_prediction.py). The availability of the epitopes on the surface of the HA molecule was analyzed using the PyMOL program. The accessibility of the epitope on the surface of the antigen globule for antibodies was assessed visually: epitopes should be completely located on the surface of the HA trimer molecule.

A set of algorithms was used to predict T-cell epitopes: (1) for CD8+, an algorithm based on a neural network was used (Nielsen et al., 2003; Andreatta and Nielsen, 2016); (2) for CD4+, a consensus method (Wang et al., 2008) and an algorithm based on a neural network were used (Jensen et al., 2018). Each method used an IC50 of less than 50 nM or a percentile less than 1 as a selection threshold (Konstantinou, 2017). The selected epitopes were tested for antigenicity using the VaxiJen 2.0 server (Doytchinova and Flower, 2001). The epitopes were also tested for allergenicity using the AllerTop 2.0 (Dimitrov et al., 2013) and AllergenFP 1.0 servers (Dimitrov et al., 2014), as well as for toxicity using the ToxinPred server (Gupta et al., 2013). A Python script was written to automatically select CD4+ epitopes (https://github.com/lioj/bioinformatics/blob/master/py/sele_MHCII.py).

2.3. Analysis of conservation

The IEDB database of immunogenic epitopes provides only an approximate estimation of the epitope conservation, so we conducted our own estimation using the method described below. The MAFFT program (http://mafft.cbrc.jp/alignment/software/) was used to align the sequences of the amino acid residues based on the FFT–NS–2 algorithm. Poorly aligned regions were additionally edited in the Jalview program (Waterhouse et al., 2009). The Shannon entropy was calculated using scripts written in Python in order to estimate the conservation of the alignment column. To calculate the number of types of k amino acid residues that could appear in some alignment position with a $1/k$

probability, the formula $k = 2S$ (S is the information entropy of this column calculated by the formula $S = -\sum(q_i \log q_i)$, where q_i is the frequency of occurrence of the i-th most common amino acid residue at a given position in a set of aligned sequences) was used. Summarization for entropy was performed for all types of residues found in the alignment column under study. Thus, the epitope was deemed to be conserved if all of the positions included therein met the criterion of $k < 1.3$ (Shenkin et al., 1991; Mintaev et al., 2014).

2.4. Docking

The immunogenicity of T-cell epitopes is determined by their affinity to MHC molecules. To estimate the affinity of CTL epitopes, a series of consequential procedures were performed for docking peptides to the predicted corresponding MHCI molecules: A*02:01 (PDB: 1B0R), A*02:06 (PDB: 3OXR), A*03:01 (PDB: 6O9B), B*07:02 (PDB: 6AT5), B*18:01 (PDB: 6MT3), B*35:01 (PDF: 4LNR), C*12:03 (PDF: 1EFX). The structures of the peptides were obtained using a script for PyMOL. In the first stage, blind rigid docking was performed using FRODOCK 2.0 (Ramírez-Aportela et al., 2016) and then flexible redocking and counting of the binding energy using FiberDock program was performed (Mashiach et al., 2010).

3. Results

3.1. Selection of B-cell epitopes

B cell epitopes can be defined as a surface accessible group of amino acid residues that are recognized by antibodies or a B cell receptor. There are two categories of B cell epitopes as follows: linear epitopes are represented by a continuous sequence of amino acid residues, and conformational epitopes are represented by 1–5 linear segments of amino acid residues, distantly separated in the protein sequence but close in space, which form the site of interaction with the antibody (Potocnakova et al., 2016). Most B cell epitopes are conformational, but there are several limitations. Conformational B cell epitopes are difficult to use without a molecular context because correct folding of their structure is important (Van Regenmortel, 2009). In contrast, linear B cell epitopes can be used without an antigen context. In addition, linear B cell epitopes are relatively small, which simplifies the discovery of conserved variants. Thus, linear B cell epitopes are optimal for developing epitope universal vaccines.

Early methods for predicting linear B-cell epitopes used a correlation between one property of an amino acid residue (hydrophobicity, flexibility, polarity) and its probability of being part of an epitope (Karplus and Schulz, 1985; Emini et al., 1985; Parker et al., 1986). Later, methods were developed that could take into account several amino acid residue properties at once in prediction models (Saha and Raghava, 2004; Larsen et al., 2006), or used machine learning to detect sequence fragments similar to those of earlier known epitopes (El-Manzalawy et al., 2008). However, with the accumulation of data from X-ray structural analyses of antigen-antibody arrays, machine learning algorithms became available that could use this interaction data to search for linear epitopes in antigen sequences. Currently BepiPred 2.0 predicts linear B-cell epitopes based on a random forest algorithm trained on epitopes annotated from antibody-antigen protein structures. Data on such epitopes are considered to be of higher quality, which makes it possible to improve predictions compared to other tools: HMM or LBTope (Jespersen et al., 2017; Larsen et al., 2006; Singh et al., 2013). This program uses a method that assigns each amino acid residue a certain antigenicity rating. Therefore, depending on the selected threshold, the lengths of the predicted epitopes, which consist of consecutive amino acid residues with individual scores higher than the given threshold, may change. However, it is known from the Bcipep database (Saha et al., 2005) that the length distribution of experimentally confirmed linear B-cell epitopes is in the range of 6–35 aa. Thus, in long sequences that the BepiPred 2.0 program

has defined as an epitope, it is possible to distinguish a subsequence that meets additional selection criteria, such as conserved residues and accessibility. Moreover, one needs to use an additional independent method to confirm the antigenic properties of the selected epitope, since the probability of predicting a nonimmunogenic epitope remains when using one method. For example, it has been shown that using the VaxiJen 2.0 program allows for the prediction of immunogenic antigens with a sensitivity of 74% and specificity of 71% when selecting an antigen score of more than 0.5 (Doytchinova and Flower, 2001). The algorithm of the given program belongs in the category of machine learning as well, but the implementation of its antigenicity estimation differs from the existing algorithms (Doytchinova and Flower, 2001), so it is often used as a tool for confirming antigenicity (this is reviewed in detail in (Nevena Zaharieva et al., 2017)).

For a universal vaccine, it is also necessary to take into account conservation. In addition, it is important to take into account epitope localization, since it is necessary that the epitope is available on the surface of the protein for direct interaction with antibodies. Thus, when choosing the final set of B-cell epitopes, several epitope properties should be taken into account simultaneously, including antigenicity, conservation, and availability for antibody binding.

The humoral immune response protects organisms from viral infection by producing antibodies against pathogens (Warrington et al., 2011). Hemagglutinin is one of the major surface proteins of influenza virus and is a known target of neutralizing antibodies. Such antibodies can block the adsorption of viral particles on the cell surface or the fusion of the cell membrane and viral particles, thereby preventing infection of new cells. In this regard, HA is an excellent candidate for the selection of B-cell epitopes (Knossow et al., 2006).

Figure 1 shows the pipeline developed for predicting linear B-cell epitopes. The BepiPred 2.0 program was given HA sequences from influenza type A subtype H1 (A/California/07/2009) and H3 (A/Victoria/208/2009), and influenza B (B/Brisbane/60/2008). A total of 41 sequences with lengths from 8 to 63 aa and potential antigenic properties were predicted by BepiPred 2.0. Then, a script was written in Python (https://github.com/lioj/bioinformatics/blob/master/py/B_Cell_epitopes_prediction.py) to select fragments (longer than 6 aa) from the predicted sequences that met the conservation criteria ($2^S < 1.3$, see methods) and had an antigen score of more than 0.5 based on the results of an analysis using the VaxiJen 2.0 server (Table 1). A total of 27 fragments obtained using the above script were mapped to the HA protein structure using the PyMOL program. The final set was made up of 10 epitopes available on the outside surface of the HA trimer (Table 2).

A total of 5 B-cell epitopes were predicted for influenza type B and 5 were predicted for influenza type A, 3 of which were for the H1 subtype and 2 of which were for the H3 subtype. Since BepiPred 2.0 evaluates each amino acid residue, the median scores were calculated for each sequence. The medians for the predicted epitopes were in the range of 0.51–0.64. According to the VaxiJen 2.0 data, high antigenicity was observed for type B in epitopes 64-TRGKLCP-71 (2.12) and 415-NSLSELEVKN-425 (1.29), and for type A subtype H1 in epitope 41-VTVTHSVNLLE-52 (0.74) and subtype H3 in epitope 378-GQAADLKSTQA-389 (1.12).

