International Journal of Peptide Research and Therapeutics
https://doi.org/10.1007/s10989-019-09980-7

In Silico Design of Epitope-Based Allergy Vaccine Against *Bellatella germanica* Cockroach Allergens

Fatemeh Ahmadi[1] · Hesam Dorosti[1,2] · Younes Ghasemi[1,2] · Navid Nezafat[1,2]

Accepted: 12 November 2019
© Springer Nature B.V. 2019

Abstract
Asthma is a chronic inflammatory airway disorder that is a subclass of IgE mediated hypersensitivity. Cockroach allergy is one of the main reasons of asthma. German cockroach (*Blattella germanica*) and American cockroach (*Periplaneta Americana*) are two principal species of cockroach. Bla g 2, Bla g 4, Bla g 5 and alpha-amylase 53 kDa are the known allergens of *Blattella germanica* that are used in this study to design the peptide allergy vaccine. In this study, with the help of immunoinformatics approaches, we try to design an efficient epitope vaccine against above-mentioned allergens. The designed vaccine consists of the common epitopes of T and B cells epitopes from the aforesaid allergens that are fused to each other by these two linkers; GGS and linker; moreover, to boost vaccine immunogenicity, the D1 of flagellin was added to N-terminal of vaccine. Afterwards, a 3D model of the final vaccine construct was generated using GalaxyWEB, and validated by ProSA-web, ERRAT and the Ramachandran plot softwares, in order to select the best 3D model. Finally, protein docking was performed between the flagellin part of vaccine and toll-like receptor 5 (TLR5). In sum, our obtained results from computational studies shows that the designed epitope vaccine can be suggested as a prophylactic or therapeutic candidate vaccine *Blattella germanica* cockroach's allergens; however, the efficacy of vaccine should be confirmed by immunological assays.

Keywords In silico design · Epitope vaccine · Allergy · *Bellatella germanica* cockroach · Immunotherapy

Introduction

Allergy is an immunoglobulin E (IgE)-mediated hypersensitivity disease. The prevalence of allergic diseases are increasing worldwide, and more than 25% of the population suffer from a variety of unpleasant manifestations in their different target organs (e.g. Rhinoconjunctivitis, asthma, dermatitis, Gastrointestinal symptoms and anaphylactic shock as a severe systemic reaction) (Valenta et al. 2017). According to the World Health Organization (WHO) statistics, the rate of asthma is estimated 300 million individuals in 2011 based on the last update report in 2013 (Organization 2014). Asthma is a chronic inflammation of airways disorder and one of the most common allergic conditions that is affected by the different factor such as the genetic background of the individual, the site of allergen contact, the dose of allergen, the nature and number of antigens presenting cells, as well as biological mediators (e.g., antibodies).

A cockroach allergy has been known as an important reason for asthma in the last 50 years. Exposing to cockroach in the environmental situation is related to cockroach sensitization and asthma (Pomés et al. 2017). The two important species of cockroach that are predominant in the both temperate and tropical regions are German cockroach (*Blattella germanica*) and American cockroach (*Periplaneta Americana*) (Pomés 2008; Pomés and Arruda 2014).

Sensitization to cockroach allergens has the main role in increasing asthma in low-income urban population throughout the world, and it is more common in children. The exposure and sensitization to cockroach in comparison with other allergens such as dust mites or pets increase the risk of morbidity and severity of asthma among children and young adults (Do et al. 2016).

Today the common form of allergy therapy is pharmacotherapy, which reduce the symptoms of allergy and control the release of mediators from basophil and mast cells;

✉ Navid Nezafat
n.nezafat@srbiau.ac.ir

[1] Pharmaceutical Sciences Research Center, Shiraz University of Medical Sciences, Shiraz, Iran

[2] Department of Pharmaceutical Biotechnology, School of Pharmacy, Shiraz University of Medical Sciences, Shiraz, Iran

Published online: 16 November 2019

moreover, the allergen avoidance is effective just for certain forms of allergy in limited extent (e.g. food and indoor allergen sources) (Murdoch and Lloyd 2010). So, none of the aforesaid medications affect the cause of allergic disease and just suppress the symptoms of disease temporarily. Currently, the only medical treatment for the severe form of IgE mediated asthma is Omalizumab, a monoclonal anti-IgE antibody, which prevents binding IgE to its receptors (Incorvaia et al. 2017; Patronov and Doytchinova 2013).

The recent immunotherapy approach for fighting allergic diseases is an epitope-based vaccine that uses immunodominant T and B cell epitopes of a certain allergen. The advantages of such vaccines are included generating long-lasting immunity, preventing unwanted immune reaction and cost-effectiveness (Negahdaripour et al. 2018; Patronov and Doytchinova 2013; Shahbazi et al. 2016). In this context, B-cell epitopes have crucial role in developing effective allergy vaccine. Allergens as an extracellular peptide stimulate T lymphocyte responses through binding to major histocompatibility complex class II (MHC II) molecules, so identification of MHC-binding molecule indirectly predicts the T-cell epitopes.

In Iran after pollen, cockroaches are the main sensitizers in asthmatic population, and the new *Bellatella germanica* cockroach allergen (alpha-amylase 53 kDa) can cause sensitivity to a relevant percentage of this population (Teifoori et al. 2014). The other important allergens of the German cockroach (*Blattella germanica*) are Bla g 2, Bla g 4, Bla g 1 and Bla g 5 (Arruda et al. 2001; Teifoori et al. 2014).

So, In this study, a peptide-allergy vaccine was designed using bioinformatics tools based on the common epitopes of T and B-cell epitopes from important allergens of *Blattella germanica* cockroach, including Bla g 2, Bla g 4, Bla g 5 and the new one that is common in Iran (alpha-amylase 53 kDa) (Teifoori et al. 2014).

Methods

The methodology that was used in this study for designing allergy vaccine is summarized in a diagram (Fig. 1).

