Original Article

Prediction of an Epitope-based Computational Vaccine Strategy for Gaining Concurrent Immunization Against the Venom Proteins of Australian Box Jellyfish

Md. Jibran Alam, Kutub Uddin Muhammad Ashraf

Department of Genetic Engineering and Biotechnology, University of Chittagong, Chittagong, Bangladesh

ABSTRACT

Background: Australian Box Jellyfish (C. fleckeri) has the most rapid acting venom known to in the arena of toxicological research and is capable enough of killing a person in less than 5 minutes inflicting painful, debilitating and potentially life-threatening stings in humans. It has been understood that C. fleckeri venom proteins CfTX-1, 2 and HSP70-1 contain cardiotoxic, neurotoxic and highly dermatonecrotic components that can cause itchy bumpy rash and cardiac arrest. **Subjects and Methods:** As there is no effective drug available, novel approaches regarding epitope prediction for vaccine development were performed in this study. Peptide fragments as nonamers of these antigenic venom proteins were analyzed by using computational tools that would elicit humoral and cell mediated immunity, were focused for attempting vaccine design. By ranking the peptides according to their proteasomal cleavage sites, TAP scores and IC50 < 250 nM, the predictions were scrutinized. Furthermore, the epitope sequences were examined by in silico docking simulation with different specific HLA receptors. **Results:** Interestingly, to our knowledge, this is the maiden hypothetical immunization that predicts the promiscuous epitopes with potential contributions to the tailored design of improved safe and effective vaccines against antigenic venom proteins of C. fleckeri which would be effective especially for the Australian population. **Conclusion:** Although the computational approaches executed here are based on concrete confidence which demands more validation and in vivo experiments to validate such in silico approach.

Key words: *C. fleckeri*, docking simulation, epitope prediction, vaccine design, venom proteins

INTRODUCTION

Chironex fleckeri (Australian box jellyfish) commonly known as sea wasp or marine stingers that inhabit and roam coastal water from northern Australia, New Guinea north to the Philippines and Vietnam.[1] It is considered the most venomous marine animal and most dangerous cubozoan jellyfish to humans and its occurrence in the tropical coastal waters of Australia is primarily a problem, particularly in summer. At least 70 deaths have been reported due to *C. fleckeri* envenoming occurred in Australia. Apart from these, numerous deaths from related species also have been reported in the South India, Malaysia, Japan, Philippines, Maldives islands, Papua New Guinea, Java, and Gulf of Thailand, but most encounters appear to result only in mild envenomation. Australian box jellyfish produces exceptionally potent and rapid-acting venom and its stings to humans cause severe localized and systemic effects that are potentially life-threatening to humans. The venom of *C. fleckeri* contains a variety of bioactive and complex mixture of venom proteins which are stored and discharged

Access this article online	
Quick Response Code:	Website: www.toxicologyinternational.com
	DOI: 10.4103/0971-6580.121677

Address for correspondence: Mr. Kutub Uddin Muhammad Ashraf, Department of Genetic Engineering and Biotechnology, Faculty of Biological Science, University of Chittagong, Chittagong - 4331, Bangladesh. E-mail: kutubuddinashraf@yahoo.com

by small, highly pressurized stinging capsules called nematocysts. Recent studies reveal that the venom proteins have cytolytic, cytotoxic, cardiotoxic (attacks the heart), and highly dermatonecrotic (destroys skin) components which are rapidly absorbed into the circulation after injection.[2] Surprisingly, the onset of symptoms has been reported extremely rapid in several case studies.[3] For its notorious sting, a massive dose of venom can cause cardiac dysfunction, cardiac arrest (arrhythmias), resulting in loss of consciousness and death within 5 min of being stung in severe cases. Children are closely vulnerable to this life-threatening venom proteins because of their smaller body mass.[4] Moreover, the box jellyfish venom has multiple effects attacking the nervous system, skin, and heart simultaneously. Symptoms of major *C. fleckeri* stings include rapid acute cutaneous inflammation, dermonecrosis, excruciating pain, permanent scarring, hypotension, shock, dyspnoea, hypertension, impaired consciousness, pulmonary edema, and cardiac dysfunction.[5] To date only two *C. fleckeri* venom proteins, CfTX-1 and -2 have been significantly identified which have potent hemolytic activity with cutaneous inflammation and may function as a pore-forming toxin. Hence, it can disrupt normal transmembrane ion concentration gradients in susceptible cells.[6] Moreover, heat shock proteins of *C. fleckeri* (HSP70-1) may play a crucial role in antigen and cross presentation.[7] In addition, heat shock protein-derived immunodominant epitopes are exploitable as therapeutic peptides in allergies have been recently reported.[8] Domestic vinegar and antivenom have been widely used as first aid treatment to neutralize against this rapid acting venom, but some people still die despite its administration. Although life-saving antivenoms have an immunoglobulin pool of unknown antigen specificity and known redundancy that necessitates the delivery of large volumes of heterologous immunoglobulin to the envenomed victim. Consequently, it increases the risk of serum sickness and anaphylactoid which has a strong adverse effect.[9] With prospects, recent developments in computational tools have paved the way to predict and to prosecute further assay of B cell and T cell epitopes from antigenic proteins in specialized tasks. This has led to peptide-based vaccines design planning that is more specific, secured, optimized, and easy to predict the peptide binding to human leuckocyte antigen (HLA) alleles using structural and modelling methodologies. Surprisingly, it has gained momentum in recent years in alleviating to some crucial immunological infections.

