The Protein Journal
https://doi.org/10.1007/s10930-022-10050-z

Design of an Epitope-Based Peptide Vaccine Against the Major Allergen Amb a 11 Using Immunoinformatic Approaches

Dzhemal Moten[1] · Desislava Kolchakova[1] · Krasimir Todorov[2] · Tsvetelina Mladenova[2] · Balik Dzhambazov[1]

Accepted: 18 March 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract
Allergic diseases are a socially significant problem of global importance. The number of people suffering from pollen allergies has increased dramatically in recent decades. Pollen allergies affect up to 30% of the world population. Pollen of the common ragweed (*Ambrosia artemisiifolia* L.) is one of the most aggressive allergens in the world. We have used a series of immunoinformatics approaches to design an effective epitope-based vaccine, which might induce a competent immunity against a major allergen Amb a 11. CD8+ and CD4+ T-cell epitopes and their corresponding MHC restricted alleles were identified by prediction tools provided by immune epitope database (IEDB). Among T-cell epitopes, MHC class I peptide (GLMEPAFTYV) and MHC class II peptide (LVCFSFSLVLILGLV) were identified as most suitable. From all predicted B-cell epitopes, only one epitope (GKLVKF**SEQQLVDC**) containing sequence from the conserved region was chosen for next processing. Selected epitopes have been validated by molecular docking analysis. These epitopes showed a very strong binding affinity to MHC I molecule and MHC II molecule with binding energy scores − 729.3 and − 725.0 kcal/mole respectively. Performed experimental validation showed that only the MHC class II peptide (LVCFSFSLVLILGLV) can stimulate T cells from ragweed allergic patients and IgE antibodies specific to the ragweed pollen do not recognize this epitope. Therefore, this peptide could be potentially used as a vaccine against the major allergen Amb a 11. The B-cell epitope GKLVKF**SEQQLVDC** forms a stable complex with the IgE molecule (energy weighted score − 695,0 kcal/mole). Tested sera from patients with ragweed allergy showed that the ragweed specific IgE antibodies can bind to the identified B-cell epitope. Population coverage analysis was performed for CD8+ and CD4+ T-cell epitopes. It was predicted that CD4+ T-cell epitope (LVCFSFSLVLILGLV) covers 90.56% of the population of Europe and 99.36% of the world population. CD8+ T-cell epitope (GLMEPAFTYV) has a population coverage of 77.37% for Europe and 71.35% for all the world.

Keywords Pollen allergy · *Ambrosia artemisiifolia* · Amb a 11 · Immunoinformatics · Peptide-based vaccine

1 Introduction

Pollen allergens are one of the major factors leading to severe type I respiratory allergies, which affect up to 30% of the human population [1]. Common ragweed (*Ambrosia artemisiifolia* L.) and its pollen cause serious losses in the economy and in various spheres of life. *Ambrosia artemisiifolia* is the most important from the allergic point of view, as its pollen is extremely allergenic and very low concentrations can trigger allergic reactions (allergic rhinitis, asthma or contact dermatitis) in sensitive patients. Allergy to pollen from *Ambrosia artemisiifolia* is a serious health issue exerting a strong negative impact on the quality of life of allergic patients [2, 3].

Ambrosia artemisiifolia belongs to genus *Ambrosia* (Asteraceae). This genus includes about 40 species and it is associated with severe and widespread allergies caused by the pollen of its representatives. Short ragweed (*Ambrosia artemisiifolia*) is native to Northern and Central America, but it has been introduced and subsequently naturalized in many countries including a large part of Europe [4, 5]. In Europe, *Ambrosia artemisiifolia* is predominant among all *Ambrosia* species. Short ragweed is mainly found in France,

✉ Dzhemal Moten
moten@uni-plovdiv.bg

1 Department of Developmental Biology, Faculty of Biology, University of Plovdiv "Paisii Hilendarski", 24 Tsar Assen Str., 4000 Plovdiv, Bulgaria

2 Department of Botany and Biological Education, Faculty of Biology, University of Plovdiv "Paisii Hilendarski", 24 Tsar Assen Str., Plovdiv 4000, Bulgaria

Published online: 01 April 2022

Germany, Hungary, Bulgaria, Czech Republic, Poland, Northern Italy, and Eastern Austria [6–10].

It is important to underline that the pollen of *Ambrosia artemisiifolia* is produced in enormous amounts (a plant may produce 100 million–3 billion pollen grains) and even low exposure (5–10 pollen/m^3 in air) is sufficient to initiate allergic reactions in sensitive individuals [11, 12]. To date, 11 ragweed allergens have been identified in the *Ambrosia artemisiifolia* pollen. Among them, the pectate lyase Amb a 1 and the cysteine protease Amb a 11 are described as major allergens. Within the ragweed pollen-allergic patients, they have a sensitization rate of 90 and 66%, respectively [13, 14]. Amb a 2 (a pectate lyase), Amb a 3 and Amb a 7 (two plastocyanins), Amb a 4 (homolog to the major mugwort allergen Art v 1), Amb a 5 (with unknown function), Amb a 6 (a nonspecific lipid transfer protein), Amb a 8 (profilin), and Amb a 9 and Amb a 10 (calcium-binding proteins) are considered as minor allergens. Sensitization rates of the minor allergens vary between 10 and 50% [15, 16]. There is significant cross-reactivity between the common ragweed and other *Ambrosia* species, mugwort, marsh elder or other grass pollens. The molecular weights of ragweed allergens range from 5 to 38 kDa, and they are grouped according to their biological differences [17].

Amb a 11 is 28 kDa glycosylated protein belonging to the peptidase family C1, subfamily C1A. It is very similar to the mite allergens Der p 1 and Der f 1, one of the most powerful known allergens. Recently, Amb a 11 was suggested as a novel major allergen with strong allergenicity (66%) [18]. Twenty different isoforms and glycoforms of the major allergen Amb a 11 were described in the WHO/IUIS allergen nomenclature database [15].

Immunoinformatics is a branch of bioinformatics that has been successfully used in the field of vaccinology. The immunoinformatics approach is the most efficient method for development of the epitope-based peptide vaccines [19, 20]. In recent years, epitope-based vaccines are preferable than the traditional vaccines because of several advantages: they are focused on the immune response, enhancing immunity and avoiding undesirable epitopes, more convenient production, and reducing costs [21].

Due to the climate change (warmer temperatures and increased carbon dioxide concentrations enable ragweed to produce more pollen), the number of individuals sensitive to ragweed pollen is increasing continuously in the United States and Europe [22, 23]. Therefore, the aim of this study was to design an efficient epitope-based peptide vaccine against the major allergen Amb a 11 using potential T-cell (CD8 + and CD4 +) and B-cell epitopes.