The Allergens and Sequences Retrieval

The allergen sequences of *Blattella germanica* cockroach, including Bla g 2 (inactive aspartic proteinase), Bla g 4 (calycin), Bla g 5 (glutathione-S-transferase) and alpha-amylase were retrieved from https://www.uniprot.org in FASTA format and were imported to CLC main Workbench 7 software (Chen et al. 2017). Moreover, the tertiary structure of human TLR5 was downloaded from the RCSB Protein Data Bank (http:// www.rcsb.org/pdb/home/home.do) (PDB ID: 3J0A).

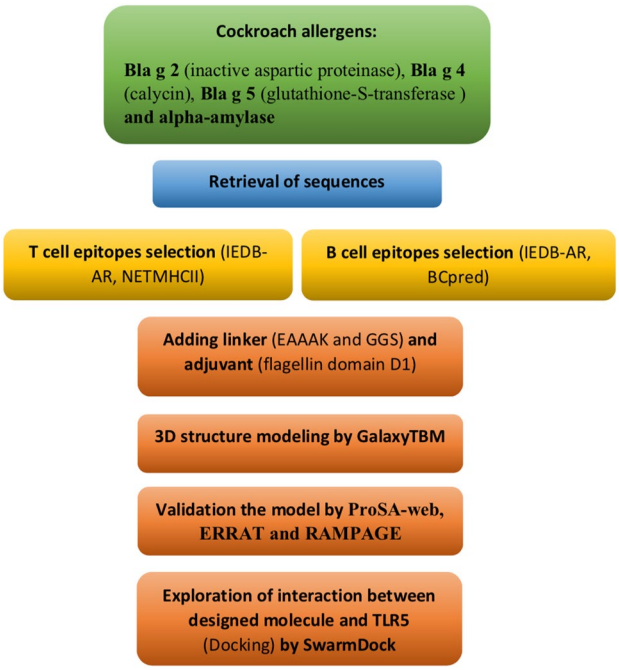


Fig. 1 Flow chart depicting methodology used for designing allergy vaccine against *Bellatella germanica* cockroach

Allergenicity and Antigenicity Evaluation

AlgPred at https://crdd.osdd.net/raghava/algpred/submission.html is one of the servers that has been developed for the allergenicity prediction. The prediction method that was chosen in our research, was "SVM method based on amino acid composition", and the threshold value -0.4 was used. The sensitivity and specificity of the abovementioned method are 88.87% and 81.86% respectively (Saha and Raghava 2006).

Antigenicity was studied by the ANTIGENpro server at https://scratch.proteomics.ics.uci.edu/. The server is a sequence-based, an alignment-free and pathogen-independent predictor applying protein antigenicity microarray data for predicting of protein antigenicity. The results of the predictions are based on multiple representations of the primary sequence and five machine learning algorithms, and a final SVM classifier (Cheng et al. 2005).

MHC Class II Binding Epitope Identification

MHC class II binding peptides were identified via two servers; IEDB-AR at https://tools.iedb.org/mhcii/ employs different methods, including a consensus approach which combines NN-align, SMM-align, and Combinatorial library methods to predict MHC Class II epitopes. The predicted output is based on units of IC_{50} nM for combinatorial library and SMM_align. IC_{50} nM values indicate the affinity of the

peptide to bind to MHC Class II, and the relation between these two is vice versa; the lower number shows higher affinity and as a guideline, high-affinity IC_{50} value < 50 nM, intermediate-affinity IC_{50} value < 500 nM and low-affinity IC_{50} value < 5000 nM (Nielsen et al. 2007).

The NetMHCII-2.2 server at https://www.cbs.dtu.dk/services/NetMHCII-2.2/ was additionally used to predict binding of peptides to HLA-DR, HLA-DQ, HLA-DP, and mouse MHC class II alleles, which applies an Artificial Neuron Networks prediction algorithm. The server predicts regions that bind to 14 HLA-DR alleles covering the nine HLA-DR supertypes, six HLA-DQ, six HLA-DP, and two mouse H2 class II alleles. The prediction of this server is based on IC_{50} nM values as well. In this server the default threshold and the epitope length are − 99.9 and 15, respectively (Nielsen and Lund 2009).

B Cell Epitope Identification

Immune epitope database analysis resource (IEDB-AR) and BCPREDS servers were employed for prediction of linear B-cell epitopes.

IEDB-AR server at https://tools.iedb.org/bcell/ is a linear B-cell prediction software that applies the collection of methods to predict linear B-cell epitopes based on sequence features of the antigen exploiting amino acid scales and hidden Markov model (HMM). The server uses a combination of HMM and a propensity scale method to predict the linear B cell epitopes in the antigen. BepiPred is one of the B-cell Linear Epitope Prediction methods in IEDB-AR server that was used in this study. The default threshold was used is 0.35 and sensitivity and specificity are 0.49 and 0.75, respectively (Jespersen et al. 2017).

The other employed server was BCPREDS at https://ailab.ist.psu.edu/bcpred/predict.html. In this server, three methods are available for users: AAP, fixed-length BCPred and flexible-length method (FBCPred). Here, the fixed-length method was chosen. This server uses subsequence kernel approach in combination with SVM for prediction of linear B cell epitopes. The default specificity and the epitope length are 75% and 20, respectively (Kringelum et al. 2012).

3D Structure Modeling and Validation on Modeled Structures

Three-dimensional structures of protein provide invaluable information for understanding the biological functions of proteins. The FASTA format of the final amino acid sequence of our protein was subjected to I-TASSER server at https://zhanglab.ccmb.med.umich.edu/I-TASSER/, and GalaxyWEB server at https://galaxy.seoklab.org/, to predict the 3D structure of the protein construct.

I-TASSER (Iterative Threading ASSEmbly Refinement) is a hierarchical method for predicting the protein structure and function prediction. In this server, at the first, structural templates are recognized from PDB by using multiple threading alignment methods, and then full-length structure models are made by iterative fragment assembly simulations (Yang and Zhang 2015).

GalaxyWEB is a protein structure prediction and refinement server which predicts the 3D protein structure by GalaxyTBM and refines the model via GalaxyREFINE.