3.2. Selection of T-cell epitopes

T-cell epitopes are linear peptides that are integrated into the MHC molecules for presentation to T-cells (Peters et al., 2020). The previous programs for predicting T-cell epitopes from a protein sequence were based on the use of patterns specific to the MHC allele. Such patterns contained information about anchor amino acid residues that determine the interaction with the binding site of the MHC molecule (Sette et al., 1989; Ruppert et al., 1993). With the accumulation of information on epitope and nonepitope sequences, programs based on machine learning have begun gaining popularity (Nielsen et al., 2003; Andreatta and Nielsen, 2016; Doytchinova and Flower 2001; Lundegaard et al., 2008; Zhang et al., 2010). According to comparative studies, the algorithm based on an artificial neural network for CD8+ epitopes had the greatest predictive power in several reports (Paul et al., 2020; Peters et al., 2020; Lundegaard et al., 2008). Predicting CD4+ epitopes is a more challenging task due to the ability of MHCII molecules to bind to long peptides, therefore, we used two algorithms at once to enhance the quality. In comparison with prediction methods, it was shown that machine learning algorithms (the consensus method (Wang et al., 2008) and the
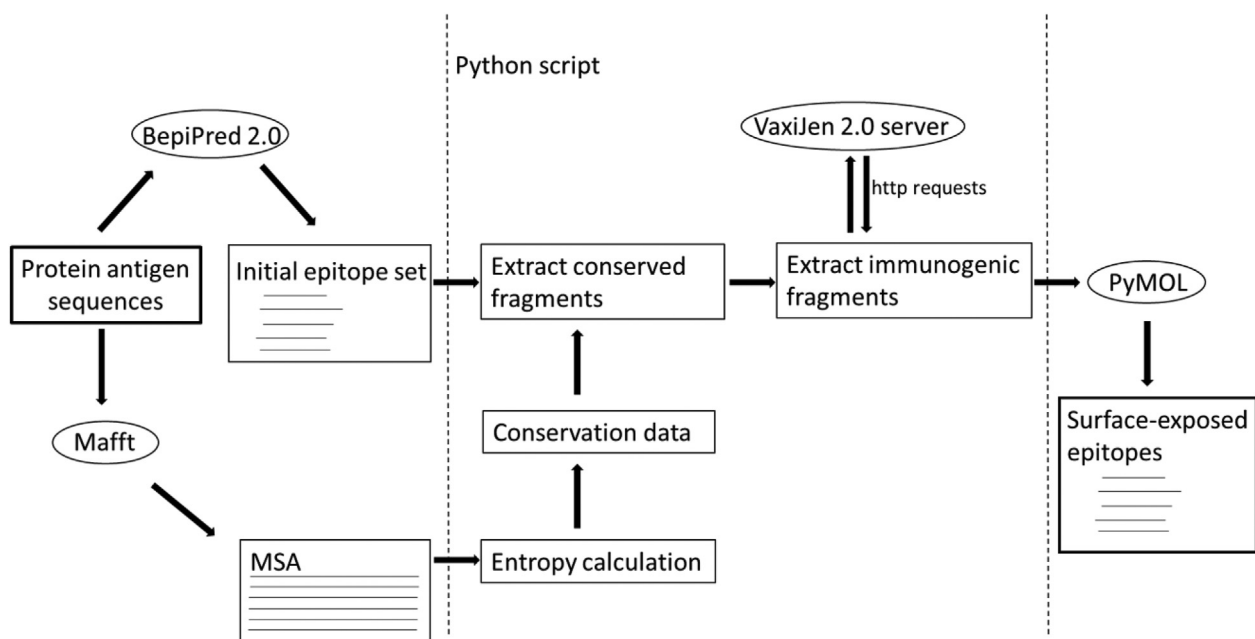
Figure 1. Pipeline for predicting B-cell epitopes. Third-party software used in the pipeline is shown in ovals. The script was fed the output results of the BepiPred 2.0 program and multiple sequence alignments (MSAs) generated using the Mafft program. The script calculated the conservation of the positions in the reference sequence, which were used to obtain the conserved fragments in the original set of epitopes. Then, immunogenic fragments of the longest length were selected using requests to the VaxiJen 2.0 server. The obtained epitopes were mapped to the structure of HA to select epitopes available on the surface of the protein.

Table 1. Epitopes from HA for subtypes H1, H3 of type A and type B, predicted using BepiPred 2.0 program. The underline shows the areas available on the surface.

#	Type/Subtype	Position	Length	Sequence	Vaxi Jen	Bepi Pred Median
1	A/H1	41	11	VTVTHSVNLLE	0.7442	0.5182
2	A/H1	117	19	ELREQLSSVSSFERFEIFP	0.661	0.6258
3	A/H1	363	18	DGWYGYHHQNEQGSGYAA	0.5173	0.6399
4	A/H1	407	8	FTAVGKEF	1.1817	0.5103
5	A/H1	493	9	MESVKNGTY	0.8339	0.5439
6	A/H3	180	7	LNVTMPN	0.724	0.509
7	A/H3	230	8	IPNIGSRP	1.3033	0.5619
8	A/H3	305	10	PNDKPFQNVN	0.6507	0.5659
9	A/H3	333	8	ATGMRNVP	0.5157	0.5733
10	A/H3	364	13	DGWYGFRHQNSEG	0.7046	0.6076
11	A/H3	378	11	GQAADLKSTQA	1.1229	0.6209
12	A/H3	412	7	EKEFSEV	0.7049	0.5234
13	A/H3	451	9	HTIDLTDSE	1.6395	0.5127
14	A/H3	473	8	ENAEDMGN	0.5454	0.5232
15	A/H3	510	12	EALNNRFQIKGV	1.2026	0.5589
16	B	56	7	FANLKGT	1.0567	0.5652
17	B	64	7	TRGKLCP	2.1198	0.5757
18	B	72	10	CLNCTDLDVA	2.0351	0.5849
19	B	102	21	VRPVTSGCFPIMHDRTKIRQL	0.7401	0.5418
20	B	153	8	GTSGSCPN	0.5347	0.5742
21	B	249	8	*QTEDGGLP*	0.6174	0.6607
22	B	301	12	SLPLIGEADCLH	0.518	0.6076
23	B	315	13	YGGLNKSKPYYTG	0.7142	0.598
24	B	349	10	GTKYRPPAKL	0.5927	0.6077
25	B	374	35	GGWEGMIAGWHGYTSHGAHGVAVAADLKSTQEAIN	0.5433	0.5548
26	B	415	35	NSLSELEVKNLQRLSGAMDELHNEILELDEKVDDL	0.5895	0.5218
27	B	468	12	GIINSEDEHLLA	0.5456	0.5723

Table 2. Conserved B-cell epitopes located on the surface of HA, predicted using the BepiPred 2.0 and VaxiJen 2.0 programs.

#	Type/Subtype	Position	Length	Sequence	VaxiJen	BepiPred 2.0
1	A/H1	41	11	VTVTHSVNLLE	0.7442	0.5182
2	A/H1	117	13	SSVSSFERFEIFP	0.61	0.6258
3	A/H1	363	18	DGWYGYHHQNEQGSGYAA	0.5173	0.6399
4	A/H3	180	7	LNVTMPN	0.724	0.509
5	A/H3	378	11	GQAADLKSTQA	1.1229	0.6209
6	B	64	7	TRGKLCP	2.1198	0.5757
7	B	153	8	GTSGSCPN	0.5347	0.5742
8	B	315	7	GLNKSKP	0.5675	0.598
9	B	374	23	HGYTSHGAHGVAVAADLKSTQEA	0.773	0.5548
10	B	415	10	NSLSELEVKN	1.29	0.5218

method based on a neural network (Jensen et al., 2018)) cope with this task best (Konstantinou, 2017; Peters et al., 2020; Lin et al., 2008; Wang et al., 2008). The selected algorithms evaluate T-cell epitopes based on an IC50 or percentile, which makes it possible to use more meaningful criteria for selecting epitopes. For additional independent confirmation of antigenicity (as with the B-cell epitopes), the VaxiJen 2.0 service was used (Doytchinova and Flower 2001).

The NP and M1 protein sequences from influenza type A (A/California/07/2009 strain), and the NP protein sequence from influenza B (B/Brisbane/60/2008 strain) were submitted as the input. Several independent algorithms were used in each case (artificial neural network, consensus method, VaxiJen 2.0 program) to improve the prediction of CD8+ and CD4+ epitopes (Doytchinova and Flower, 2001; Nielsen et al., 2003; Andreatta and Nielsen, 2016; Wang et al., 2008; Jensen et al., 2018). The overlap in the predicted epitope sets was used for further selection procedures. When using an artificial neural network and a consensus method, the score was generated based on the parameter

defined by the binding strength of the peptide, which can be expressed as the IC50 or percentile in prediction programs. The IC50 is the concentration of a peptide in nM required to bind 50% of MHC molecules. The percentile is a derived function from the IC50, which compares the value to the IC50 of a random set of peptides from the UniProt database. In both cases, the lower the parameter value is, the stronger the expected binding force of the peptide is (Konstantinou, 2017). A score of higher than 0.5 was used as the selection criterion when using VaxiJen 2.0 server (Doytchinova and Flower, 2001).