2 Materials and Methods

2.1 Retrieval of Amb a 11 and Identification of Conserved Regions Among Homologous Pollen Allergen of Amb a 11

The amino acids sequence and 3D structure of Amb a 11 allergenic protein was extracted from UniProt database (https://www.uniprot.org/) [24] in FASTA and PDB formats. In structural database of allergenic proteins (SDAP) [25], FAO/WHO Allergenicity Rules, which are based on sequence homology for full FASTA alignment, showed that the sequence of "V5LU01" contains 22 sequences of allergens with E-values < 0.01. The protein sequences of selected pollen allergens were extracted from Swiss-prot in FASTA format. Conservancy can be defined as the portion of protein sequences that restrain the epitope measured at or exceeding a specific level of identity. To identify the conserved regions of Amb a 11 protein, the sequence was aligned with retrieved sequences of selected pollen allergens by using the ClustalW program of the phylogenetic software MEGA-7 [26]. Phylogenetic tree of the retrieved sequences of the Amb a 11 and the pollen allergens was created using MEGA-7 software [26]. The evolutionary distances were calculated using the Maximum Composite Likelihood method within the software.

2.2 Prediction of CD8 + T-cell Epitopes and Their MHC Binding HLA-Alleles

MHC I prediction tool of the immune epitope database (IEDB) (http://tools.iedb.org/mhc1) [27] was used to determine the ability of the selected sequence to bind to a specific MHC class I molecule. Linear CD8 + T-cell epitopes were recognized by consensus approaches using various methods such as artificial neural network (ANN) [28], which was used to calculate IC_{50} values of peptide binding to MHC I molecules and determination of the HLA alleles; Stabilized matrix method (SMM) [29], which was applied to predict peptide binding to MHC molecules, proteasomal cleavage of protein sequences and peptide transport by the transporter associated with antigen presentation (TAP). For both, frequent and non-frequent alleles, peptide length was set to 10 amino acids. On the basis of low IC_{50} values, only 10 potential peptides were chosen for next processing.

2.3 Prediction of CD4 + T-cell Epitopes and Their MHC Binding HLA-Alleles

The MHC II groove has the ability to bind peptides of different lengths, making the prediction more difficult and less

accurate. MHC II prediction tool provided by the immune epitope database (IEDB) was used to predict 15 amino acid long CD4 + T-cell epitopes. MHC II binding prediction was achieved using a consensus approach, which combines NN-align, SMM-align [30], and Combinatorial library methods. All epitopes that bind to many alleles at score equal or less than 1000 half-maximal inhibitory concentration (IC_{50}) were selected for further analysis.

2.4 Population Coverage Prediction for CD8 + and CD4 + T-cell Epitopes and Their Alleles

The population coverage for identified T-cell epitopes was calculated by using epitope analysis tools of IEDB analysis resource server (http://tools.iedb.org/population/) [31]. All finally predicted CD8+ and CD4 + T-cell epitopes and their MHC I and MHC II molecules were assessed for the European and world population.

2.5 Identification of B-cell Epitopes

The prediction of B-cell epitopes is a main step for epitope-based vaccine development as it produces antibodies that assure humoral immunity. Recognition of B-cell epitopes was depended on various physico-chemical properties such as antigenicity, accessibility of surface, flexibility, turns, exposed surface, polarity hydrophilicity and predictions of linear epitope [32]. Potential B-cell epitopes were characterized by B-cell epitope prediction tools (http://tools.iedb.org/bcell/) of the immune epitope database using B-cell prediction methods such as Kolaskar and Tongaonkar antigenicity scale [33], Bepipred linear epitope prediction [34] Emini surface accessibility prediction [35], Karplus and Schulz flexibilty prediction [36], Parker hydrophilicity prediction [37] and Chou and Fashman beta turn prediction tool [38].

2.6 In Silico Molecular Docking

Peptide structure prediction server (PEP-FOLD 3) (https://bioserv.rpbs.univ-paris-diderot.fr/services/PEP-FOLD3/) was utilized to predict the three dimensional structure of the selected epitopes [39]. The crystal structures of HLA-A0201 (PDB ID: 4UQ3) and DQA1*03:01/DQB1*02:01 (PDB ID: 4D8P) were downloaded from the RCSB protein data bank (http://www.rcsb.org/pdb/home/home.do) as PDB format [40, 41]. Crystal structures were in the complex form of protein and ligand, hence the discovery studio v16.1.0 [42]. was used to separate the protein and ligand from the complex structures. All water molecules and heteroatoms were then removed. In silico docking simulation was done in order to estimate the binding affinities between the epitopes and specific HLA proteins (MHC I & MHC II) and immunoglobulin E (IgE). Docking simulation study was performed through automated online server ClusPro 2.0 (https://cluspro.bu.edu/home.php) [43].

2.7 Experimental Validation

Ten subjects (7 females and 3 males, 30–45 years old) with proven allergy to ragweed pollen and 10 healthy age-matched controls (4 females and 6 males, 32–43 years old) were enrolled in this study. The selection of allergic patients was performed in collaboration with an experienced allergist based on a positive prick test to ragweed pollen (ALK-Abelló A/S Hørsholm, Denmark) and presence of IgE antibodies specific to ragweed pollen extract (ImmunoCAP, ImmunoDiagnostics Thermo Fisher Scientific, Uppsala, Sweden). Recruited healthy individuals had negative prick tests to ragweed pollen and no clinical histories of pollen allergies. None of the participants had received corticosteroids or allergen-specific immunotherapy during the last 2 months before taking blood samples.

Peripheral venous blood samples (3 mL) were collected from each participant into BD Vacutainer® K2EDTA tubes (Becton, Dickinson and Company, Oakville, ON, USA) and centrifuged at 1500 g for 15 min. The sera were transferred to 2 mL cryotubes and stored at − 80 °C until further analyses. Then, red blood cells were lysed with 0.84% NH_4Cl, and after washing with Dulbecco's phosphate-buffered saline (D-PBS, Gibco®, Life Technologies™, Paisley, Scotland, UK), the remaining cells were resuspended at a concentration of 2.5×10^6 cells/mL in RPMI-1640 medium, supplemented with a stabilized antibiotic antimycotic solution (penicillin, streptomycin, amphotericin B) and 10% fetal bovine serum (all from Sigma-Aldrich Chemie GmbH, Steinheim, Germany).

Cells were incubated at 37 °C in a humidified incubator containing 5% CO_2 in the presence of 5 µg/mL of the identified T-cell (MHC class-I peptide GLMEPAFTYV and MHC class-II peptide LVCFSFSLVLILGLV) and B-cell (GKLVKFSEQQLVDC) epitopes in a volume of 200 µL in 96-well plates (TPP, Trasadingen, Switzerland). Peptides representing T- and B-cell epitopes were synthesized by Schafer-N (Copenhagen, Denmark).

After 72 h, the production of IL-2 in the culture supernatant was measured by LEGEND MAX™ Human IL-2 kit (BioLegend Inc., San Diego, CA, USA), according to the manufacturer's instructions. The sensitivity of the assay was 4 pg/mL.