GalaxyTBM server generates the 3D structure of the protein from the primary sequence by template-based modeling (TBM) when experimental structures of homologous proteins are available as templates (Shin et al. 2014).

To validate the overall quality of the generated models, three servers were used: the first one is Protein Structural Analysis tools (PROSA-web) at https://prosa.services.came.sbg.ac.at/prosa.php, The ProSA-web evaluate the interaction energy of each residue with the rest of the protein to characterize if it achieves certain energy criteria. Its Z-score is shown in a plot that includes the Z-score of experimentally defined structures deposited in PDB (Wiederstein and Sippl 2007). The second server is ERRAT at https://servicesn.mbi.ucla.edu/SAVES/ that was used to recognize the incorrect statistics of bad non-bonded interactions within the structure model; in this server the protein structures are verified based on an empirical atom-based approach (Colovos and Yeates 1993). The third server is RAMPAGE at https://mordred.bioc.cam.ac.uk/~rapper/rampage.php. The program classifies the residues in the favored, allowed, and outlier regions in the plot by analyzing the backbone conformation of amino acid residues in structure (Lovell et al. 2003).

Molecular Docking Analysis

Swarm dock at https://bmm.crick.ac.uk/~svc-bmm-swarmdock/submit.cgi builds the 3D structure of the protein–protein complex via SwarmDock algorithm. The server generates low energy conformations based on the uploading PDB files of binding partners and produces a ranked list of clustered docking poses and their corresponding structure. In this server to submit a job, the user can choose the docking types between full-blind docking or local docking with restraints. In this study, the flagellin part of the designed construct was applied as ligand, and the TLR5 used as a receptor to choose interface residues (Torchala and Bates 2014).

The refinement of the selected model was done by FireDock server at https://bioinfo3d.cs.tau.ac.il/FireDock/php.php. The server refines the protein–protein docked complex by restricted interface side-chain rearrangement and by soft rigid-body optimization. The refined models are ranked via the binding score, which includes Atomic Contact Energy,

softened van der Waals interactions, partial electrostatics and additional estimations of the binding-free energy.

Results

T Cell Epitope Prediction

In this study, the high-affinity MHC class II binding epitopes of Bla g 2 (inactive aspartic proteinase), Bla g 4 (calycin), Bla g 5 (glutathione-S-transferase) and alpha-amylase allergens were mapped based on the same alleles by IEDB-AR and NetMHCII-2.2 servers.

In the IEDB-AR method, The MHC binding affinity of selected epitopes was chosen based on Stabilized Matrix Method (SMM) and IC_{50} value ($IC_{50} < 50$ nM) as an output. The results are shown in Table 1. Moreover, the results of NetMHCII-2.2 server which are shown in Table 2, are based on the alleles that found in the output of IEDB-AR server.

B Cell Epitope Prediction

Surface accessibility and fragment flexibility are significant features for predicting antigenic epitopes. In addition, the existence of high hydrophobic zones also provides strong evidence for B-cell epitope identification.

Table 1 The immunodominant epitope selected from Bla g 4, Alpha-amylase, Bla g 2, Bla g 5 allergens via IEDB-AR server, according to IC_{50} value

Allergen name	Start	End	Best smm_align_ic50	Epitopes	Allele
Bla g 4	113	131	39	HVIYVQIRFSVRRFHPKLG	HLA-DRB1*11:01
Bla g 4	16	38	18	RFRGSWIIAAGTSEALTQYKC	HLA-DQA1*05:01/DQB1*03:01
Bla g 4	97	116	25	YENYAIVEGCPAAANGHVIY	HLA-DQA1*05:01/DQB1*03:01
Bla g 4	80	95	42	NDKGKAFSAPYSVLAT	HLA-DRB1*01:01
Alpha-amylase	306	324	26	YKTSKLYKMAVAFMLAYPY	HLA-DRB1*09:01
Alpha-amylase	180	200	12	DCGVAGFRVDAAKHMWPADLQ	HLA-DRB3*01:01
Alpha-amylase	306	332	14	YKTSKLYKMAVAFMLAYPYGYPRVMSS	HLA-DRB5*01:01
Alpha-amylase	195	213	32	WPADLQYIYSKVNNLNTDH	HLA-DRB1*07:01
Alpha-amylase	306	322	11	YKTSKLYKMAVAFMLAY	HLA-DRB1*07:01
Alpha-amylase	291	309	8	DNQRGHGAGGASILTYKTS	HLA-DQA1*05:01/DQB1*03:01
Alpha-amylase	374	393	14	IFNMVGFRNAVAGTAVSNWW	HLA-DQA1*05:01/DQB1*03:01
Alpha-amylase	454	473	18	GSDGKAYIEILSSADDGVLA	HLA-DRB1*04:05
Alpha-amylase	217	234	45	SGARPFFYQEVIDLGGEA	HLA-DPA1*03:01/DPB1*04:02
Alpha-amylase	217	233	20	SGARPFFYQEVIDLGGE	HLA-DPA1*02:01/DPB1*01:01
Bla g 2	293	311	32	GHSDHFFIGDFFVDHYYSE	HLA-DPA1*01:03/DPB1*02:01
Bla g 2	179	197	41	EFTYVPLVGDDSWKFRLDG	HLA-DRB3*01:01
Bla g 2	258	276	48	LPDVTFVINGRNFNISSQY	HLA-DRB1*13:02
Bla g 2	51	66	49	CPNLQKYEKLKPKYIS	HLA-DRB1*07:01
Bla g 2	209	226	32	GTQAIIDTSKAIIVGPKA	HLA-DRB1*07:01
Bla g 2	233	251	44	EAIGCVVEKTTTRRICKLD	HLA-DRB1*07:01
Bla g 2	264	282	40	VINGRNFNISSQYYIQQNG	HLA-DRB1*07:01
Bla g 2	112	132	28	CILSADVVVGIAAPGCPNALK	HLA-DQA1*05:01/DQB1*03:01
Bla g 2	51	68	20	CPNLQKYEKLKPKYISDG	HLA-DRB1*01:01
Bla g 2	105	123	39	DELSQEVCILSADVVVGIA	HLA-DRB1*01:01
Bla g 2	51	69	44	CPNLQKYEKLKPKYISDGN	HLA-DRB5*01:01
Bla g 2	51	68	48	CPNLQKYEKLKPKYISDG	HLA-DRB1*11:01
Bla g 5	95	110	16	ISDFRAAIANYHYDAD	HLA-DRB5*01:01
Bla g 5	145	161	36	AAGKLTWADFYFVAILD	HLA-DPA1*01:03/DPB1*02:01
Bla g 5	150	168	21	TWADFYFVAILDYLNHMAK	HLA-DRB1*04:05
Bla g 5	179	199	14	KALREKVLGLPAIKAWVAKRP	HLA-DRB1*01:01
Bla g 5	153	168	45	DFYFVAILDYLNHMAK	HLA-DRB1*01:01
Bla g 5	149	166	39	LTWADFYFVAILDYLNHM	HLA-DPA1*02:01/DPB1*01:01
Bla g 5	133	153	20	FDEVVKANGGYLAAGKLTWAD	HLA-DQA1*05:01/DQB1*03:01