T-cells recognize epitopes in the peptides presented by MHC molecules, which represent information about the actual protein composition of the cell expressing MHC. MHC molecules have extreme variability at the population level, which leads to different peptide affinities. In humans, MHC molecules are called human leukocyte antigens (HLA), and approximately one thousand allelic variants are known. Any individual has a unique set of 8 genes with two alleles, i.e., genome can have up to 16 different variants; at the population level, a

separate variant appears with different frequency in different ethnic groups. However, some MHC molecule variants have similar specificity for peptides due to physical and chemical similarity in the binding site, for which they can be grouped into subtypes (Reche and Reinherz, 2007). Thus, when selecting epitopes, it is possible to focus on the distribution of allele subtypes in populations to create an optimal set. For the given purposes, the AFND database was used, from which allele frequencies for the Moscow region were obtained. The top alleles were selected for the set according to their prevalence, and together they made up more than half of the distribution.

An artificial neural network was used to predict CD8+ epitopes (Nielsen et al., 2003; Andreatta and Nielsen, 2016). The prediction was made for MHCI alleles from the Moscow region population as follows: A*01, A*02, A*03, A*24, B*07, B*08, B*13, B*18, B*35, B*44, C*04, C*06, C*07, C*12. A total of 57 epitopes of 9 aa in length with an IC50 of less than 50 nM were predicted based on the results of the above-mentioned program. It has been experimentally shown in the work of Sette et al. that peptides with an IC50 of less than 50 nM have a high affinity for MHCI molecules (Sette et al., 1994). Further, epitopes were tested for antigenicity using the VaxiJen 2.0 server (Doytchinova and Flower, 2001). A total of 31 epitopes had an antigenicity score higher than 0.5 (Table 3). After excluding variable epitopes containing positions with entropies above the threshold, 16 epitopes remained. Finally, toxicity was calculated for each epitope using ToxinPred (Gupta et al., 2013), as well as allergenicity using the AllerTop 2.0 (Dimitrov et al., 2013) and AllergenFP 1.0 programs (Dimitrov et al., 2014). Because AllerTop 2.0 and AllergenFP 1.0 showed opposite results for some epitopes, only the epitopes for which the allergenicity prediction coincided were removed. After evaluating the toxicity and allergenicity, 10 CD8+ epitopes were selected: 4 from the NP protein of influenza A, 5 from the NP protein of influenza B, and 1 from the M1 protein of influenza A (Table 4).

The consensus method (Wang et al., 2008) and the neural network-based method (Jensen et al., 2018) were used to predict CD4+ epitopes. The following MHCII alleles were used: DQB1*06:02, DRB1*15:01, DPB1*02:01, DQB1*03:02, DPB1*05:01, DRB1*13:01, DPA1*01, DPB1*04:02, DPA1*01:03, DRB1*11:01, DPB1*04:01, DQA1*01:02, DPB1*01:01, DPA1*03:01, DQA1*05:01, DQB1*03:01, DPA1*02:01, DQA1*03:01, DRB1*07:01. Epitopes of 13–18 aa in length were predicted. According to Southwood et al. (1998), epitopes with an IC50 less than 1000 nM or a percentile less than 10 have a high affinity for corresponding MHCII molecules, however, due to a large quantity of predicted peptides, the threshold was reduced to an IC50 of 50 nM Figure 2 shows the developed pipeline. Based on the results of the programs, 1,324 epitopes and 1,758 estimates for their interactions with different selected MHCII alleles that met the program threshold criteria were obtained.

Among the predicted epitopes, variable epitopes were first discarded, then a request to the VaxiJen 2.0 server was made for each remaining epitope to verify their antigenicity. Further, the epitopes were grouped based on similar properties to highlight the optimal properties among the predicted epitopes.

The MHCII molecule is able to bind to long peptides (lengths from 13 to 25 aa) because the ends of its binding sites are more 'open' than those of MHCI (Brown et al., 1993). However, it has been shown that the main contribution to the interaction is made by a fragment of 9 aa within the peptide, the so-called interaction core (Jones et al., 2006). Algorithms take the interaction core into account and specify it for each prediction. Therefore, the epitopes were divided into groups based on a common interaction core and sorted both within these groups based on the IC50 value and between the groups based on the average IC50 of the group. A Python script was written to implement the given algorithm, including conservative selection and analysis using VaxiJen 2.0 (https://github.com/lioj/bioinformatics/blob/master/py/sele_MHCII.py).

Of the 649 conserved epitopes, 355 had antigenicity above the threshold of 0.5 according to VaxiJen 2.0. After grouping the epitopes

Table 3. MHC I epitopes predicted using artificial neural network and VaxiJen 2.0

| #	Type: Protein	Position	Peptide	allele	ANN, IC50	VaxiJen
1	A:NP	44	CTELKLSDY	HLA-A*01:01	11,02	1,7
2	A:NP	140	HSNLNDATY	HLA-B*35:01	25,4	0,53
3	A:NP	146	ATYQRTRAL	HLA-B*07:02	40,75	0,59
4	A:NP	172	LPRRSGAAG	HLA-B*07:02	35,29	1,19
5	A:NP	217	VAYERMCNI	HLA-C*12:03	10,25	0,95
6	A:NP	221	RMCNILKGK	HLA-A*03:01	40,75	0,51
7	A:NP	251	AEIEDLIFL	HLA-A*02:06	44,07	1,03
8	A:NP	258	FLARSALIL	HLA-A*02:01	43,19	0,57
9	A:NP	283	LAVASGHDF	HLA-B*35:01	34,83	0,62
10	A:NP	342	RVSSFIRGK	HLA-A*03:01	34,31	0,75
11	A:NP	380	ELRSRYWAI	HLA-B*08:01	6	1,67
12	A:NP	412	FSVQRNLPF	HLA-B*35:01	39	0,88
13	A:NP	418	LPFERATVM	HLA-B*35:01	6,39	0,79
14	A:NP	458	FQGRGVFEL	HLA-A*02:06	7,97	1,28
				HLA-A*02:01	36,2	
15	A:NP	473	NPIVPSFDM	HLA-B*35:01	43,9	0,58
16	B:NP	36	ATLAPPSNK	HLA-A*03:01	17,66	0,56
17	B:NP	82	MVVKLGEFY	HLA-B*35:01	42,41	0,6
18	B:NP	165	YFSPIRITF	HLA-C*12:03	25,34	1,7
19	B:NP	242	VAIKGGGTL	HLA-C*12:03	44,39	0,7
20	B:NP	325	RPSVASKVV	HLA-B*07:02	23,37	0,74
21	B:NP	352	YSMVGYEAM	HLA-B*35:01	22,06	1
22	B:NP	357	YEAMALYNM	HLA-B*18:01	6,18	0,69
23	B:NP	363	YNMATPVSI	HLA-C*12:03	28,68	0,52
24	B:NP	392	AAYEDLRVL	HLA-C*12:03	32,71	0,59
25	B:NP	429	GMGAALMSI	HLA-A*02:01	39,58	0,58
26	A: M1	58	GILGFVFTL	HLA-A*02:01	15,71	0,76
				HLA-A*02:06	41,19	
27	A: M1	130	LIYNRMGTV	HLA-C*12:03	37,67	0,6
				HLA-A*02:06	55,38	0,6
28	A: M1	138	VTTEAAFGL	HLA-A*02:06	29,53	1,28
29	A: M1	173	IRHENRMVL	HLA-C*06:02	17,89	1,07
				HLA-C*07:01	26,63	1,07
30	A: M1	179	MVLASTTAK	HLA-A*03:01	15,17	0,96
31	A: M1	181	LASTTAKAM	HLA-B*35:01	21,08	0,75

based on common interaction cores, 45 groups were obtained. In each group, the epitopes interacting with the largest quantity of different MHCII alleles were selected and then from that set, the epitope with the minimum IC50 value was selected. For each selected epitope, allergenicity was calculated using the AllerTop 2.0 (Dimitrov et al., 2013) and AllergenFP 1.0 (Dimitrov et al., 2014) programs and toxicity was tested using the ToxinPred (Gupta et al., 2013) program. As with MHCI, only those epitopes for which the allergenicity predictions from the two programs coincided were removed. Based on these selection criteria, 6 epitopes were obtained (Table 5).