SUBJECTS AND METHODS

Protein sequence retrieval

The toxin protein sequences of *C. fleckeri* were retrieved from protein database of National Center for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov/protein/) by GenBank accession no. ABS30940.1 for Toxin-1, ABS30941.1 for Toxin-2 and ACS12895.1 for HSP70-1. The sequences were analyzed with a view to recognizing the immunologically relevant regions, which was done by studying antigenecity, solvent accessible regions, and Major Histocompatibility Complex (MHC) class I and II binding sites.

Antigenic peptide prediction

In this method, the potential hydrophilic regions of the proteins were found out in order to identify the antigenic determinants. Antibody epitope prediction of Immune Epitope Database (IEDB) analysis resource server (http://tools.immuneepitope.org/tools/bcell/tutorial.jsp) was used which predicted the sites that produce antigenic response against an antigenic protein. Antigenic epitopes are determined using several prediction methods, for example, Kolaskar and Tongaonkar antigenicity,[10] and Hopp and Woods hydrophobicity[11] methods.

Secondary structure prediction

We used ExPASy's secondary structure prediction server (http://web.expasy.org/protparam/)[12] to get an idea about the secondary structure of the venom proteins. Several parameters given by ProtParam tool were studied, for example, molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index, and grand average of hydropathicity.[13-15]

Solvent accessible regions and hydrophilicity estimation

Hydrophilicity of the proteins was estimated by using ProtScale (http://web.expasy.org/protscale/) server of ExPASy. By using Wolfenden *et al*., and Eisenberg *et al*.,[16-18] hydrophobicity scales, we found different hydrophobicity plots, which were then analyzed to predict the solvent accessible regions and to estimate the hydrophobic sites on the protein.

Beta turn prediction

Beta turn in the selected protein structures for epitope prediction was determined by using Levitt scale[19] by ExPASy's ProtScale server.

Prediction of major histocompatibility complex binding peptide

To predict the MHC binding peptides for the venom proteins of *C. fleckeri,* we used two options provided by IEDB analysis resource. For MHC class I peptide prediction, we used Proteasomal cleavage/transporter of antigenic peptides transporter/MHC class I combined prediction server (http://tools.immuneepitope.org/processing/) and for MHC class II peptide, we used MHC II-binding prediction (http://tools.immuneepitope.org/mhcii).

In both the prediction servers, we used the artificial neural network prediction method to predict the potential nonamers that may efficiently bind to the binding groove of the MHC molecules.

Allergenicity assessment

In order to assay the degrees of allergenicity we operated AllerHunter (http://tiger.dbs.nus.edu.sg/AllerHunter/index.html). A combinational prediction by using both Support Vector Machine (SVM) and pair-wise sequence similarity makes AllerHunter a very useful program for cross-reactive allergen prediction. Cross-reactivity is a phenomenon which is based on similarity among proteins and allergens, whereas allergenecity means the ability of an allergen to induce immunoglobulin E antibody production. AllerHunter predicts allergens as well as nonallergens with high specificity. Moreover, it does not compromise its efficiency while classifying proteins with similar sequence to known allergens.

Docking simulation

We performed *in silico* docking simulation to find out whether or not the predicted peptides will bind to the MHC molecules when these will be applied for further *in vivo* experiments. For docking simulation study, we used Autodock Vina[20] developed by The Scripps research Institute. To run the docking simulations, three MHC I molecules (PDB ID: 1A1O, 1DUZ and 1JHT) and three MHC II molecules (PDB ID: 1AQD, 1DLH and 1H15) were selected. Protein Data Bank (PDB) files for the predicted epitopes were prepared by using HHPred to use them as ligands. Autodock tools were used for preparation of receptor and ligand molecules for docking simulations at the binding groove of the MHC molecules.

RESULTS

Secondary structure analysis

The secondary structural features of *C. fleckeri* toxin-1, toxin-2, and HSP70-1 are summarized in Table 1. All of these proteins were found to be rich in leucine, isoleucine, and glycine residues. Toxin-1 and toxin-2 were found to be alkaline in nature, while HSP70-1 was found to be acidic.

Solvent accessible regions

For *C. fleckeri* toxin-1, the minimal value in Eisenberg hydrophobicity scale was -1.154 and maximal value was 0.993. Eisenberg scale puts negative values for hydropathic residues in protein. According to Eisenberg's scale, the most hydrophilic regions of toxin-1 were 70-83, 88-128, 134-141, 164-170, 181-204, 223-231, 235-260, 305-311, 325-341, 345-352, 399-404, and 431-444 [Figure 1]. For toxin-2, the predicted hydrophilic regions were 67-80, 85-125, 132-138, 181-201, 220-260, 282-289, 303-310, 319-337, 361-367, 395-402 and 444-454 [Figure 1]. Again, for HSP70-1 the hydrophilic regions were 26-31, 60-67, 112-125, 141-150, 162-184, 194-227, 281-296, 332-340, 346-353, 366-381, 389-412, 436-443, 456-469, 473-489, and 601-609 [Figure 1]. From the analysis, it was found that toxin-1, toxin-2, and HSP70-1 of *C. fleckeri* were hydrophilic in nature with high flexibility and low complexity segments. In a vaccine design program, it is the first step to make sure that the predicted antigenic fragments can bind to MHC molecules.