Table 2 The immunodominant epitope selected from Bla g 4, Alpha-amylase, Bla g 2, Bla g 5 allergens via NetMHCII-2.2 server, according to IC_{50} value

Allergen name	Start	End	Epitopes
Bla g 4	13	38	NLDYERFRGSWIIAAGTSEALTQYKC
Bla g 4	99	133	NYAIVEGCPAAANGHVIYVQIRFSVRRFHPKLGDK
Alpha-amylase	180	209	DCGVAGFRVDAAKHMWPADLQYIYSKVNNL
Alpha-amylase	272	326	GVGWGYIPDGDALVFVDNHDNQRGHGAGGASILTYKT-SKLYKMAVAFMLAYPYGY
Alpha-amylase	375	391	FNMVGFRNAVAGTAVSN
Bla g 2	2	19	PLYKLVHVFINTQYAGIT
Bla g 2	58	82	EKLKPKYISDGNVQVKFFDTGSAVG
Bla g 2	143	158	ENLIAPVFSIHHARFQ
Bla g 2	181	196	TYVPLVGDDSWKFRLD
Bla g 2	208	225	AGTQAIIDTSKAIIVGPK
Bla g 2	259	290	PDVTFVINGRNFNISSQYYIQQNGNLCYSGFQ
Bla g 2	294	311	HSDHFFIGDFFVDHYYSE
Bla g 5	92	112	VDTISDFRAAIANYHYDADEN
Bla g 5	132	168	KFDEVVKANGGYLAAGKLTWADFYFVAILDYLNHMAK
Bla g 5	179	199	KALREKVLGLPAIKAWVAKRP

The B-cell epitopes that were determined via BCPREDS and IEDB-AR BepiPred from the desired allergens are presented in Tables 3 and 4, respectively.

The overlapping high immunogenic regions of identified B-cell and MHC II binding epitopes of each allergen (Bla g 2, Bla g 4, Bla g 5 and alpha-amylase 53 kDa) are joined together by GGS linkers.

The schematic diagram of ultimate construct includes of vaccine domains and linkers are shown in Fig. 2.

Homology Modeling and Quality Assessment of Protein 3D Structure

The initial 3D structure of selected protein was generated by GalaxyWeb and I-TASSER; each of them provided 5 models. RAMPAGE, ERRAT and Prosa-Web were used for validation of the obtained models. Based on the validation results, model 5 of GalaxyWeb was chosen as the best 3D model (Table 5).

In ProsA-Web the z-score of the best model was −4.18 (Fig. 3). According to the Ramachandran plot analysis, the amino acid residues of the input model are located into three regions: (1) a number of residues in the favored region (expected value ~ 98.0%), (2) a number of residues in the allowed region (~ 2.0% expected), and 3) a number of residues in the outlier region. In our final 3D model, 92.2%, 5.8%, and 2.2% of amino acid residues were placed in the favored regions, allowed regions, and outlier region, respectively (Fig. 4). The overall quality score of ERRAT in this model was 77.178 (Fig. 5), the standard high-resolution structures of ERRAT generally produce the overall quality factor around 95% or higher.

Docking Analysis

Based on the knowledge on binding sites between TLR5 and the flagellin part of vaccine, pose of interaction, the lowest energy, and the highest score of models, the best docked model was selected (Fig. 6). The global energy of docked complex after refinement via FireDock server was − 117.20.

Discussion

In recent decades, allergic diseases such as asthma and rhinitis have increased especially in children; this makes it as one of the major concerns of the medical community. Various strategies like allergen avoidance and pharmacotherapy are used to control the releasing of mediators from basophil and mast cells in order to reduce the symptoms of allergy. However, the aforesaid strategies have some flaws such as being temporary or their limited extent. In this regard, various types of allergy vaccines can be engineered based on the DNA and protein structure of allergen in both form proteins/peptides or DNA. Generally, six main approaches, including genetic vaccination based on the administration of allergen-encoding nucleic acids, recombinant wild-type allergen, carrier-bound peptide, T cell epitope-based vaccines, hypoallergens, and gene therapy are used for allergy vaccination (Valenta et al. 2009).