3.3. Investigation of the interaction between CD8+ epitopes and MHCI molecules

The interaction of proteins with each other, as well as proteins with peptides or low-molecular compounds, is an important part of any biological process. Predicting and modeling such interactions can save time and effort when creating drugs and epitope vaccines, since it is possible to choose the optimal components. The procedure of modeling the interactions between molecules is called molecular docking, which predicts the optimal location of the 3D structures of two interacting molecules relative to each other based on a calculation of the conformation with minimal interaction energy (Antunes et al., 2018).

Table 4. Conserved CD8+ T-cell influenza A and B epitopes. The binding energy was calculated in the FiberDock program (Mashiach et al., 2010).

#	Type: Protein	Position	Allele	Peptide	ANN, IC50	VaxiJen score	Allergen FP v.1.0	Aller TOP v.2.0
1	A:NP	140	HLA-B*35:01	HSNLNDATY	25.40	0.53	Y	N
2	A:NP	221	HLA-A*03:01	RMCNILKGK	40.75	0.51	N	Y
3	A:NP	251	HLA-A*02:06	AEIEDLIFL	44.07	1.03	N	Y
4	A:NP	258	HLA-A*02:01	FLARSALIL	43.19	0.57	Y	N
5	A:M1	58	HLA-A*02:01	GILGFVFTL	15.71	0.76	Y	N
			HLA-A*02:06	GILGFVFTL	41.19	0.76		
6	B:NP	242	HLA-C*12:03	VAIKGGGTL	44.39	0.7	N	N
7	B:NP	325	HLA-B*07:02	RPSVASKVV	23.37	0.74	N	N
8	B:NP	352	HLA-B*35:01	YSMVGYEAM	22.06	1	Y	N
9	B:NP	392	HLA-C*12:03	AAYEDLRVL	32.71	0.59	N	Y
10	B:NP	429	HLA-A*02:01	GMGAALMSI	39.58	0.58	N	N
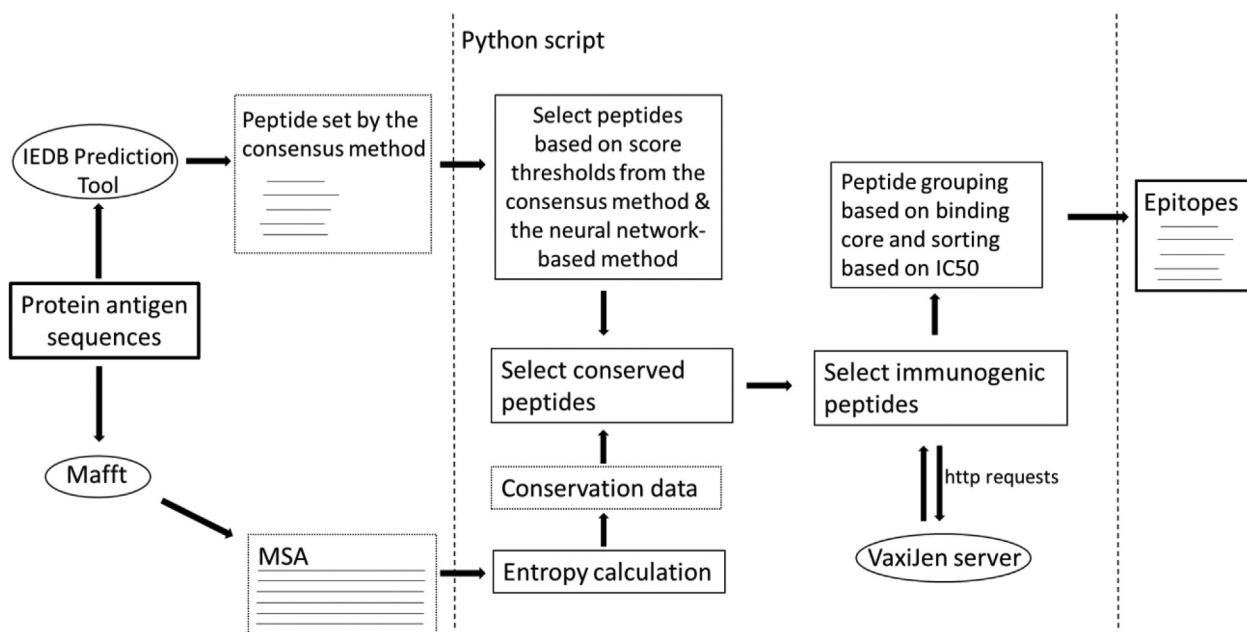
Figure 2. Pipeline for predicting CD4+ T-cell epitopes. Third-party software used in the pipeline is shown in ovals. The script was fed the results from the IEDB prediction service using a consensus algorithm and multiple alignments (MSAs) generated using the Mafft program. The output results from the IEDB service contain the predictions of several algorithms at once, so the script selected peptides that were included in the overlapping sets of peptides from several algorithms.

For docking, one needs to have 3D structures of all interaction components. X-ray structural analysis data are available for different MHCI molecule alleles. The structures of the predicted CD8+ peptides were obtained using a script in Python for the PyMol program. The dihedral Phi (−139) and Psi (−135) angles, typical for the antiparallel beta layer, were used in the construction. The spatial structure of stablepeptide conformations have minimal energy, so for each peptide structure, a procedure to minimize the energy was performed using the "optimize" command in the PyMol program, which identifies the optimal conformation with minimal energy.

The docking procedure was carried out in several stages. In the first stage, blind rigid docking was performed, which enables a rapid identification of the approximate areas of the ligand binding sites. It was previously shown that (Agrawal et al., 2019) when comparing different docking methods by CAPRI parameters, the FRODOCK 2.0 program (Ramírez-Aportela et al., 2016) was the most effective for the task. The structures of the peptides and corresponding MHCI molecules were given as the input. All 10 variants of the predicted epitope structures produced by the program were tested for the correct location in the binding site and incorrectly positioned peptides were excluded (Wieczorek et al., 2017; Zacharias and Springer, 2004; Mage et al., 2012; Hafstrand et al., 2019). Then, for the best structure with the highest prediction score, flexible

redocking was performed using the FiberDock program (Mashiach et al., 2010), which also calculates the binding energy considering van der Waals forces, electrostatic and hydrophobic interactions, hydrogen bond energy, and π-stacking interactions. Docking of peptides with negative binding energy was considered successful (Mashiach et al., 2010).

All of the predicted epitopes had a negative binding energy (Table 4). When examining the results of the CD8+ peptide docking to MHCI molecules, the distance between the amino acid residues of MHCI and the peptide were found to be within 4 Å. When analyzing 6444 protein-ligand complex structures in the 3D structure data bank (PDB), for which there was experimental confirmation of the interaction, it was shown that most interactions are realized within a distance of 4 Å (Ferreira de Freitas et al., 2017). Moreover, the majority of complexes were formed by hydrophobic interactions at a distance between carbon atoms of up to 4 Å, followed by hydrogen bonds, various kinds of stacking interactions and ion bonds. Table 6 shows the amino acid residues from the MHCI molecule that can potentially be involved in interactions with each peptide. Figure 3 shows the result of docking the A/M1:58-GILGFVFTL-66 peptide to the HLA-a*02:01 molecule (this interaction has the lowest binding energy (−79.18)). The MHCI amino acid residues that participate in the interaction are shown by the electron density surface. For other predicted structures, most of the interaction interface is formed by hydrophobic amino

Table 5. Conserved CD4+ T-cell influenza A and B epitopes.

#	Type: Protein	Position	Peptide	Allele	Consensus, percentile rank	ANN, IC50	Vaxi Jen	Allergen FP v.1.0	Aller Top v.2.0	Toxin Pred
1	A:NP	250	NAEIEDLIFLARSALILR	DRB1*07:01	1,5	21,1	0,88	N	Y	N
				DRB1*15:01	3,3	29,4				
				DPA1*02:01/DPB1*05:01	2,3	48,2				
2	A:NP	254	EDLIFLARSALILRGSVA	DQA1*01:02/DQB1*06:02	5,5	26,7	0,62	N	N	N
				DRB1*07:01	1,6	36				
				DRB1*15:01	3,3	27,9				
				DPA1*02:01/DPB1*05:01	4,8	47,3				
3	B:NP	161	DKTIYFSPIRITFLK	DRB1*07:01	3,7	14,8	0,96	N	N	N
				DRB1*11:01	8,2	45,2				
				DPA1*02:01/DPB1*01:01	5,7	17,6				
				DPA1*03:01/DPB1*04:02	5,8	13,6				
4	B:NP	164	IYFSPIRITFLKEEVKTM	DPA1*02:01/DPB1*05:01	5,5	22,1	0,77	N	N	N
				DPA1*03:01/DPB1*04:02	7,8	17,3				
				DPA1*02:01/DPB1*01:01	7,9	19,8				
5	B:NP	424	KEQVEGMGAALMSIKLQF	DQA1*01:02/DQB1*06:02	4	31,5	0,94	N	Y	N
				DQA1*05:01/DQB1*03:01	8	38,6				
6	A:M1	60	LGFVFTLTVPSER	DRB1*11:01	1,7	12,1	0,85	N	N	N
				DRB1*07:01	2,1	17,7				

acid residues. There are also hydrogen bonds with MHCI amino acid residues GLU-63, TYR-84, ARG-97, THR-143, and TRP-167. The low binding energy, as well as the identified peptide interactions with the MHCI binding site package, indicate a high probability of their presentation on the cell surface. The differences in binding energy between peptides indicate their different immunogenicity. Because MHCI molecules on the cell surface are constantly recycled, the lifetime of a particular complex depends on the binding strength of the epitope peptide with MHC (Wu et al., 2011). The higher the binding force, the greater the probability of cell surface presentation of the peptide. Therefore, this increases the probability of an interaction with CD8+ T-lymphocytes to produce a CD8+ T-cell response. Thus, epitopes with low binding energy can potentially have immunodominant properties. It is known from the literature that the A/M1:58-GILGFVFTL-66 epitope is immunodominant for HLA-A2 alleles