Table 1: Secondary structural analysis of *Chironex fleckeri* dermatonecrotic proteins by ProtParam tool

Criteria	Toxin 1	Toxin 2	HSP70-1
Number of amino acids	456	462	652
Molecular weight	51390.3 Da	51684.3 Da	71391.6
Isoelectric pH	8.47	7.57	5.25
No. of negatively charged residues (Asp+Glu)	55	56	99
No. of positively charged residues (Arg+lys)	59	57	85
Formula	$C_{2335}H_{3673}N_{589}O_{680}S_{16}$	$C_{2337}H_{3691}N_{599}O_{695}S_{12}$	$C_{3121}H_{5071}N_{867}O_{1002}S_{21}$
Estimated half-life	30 Hours	30 Hours	30 Hours
Extinction coefficient	37040	43570	33600
Instability index	34.25	34.06	40.30
Aliphatic index	91.69	92.64	78.24
GRAVY	−0.165	−0.189	-0.492

GRAVY = Grand average of hydropathicity

Antigenic peptide evaluation

By analyzing numerical and graphical data, it was found that according to Hopp and Woods scale the regions 19-24, 31-42, 48-69, 84-91, 105-111, 129-135, 138-152, 158-164, 171-181, 208-220, 258-285, 294-303, 310-324, 357-364, 371-380, 390-398, and 415-421 contained the potential hydrophilic regions for toxin-1 [Figure 2]. For toxin-2, the predicted hydrophilic regions were 28-66, 81-87, 102-108, 126-132, 139-149, 168-180, 205-217, 255-282, 291-301, 308-320, 348-360, 381-395, and 403-416 [Figure 2]. Again, for HSP70-1, the predicted regions were 24-38, 45-57, 68-88, 94-111, 126-139,