IgE antibodies specific to the identified epitopes were assayed by ELISA. Briefly, plates were coated overnight at 4 °C with 10 µg/mL of the synthetic peptides that represent the epitopes (50 µL/well). After blocking with D-PBS, 0.05% Tween 20 and 1% BSA, the plates were washed and sera diluted 1:5 (100 µL/well) were applied onto the coated plates in triplicates and incubated overnight at 4 °C. IgE

was detected with peroxidase-conjugated monoclonal anti-human IgE (clone GE-1, Sigma-Aldrich, St. Louis, MO, USA). 2,2-azino-di-[2-ethylbenzthiazoline sulfonate] diammonium salt (ABTS tablets, Roche Diagnostics) was used as substrate, and OD values were determined by a microplate reader (ELx800™, BioTek Instruments, Winooski, VT, USA) at 405 nm.

All tests and analyses were performed in adherence to the Declaration of Helsinki ethical guidelines and was approved by the local Ethics committee at the Paisii Hilendarski University of Plovdiv. Written informed consent was obtained from all participants.

2.8 Statistics

All experiments related to the experimental validation were conducted in triplicate and results are presented as mean ± SD (n = 10). To compare non-parametric data for statistical significance, the Kruskal–Wallis test was applied using the StatView program (SAS Institute Inc., Cary, NC, USA). Values of $p < 0.05$ were considered significant (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$).

3 Results

3.1 Retrieving of Amb a 11 Protein Sequences and Identification of Conserved Domains

In this study, the amino acids sequence of the Amb a 11 was retrieved from the UniProt database and then applied FAO/WHO Allergenicity Rules based on sequence homology in the SDAP allergen database. We identified 22 homologous allergens of the pollen allergen Amb a 11 as shown in Table 1. Multiple sequence alignment was performed to identify conserved sequences with varying lengths (Table 2). The conserved region analysis helps us to identify the essential functional parts of the target protein sequences. A phylogenetic tree was created to indicate the evolutionary relationship of the pollen allergen Amb a 11 and the identified 22 allergens (Fig. 1).

Table 1 List of homologous allergens of Amb a 11

No.	Allergen	Sequence ID in SwissProt/NCBI/PIR	Sequence length	E score
1	Act d 1	P00785	380	2.2e−41
2	Act d 1	AAA32629	380	4.4e−41
3	Ana c 2	BAA21849	351	6.3e−41
4	Car p papain	AAB02650	345	7.0e−39
5	Gly m 1	AAB09252	379	1.0e−35
6	Gly m 1	P22895	379	1.5e−35
7	Der f 1.0109	ABL84750	321	5.2e−25
8	Der f 1.0110	ABL84751	321	5.2e−25
9	Der f 1.0101	BAC53948	321	5.2e−25
10	Der f 1.0108	ABL84749	321	1.1e−24
11	Der f 1	P16311	321	1.6e−24
12	Eur m 1.0102	AAC82352	321	1.9e−24
13	Eur m 1.0101	P25780	321	2.9e−24
14	Eur m 1.0101	AAC82351	321	2.9e−24
15	Der f 1	CAA46316	212	6.9e−23
16	Der p 1.0124	CAQ68250	302	3.6e−22
17	Der p 1.0113	ABA39435	302	5.5e−22
18	Der p 1	P08176	320	7.8e−22
19	Eur m 1.0101	CAA42677	211	7.9e−22
20	Blo t 1.0201	AAQ24541	333	5.0e−16
21	Blo t 1.0101	AAK58415	221	6.8e−14
22	Der p 1	AAA28296	96	4.4e−08

Table 2 Conserved sequences of homologous pollen allergens with their positions

Conserved sequences of pollen allergen	Position of the selected sequences
QGGCGSCWAF	164–173
SEQELVDC	194–201
YPYV	239–242
IVGYG	326–330
YWIVRNSWDTTWG	338–350

3.2 Prediction of CD8 + T-cell Epitopes and Their MHC Binding HLA-Alleles

In this experiment, we utilized the MHC I binding prediction tool from the immune epitope database (IEDB) to calculate IC_{50} value and percentile rank for peptides binding to specific MHC I molecules. Peptides that have the ability to bind to a larger number of alleles are identified as the most suitable because of their potential to elicit a potent protective response. Only ten CD8 + T-cell epitopes were selected as potential vaccine candidates according to the low percentile rank and IC_{50} values for their respective alleles (Table 3).

3.3 Prediction of CD4 + T-cell Epitopes and Their MHC Binding HLA-Alleles

In this study, the amino acids sequence of Amb a 11 allergen was analyzed using IEDB MHC II binding prediction tool to determine the ability of the submitted sequence to bind to a specific MHC II molecule. A consensus approach, which combines NN-align, SMM-align, and Combinatorial library methods were used for the prediction of peptides, which can

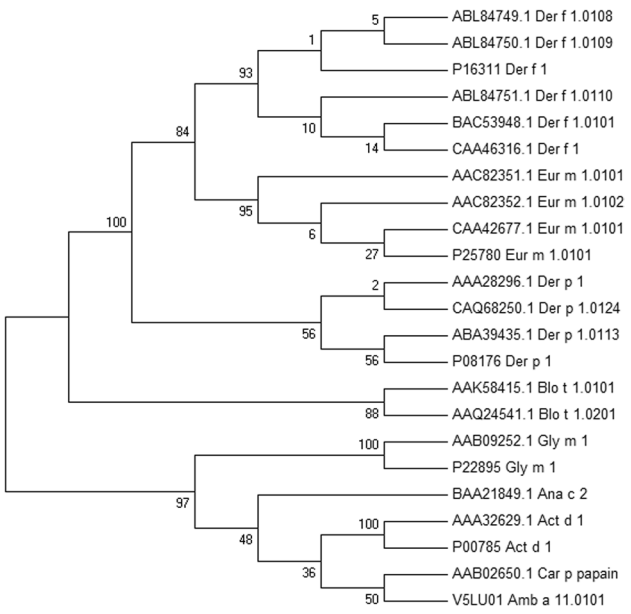

Fig. 1 Phylogenetic tree based on amino acid sequences of the pollen allergen Amb a 11 and the identified 22 homologous allergens. The reconstruction has been performed by using ML analysis applying the GTR + I + G evolutionary model. The numbers above branches indicate the bootstrap support from 1000 replicates

interact with MHC class II alleles. The list of the most promising CD4 + T-cell epitopes and their correspondent binding MHC II alleles are shown in Table 4.

3.4 Analysis of Population Coverage

The population coverage analysis is a vital factor in the design of epitope-based vaccines. High population coverage of vaccine compounds is significant due to a lot of people can be benefited by only one vaccine preparation. In this study, IEDB population coverage tool (http://tools.iedb.org/population/) was applied to determine the population coverage of the predicted CD8 + and CD4 + T-cell epitopes. The population coverage analysis results showed a great coverage of the European and world population (Table 5).