(Boon et al., 2002). The other epitopes, A/NP:140-HSNLNDATY-148, A/NP:258-FLARSALIL-266, and B/NP:392-AAYEDLRVL-400, also may have high potential immunogenicity, so they are excellent candidates for experimental verification.

4. Discussion

Recently, bioinformatics has been used for a selection of conserved CD4+ and CD8+ epitopes of dengue virus (Chong and Khan, 2019), CD8+ epitopes of hepatitis C virus (Molero-Abraham et al., 2013), and CD4+ and CD8+ epitopes of influenza virus (Sheikh et al., 2016). However, in these studies, epitopes were not predicted but were selected only from the set of experimentally confirmed epitopes contained in the IEDB, using information entropy to analyze their conservation. Using

Table 6. The molecular docking of the predicted CD8+ peptides to MHCI molecules.

#	Type/Protein:pos-peptide	MHCI allele	Binding energy	Interacting residues
1	A/NP:140-HSNLNDATY	HLA-B*35:01	-49,93	TYR-9, ARG-62, ILE-66, ASN-70, THR-73, SER-77, ASN-80, LEU-81, TYR-84, ILE-95, ARG-97, TYR-99, SER-116, TYR-123, ILE-124, THR-143, LYS-146, TRP-147, GLN-155, LEU-156, TYR-159, LEU-163, GLU-166, TRP-167, ARG-170
2	A/NP:221-RMCNILKGK	HLA-A*03:01	-9,06	TYR-7, GLN-54, GLU-55, GLY-56, TYR-59, GLU-63, ASN-66, ALA-69, GLN-70, THR-73, TYR-99, ARG-114, LEU-126, TRP-133, TRP-147, GLU-152, ALA-153, GLN-155, LEU-156, TYR-159, THR-163, TRP-167, ARG-170, TYR-171, ASN-174
3	A/NP:251-AEIEDLIFL	HLA-A*02:06	-31,16	GLU-63, LYS-66, ALA-69, HIS-70, THR-73, VAL-76, ASP-77, THR-80, TYR-84, TYR-99, TYR-123, THR-143, LYS-146, TRP-147, VAL-152, TYR-159, THR-163, TRP-167
4	A/NP:258-FLARSALIL	HLA-A*02:01	-51,32	THR-73, VAL-76, ASP-77, THR-80, LEU-81, TYR-84, TYR-123, THR-142, THR-143, LYS-146, TRP-147, ALA-150, VAL-152, GLN-155, LEU-156, ALA-158, TYR-159, THR-163, GLU-166
5	A/M1:58-GILGFVFTL	HLA-A*02:01	-79,18	GLU-58, TYR-59, GLY-62, GLU-63, ARG-65, LYS-66, HIS-70, THR-73, ASP-77, THR-80, LEU-81, TYR-84, ARG-97, TYR-99, HIS-114, TYR-116, TYR-123, THR-142, THR-143, LYS-146, TRP-147, VAL-152, GLN-155, LEU-156, TYR-159, THR-163, TRP-167
6	A/M1:58-GILGFVFTL	HLA-A*02:06	-72,67	MET-5, TYR-7, TYR-9, PHE-33, MET-45, TYR-59, GLU-63, LYS-66, VAL-67, HIS-70, THR-73, VAL-76, ASP-77, ARG-97, TYR-99, LYS-146, TRP-147, ALA-150, GLN-155, LEU-156, TYR-159, THR-163, TRP-167, TYR-171,
7	B/NP:242-VAIKGGGTL	HLA-C*12:03	-16,98	TYR-9, GLN-70, THR-73, ASP-74, SER-77, LEU-78, LEU-81, ILE-95, ARG-97, TYR-116, TYR-123, ILE-124, THR-143, TRP-147, GLU-152, GLN-155, LEU-156, ALA-158, TYR-159
8	B/NP:325-RPSVASKVV	HLA-B*07:02	-39,39	TYR-7, GLN-54, GLU-55, GLY-56, TYR-59, GLU-63, ASN-66, ALA-69, GLN-70, THR-73, TYR-99, ARG-114, LEU-126, TRP-133, TRP-147, GLU-152, ALA-153, GLN-155, LEU-156, TYR-159, THR-163, TRP-167, ARG-170, TYR-171, ASN-174
9	B/NP:352-YSMVGYEAM	HLA-B*35:01	-27,44	TYR-9, ARG-62, ILE-66, THR-69, ASN-70, THR-73, TYR-74, GLU-76, SER-77, ASN-80, TYR-84, ARG-97, THR-143, LYS-146, TRP-147, TYR-159, LEU-163, TRP-167, ARG-170
10	B/NP:392-AAYEDLRVL	HLA-C*12:03	-75,58	TYR-9, ARG-69, GLN-70, THR-73, ASP-74, SER-77, LEU-78, LEU-81, ILE-95, ARG-97, ASP-114, TYR-116, TYR-123, ILE-124, THR-143, TRP-147, GLU-152, LEU-156, LEU-163
11	B/NP:429-GMGAALMSI	HLA-A*02:01	-44,8	TYR-7, PHE-9, GLU-63, LYS-66, VAL-67, HIS-70, THR-73, ASP-77, THR-80, ARG-97, TYR-99, TYR-116, LYS-146, TRP-147, ALA-150, TRP-167, TYR-171
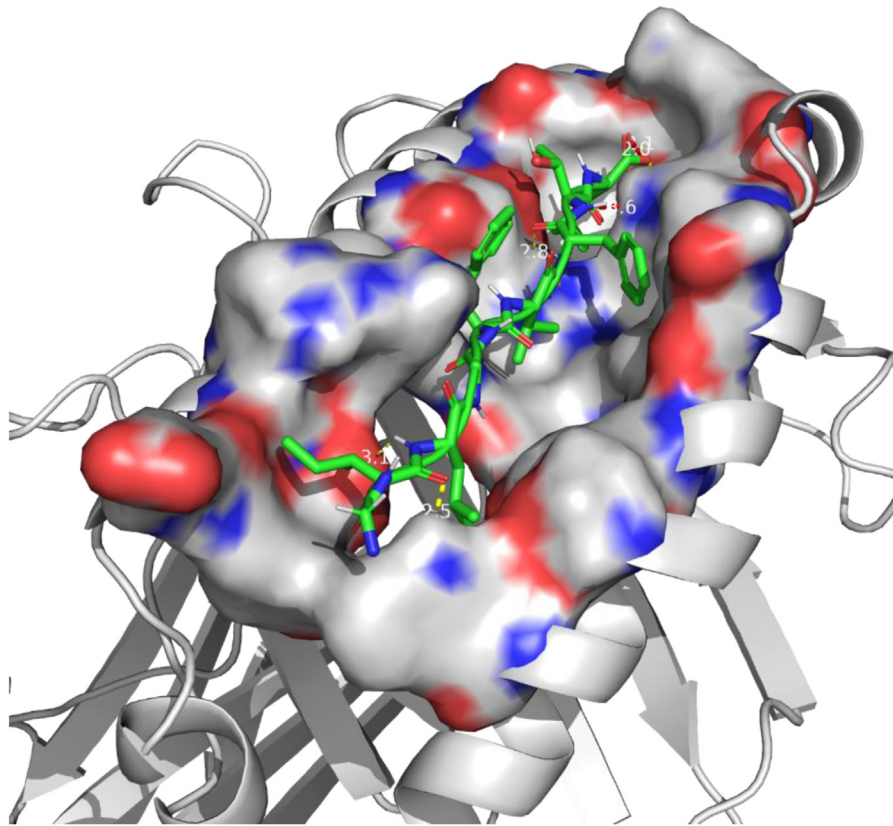
Figure 3. The result of docking the A/M1:58-GILGFVFTL-66 peptide to the HLA-A*02:01 molecule. The surface shows the electron density of MHCI amino acid residues located no further than 4 Å from the amino acid residues of the peptide. The carbon atoms on the surface of the MHCI molecule are shown in gray. Green color indicates the carbon atoms of the peptide. Blue color indicates nitrogen atoms and red indicates oxygen atoms.

prediction programs, B and T cells (Jain et al., 2021), or only T cell epitopes of the new coronavirus (Sirohi et al., 2020; Montes-Grajales and Olivero-Verbel, 2021) were selected, but their conservation was not assessed. For the vaccine against Chagas disease, B and T cell epitopes from L. cruzi were selected using methods for predicting epitopes and assessing their conservation (Michel-Todó et al., 2020). Our approach made it possible to predict new epitopes for the influenza virus, as well as to increase the severity of selection in the selection of epitopes, due to a systematic review of modern algorithms of prediction in conjunction with the use of several prediction programs simultaneously.