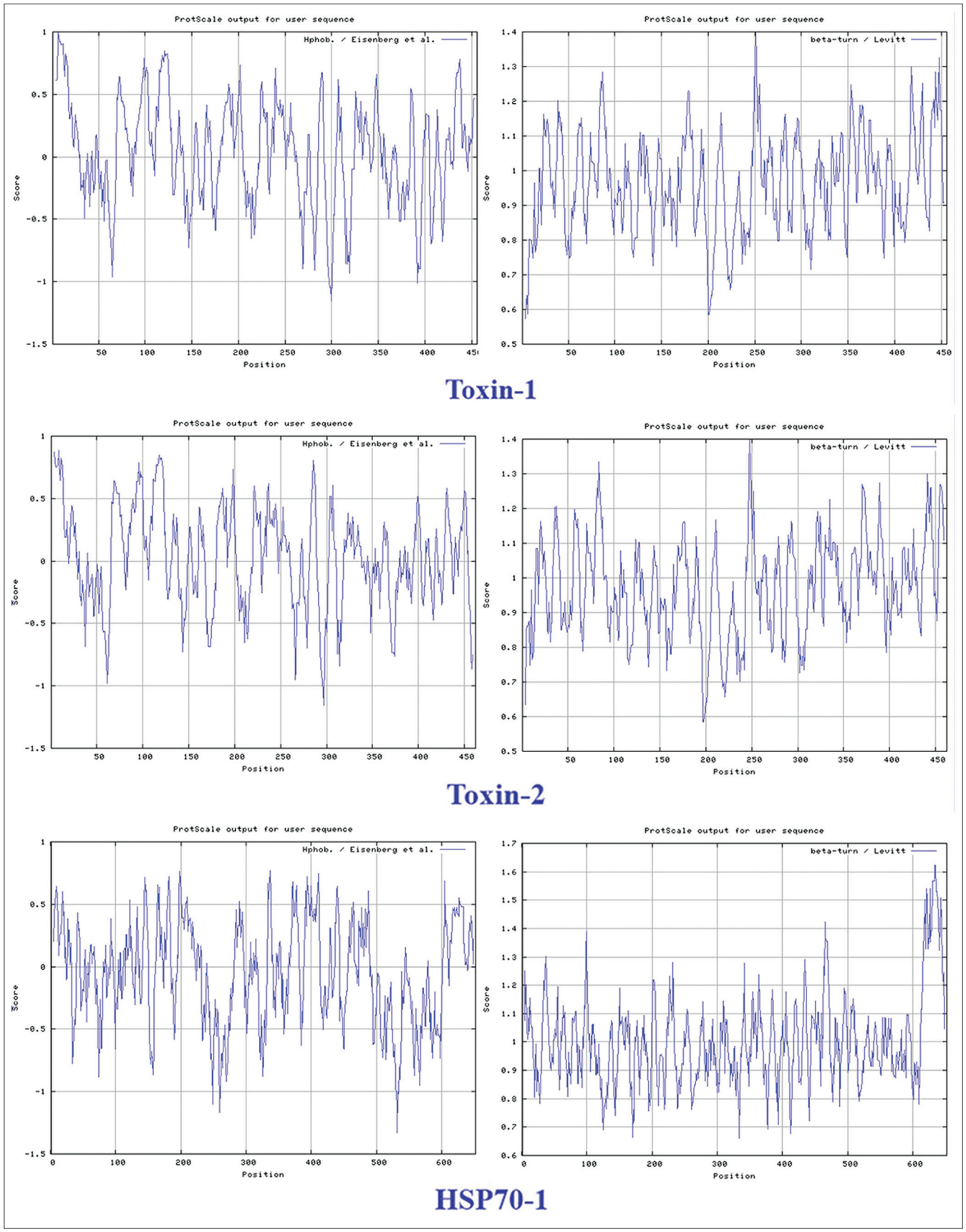

Figure 1: Graphical representation of solvent accessible region (Left - eisenberg) and beta turn region (Right - levitt) analysis of Chironex fleckeri venoms

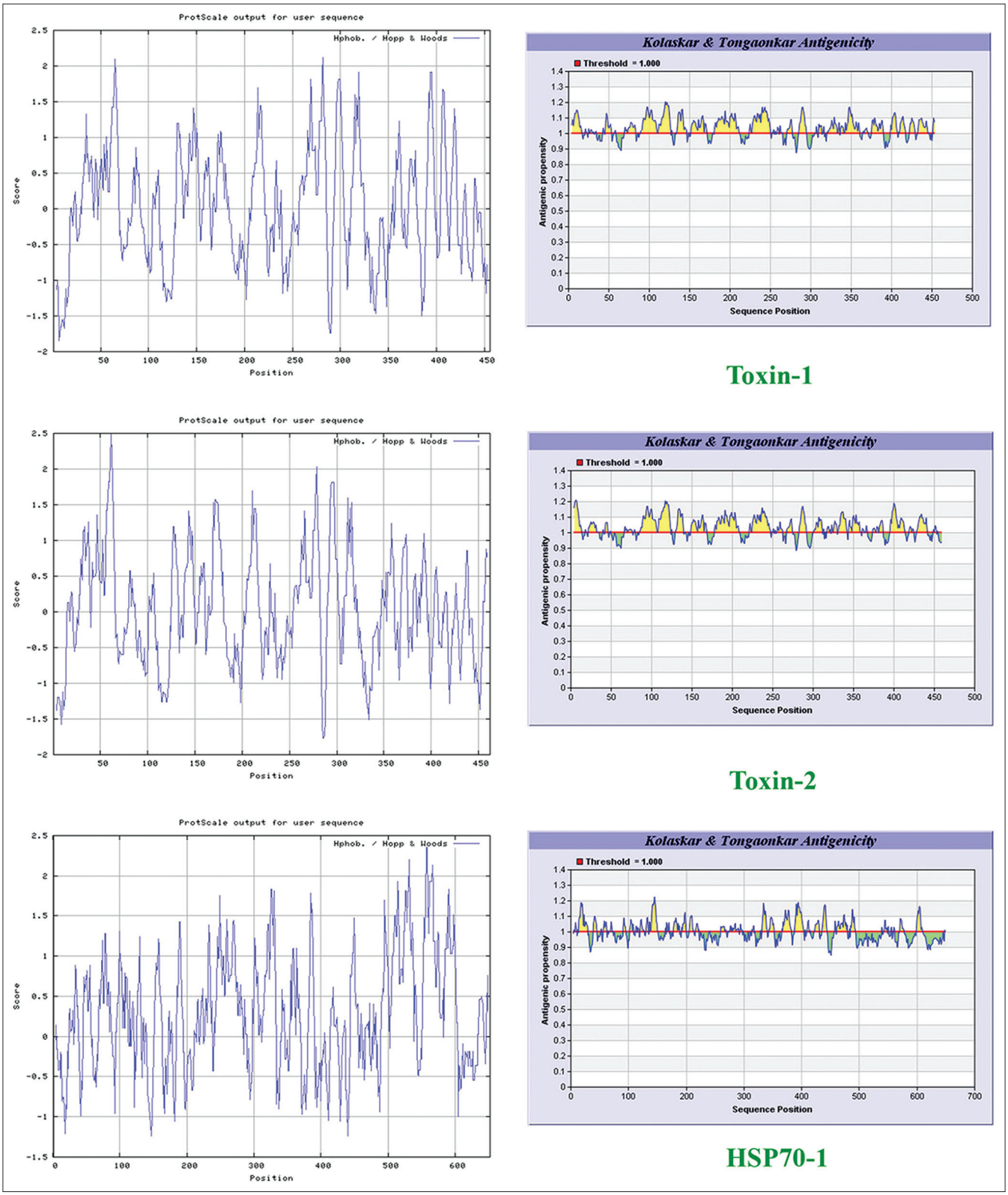

Figure 2: Graphical representation of antigenic peptide evaluation by Hopp and Woods (left) and Kolaskar and Tangaonkar (right) of Chironex fleckeri venoms