In the recent approaches, peptide vaccines, including B- and T-cell epitopes have been proposed instead of applying the crude allergen extracts, to decrease the high risk of anaphylactic shock (Chen et al. 2014). The concept of epitope peptide vaccines is based on the identification of B-cell and

Table 3 The linear B cell epitopes selected from the Bla g 4, Alpha-amylase, Bla g 2, Bla g 5 allergens via BCPred server

Allergen name	Score	Start	End	Epitopes
Bla g 4	0.983	54	73	TDSQGKNRTTIRGRTKFEGN
Bla g 4	0.949	93	112	LATDYENYAIVEGC-PAAANG
Bla g 4	0.865	135	154	MIQHYTLDQVNQHK-KAIEED
Alpha-amylase	1	331	350	SSFSFDNSDQGPPQDGNGNI
Alpha-amylase	1	123	142	QYPAVPYGPGDFHSTCTVSN
Alpha-amylase	0.998	476	495	VNSKVGSKSQTTTTQSSHCT
Alpha-amylase	0.991	30	49	RFLGPKGFAGVQVS-PVHENV
Alpha-amylase	0.988	283	302	ALVFVDNHDNQRGHGAG-GAS
Alpha-amylase	0.98	101	120	QMSGSWPDAH-GQGGSTADTY
Alpha-amylase	0.938	205	224	KVNNLNTDH-GFPSGARPFFY
Alpha-amylase	0.921	387	406	TAVSNWWDNGDKQISF-CRGN
Alpha-amylase	0.917	262	281	NNAIKWLVNFGVGWGY-IPDG
Alpha-amylase	0.817 0.786	352	371	SPSINADGTCGNG-WVCEHRW
Alpha-amylase		71	90	SRSGDENAFRDMVRRCN-NVG
Alpha-amylase	0.785	429	448	PAGDYCDVISGSYENGSCTG
Bla g 5	0.981	35	54	QEGDWPNLKPSMPFGKT-PVL
Bla g 5	0.978	116	135	KKWDPLKKETIPYYT-KKFDE
Bla g 2	0.998	116	135	ADVVVGIAAPGCPNALK-GKT
Bla g 2	0.991	199	218	KIGDTTVAPAGTQAIIDTSK
Bla g 2	0.976	36	55	NVVVASQECVG-GACVCPNLQ
Bla g 2	0.946	270	289	FNISSQYYIQQNGNLCYSGF
Bla g 2	0.912	309	138	YSEFNWENKTMGFGRS-VESV
Bla g 2	0.892	7	26	VHVFINTQYAGITKIGNQNF
Bla g 2	0.826	220	239	IIVGPKAYVNPINEAIGCVV
Bla g 2	0.764	168	187	FGGSDWKYVDGEFTYV-PLVG

T-cell epitopes to induce specific immune responses (Meloen et al. 2001); in this context, the immunoinformatics tools are useful for selecting epitopes from protective antigens. The immunoinformatics strategies emphasize mostly on the design and developing algorithms for mapping potential B- and T-cell epitopes, which accelerate the vaccine projects, saving the time and reduce the cost of laboratory analysis. (Tomar and De 2010). So, in this research, it has been tried to suggest a method to design an efficient epitope-based vaccine based on B and T cell epitopes against *Blattella germanica* cockroach's allergy. The selected allergens are an extracellular peptide presented on the antigen presenting cell (APC) through MHC II molecules. The production of IgE antibody is the hallmark of allergic diseases; therefore, to eliminate IgE activation, T cell immune responses can be activated by T cell epitopes whenever IgE recognizes B cell epitopes.

Although the epitope-based vaccines have different advantages such as high specificity, safety and cost-effectiveness, but the low immunogenicity is the most important flaw of such vaccines. To conquer this drawback adding proper linkers and strong adjuvant to vaccine are suggested as the important strategies. In our study, the final construct of vaccine was generated via fusing the common epitopes of T- and B-cell epitopes of each allergen to each other via short flexible linker amino acids linker, GGS; moreover, the part of flagellin as a natural adjuvant were linked to the N-terminal of vaccine via EAAAK linker. The roles of linkers in epitope-based vaccine structure are to prevent the generation of neoepitopes as an important concern in designing these kinds of vaccine and facilitating the presentation of MHC-II binding epitopes and the process of immune system (Livingston et al. 2002). In line with other reports, coupling adjuvant such as TLR ligands with antigens has the critical role in enhancing immune responses (Schülke et al. 2011) through boosting antigen uptake and DC activation (Cuadros et al. 2004; Khan et al. 2007). The flagellin, TLR5 ligand, is broadly used with different antigens to strengthen specific immune responses by activating TLR5-expressing cells. In this regard, recently, flagellin as an adjuvant has obtained much attention for preventing type I allergy (Negahdaripour et al. 2017). In this study, domain 1 (D1) of flagellin that plays an important role in binding and signaling of TLR5 was utilized as the adjuvant (Schülke et al. 2011). After designing the epitope vaccine construct, the 3D structure was generated by homology modeling via the GalexyWeb, and validated through ProSA-web, ERRAT and Ramachandran plot to select the quality of model. Afterward, the interaction between receptor (TLR5 molecule) and ligand (flagellin D1 domain) was performed by Swarmdock server (Fig. 5). Until now, few researches have been done on cockroach allergy, and there is no vaccine for *Blattella germanica* cockroach allergens. Recently, epitope-based allergy vaccine has gained considerable attention; in some studies (Chen et al. 2014; Pascal et al. 2013) only T and B cells investigated separately and mostly for one allergen; however, in this study we try to design the multi-epitope based vaccine that was constructed through the combination of the overlapping T-cell and B-cell epitopes from four allergens of German cockroach (*Blattella germanica*) by means of the different immunoinformatics tools. In sum, computational studies suggest that our designed vaccine could be proper

Table 4 The linear B cell epitopes selected from the Bla g 4, Alpha-amylase, Bla g 2, Bla g 5 allergens via *IEDB-AR server*