3.5 B-cell Epitope Prediction

B-cell epitope identification is an essential characteristic in the design of an epitope-based peptide vaccine because an ideal epitope-based vaccine contains T- and B-cell epitopes. The amino acid sequence of Amb a 11 allergenic protein was subjected to Parker Hydrophilicity, Chou & Fasman Beta-Turn, Karplus & Schulz Flexibility, Emini surface accessibility, and Kolaskar and Tongaonkar antigenicity methods in IEDB, to determine the binding to B cell, being in the surface and to test the immunogenicity. From all predicted B-cell epitopes, only one epitope ("GKLVKF**SEQQLVDC**") was chosen for next processing (Table 6). Predicted linear B-cell epitope and the accessibility, antigenicity, hydrophilicity, flexibility and beta-turn prediction score for each residue are summarized in Table 6. This peptide contains the sequence "SEQQLVDC", which is presented in the conserved region of pollen allergen Amb a 11 (Table 2).

3.6 Docking Study of B-Cell and T-Cell Epitopes

A web-based server, PEP-FOLD 3 was used to create 3D structures of the predicted T- and B-cell epitopes ("GLMEPAFTYV", "LVCFSFSLVLILGLV" and "GKLVKF**SEQQLVDC**") (Fig. 2). Five models were created for each peptide and the structure with the lowest energy was chosen for further analysis. The selected models were docked through ClusPro v2.0 with the MHC receptors and the immunoglobulin protein E, respectively. CD8 + T-cell peptide showed significant binding affinity to the MHC I molecule (Fig. 3B). The 3D structures of the best docked-complexes are illustrated in Fig. 3.

3.7 Functional Validation of Predicted Epitopes

In order to validate the functional activity of the predicted epitopes we have tested whether the T-cells and antibodies from patients allergic to ragweed pollen recognize these epitopes. As can be seen in Fig. 4, T-cell response was observed only for the cells isolated from allergic patients against T-cell epitope 2 (LVCFSFSLVLILGLV) presented from MHC class II molecules. Stimulation of T cells with this peptide significantly increased production of IL-2 in the group of patients with ragweed allergy compared to the healthy controls (Fig. 4). There was no response against predicted T-cell epitope 1 (GLMEPAFTYV) and the B-cell epitope (GKLVKFSEQQLVDC).

When we evaluated the sera collected from the same allergic and healthy subjects for presence of specific IgE antibodies, we detected in the allergic group IgE that specifically recognized the identified B-cell epitope (Fig. 5). Compared to the control group, levels of these antibodies were significantly higher.

Our data showed that predicted T-cell epitope 2 (LVCFSFSLVLILGLV) and the B-cell epitope (GKLVKFSEQQLVDC) are functional. They can be recognized by T-cells and antibodies, respectively.

4 Discussion

Nowadays, the development of new vaccines mainly depends on the design of peptide-based vaccines. The synthetic peptide-based vaccines are an alternative solution to overcome all the adverse effects of the classical vaccines. The

Table 3 Most potential CD8 + T-cell epitopes with interacting MHC-I alleles, their positions, length of epitope, Ann IC_{50}, Smm IC_{50}, and percentile rank

No.	T-cell epitopes	Interacting MHC-I alleles	Start position	End position	Length of epitope	Percentile rank	Ann IC_{50}	Smm IC_{50}
1	LQFYSEGVY	HLA-B*15:01	271	279	9	0.1	7.97	9.44
		HLA-A*30:02				0.22	22	93.64
2	NVFKYNVRR	HLA-A*68:01	58	66	9	0.12	5.59	7.47
		HLA-A*33:01				0.44	41.77	106.67
3	**GLMEPAFTYV**	HLA-A*02:01	196	205	10	0.12	5.24	2.23
		HLA-A*02:03				0.12	3.32	2.31
		HLA-A*02:06				0.17	4.1	19.01
		HLA-A*68:02				0.2	19.67	32.95
		HLA-B*15:01				0.72	21.51	175.15
		HLA-A*03:01				1.24	22.66	281.39
		HLA-A*31:01				2.15	23.4	1755.17
4	MTNLEFVNTY	HLA-A*26:01	87	96	10	0.15	169.65	287.53
		HLA-A*01:01				0.28	121.96	416.68
5	TYANSKISHF	HLA-A*23:01	95	104	10	0.16	33.13	18.82
		HLA-A*24:02				0.2	65.19	68.98
6	RFNVFKYNVR	HLA-A*31:01	56	65	10	0.16	4.51	21.55
7	GENEKGIKFW	HLA-B*44:02	297	306	10	0.17	29.91	85.68
		HLA-B*44:03				0.28	51.23	269.2
8	MEINKLVCF	HLA-B*44:02	1	9	9	0.18	42.26	248.66
		HLA-B*44:03				0.18	95.9	110.53
9	ESEEGFMGMY	HLA-A*26:01	31	40	10	0.19	86.03	542.86
		HLA-A*01:01				0.2	37.78	351.4
10	FADMTNLEF	HLA-B*53:01	84	92	9	0.2	20.49	39.53
		HLA-A*01:01				0.22	50.8	313.73
		HLA-B*35:01				0.4	8.44	15.76

improvement in the field of immunoinformatics has become a potential field for predicting the efficacy of the peptide-based vaccines [44].

In the current study, potential T- and B-cell epitopes were evaluated as possible epitope-based peptide vaccines against the allergen Amb a 11. The obtained results of FAO/WHO Allergenicity Rules based on Sequence Homology showed that Amb a 11 shares homology with other major allergens belonging to the same protease family, such as Act d 1 from *Actinidia deliciosa* (37.0%), Ana c 2 from *Ananas comosus* (34.1%), and Der f/Der p 1 from *Dermatophagoides farinae/Dermatophagoides pteronyssinus* (26.0/24.4%;). These results correspond to the results obtained by Bouley et al. [18]. Car p papain from *Carica papaya* (34.6%), Blo t 1 from *Blomia tropicalis* (34.6%), Eur m 1 from *Euroglyphus maynei*, and Gly m 1 from *Glycine max* were newly established. Multiple sequence alignment with other allergens was performed to identify the conserved sequences (Table 2). A phylogenetic tree was created to indicate the evolutionary relationship of the pollen allergen Amb a 11 and the identified twenty two homologous allergens (Fig. 1). The phylogenetic analysis showed a close evolutionary relationship between the homologous allergens.

The binding between T-cell epitopes and the MHC molecules is crucial for the cellular immune response. The selection of T-cell epitopes was performed according to the binding affinity of the highest numbers of alleles and the lowest IC_{50} value. The CD8 + T-cell peptide "GLMEPAFTYV" was found to interact with seven alleles (HLA-A*02:01, HLA-A*02:03, HLA-A*02:06, HLA-A*68:02, HLA-B*15:01, HLA-A*03:01, HLA-A*31:01), while the CD4 + T-cell peptide "LVCFSFSLVLILGLV" interacts with a large number of alleles including HLA-DQA1*03:01/DQB1*02:01, HLA-DPA1*03:01/DPB1*04:02, HLA-DPA1*01/DPB1*04:01, HLA-DPA1*01:03/DPB1*02:01, HLA-DPA1*02:01/DPB1*01:01, respectively.