10 B-cell epitopes, 10 CD8+ and 6 CD4+ T-cell epitopes were selected. Interestingly, experimental data was not found for all of the epitopes in the IEDB database. Only three B-cell epitopes that were conserved among the H1 subtype viruses were found to have experimental confirmation of immunogenicity. It was previously shown (Klausberger et al., 2016) those antibodies against HA stem epitopes, including epitope A/H1:41-51, were found in patients after their immunization with a live vaccine. Additionally, during the search for immune-dominant B-cell HA epitopes using ELISA (Zhao et al., 2011), it was shown that fragments of A/H1:118-132 reacted with sera from patients who had been ill with pandemic swine flu. In addition, sera from mice immunized with the A/H1:38-52 peptide reacted on the immunoblot and in ELISAs with the HA0 fragment. For fragment A/H1:364-370, it was shown that it binds to broadly neutralizing antibodies (Guo et al., 2018). Thus, all of the identified conserved epitopes for the H1 subtype are immunogens, and they may induce the formation of broadly neutralizing antibodies. However, the value of the remaining predicted A/H3 B-cell epitopes should not be underestimated, when analyzing the IEDB database, we observed an uneven representation of epitopes among the subtypes in favor of H1, which may reflect a trend in the scientific world to choose primarily H1 strains. Nevertheless, the danger and

circulation of the H3 subtypes of influenza A and B remains, so the identified conserved epitopes are good candidates for experimental research into their immunogenic properties.

For the 10 CD8+ and 6 CD4+ T-cell epitopes selected, their previous studies in the literature were analyzed using IEDB. As was the case for the B-cell epitopes, the database has a preference in favor of influenza A. The most well-studied epitope turned out to be CD8+ epitope A/M1:58-GILGFVFTL-66, which was characterized in previous studies by immunodominant properties capable of causing a resistant cross-reactive T-cell response (Choo et al., 2014; van de Sandt et al., 2018; Sant et al., 2018; Boon et al., 2002). In addition, there are experimental data for epitopes A/NP:140-HSNLNDATY-148, A/NP:251-AEIED-LIFL-259, and A/M1:60-LGFVFTLTVPSER-72, which provides evidence for their immunogenicity. The absence of some epitopes in the database does not indicate that they should be considered poorly effective. In contrast, given the high immunogenicity scores of these epitopes, these data advocate for the experimental verification of their properties.

T cell epitopes were selected for the MHC variants from the Moscow region population according to AFND data. However, there is a diversity of ethnicities in this region, which may affect the broader profile of MHC variants. Analysis of the AFND database demonstrated that the selected MHCI and MHCII variants have a frequency of more than 10% in populations of 68 131 countries in the world, respectively. Thus, the selected T cell epitopes are more relevant for the world population than the Moscow region.

Universal vaccines should induce immunity against all or most of the influenza virus strains. The main contribution to the universal properties of such vaccines should be made by T cell epitopes from the internal proteins of the virus due to their high conservation among all strains. B cell epitopes are selected from HA sequences of the H1 and H3 subtypes of type A and type B because viruses of these groups mainly circulate in

the human population. The use of many B cell epitopes from all 18 subtypes of HA to induce an immune response can be complicated. Epitopes in the form of peptides are weakly immunogenic, but difficulties with gene expression may arise when epitopes are combined in a polypeptide, such as in the M001 vaccine. In addition, a hierarchy of immunogenicity arises when epitopes are introduced. In this case, the immune response against epitopes of more significant strains circulating in the human population may deteriorate. In contrast, Joyce et al., (2016) reported that antibodies on the region of the HA trunk exhibit a cross-reactive effect due to the antigenic similarity of epitopes within Groups 1 and 2. Because subtypes H1 and H3 are phylogenetically distant from each other and are representatives of Groups 1 and 2, respectively, they may stimulate a cross-reactive antibody response within this groups. However, this issue is a topic for further research. Nevertheless, the use of B cell epitopes in conjunction with T cell epitopes from conserved influenza proteins may provide a good basis for the design of a universal influenza vaccine.

5. Conclusion

In this work, the epitopes of influenza virus antigens that could potentially induce humoral and T-cell immunity were predicted using bioinformatics approaches. Epitopes were selected from conserved regions of the immunogenic HA, M1 and NP proteins from the H1 and H3 subtypes of influenza A and B viruses. A pipeline that automates the selection of an optimal set of epitopes based on conservation and antigenicity criteria was created for these purposes. Regarding the results, 10 B-cell epitopes, 10 CD8+ and 6 CD4+ T-cell epitopes characterized by high conservation and antigenicity were selected. The predictive power of the developed approach was determined based on the algorithms included therein. The best algorithms for predicting linear epitopes were selected based on the results of a comparative analysis of the algorithms. The scripts used in the pipeline are universal and can be used to select epitopes from other pathogens.

Declarations

Author contribution statement

R.R. Mintaev: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Wrote the paper.

D.V. Glazkova: Conceived and designed the experiments; Analyzed and interpreted the data.

E.V. Bogoslovskaya, G.A. Shipulin: Conceived and designed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data.

Funding statement

This work was supported by Federal State Budgetary Institution «Center for Strategic Planning and Management of Medical and Biological Health Risks», Russia.

Data availability statement

Data included in article/supp. material/referenced in article.

Declaration of interest's statement

The authors declare no competing interests.

Additional information

No additional information is available for this paper.

References

Agrawal, P., Singh, H., Srivastava, H.K., Singh, S., Kishore, G., Raghava, G.P.S., 2019. Benchmarking of different molecular docking methods for protein-peptide docking. BMC Bioinf. 19.

Andreatta, M., Nielsen, M., 2016. Gapped sequence alignment using artificial neural networks: application to the MHC class I system. Bioinformatics 32 (4), 511–517.

Antunes, D.A., Abella, J.R., Devaurs, D., Rigo, M.M., Kavraki, L.E., 2018. Structure-based methods for binding mode and binding affinity prediction for peptide-MHC complexes. Curr. Top. Med. Chem. 18 (26), 2239–2255.

Boon, A.C., Vos, A.P., Graus, Y.M., Rimmelzwaan, G.F., Osterhaus, A.D., 2002a. In vitro effect of bioactive compounds on influenza virus specific B- and T-cell responses. Scand. J. Immunol. 55 (1), 24–32.

Boon, A.C., de Mutsert, G., Graus, Y.M., Fouchier, R.A., Sintnicolaas, K., Osterhaus, A.D., Rimmelzwaan, G.F., 2002b. The magnitude and specificity of influenza A virus-specific cytotoxic T-lymphocyte responses in humans is related to HLA-A and -B phenotype. J. Virol. 76 (2), 582–590.

Brown, J.H., Jardetzky, T.S., Gorga, J.C., Stern, L.J., Urban, R.G., Strominger, J.L., Wiley, D.C., 1993. Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. Nature 364 (6432), 33–39.

Chong, L.C., Khan, A.M., 2019. Identification of highly conserved, serotype-specific dengue virus sequences: implications for vaccine design. BMC Genom. 20 (Suppl 9), 921.

Choo, J.A., Liu, J., Toh, X., Grotenbreg, G.M., Ren, E.C., 2014. The immunodominant influenza A virus M158-66 cytotoxic T lymphocyte epitope exhibits degenerate class I major histocompatibility complex restriction in humans. J. Virol. 88 (18), 10613–10623.