152-162, 185-194, 211-217, 227-237, 241-289, 298-310, 313-332, 354-369, 381-389, 413-424, 443-455, 490-501, and 551-603 [Figure 2]. According to Kolaskar and Tangaonkar antigenicity scale, at 1.0 as the threshold level, the most likely antigenic determinants for toxin-1 were at 89-109, 134-146, 181-211, 224-250, 285-294,

342-360, 399-407, 409-417, and 422-429 [Figure 2]. For toxin-2, the antigenicity plot was predicted as 85-106, 129-139, 178-208, 221-247, 283-291, 344-359, 374-386, and 395-413 were the potential antigenic sites [Figure 2]. Antigenic sites for HSP70-1 were predicted at the regions 136-150, 192-199, 205-213, 332-340, 366-382, 388-403, 423-430, 435-444, and 485-493 [Figure 2]. It was found that according to Levitt scale, amino acid residues 23-30, 38-45, 60-65, 83-91, 126-135, 172-184, 247-256, 260-269, 278-284, 291-300, 315-324, 352-368, 373-383, and 416-432 fall inside the beta turn region for toxin-1 by considering 1.000 as the threshold level and the upper section of the graph was analyzed as beta turn region [Figure 1]. For toxin-2 and HSP70-1 the beta turn regions are shown in Figure 1. These predicted fragments are assumed to bind with MHC molecules of immune system, which is the first step toward vaccine design.

Allergenecity evaluation

The query sequences did not meet the criteria set by the Food and Agriculture Organization (FAO)/World Health Organization (WHO) evaluation scheme for cross-reactive allergen prediction. So, the query sequences were classified as a nonallergen by the FAO/WHO evaluation scheme. Both toxin-1 and toxin-2 were predicted as a potential nonallergen with a prediction score of 0.0 (Sensitivity, SE = 91.6%; Specificity, SP = 89.3%). HSP70-1 was also a nonallergen with a score of 0.04 (Sensitivity, SE = 96.0%; Specificity, SP = 45.9%).

Prediction of major histocompatibility complex-binding peptide

A total of 58 alleles were analyzed for MHC class I peptide prediction by using artificial neural network method.[21,22] Again 26 MHC class II alleles were analyzed for prediction of MHC II-binding peptides from the selected venom proteins. We predicted three nonamers which showed sufficiently high results in the prediction methods that were used in this study. The predicted nonamers were "ILLDLYQLV" for toxin-1, "FIAMVVQRI" for toxin-2, and "FQHGKVEII" for HSP70-1. Theses peptides showed interaction with multiple MHC class I and MHC class II alleles. Interaction among different alleles with these peptides is summarized in Table 2 [Supplementary materials 1-3] and Table 3 [Supplementary materials 4-6]. In case of MHC class II prediction, artificial neural network method was used.[23] For selection of all the MHC-binding peptides, MHC IC50 score was below 250 nM. The predicted epitope for toxin-1 interacted with three MHC I alleles (belong to two supertypes A, C) and 12 MHC II alleles (belong to three supertypes and six complexes). The epitope FIAMVVQRI interacted with five MHC I alleles (belong to two supertypes) and 15 MHC II alleles (belong to two supertypes and six complexes). Epitope for HSP70-1 interacted with five MHC I alleles (belong to three supertypes A, B, and C) and five MHC II alleles.

Docking simulation results

The area that were selected on the receptor molecules for docking with the epitopes are summarized in Table 4. One angstrom, spacing was used to select the binding site. The center box area was positioned carefully to make the docking of ligands at the binding groove of the receptors. The three predicted peptides showed significant binding affinity to the MHC receptors [Table 5], except for a few, compared to the Epstein-Barr virus epitope-binding energy with 1H15 (5.9 Kcal/moL). Strong binding affinity gives a clear idea that peptide vaccine designed by using these epitopes may efficiently work *in vivo* to elicit humoral and cell-mediated immunity [Figures 3 and 4].

Table 2: Prediction of MHC class I peptides of *Chironex fleckeri* venom by using proteasome/transporter of antigenic peptides/MHC-combined method

Venom protein	Allele	Start	End	Length	Sequence	Proteasome score	TAP score	MHC score	Processing score	MHC IC50 (nM)
Toxin-1	HLA-A*02:01	238	246	9	ILLDLYQLV	1.25	0.18	−0.70	1.44	5.00
	HLA-A*02:06	238	246	9	ILLDLYQLV	1.25	0.18	−0.60	1.44	4.00
	HLA-C*12:03	238	246	9	ILLDLYQLV	1.25	0.18	−2.14	1.44	138.00
Toxin-2	HLA-A*02:01	196	204	9	FIAMVVQRI	1.19	0.31	−2.08	1.50	120.00
	HLA-A*02:06	196	204	9	FIAMVVQRI	1.19	0.31	−1.99	1.50	97.00
	HLA-A*68:02	196	204	9	FIAMVVQRI	1.19	0.31	−1.70	1.50	50.00
	HLA-C*12:03	196	204	9	FIAMVVQRI	1.19	0.31	−1.86	1.50	73.00
	HLA-C*14:02	196	204	9	FIAMVVQRI	1.19	0.31	−2.27	1.50	186.00
HSP70-1	HLA-A*02:06	21	29	9	FQHGKVEII	1.37	0.26	−1.81	1.62	65.00
	HLA-B*39:01	21	29	9	FQHGKVEII	1.37	0.26	−1.95	1.62	89.00
	HLA-C*06:02	21	29	9	FQHGKVEII	1.37	0.26	−2.10	1.62	125.00
	HLA-C*07:01	21	29	9	FQHGKVEII	1.37	0.26	−2.18	1.62	153.00
	HLA-C*12:03	21	29	9	FQHGKVEII	1.37	0.26	−1.28	1.62	19.00