B-cell epitope prediction is an essential characteristic in the design of an epitope-based vaccine [45]. The prediction scores of Emini surface accessibility [35], Parker hydrophilicity [37], Karplus and Schulz flexibility [36], and Chou and Fashman beta turn [38] for each residue of the peptide "GKLVKF**SEQQLVDC**" as a part of the amino acid

Table 4 Most potential CD4 + T-cell epitopes with interacting MHC-II alleles, their positions, length of epitope, Smm IC_{50}, and percentile rank

No.	T-cell epitopes	Interacting MHC-II alleles	Start position	End position	Length	Percentile rank	Smm IC_{50}
1	CGSCWAFAAVVALEG	HLA-DQA1*04:01/DQB1*04:02	152	166	15	0.06	112.00
		HLA-DQA1*01:02/DQB1*06:02				1.05	123.00
		HLA-DQA1*05:01/DQB1*03:01				1.45	34.00
2	GCGSCWAFAAVVALE	HLA-DQA1*04:01/DQB1*04:02	151	165	15	0.06	110.00
		HLA-DQA1*01:02/DQB1*06:02				1.15	133.00
3	GSCWAFAAVVALEGI	HLA-DQA1*04:01/DQB1*04:02	153	167	15	0.06	112.00
		HLA-DQA1*01:02/DQB1*06:02				1.05	121.00
		HLA-DQA1*05:01/DQB1*03:01				1.65	37.00
4	KLVCFSFSLVLILGL	HLA-DPA1*03:01/DPB1*04:02	5	19	15	0.11	17.00
		HLA-DPA1*01/DPB1*04:01				0.51	80.00
5	**LVCFSFSLVLILGLV**	HLA-DPA1*03:01/DPB1*04:02	6	20	15	0.11	17.00
		HLA-DPA1*01/DPB1*04:01				0.51	80.00
		HLA-DPA1*01:03/DPB1*02:01				0.66	54.00
		HLA-DPA1*02:01/				0.87	102.00
		DPB1*01:01				1.30	284.00
		HLA-DQA1*01:02/					
		DQB1*06:02					
6	VCFSFSLVLILGLVE	HLA-DPA1*03:01/DPB1*04:02	7	21	15	0.12	18.00
		HLA-DPA1*01/DPB1*04:01				0.60	95.00
		HLA-DPA1*02:01/DPB1*01:01				0.87	43.00
7	INKLVCFSFSLVLIL	HLA-DPA1*03:01/DPB1*04:02	3	17	15	0.18	23.00
		HLA-DPA1*01/DPB1*04:01				0.35	62.00
		HLA-DPA1*01:03/DPB1*02:01				0.39	36.00
8	NKLVCFSFSLVLILG	HLA-DPA1*03:01/DPB1*04:02	4	18	15	0.18	23.00
		HLA-DPA1*01/DPB1*04:01				0.35	62.00
		HLA-DPA1*01:03/DPB1*02:01				0.37	35.00
9	CFSFSLVLILGLVES	HLA-DPA1*03:01/DPB1*04:02	8	22	15	0.26	32.00
		HLA-DPA1*02:01/DPB1*01:01				1.25	65.00
10	FSFSLVLILGLVESF	HLA-DPA1*03:01/DPB1*04:02	9	23	15	0.27	33.00
		HLA-DQA1*05:01/DQB1*02:01				0.85	428.00

sequence of Amb a 11, reveal that this peptide is the most possible B-cell epitope presented in this pollen allergen.

The population coverage is an important analysis step that might reflect to the possibility of the epitope to be used in vaccine design [46]. The selected T-cell peptides were subjected to MHC I and MHC II based population coverage analysis in Europe and worldwide as shown in Table 5. It was predicted that CD4 + T-cell epitope "LVCFSFSLVLILGLV" covers 90.56% of the population of Europe and 99.36% of the world population. CD8 + T-cell epitope "GLMEPAFTYV" has a population coverage of 77.37% for Europe and 71.35% worldwide. So, it may be concluded that these epitopes may cover a high amount of population when applied as a vaccine.

Our validation experiments demonstrate that T-cell epitope 2 (LVCFSFSLVLILGLV) could be used as a potential vaccine against the major allergen Amb a 11, because it can activate T cells leading to production of IgG antibodies against the ragweed pollen (neutralizing antibodies), but at the same time it is not recognized by the IgE antibodies. Thus, this peptide should not cause an allergic reaction.

An in silico molecular docking was done to explore the binding affinity between the selected peptides and their immune receptors (MHC-I and MHC-II molecules and IgE). The HLA-A0201, HLA-DQA1*03:01/DQB1*02:01 and immunoglobulin E "4poz" were selected as the macromolecules for docking. Docking study of CD8 + T-cell epitope "GLMEPAFTYV" with HLA-A0201, showed lowest binding energy − 729.3 kcal/mole. Interaction between CD4 + T-cell epitope "LVCFSFSLVLILGLV" with MHC II allelic protein HLA-DQA1*03:01/DQB1*02:01 with binding energy − 725.0 kcal/mole is shown in Fig. 3A. The B-cell epitope GKLVKF**SEQQLVDC** forms a stable complex with the IgE molecule (energy weighted score − 695,0) (Fig. 3C). The 3D structures of the best docked-complexes are illustrated in Fig. 3.