Dimitrov, I., Flower, D.R., Doytchinova, I., 2013. AllerTOP-a server for in silico prediction of allergens. BMC Bioinf. 14 (Suppl 6 S4).

Dimitrov, I., Naneva, L., Doytchinova, I., Bangov, I., 2014. AllergenFP: allergenicity prediction by descriptor fingerprints. Bioinformatics 30 (6), 846–851.

Doytchinova, I.A., Flower, D.R., 2001. Toward the quantitative prediction of T-cell epitopes: coMFA and coMSIA studies of peptides with affinity for the class I MHC molecule HLA-A*0201. J. Med. Chem. 44 (22), 3572–3581.

El-Manzalawy, Y., Dobbs, D., Honavar, V., 2008. Predicting linear B-cell epitopes using string kernels. J. Mol. Recogn. : JMR (J. Mol. Recognit.) 21 (4), 243–255.

Emini, E.A., Hughes, J.V., Perlow, D.S., Boger, J., 1985. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. J. Virol. 55 (3), 836–839.

Ferreira de Freitas, R., Schapira, M., 2017. A systematic analysis of atomic protein-ligand interactions in the PDB. MedChemComm 8 (10), 1970–1981.

Guo, C., Zhang, H., Xie, X., Liu, Y., Sun, L., Li, H., Yu, P., Hu, H., Sun, J., Li, Y., Feng, Q., Zhao, X., Liang, D., Wang, Z., Hu, J., 2018. H1N1 influenza virus epitopes classified by monoclonal antibodies. Exp. Ther. Med. 16 (3), 2001–2007.

Gupta, S., Kapoor, P., Chaudhary, K., Gautam, A., Kumar, R., Raghava, G.P., 2013. In silico approach for predicting toxicity of peptides and proteins. PLoS One 8 (9), e73957.

Hafstrand, I., Sayitoglu, E.C., Apavaloaei, A., Josey, B.J., Sun, R., Han, X., Pellegrino, S., Ozkazanc, D., Potens, R., Janssen, L., Nilvebrant, J., Nygren, P.Å., Sandalova, T., Springer, S., Georgoudaki, A.M., Duru, A.D., Achour, A., 2019. Successive crystal structure snapshots suggest the basis for MHC class I peptide loading and editing by tapasin. Proc. National Acad. Sci. U. S. A. 116 (11), 5055–5060.

Jain, R., Jain, A., Verma, S.K., 2021. Prediction of epitope based peptides for vaccine development from complete proteome of novel corona virus (SARS-COV-2) using immunoinformatics. Int. J. Pept. Res. Therapeut. 27, 1729–1740.

Jensen, K.K., Andreatta, M., Marcatili, P., Buus, S., Greenbaum, J.A., Yan, Z., Sette, A., Peters, B., Nielsen, M., 2018. Improved methods for predicting peptide binding affinity to MHC class II molecules. Immunology 154 (3), 394–406.

Jespersen, M.C., Peters, B., Nielsen, M., Marcatili, P., 2017. BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. Nucleic Acids Res. 45.

Jones, E.Y., Fugger, L., Strominger, J.L., Siebold, C., 2006. MHC class II proteins and disease: a structural perspective. Nat. Rev. Immunol. 6 (4), 271.

Joyce, M.G., Wheatley, A.K., Thomas, P.V., Chuang, G.Y., Soto, C., Bailer, R.T., Druz, A., Georgiev, I.S., Gillespie, R.A., Kanekiyo, M., Kong, W.P., Leung, K., Narpala, S.N., Prabhakaran, M.S., Yang, E.S., Zhang, B., Zhang, Y., Asokan, M., Boyington, J.C., Bylund, T., Darko, S., Lees, C.R., Ransier, A., Shen, C.H., Wang, L., Whittle, J.R., Wu, X., Yassine, H.M., Santos, C., Matsuoka, Y., Tsybovsky, Y., Baxa, U., Mullikin, J.C., Subbarao, K., Douek, D.C., Graham, B.S., Koup, R.A., Ledgerwood, J.E., Roederer, M., Shapiro, L., Kwong, P.D., Mascola, J.R., McDermott, A.B., 2016. Vaccine-induced antibodies that neutralize group 1 and group 2 influenza A viruses. Cell 166 (3), 609–623.

Karplus, P.A., Schulz, G.E., 1985. Prediction of chain flexibility in proteins. Naturwissenschaften 72, 212–213.

Klausberger, M., Tscheliessnig, R., Neff, S., Nachbagauer, R., Wohlbold, T.J., Wilde, M., Palmberger, D., Krammer, F., Jungbauer, A., Grabherr, R., 2016. Globular head-displayed conserved influenza H1 hemagglutinin stalk epitopes confer protection against heterologous H1N1 virus. PLoS One 11 (4), e0153579.

Knossow, M., Skehel, J.J., 2006. Variation and infectivity neutralization in influenza. Immunology 119 (1), 1–7.

Konstantinou, G.N., 2017. T-cell epitope prediction. Methods Mol. Biol. 1592, 211–222.

Larsen, J.E., Lund, O., Nielsen, M., 2006. Improved method for predicting linear B-cell epitopes. Immunome Res. 2, 2.

Lin, H.H., Zhang, G.L., Tongchusak, S., Reinherz, E.L., Brusic, V., 2008. Evaluation of MHC-II peptide binding prediction servers: applications for vaccine research. BMC Bioinf. 9 (Suppl 12), S22.

Lowell, G.H., Ziv, S., Bruzil, S., Babecoff, R., Ben-Yedidia, T., 2017. Back to the future: immunization with M-001 prior to trivalent influenza vaccine in 2011/12 enhanced protective immune responses against 2014/15 epidemic strain. Vaccine 35 (5), 713–715.

Lundegaard, C., Lamberth, K., Harndahl, M., Buus, S., Lund, O., Nielsen, M., 2008. NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8-11. Nucleic Acids Res. 36, W509–W512 (Web Server issue).

Mage, M.G., Dolan, M.A., Wang, R., Boyd, L.F., Revilleza, M.J., Robinson, H., Natarajan, K., Myers, N.B., Hansen, T.H., Margulies, D.H., 2012. The peptide-receptive transition state of MHC class I molecules: insight from structure and molecular dynamics. J. Immunol. 189 (3), 1391–1399.

Mashiach, E., Nussinov, R., Wolfson, H.J., 2010. FiberDock: flexible induced-fit backbone refinement in molecular docking. Proteins 78 (6), 1503–1519.

Michel-Todó, L., Reche, P.A., Bigey, P., Pinazo, M.J., Gascón, J., Alonso-Padilla, J., 2020. Corrigendum: in silico design of an epitope-based vaccine ensemble for Chagas disease. Front. Immunol. 10, 3124.

Mintaev, R.R., Alexeevski, A.V., Kordyukova, L.V., 2014. Co-evolution analysis to predict protein-protein interactions within influenza virus envelope. J. Bioinf. Comput. Biol. 12 (2), 1441008.

Molero-Abraham, M., Lafuente, E.M., Flower, D.R., Reche, P.A., 2013. Selection of conserved epitopes from hepatitis C virus for pan-populational stimulation of T-cell responses. Clin. Dev. Immunol. 601943.

Montes-Grajales, D., Olivero-Verbel, J., 2021. Bioinformatics prediction of SARS-CoV-2 epitopes as vaccine candidates for the Colombian population. Vaccines 9 (7), 797.

Nachbagauer, R., Krammer, F., 2017. Universal influenza virus vaccines and therapeutic antibodies. Clin. Microbiol. Infect. : Off. Publ. Europ. Soc. Clin. Microbiol. Infect. Dise. 23 (4), 222–228.

Nair, H., Brooks, W.A., Katz, M., Roca, A., Berkley, J.A., Madhi, S.A., Simmerman, J.M., Gordon, A., Sato, M., Howie, S., Krishnan, A., Ope, M., Lindblade, K.A., Carosone-Link, P., Lucero, M., Ochieng, W., Kamimoto, L., Dueger, E., Bhat, N., Vong, S., Theodoratou, E., Chittaganpitch, M., Chimah, O., Balmaseda, A., Buchy, P., Harris, E., Evans, V., Katayose, M., Gaur, B., O'Callaghan-Gordo, C., Goswami, D., Arvelo, W., Venter, M., Briese, T., Tokarz, R., Widdowson, M.A., Mounts, A.W., Breiman, R.F., Feikin, D.R., Klugman, K.P., Olsen, S.J., Gessner, B.D., Wright, P.F., Rudan, I., Broor, S., Simões, E.A., Campbell, H., 2011. Global burden of respiratory infections due to seasonal influenza in young children: a systematic review and meta-analysis. Lancet 378 (9807), 1917–1930.