*TAP = Transporter of antigenic peptides, MHC = Major histocompatibility complex

Supplementary material 1: Toxin 1 MHC 1 allele interaction

Allele	Start	End	Length	Sequence	Proteasome score	TAP score	MHC score	Processing score	MHC IC50
HLA-A*02:01	238	246	9	ILLDLYQLV	1.25	0.18	−0.70	1.44	5.00
HLA-A*02:06	238	246	9	ILLDLYQLV	1.25	0.18	−0.60	1.44	4.00
HLA-C*12:03	238	246	9	ILLDLYQLV	1.25	0.18	−2.14	1.44	138.00
HLA-A*02:01	117	125	9	SILSLVVGL	1.53	0.46	−1.65	1.99	45.00
HLA-A*02:06	117	125	9	SILSLVVGL	1.53	0.46	−1.81	1.99	65.00
HLA-A*32:01	117	125	9	SILSLVVGL	1.53	0.46	−2.35	1.99	225.00
HLA-A*02:01	237	245	9	LILLDLYQL	1.56	0.42	−1.95	1.98	89.00
HLA-A*02:06	237	245	9	LILLDLYQL	1.56	0.42	−1.73	1.98	54.00
HLA-A*02:01	199	207	9	FIAMVVQRI	1.19	0.31	−2.08	1.50	120.00
HLA-A*02:06	199	207	9	FIAMVVQRI	1.19	0.31	−1.99	1.50	97.00
HLA-A*68:02	199	207	9	FIAMVVQRI	1.19	0.31	−1.70	1.50	50.00
HLA-C*12:03	199	207	9	FIAMVVQRI	1.19	0.31	−1.86	1.50	73.00
HLA-C*14:02	199	207	9	FIAMVVQRI	1.19	0.31	−2.27	1.50	186.00
HLA-A*02:01	98	106	9	ILVGISSVL	1.70	0.41	−2.37	2.11	236.00
HLA-C*03:03	98	106	9	ILVGISSVL	1.70	0.41	−0.70	2.11	5.00
HLA-A*02:06	194	202	9	YQGVRFIAM	0.99	0.09	−1.57	1.08	37.00
HLA-B*08:01	194	202	9	YQGVRFIAM	0.99	0.09	−1.67	1.08	47.00
HLA-B*15:01	194	202	9	YQGVRFIAM	0.99	0.09	−1.57	1.08	37.00
HLA-B*39:01	194	202	9	YQGVRFIAM	0.99	0.09	−2.03	1.08	107.00
HLA-A*02:06	185	193	9	SALAANVPI	0.98	0.31	−1.67	1.29	47.00
HLA-A*32:01	185	193	9	SALAANVPI	0.98	0.31	−2.00	1.29	99.00
HLA-A*68:02	185	193	9	SALAANVPI	0.98	0.31	−2.00	1.29	100.00
HLA-C*03:03	185	193	9	SALAANVPI	0.98	0.31	−1.18	1.29	15.00
HLA-C*12:03	185	193	9	SALAANVPI	0.98	0.31	−1.66	1.29	46.00
HLA-C*14:02	185	193	9	SALAANVPI	0.98	0.31	−2.22	1.29	166.00
HLA-C*15:02	185	193	9	SALAANVPI	0.98	0.31	−2.30	1.29	198.00
HLA-A*02:06	202	210	9	MVVQRIKYI	1.06	0.41	−2.32	1.47	209.00
HLA-A*68:02	202	210	9	MVVQRIKYI	1.06	0.41	−2.31	1.47	206.00
HLA-C*12:03	202	210	9	MVVQRIKYI	1.06	0.41	−1.74	1.47	55.00
HLA-C*15:02	202	210	9	MVVQRIKYI	1.06	0.41	−2.18	1.47	153.00
HLA-A*03:01	154	162	9	ALYGVKREY	1.44	1.35	−2.33	2.79	213.00
HLA-B*15:01	154	162	9	ALYGVKREY	1.44	1.35	−2.24	2.79	175.00
HLA-B*15:02	154	162	9	ALYGVKREY	1.44	1.35	−1.92	2.79	83.00
HLA-A*11:01	203	211	9	VVQRIKYIK	0.70	0.29	−1.57	0.99	37.00
HLA-A*11:01	74	82	9	GSLSTAVGK	0.90	0.15	−1.61	1.05	41.00
HLA-A*11:01	99	107	9	LVGISSVLK	0.88	0.20	−1.68	1.08	48.00
HLA-A*68:01	99	107	9	LVGISSVLK	0.88	0.20	−2.37	1.08	237.00
HLA-A*11:01	200	208	9	IAMVVQRIK	1.03	0.22	−2.20	1.25	159.00
HLA-A*68:01	200	208	9	IAMVVQRIK	1.03	0.22	−2.39	1.25	243.00
HLA-A*23:01	285	293	9	TYLFLSYLY	1.09	1.31	−1.83	2.40	68.00
HLA-A*24:02	285	293	9	TYLFLSYLY	1.09	1.31	−2.01	2.40	102.00
HLA-A*29:02	285	293	9	TYLFLSYLY	1.09	1.31	−0.30	2.40	2.00
HLA-C*14:02	285	293	9	TYLFLSYLY	1.09	1.31	−2.29	2.40	194.00
HLA-A*24:02	111	119	9	KFSPIFSIL	1.54	0.60	−2.18	2.14	153.00
HLA-C*14:02	111	119	9	KFSPIFSIL	1.54	0.60	−1.86	2.14	72.00
HLA-A*29:02	186	194	9	ALAANVPIY	1.34	1.37	−2.01	2.72	103.00
HLA-B*15:01	186	194	9	ALAANVPIY	1.34	1.37	−1.79	2.72	62.00
HLA-A*29:02	201	209	9	AMVVQRIKY	1.14	1.42	−2.35	2.56	224.00
HLA-A*30:01	196	204	9	GVRFIAMVV	0.96	0.12	−1.46	1.08	29.00
HLA-A*68:01	227	235	9	LFTDLCSLR	0.90	0.63	−2.00	1.53	100.00
HLA-A*68:02	183	191	9	EVSALAANV	1.05	0.11	−0.60	1.16	4.00
HLA-A*68:02	161	169	9	EYAVSKAFL	1.38	0.51	−2.19	1.88	156.00
HLA-B*07:02	113	121	9	SPIFSILSL	1.48	0.29	−1.18	1.77	15.00
HLA-B*35:01	113	121	9	SPIFSILSL	1.48	0.29	−2.27	1.77	188.00