Table 5 Population coverage for predicted T-cell epitopes

Predicted CD8 + T-cell epitopes	Interacting MHC I alleles	Predicted CD4 + T-cell epitopes	Interacting MHC II alleles	Population coverage	
				World	Europe
LQFYSEGVY	HLA-B*15:01 HLA-A*30:02	CGSCWAFAAVVALEG	HLA-DQA1*04:01/ DQB1*04:02 HLA-DQA1*01:02/ DQB1*06:02 HLA-DQA1*05:01/ DQB1*03:01	CD8 + : 10.67% CD4 + : 80.52%	CD8 + : 11.42% CD4 + : 79.36%
NVFKYNVRR	HLA-A*68:01 HLA-A*33:01	GCGSCWAFAAVVALE	HLA-DQA1*04:01/ DQB1*04:02 HLA-DQA1*01:02/ DQB1*06:02	CD8 + : 7.56% CD4 + : 47.25%	CD8 + : 7.64% CD4 + : 45.43%
GLMEPAFTYV	HLA-A*02:01 HLA-A*02:03 HLA-A*02:06 HLA-A*68:02 HLA-B*15:01 HLA-A*03:01 HLA-A*31:01	GSCWAFAAVVALEGI	HLA-DQA1*04:01/ DQB1*04:02 HLA-DQA1*01:02/ DQB1*06:02 HLA-DQA1*05:01/ DQB1*03:01	CD8 + : 71.35% CD4 + : 68.82%	CD8 + : 77.37% CD4 + : 79.36%
MTNLEFVNTY	HLA-A*26:01 HLA-A*01:01	KLVCFSFSLVLILGL	HLA-DPA1*03:01/ DPB1*04:02 HLA-DPA1*01/ DPB1*04:01	CD8 + : 22.62% CD4 + : 68.91%	CD8 + : 32.31% CD4 + : 86.60%
TYANSKISHF	HLA-A*23:01 HLA-A*24:02	**LVCFSFSLVLILGLV**	HLA-DQA1*03:01/ DQB1*02:01 HLA-DPA1*03:01/ DPB1*04:02 HLA-DPA1*01/ DPB1*04:01 HLA-DPA1*01:03/ DPB1*02:01 HLA-DPA1*02:01/ DPB1*01:01	CD8 + : 26.18% CD4 + : 90.56%	CD8 + : 21.31% CD4 + : 99.36%
RFNVFKYNVR	HLA-A*31:01	VCFSFSLVLILGLVE	HLA-DPA1*03:01/ DPB1*04:02 HLA-DPA1*01/ DPB1*04:01 HLA-DPA1*02:01/ DPB1*01:01	CD8 + : 5.36% CD4 + : 67.19%	CD8 + : 4.72% CD4 + : 79.13%
GENEKGIKFW	HLA-B*44:02 HLA-B*44:03	INKLVCFSFSLVLIL	HLA-DPA1*03:01/ DPB1*04:02 HLA-DPA1*01/ DPB1*04:01 HLA-DPA1*01:03/ DPB1*02:01	CD8 + : 13.63% CD4 + : 84.60%	CD8 + : 20.13% CD4 + : 99.36%

Table 5 (continued)

Predicted CD8 + T-cell epitopes	Interacting MHC I alleles	Predicted CD4 + T-cell epitopes	Interacting MHC II alleles	Population coverage	
				World	Europe
MEINKLVCF	HLA-B*44:02	NKLVCFSFSLVLILG	HLA-DPA1*03:01/	CD8 +:	CD8 +:
	HLA-B*44:03		DPB1*04:02	13.63%	20.13%
			HLA-DPA1*01/	CD4 +:	CD4 +:
			DPB1*04:01	84.60%	99.36%
			HLA-DPA1*01:03/		
			DPB1*02:01		
ESEEGFMGMY	HLA-A*26:01	CFSFSLVLILGLVES	HLA-DPA1*03:01/	CD8 +:	CD8 +:
	HLA-A*01:01		DPB1*04:02	22.62%	32.31%
			HLA-DPA1*02:01/	CD4 +:	CD4 +:
			DPB1*01:01	54.34%	41.00%
FADMTNLEF	HLA-B*53:01	FSFSLVLILGLVESF	HLA-DPA1*03:01/	CD8 +:	CD8 +:
	HLA-A*01:01		DPB1*04:02	26.30%	33.82%
	HLA-B*35:01		HLA-DQA1*05:01/	CD4 +:	CD4 +:
			DQB1*02:01	63.48%	62.03%

Table 6 Surface accessibility (threshold = 1.000), hydrophilicity (threshold = 1.916), flexibility (threshold = 1.004), beta turn (threshold = 1.016) and antigenicity (threshold = 1.013) prediction score for each residue of B-cell epitope

B-cell epitope	Emini surface accessibility	Parker hydrophilicity	Karplus and Schulz flexibility	Chou and Fashman beta turn	Kolaskar and Tongaonkar antigenicity
G	0.949	1.714	1.052	1.063	1.053
K	0.863	1.943	1.045	1.094	1.021
L	0.998	1.829	1.017	1.089	1.052
V	1.039	1.614	0.998	1.057	1.067
K	1.062	1.719	1.0	1.084	1.064
F	1.379	1.957	1.013	1.036	1.076
S	2.285	2.729	1.041	1.091	1.042
E	1.942	1.943	1.059	1.104	1.023
Q	1.808	2.6	1.056	1.131	1.088
Q	1.007	3.343	1.032	1.054	1.056
L	1.312	2.614	1.097	1.192	1.113
V	1.3	2.929	1.066	1.023	1.115
D	1.172	1.471	1.047	1.069	1.088
C	1.386	1.357	1.05	1.066	1.073

5 Conclusion

At the moment, in Europe, short ragweed (*Ambrosia artemisiifolia* L.) is of great concern because it is one of the most prominent invasive species. The development of allergy to the short ragweed pollen is a major health problem. Therefore, in this study we designed an epitope-based peptide vaccine against Amb a 11 (a new major allergen belonging to the cysteine protease family) by using an immunoinformatics approach. The proposed T- and B-cell epitopes ("GLMEPAFTYV", "LVCFSFSLVLILGLV" and "GKLVKFSEQQLVDC") showed high susceptibility and can be proceeded for further experimental verification for vaccine designing against the allergen Amb a 11. The MHC class II peptide (LVCFSFSLVLILGLV) is a promising candidate as a specific vaccine against the major allergen Amb a 11 because it can stimulate T cells from ragweed allergic patients.

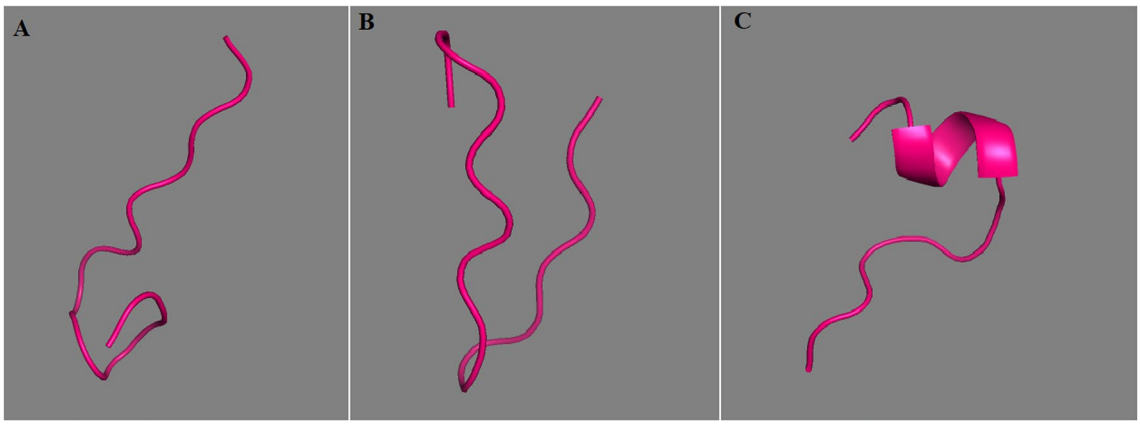

Fig. 2 3D structures of the predicted T and B-cell epitopes. **A** CD4 + T-cell epitope; **B** CD8 + T-cell epitope; **C** B-cell epitope