Nielsen, M., Lundegaard, C., Worning, P., Lauemøller, S.L., Lamberth, K., Buus, S., Brunak, S., Lund, O., 2003. Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. Protein Sci. : Publ. Protein Soc. 12 (5), 1007–1017.

Parker, J.M., Guo, D., Hodges, R.S., 1986. New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. Biochemistry 25 (19), 5425–5432.

Paul, S., Croft, N.P., Purcell, A.W., Tscharke, D.C., Sette, A., Nielsen, M., Peters, B., 2020. Benchmarking predictions of MHC class I restricted T cell epitopes in a comprehensively studied model system. PLoS Comput. Biol. 16 (5), e1007757.

Peters, B., Nielsen, M., Sette, A., 2020. T cell epitope predictions. Annu. Rev. Immunol. 38, 123–145.

Pleguezuelos, O., James, E., Fernandez, A., Lopes, V., Rosas, L.A., Cervantes-Medina, A., Cleath, I., Edwards, K., Neitzey, D., Gu, W., Hunsberger, S., Taubenberger, J.K., Stoloff, G., Memoli, M.J., 2020. Efficacy of FLU-v, a broad-spectrum influenza vaccine, in a randomized phase IIb human influenza challenge study. NPJ Vaccines 5, 22.

Potocnakova, L., Bhide, M., Pulzova, L.B., 2016. An introduction to B-cell epitope mapping and in silico epitope prediction. J. Immunol. Res. 6760830.

Ramírez-Aportela, E., López-Blanco, J.R., Chacón, P., 2016. FRODOCK 2.0: fast protein-protein docking server. Bioinformatics 32 (15), 2386–2388.

Reche, P.A., Reinherz, E.L., 2007. Definition of MHC supertypes through clustering of MHC peptide-binding repertoires. Methods Mol. Biol. 409, 163–173.

Ruppert, J., Sidney, J., Celis, E., Kubo, R.T., Grey, H.M., Sette, A., 1993. Prominent role of secondary anchor residues in peptide binding to HLA-A2.1 molecules. Cell 74 (5), 929–937.

Saha, Sudipto, Raghava, G.P.S., 2004. BcePred: prediction of continuous B-cell epitopes in antigenic sequences using physico-chemical properties. In: International Conference on Artificial Immune Systems, pp. 197–204.

Saha, S., Bhasin, M., Raghava, G.P., 2005. Bcipep: a database of B-cell epitopes. BMC Genom. 6, 79.

Sant, S., Grzelak, L., Wang, Z., Pizzolla, A., Koutsakos, M., Crowe, J., Loudovaris, T., Mannering, S.I., Westall, G.P., Wakim, L.M., Rossjohn, J., Gras, S., Richards, M.,

Xu, J., Thomas, P.G., Loh, L., Nguyen, T.H.O., Kedzierska, K., 2018. Single-cell approach to influenza-specific CD8+ T cell receptor repertoires across different age groups, tundefineds, and following influenza virus infection. Front. Immunol. 9, 1453.

Sette, A., Buus, S., Appella, E., Smith, J.A., Chesnut, R., Miles, C., Colon, S.M., Grey, H.M., 1989. Prediction of major histocompatibility complex binding regions of protein antigens by sequence pattern analysis. Proc. National Acad. Sci. U. S. A. 86 (9), 3296–3300.

Sette, A., Vitiello, A., Reherman, B., Fowler, P., Nayersina, R., Kast, W.M., Melief, C.J., Oseroff, C., Yuan, L., Ruppert, J., Sidney, J., del Guercio, M.F., Southwood, S., Kubo, R.T., Chesnut, R.W., Grey, H.M., Chisari, F.V., 1994. The relationship between class I binding affinity and immunogenicity of potential cytotoxic T cell epitopes. J. Immunol. Baltimore, Md. : 1950 153 (12), 5586–5592.

Sheikh, Q.M., Gatherer, D., Reche, P.A., Flower, D.R., 2016. Towards the knowledge-based design of universal influenza epitope ensemble vaccines. Bioinformatics (Oxford, England) 32 (21), 3233–3239.

Shenkin, P.S., Erman, B., Mastrandrea, L.D., 1991. Information-theoretical entropy as a measure of sequence variability. Proteins 11 (4), 297–313.

Singh, H., Ansari, H.R., Raghava, G.P., 2013. Improved method for linear B-cell epitope prediction using antigen's primary sequence. PLoS One 8 (5), e62216.

Sirohi, P.R., Gupta, J., Somvanshi, P., Prajapati, V.K., Grover, A., 2020. Multiple epitope-based vaccine prediction against SARS-CoV-2 spike glycoprotein. J. Biomol. Struct. Dynam. 1–12.

Southwood, S., Sidney, J., Kondo, A., del Guercio, M.F., Appella, E., Hoffman, S., Kubo, R.T., Chesnut, R.W., Grey, H.M., Sette, A., 1998. Several common HLA-DR types share largely overlapping peptide binding repertoires. J. Immunol. Baltimore, Md. : 1950 160 (7), 3363–3373.

Takeshita, L.Y., Jones, A.R., Gonzalez-Galarza, F.F., Middleton, D., 2014. Allele frequencies database. Transfus. Med. Hemother.: Off. Organ der Dtsch. Ges. fur Transfusionsmed. Immunhamatol. 41 (5), 352–355.

Trucchi, C., Paganino, C., Amicizia, D., Orsi, A., Tisa, V., Piazza, M.F., Icardi, G., Ansaldi, F., 2019. Universal influenza virus vaccines: what needs to happen next? Expet Opin. Biol. Ther. 19 (7), 671–683.

van de Sandt, C.E., Sagong, K.A., Pronk, M.R., Bestebroer, T.M., Spronken, M.I., Koopmans, M.P.G., Fouchier, R.A.M., Rimmelzwaan, G.F., 2018. H1N1pdm09 influenza virus and its descendants lack extra-epitopic amino acid residues associated with reduced recognition by M158-66-specific CD8+ T cells. J. Infect. Dis. 218 (4), 581–585.

Van Regenmortel, M.H., 2009. What is a B-cell epitope? Methods Mol. Biol. 524, 3–20.

Vemula, S.V., Sayedahmed, E.E., Sambhara, S., Mittal, S.K., 2017. Vaccine approaches conferring cross-protection against influenza viruses. Expet Rev. Vaccine 16 (11), 1141–1154.

Wang, P., Sidney, J., Dow, C., Mothé, B., Sette, A., Peters, B., 2008. A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach. PLoS Comput. Biol. 4 (4), e1000048.

Warrington, R., Watson, W., Kim, H.L., Antonetti, F.R., 2011. An introduction to immunology and immunopathology. Allergy, asthma, and clinical immunology. Off. J. Canad. Soc. Allergy Clin. Immunol. 7, 1.

Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M., Barton, G.J., 2009. Jalview Version 2-a multiple sequence alignment editor and analysis workbench. Bioinformatics (Oxford, England) 25 (9), 1189–1191.

Wieczorek, M., Abualrous, E.T., Sticht, J., Álvaro-Benito, M., Stolzenberg, S., Noé, F., Freund, C., 2017. Major histocompatibility complex (MHC) class I and MHC class II proteins: conformational plasticity in antigen presentation. Front. Immunol. 8, 292.

Wu, C., Zanker, D., Valkenburg, S., Tan, B., Kedzierska, K., Zou, Q.M., Doherty, P.C., Chen, W., 2011. Systematic identification of immunodominant CD8+ T-cell responses to influenza A virus in HLA-A2 individuals. Proc. National Acad. Sci. U. S. A. 108 (22), 9178–9183.

Zacharias, M., Springer, S., 2004. Conformational flexibility of the MHC class I alpha1-alpha2 domain in peptide bound and free states: a molecular dynamics simulation study. Biophys. J. 87 (4), 2203–2214.

Zaharieva, Nevena, Dimitrov, Ivan, Flower, Darren R., Doytchinova, Irini, 2017. Immunogenicity prediction by VaxiJen: a ten year overview. J. Proteonomics Bioinf. 10, 11.

Zhang, H., Wang, P., Papangelopoulos, N., Xu, Y., Sette, A., Bourne, P.E., Lund, O., Ponomarenko, J., Nielsen, M., Peters, B., 2010. Limitations of Ab initio predictions of peptide binding to MHC class II molecules. PLoS One 5 (2), e9272.

Zhao, R., Cui, S., Guo, L., Wu, C., Gonzalez, R., Paranhos-Baccalà, G., Vernet, G., Wang, J., Hung, T., 2011. Identification of a highly conserved H1 subtype-specific epitope with diagnostic potential in the hemagglutinin protein of influenza A virus. PLoS One 6 (8), e23374.