Allergen name	Start	End	Epitopes
Bla g 4	30	33	SEAL
Bla g 4	53	65	YTDSQGKNRTTIR
Bla g 4	68	72	TKFEG
Bla g 4	77	87	IDYNDKGKAFS
Bla g 4	106	110	CPAAA
Bla g 4	129	131	KLG
Bla g 4	144	146	VNQ
Bla g 4	148	151	KKAI
Alpha-amylase	1	8	QKDPHVWD
Alpha-amylase	36	40	GFAGV
Alpha-amylase	42	45	VSPV
Alpha-amylase	72	78	RSGDENA
Alpha-amylase	103	135	SGSWPDAHGQGGSTADTYNLQYPAVPYGPGDFH
Alpha-amylase	140	149	VSNYQDPSNV
Alpha-amylase	161	167	NQGSDYV
Alpha-amylase	208	221	NLNTDHGFPSGARP
Alpha-amylase	233	238	EAIHST
Alpha-amylase	240	243	YTGF
Alpha-amylase	256	261	GDAFRG
Alpha-amylase	276	281	GYIPDG
Alpha-amylase	290	301	HDNQRGHGAGGA
Alpha-amylase	325	328	GYPR
Alpha-amylase	335	352	FDNSDQGPPQDGNGNIIS
Alpha-amylase	354	362	SINADGTCG
Alpha-amylase	385	388	AGTA
Alpha-amylase	390	398	SNWWDNGDK
Alpha-amylase	415	419	EFNND
Alpha-amylase	439	458	GSYENGSCTGKTVTVGSDGK
Alpha-amylase	479	493	KVGSKSQTTTTQSSH
Bla g 5	26	51	EKDFEDYRFQEGDWPNLKPSMPFGKT
Bla g 5	53	63	VLEIDGKQTHQ
Bla g 5	76	85	GLSGKDDWEN
Bla g 5	106	128	HYDADENSKQKKWDPLKKETIPY
Bla g 2	63	67	KYISD
Bla g 2	79	86	SAVGRGIE
Bla g 2	124	132	APGCPNALK
Bla g 2	157	162	FQDGEH
Bla g 2	173	176	WKYV
Bla g 2	185	188	LVGD
Bla g 2	200	214	IGDTTVAPAGTQAII
Bla g 2	226	230	AYVNP
Bla g 2	255	258	IPSL
Bla g 2	290	294	QPCGH

candidate against *Blattella germanica* cockroach's allergens; however, the experimental immunological assays should be performed to confirm its efficacy.

Conclusion

In this study, an epitope-based potent vaccine against *Bellatella germanica* cockroach allergens was designed by the employment of immunoinformatics tools for finding T-cell and B-cell epitopes to save the time and cost before

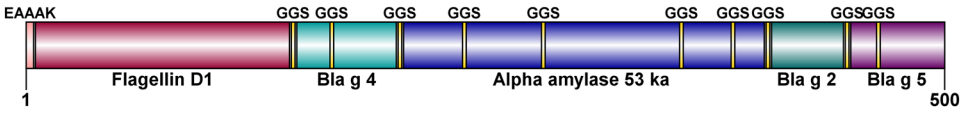


Fig. 2 Schematic diagram of vaccine construct consist of amino acids residues of TLR2 agonist (flagellin D1) and high immunogenic regions of main *Bellatella germanica* cockroach's allergens (Bla g 4, Alpha amylase 53 ka, Bla g 2, Bla g 5) which are fused together by appropriate linker (EAAAk, GGS)

Table 5 The validation result of five models obtained from Galexy-Web, via Ramachandran, ERRAT and PROSA-web servers

Servers	Model 1	Model 2	Model 3	Model 4	Model 5
Ramachandran					
Favored region	92.0	91.6	92.0	92.0	92.2
Allowed region	6.0	6.4	5.8	6.2	5.8
Outliner region	2.0	2.0	2.2	1.8	2.0
ERRAT[a]	71	72.78	67	73.3	77.1
PROSA-web[b]	3.99	3.75	3.88	3.74	4.18

[a]The overall quality factor

[b]Z-score

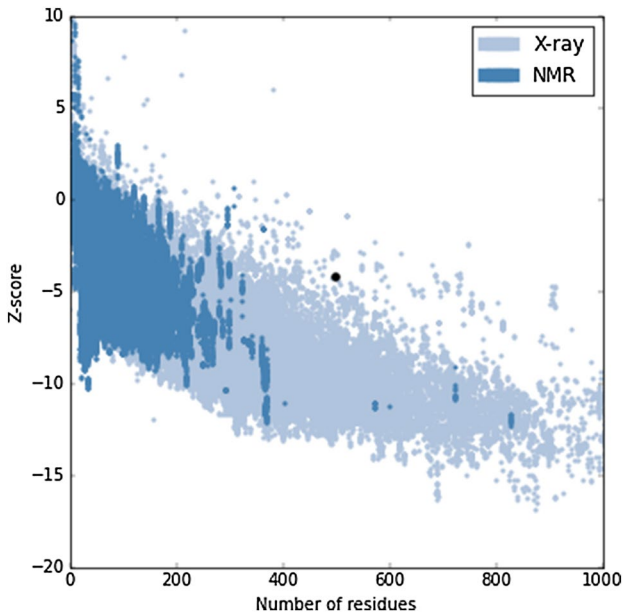


Fig. 3 ProSA-web z-score plot for 3D structure of protein vaccine. The z-score of the model is −4.18, that slightly in range of native protein conformation. The z-score of protein vaccine is indicated in black spot. z-Score plot includes the z-scores of all experimentally protein chains in PDB that determined by NMR spectroscopy in dark blue and X-ray crystallography in light blue (Color figure online)

experimental studies. An allergenic peptide was designed as an epitope-based vaccine candidate.

The previous studies indicated the prediction of B-cell or T-cell epitopes of one allergen alone, but in the present study, the goal is designing the multiple-peptide vaccine based on the overlapping epitopes of potential B- and T-cell epitopes of each allergen (Bla g 2, Bla g 4, Bla g 5 and alpha-amylase 53 kDa) which are joined together by linkers to increase efficiency of this kind of vaccine. Designing an epitope-based vaccine for each of the allergens alone or based on the B-cell or T-cell epitopes, separately also is feasible by following the methods that mentioned above.

The selected epitope peptide from aforesaid allergens in combination or alone could be applied to prioritize in experimental validation models. Overall, resulting peptides analysis from this study could be suggested them as vaccine candidates or allergy immunotherapy against cockroach-mediated allergy in experimental prioritization.