Contd...

Supplementary material 1: Contd...

Allele	Start	End	Length	Sequence	Proteasome score	TAP score	MHC score	Processing score	MHC IC50
HLA-B*39:01	113	121	9	SPIFSILSL	1.48	0.29	−1.93	1.77	86.00
HLA-B*07:02	191	199	9	VPIYQGVRF	1.36	1.09	−2.16	2.45	143.00
HLA-B*35:01	191	199	9	VPIYQGVRF	1.36	1.09	−1.80	2.45	63.00
HLA-B*35:03	191	199	9	VPIYQGVRF	1.36	1.09	−2.28	2.45	189.00
HLA-B*53:01	191	199	9	VPIYQGVRF	1.36	1.09	−2.34	2.45	218.00
HLA-B*15:01	160	168	9	REYAVSKAF	1.48	1.27	−1.70	2.75	50.00
HLA-B*18:01	160	168	9	REYAVSKAF	1.48	1.27	−1.11	2.75	13.00
HLA-B*40:01	160	168	9	REYAVSKAF	1.48	1.27	−1.60	2.75	40.00
HLA-B*40:02	160	168	9	REYAVSKAF	1.48	1.27	−1.15	2.75	14.00
HLA-B*44:02	160	168	9	REYAVSKAF	1.48	1.27	−1.64	2.75	44.00
HLA-B*44:03	160	168	9	REYAVSKAF	1.48	1.27	−1.53	2.75	34.00
HLA-B*48:01	160	168	9	REYAVSKAF	1.48	1.27	−2.28	2.75	192.00
HLA-C*14:02	160	168	9	REYAVSKAF	1.48	1.27	−2.28	2.75	189.00
HLA-B*15:01	75	83	9	SLSTAVGKF	1.22	1.12	−2.34	2.34	220.00
HLA-B*15:02	75	83	9	SLSTAVGKF	1.22	1.12	−2.37	2.34	232.00
HLA-B*18:01	283	291	9	KETYLFLSY	1.21	1.18	−1.32	2.39	21.00
HLA-C*03:03	91	99	9	IASGCLDIL	1.47	0.49	−0.85	1.96	7.00
HLA-C*12:03	91	99	9	IASGCLDIL	1.47	0.49	−1.82	1.96	66.00
HLA-C*12:03	192	200	9	PIYQGVRFI	1.31	0.18	−2.11	1.49	128.00
HLA-C*12:03	97	105	9	DILVGISSV	0.68	0.08	−2.36	0.76	227.00
HLA-C*14:02	161	169	9	EYAVSKAFL	1.38	0.51	−1.97	1.88	93.00