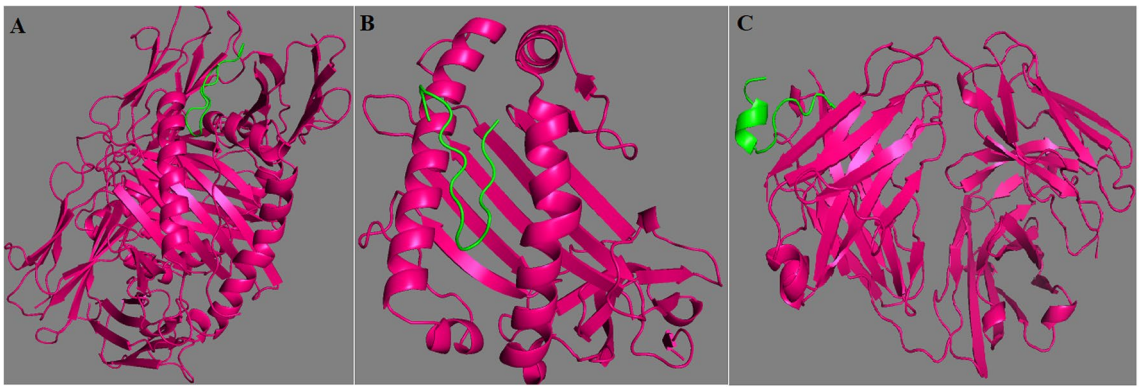

Fig. 3 **A** Peptide-allele docked complex of the predicted CD4 + T-cell epitope "LVCFSFSLVLILGLV" with respective frequent allele (energy weighted score − 725.0 kcal/mole); **B** Molecular docking of CD8 + T-cell epitope "GLMEPAFTYV" of Amb a 11 allergenic protein docked in HLA-A0201 (energy weighted score − 729.3 kcal/mole); **C** B-cell epitope "GKLVKFSEQQLVDC" is bound with immunoglobulin E "4poz" with lowest energy weighted score − 695.0 kcal/mole

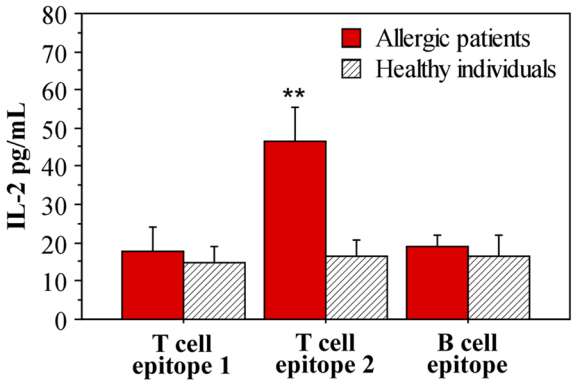

Fig. 4 T-cell response to identified T- and B-cell epitopes in patients with ragweed pollen allergy (n = 10) and healthy individuals (n = 10). IL-2 production in the supernatants was assayed by ELISA after stimulation of the cells with the synthetic peptides for 72 h. T-cell epitope 1: GLMEPAFTYV; T-cell epitope 2: LVCFSFSLVLILGLV; B-cell epitope: GKLVKFSEQQLVDC. **$p < 0.01$ as determined by Kruskal–Wallis test. Results are presented as means ± SD (n = 10)

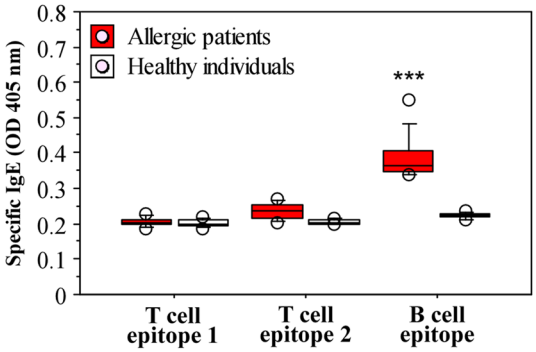

Fig. 5 Serum levels of specific IgE antibodies to identified T- and B-cell epitopes in patients with ragweed pollen allergy (n = 10) and healthy individuals (n = 10). T-cell epitope 1: GLMEPAFTYV; T-cell epitope 2: LVCFSFSLVLILGLV; B-cell epitope: GKLVKFSEQQLVDC. ***$p < 0.001$ as determined by Kruskal–Wallis test

Author Contributions Concept and planning were done by DM and BD; data preparing and computational analysis were done by DM, DK, KT and TM; the manuscript was written by DM and BD; the manuscript was reviewed and rewritten by DM, DK, KT, TM and BD.

Funding This work was supported by the Bulgarian National Research Fund; contract number KP-06-M31/2. The National program "Young Researchers and Postdocs, 2021" financed by the Ministryof Education and Science, Bulgaria, is highly appreciated.

Declarations

Conflict of interest The authors declare that they have no potential conflict of interest.