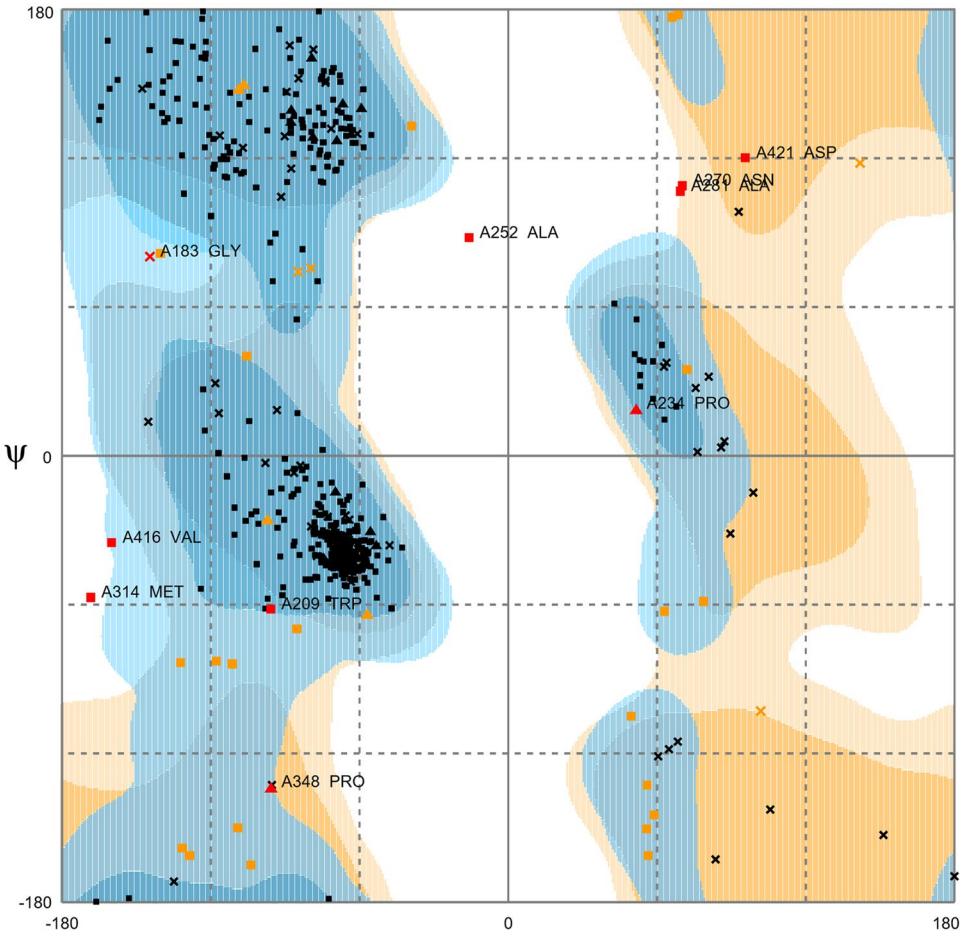


Fig. 4 Validation of protein 3D model, using Ramachandran plot. In the model, 459 (92.2%), 29 (5.8%) and 10 (2.0%) of residues are located in favored, allowed and outlier regions, respectively

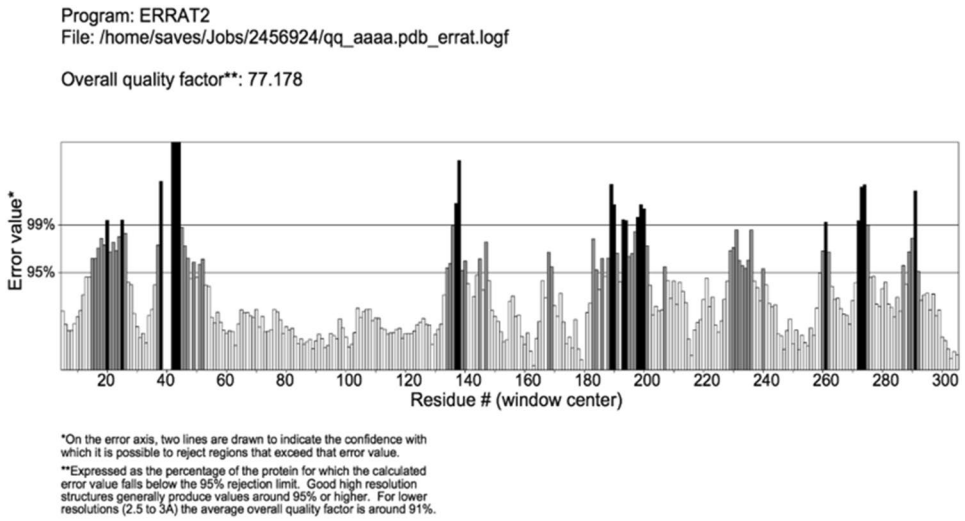


Fig. 5 The overall quality factor plot of the vaccine model is 77.17. In ERRAT plot, regions of the 3D model that can be rejected at the 95% confidence level are shown in gray lines and regions that can be rejected at the 99% level are depicted in black lines. The overall quality factor manifests the percentage of protein for which the computed error value falls under the 95% rejection limit. Good high-resolution structures usually generate values about 95% or higher

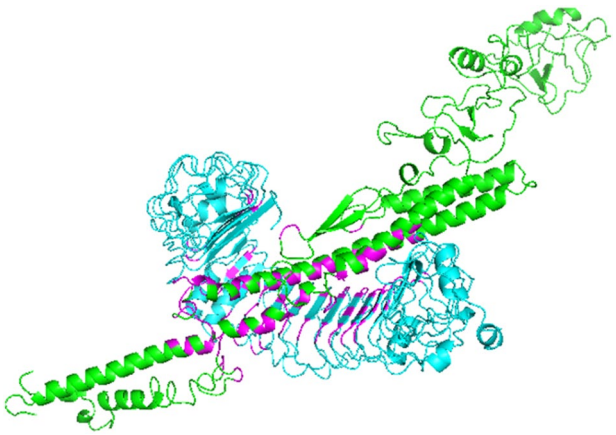


Fig. 6 Docking model of TLR5 and the flagellin part of vaccine. TLR5 molecule, vaccine, and the interaction sites are shown in cyan, green, and magenta respectively (Color figure online)

Acknowledgements The authors wish to thank Shiraz University of Medical Sciences (16878) for supporting the conduct of this research.

Compliance with Ethical Standards

Conflict of interest All the authors declare that they have none conflict of interest.

Research Involving Human and Animal Participants This article does not contain any studies with human participants or animals performed by any of the authors.

References

Arruda LK, Vailes LD, Ferriani VP, Santos ABR, Pomés A, Chapman MD (2001) Cockroach allergens and asthma. J Allergy Clin Immunol 107:419–428

Chen H, Yang HW, Wei JF, Tao AL (2014) In silico prediction of the T-cell and IgE-binding epitopes of Per a 6 and Bla g 6 allergens in cockroaches. Mol Med Rep 10:2130–2136

Chen C, Huang H, Wu CH (2017) Protein bioinformatics databases and resources. In: Frishman D, Valencia A (eds) Protein Bioinformatics. Springer, New York, pp 3–39

Cheng J, Randall AZ, Sweredoski MJ, Baldi P (2005) SCRATCH: a protein structure and structural feature prediction server. Nucleic Acids Res 33:W72–W76

Colovos C, Yeates TO (1993) Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci 2:1511–1519

Cuadros C, Lopez-Hernandez FJ, Dominguez AL, McClelland M, Lustgarten J (2004) Flagellin fusion proteins as adjuvants or vaccines induce specific immune responses. Infect Immun 72:2810–2816

Do DC, Zhao Y, Gao P (2016) Cockroach allergen exposure and risk of asthma. Allergy 71:463–474

Incorvaia C, Riario-Sforza GG, Ridolo E (2017) IgE depletion in severe asthma: what we have and what could be added in the near future. EBioMedicine 17:16–17

Jespersen MC, Peters B, Nielsen M, Marcatili P (2017) BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. Nucleic Acids Res 45:W24–W29

Khan S et al (2007) Distinct uptake mechanisms but similar intracellular processing of two different toll-like receptor ligand-peptide conjugates in dendritic cells. J Biol Chem 282:21145–21159

Kringelum JV, Lundegaard C, Lund O, Nielsen M (2012) Reliable B cell epitope predictions: impacts of method development and improved benchmarking. PLoS Comput Biol 8:e1002829

Livingston B, Crimi C, Newman M, Higashimoto Y, Appella E, Sidney J, Sette A (2002) A rational strategy to design multiepitope immunogens based on multiple Th lymphocyte epitopes. J Immunol 168:5499–5506

Lovell SC et al (2003) Structure validation by Cα geometry: ϕ, ψ and Cβ deviation. Proteins 50:437–450

Meloen RH, Langeveld JP, Schaaper WM, Slootstra JW (2001) Synthetic peptide vaccines: unexpected fulfillment of discarded hope? Biologicals 29:233–236

Murdoch JR, Lloyd CM (2010) Chronic inflammation and asthma. Mutat Res 690:24–39. https://doi.org/10.1016/j.mrfmmm.2009.09.005

Negahdaripour M et al (2017) A novel HPV prophylactic peptide vaccine, designed by immunoinformatics and structural vaccinology approaches. Infect Genet Evol 54:402–416

Negahdaripour M et al (2018) Structural vaccinology considerations for in silico designing of a multi-epitope vaccine. Infect Genet Evol 58:96–109

Nielsen M, Lund O (2009) NN-align. An artificial neural network-based alignment algorithm for MHC class II peptide binding prediction. BMC Bioinform 10:296

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

Organization WA (2014) WAO white book on allergy update 2013. World Allergy Organization, Milwaukee

Pascal M, Konstantinou G, Masilamani M, Lieberman J, Sampson H (2013) In silico prediction of Ara h 2 T cell epitopes in peanut-allergic children. Clin Exp Allergy 43:116–127

Patronov A, Doytchinova I (2013) T-cell epitope vaccine design by immunoinformatics. Open Biol 3:120139

Pomés A (2008) Cockroach and other inhalant insect allergens. Clin Allergy Immunol 21:183

Pomés A, Arruda LK (2014) Investigating cockroach allergens: aiming to improve diagnosis and treatment of cockroach allergic patients. Methods 66:75–85

Pomés A, Mueller GA, Randall TA, Chapman MD, Arruda LK (2017) New insights into cockroach allergens. Curr Allergy Asthma Rep 17:25

Saha S, Raghava G (2006) AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. Nucleic Acids Res 34:W202–W209

Schülke S et al (2011) A fusion protein of flagellin and ovalbumin suppresses the TH2 response and prevents murine intestinal allergy. J Allergy Clin Immunol 128(1340–1348):e1312

Shahbazi M, Haghkhah M, Rahbar MR, Nezafat N, Ghasemi Y (2016) In silico sub-unit hexavalent peptide vaccine against an Staphylococcus aureus biofilm-related infection. Int J Pept Res Ther 22:101–117

Shin W-H, Lee GR, Heo L, Lee H, Seok C (2014) Prediction of protein structure and interaction by GALAXY protein modeling programs Bio Des 2:1–11

Teifoori F et al (2014) Identification of the main allergen sensitizers in an Iran asthmatic population by molecular diagnosis. Allergy Asthma Clin Immunol 10:41

Tomar N, De RK (2010) Immunoinformatics: an integrated scenario. Immunology 131:153–168

Torchala M, Bates PA (2014) Predicting the structure of protein–protein complexes using the SwarmDock web server. In: Kihar D (ed) Protein structure prediction. Springer, New York, pp 181–197

Valenta R, Ferreira F, Focke-Tejkl M, Linhart B, Niederberger V, Swoboda I, Vrtala S (2009) From allergen genes to allergy vaccines. Ann Rev Immunol 28:211–241

Valenta R, Campana R, Niederberger V (2017) Recombinant allergy vaccines based on allergen-derived B cell epitopes. Immunol Lett 189:19–26. https://doi.org/10.1016/j.imlet.2017.04.015

Wiederstein M, Sippl MJ (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res 35:W407–W410

Yang J, Zhang Y (2015) I-TASSER server: new development for protein structure and function predictions. Nucleic Acids Res 43:W174–W181. https://doi.org/10.1093/nar/gkv342

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