References

1. Smith M, Cecchi L, Skjoth CA, Karrer G, Sikoparija B (2013) Common ragweed: a threat to environmental health in Europe. Environ Int 61:115–126
2. Taramarcaz P, Lambelet B, Clot B, Keimer C, Hauser C (2005) Ragweed (*Ambrosia*) progression and its health risks: will Switzerland resist this invasion? Swiss Med Wkly 135:538–548
3. Lehoczky E, Szabo R, Nelima MO, Nagy P, Beres I (2010) Examination of common ragweed's (*Ambrosia artemisiifolia* L.) allelopathic effect on some weed species. Commun Agric Appl Biol Sci 75:107–111
4. D'Amato G, Cecchi L, Bonini S, Nunes C, Annesi-Maesano I, Behrendt H et al (2007) Allergenic pollen and pollen allergy in Europe. Allergy 62:976–990
5. Quimby PC Jr, DeLoach CJ, Wineriter SA, Goolsby JA, Sobhian R, Boyette CD et al (2003) Biological control of weeds: research by the United States Department of Agriculture-Agricultural Research Service: selected case studies. Pest Manage Sci 59:671–680
6. Fumanal B, Chauvel B, Bretagnolle F (2007) Estimation of pollen and seed production of common ragweed in France. Ann Agric Environ Med 14:233–236
7. Burbach GJ, Heinzerling LM, Rohnelt C, Bergmann KC, Behrendt H, Zuberbier T et al (2009) Ragweed sensitization in Europe—GA(2)LEN study suggests increasing prevalence. Allergy 64:664–665
8. Sikoparija B, Smith M, Skjoth CA, Radisic P, Milkovska S, Simic S et al (2009) The Pannonian plain as a source of *Ambrosia* pollen in the Balkans. Int J Biometeorol 53:263–272
9. Cunze S, Leiblein MC, Tackenberg O (2013) Range expansion of *Ambrosia artemisiifolia* in Europe is promoted by climate change. ISRN Ecol 2013:610126
10. Chapman DS, Haynes T, Beal S, Essl F, Bullock JM (2014) Phenology predicts the native and invasive range limits of common ragweed. Glob Chang Biol 20:192–202
11. Emberlin J (1994) The effects of patterns in climate and pollen abundance on allergy. Allergy 49:15–20
12. DellaValle CT, Triche EW, Leaderer BP, Bell ML (2012) Effects of ambient pollen concentrations on frequency and severity of asthma symptoms among asthmatic children. Epidemiology 23:55–63
13. Asero R, Wopfner N, Gruber P, Gadermaier G, Ferreira F (2006) *Artemisia* and *Ambrosia* hypersensitivity: co-sensitization or co-recognition? Clin Exp Allergy 36:658–665
14. Gadermaier G, Wopfner N, Wallner M, Egger M, Didierlaurent A, Regl G et al (2008) Array-based profiling of ragweed and mugwort pollen allergens. Allergy 63:1543–1549
15. Goodman RE, Breiteneder H (2019) The WHO/IUIS allergen nomenclature. Allergy 74:429–431
16. Bordas-Le Floch V, Groeme R, Chabre H, Baron-Bodo V, Nony E, Mascarell L et al (2015) New insights into ragweed pollen allergens. Curr Allergy Asthma Rep 15:63
17. Stemeseder T, Hemmer W, Hawranek T, Gadermaier G (2014) Marker allergens of weed pollen—basic considerations and diagnostic benefits in the clinical routine: part 16 of the series molecular allergology. Allergo J Int 23:274–280
18. Bouley J, Groeme R, Le Mignon M, Jain K, Chabre H, Bordas-Le Floch V et al (2015) Identification of the cysteine protease Amb a 11 as a novel major allergen from short ragweed. J Allergy Clin Immunol 136:1055–1064
19. Mahendran R, Jeyabaskar S, Sitharaman G, Michael RD, Paul AV (2016) Computer-aided vaccine designing approach against fish pathogens *Edwardsiella tarda* and *Flavobacterium columnare* using bioinformatics softwares. Drug Des Dev Ther 10:1703–1714
20. Zahroh H, Ma'rup A, Tambunan US, Parikesit AA (2016) Immunoinformatics approach in designing epitope-based vaccine against meningitis-inducing bacteria (*Streptococcus pneumoniae*, *Neisseria meningitidis*, and *Haemophilus influenzae* Type b). Drug Target Insights 10:19–29
21. Hirayama M, Nishimura Y (2016) The present status and future prospects of peptide-based cancer vaccines. Int Immunol 28:319–328
22. Cunze S, Heydel F, Tackenberg O (2013) Are plant species able to keep pace with the rapidly changing climate? PLoS ONE 8:e67909
23. Ring J, Kramer U, Schafer T, Behrendt H (2001) Why are allergies increasing? Curr Opin Immunol 13:701–708
24. Groeme R, Airouche S, Kopecny D, Jaekel J, Savko M, Berjont N et al (2016) Structural and functional characterization of the major allergen Amb a 11 from short ragweed pollen. J Biol Chem 291:13076–13087
25. Ivanciuc O, Schein CH, Braun W (2003) SDAP: database and computational tools for allergenic proteins. Nucleic Acids Res 31:359–362
26 Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol 33:1870–1874
27. Vita R, Mahajan S, Overton JA, Dhanda SK, Martini S, Cantrell JR et al (2019) The immune epitope database (IEDB): 2018 update. Nucleic Acids Res 47:D339–D343
28. Nielsen M, Lundegaard C, Worning P, Lauemoller SL, Lamberth K, Buus S et al (2003) Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. Protein Sci 12:1007–1017
29. Peters B, Sette A (2005) Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. BMC Bioinform 6:132
30. Nielsen M, Lundegaard C, Lund O (2007) Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method. BMC Bioinform 8:238
31. Bui HH, Sidney J, Dinh K, Southwood S, Newman MJ, Sette A (2006) Predicting population coverage of T-cell epitope-based diagnostics and vaccines. BMC Bioinform 7:153
32. Fieser TM, Tainer JA, Geysen HM, Houghten RA, Lerner RA (1987) Influence of protein flexibility and peptide conformation on reactivity of monoclonal anti-peptide antibodies with a protein alpha-helix. Proc Natl Acad Sci U S A 84:8568–8572
33. Kolaskar AS, Tongaonkar PC (1990) A semi-empirical method for prediction of antigenic determinants on protein antigens. FEBS Lett 276:172–174
34. Larsen JE, Lund O, Nielsen M (2006) Improved method for predicting linear B-cell epitopes. Immunome Res 2:2

35. Emini EA, Hughes JV, Perlow DS, Boger J (1985) Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. J Virol 55:836–839
36. Karplus PA, Schulz GE (1985) Prediction of chain flexibility in proteins. Naturwissenschaften 72:212–213
37. Parker JM, Guo D, Hodges RS (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 25:5425–5432
38. Yang JT (1996) Prediction of protein secondary structure from amino acid sequence. J Protein Chem 15:185–191
39. Shen Y, Maupetit J, Derreumaux P, Tuffery P (2014) Improved PEP-FOLD approach for peptide and miniprotein structure prediction. J Chem Theory Comput 10:4745–4758
40 Tollefsen S, Hotta K, Chen X, Simonsen B, Swaminathan K, Mathews II et al (2012) Structural and functional studies of trans-encoded HLA-DQ2.3 (DQA1*03:01/DQB1*02:01) protein molecule. J Biol Chem 287:13611–13619
41. Choo JA, Thong SY, Yap J, van Esch WJ, Raida M, Meijers R et al (2014) Bioorthogonal cleavage and exchange of major histocompatibility complex ligands by employing azobenzene-containing peptides. Angew Chem Int Ed Engl 53:13390–13394
42. Gao YD, Huang JF (2011) An extension strategy of discovery studio 2.0 for non-bonded interaction energy automatic calculation at the residue level. Dongwuxue Yanjiu 32:262–266
43. Kozakov D, Hall DR, Xia B, Porter KA, Padhorny D, Yueh C et al (2017) The ClusPro web server for protein-protein docking. Nat Protoc 12:255–278
44. Pandey RK, Bhatt TK, Prajapati VK (2018) Novel immunoinformatics approaches to design multi-epitope subunit vaccine for malaria by investigating anopheles salivary protein. Sci Rep 8:1125
45. Bibi S, Ullah I, Zhu B, Adnan M, Liaqat R, Kong W-B et al (2021) In silico analysis of epitope-based vaccine candidate against tuberculosis using reverse vaccinology. Sci Rep 11:1249
46 Fleri W, Paul S, Dhanda SK, Mahajan S, Xu X, Peters B et al (2017) The immune epitope database and analysis resource in epitope discovery and synthetic vaccine design. Front Immunol. https://doi.org/10.3389/fimmu.2017.00278

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.