TAP = Transporter of antigenic peptides, MHC = Major histocompatibility complex

Supplementary material 2: Toxin 2 MHC 1 allele interaction

Allele	Start	End	Length	Sequence	Proteasome score	TAP score	MHC score	Processing score	MHC IC50 (nM)
HLA-A*02:01	235	243	9	ILLDLHQLI	1.36	0.28	−1.23	1.64	17.00
HLA-A*02:06	235	243	9	ILLDLHQLI	1.36	0.28	−1.23	1.64	17.00
HLA-C*12:03	235	243	9	ILLDLHQLI	1.36	0.28	−2.35	1.64	223.00
HLA-A*02:01	114	122	9	SILSLVVGL	1.53	0.46	−1.65	1.99	45.00
HLA-A*02:06	114	122	9	SILSLVVGL	1.53	0.46	−1.81	1.99	65.00
HLA-A*32:01	114	122	9	SILSLVVGL	1.53	0.46	−2.35	1.99	225.00
HLA-A*02:01	182	190	9	SALAANIPV	0.82	0.21	−1.87	1.03	74.00
HLA-A*02:06	182	190	9	SALAANIPV	0.82	0.21	−1.08	1.03	12.00
HLA-A*30:01	182	190	9	SALAANIPV	0.82	0.21	−2.28	1.03	190.00
HLA-A*68:02	182	190	9	SALAANIPV	0.82	0.21	−1.63	1.03	43.00
HLA-C*03:03	182	190	9	SALAANIPV	0.82	0.21	-1.81	1.03	64.00
HLA-C*12:03	182	190	9	SALAANIPV	0.82	0.21	−1.94	1.03	88.00
HLA-C*15:02	182	190	9	SALAANIPV	0.82	0.21	−2.05	1.03	113.00
HLA-A*02:01	196	204	9	FIAMVVQRI	1.19	0.31	−2.08	1.50	120.00
HLA-A*02:06	196	204	9	FIAMVVQRI	1.19	0.31	−1.99	1.50	97.00
HLA-A*68:02	196	204	9	FIAMVVQRI	1.19	0.31	−1.70	1.50	50.00
HLA-C*12:03	196	204	9	FIAMVVQRI	1.19	0.31	−1.86	1.50	73.00
HLA-C*14:02	196	204	9	FIAMVVQRI	1.19	0.31	−2.27	1.50	186.00
HLA-A*02:01	95	103	9	ILVGISSVL	1.70	0.41	−2.37	2.11	236.00
HLA-C*03:03	95	103	9	ILVGISSVL	1.70	0.41	−0.70	2.11	5.00
HLA-A*02:01	234	242	9	LILLDLHQL	1.43	0.42	−2.38	1.85	242.00
HLA-A*02:06	234	242	9	LILLDLHQL	1.43	0.42	−1.69	1.85	49.00
HLA-A*02:06	191	199	9	YQGVRFIAM	0.99	0.11	−1.57	1.10	37.00
HLA-B*08:01	191	199	9	YQGVRFIAM	0.99	0.11	−1.67	1.10	47.00
HLA-B*15:01	191	199	9	YQGVRFIAM	0.99	0.11	−1.57	1.10	37.00
HLA-B*39:01	191	199	9	YQGVRFIAM	0.99	0.11	−2.03	1.10	107.00
HLA-A*02:06	199	207	9	MVVQRIKYI	1.06	0.41	−2.32	1.47	209.00

Contd...

Supplementary material 2: Contd...

Allele	Start	End	Length	Sequence	Proteasome score	TAP score	MHC score	Processing score	MHC IC50 (nM)
HLA-A*68:02	199	207	9	MVVQRIKYI	1.06	0.41	−2.31	1.47	206.00
HLA-C*12:03	199	207	9	MVVQRIKYI	1.06	0.41	−1.74	1.47	55.00
HLA-C*15:02	199	207	9	MVVQRIKYI	1.06	0.41	−2.18	1.47	153.00
HLA-A*02:06	111	119	9	PVFSILSLV	0.88	0.06	−2.39	0.94	246.00
HLA-A*68:02	111	119	9	PVFSILSLV	0.88	0.06	−1.91	0.94	81.00
HLA-A*03:01	96	104	9	LVGISSVLK	0.88	0.20	−2.31	1.08	202.00
HLA-A*11:01	96	104	9	LVGISSVLK	0.88	0.20	−1.68	1.08	48.00
HLA-A*68:01	96	104	9	LVGISSVLK	0.88	0.20	−2.37	1.08	237.00
HLA-A*11:01	200	208	9	VVQRIKYIK	0.70	0.29	−1.57	0.99	37.00
HLA-A*31:01	200	208	9	VVQRIKYIK	0.70	0.29	−1.91	0.99	82.00
HLA-A*11:01	197	205	9	IAMVVQRIK	1.03	0.22	−2.20	1.25	159.00
HLA-A*68:01	197	205	9	IAMVVQRIK	1.03	0.22	−2.39	1.25	243.00
HLA-A*24:02	108	116	9	KFSPVFSIL	1.54	0.60	−2.29	2.14	193.00
HLA-C*07:02	108	116	9	KFSPVFSIL	1.54	0.60	−2.11	2.14	130.00
HLA-C*14:02	108	116	9	KFSPVFSIL	1.54	0.60	−1.95	2.14	90.00
HLA-A*29:02	183	191	9	ALAANIPVY	1.45	1.37	−2.15	2.82	141.00
HLA-A*30:02	183	191	9	ALAANIPVY	1.45	1.37	−1.23	2.82	17.00
HLA-B*15:01	183	191	9	ALAANIPVY	1.45	1.37	−1.59	2.82	39.00
HLA-A*29:02	198	206	9	AMVVQRIKY	1.14	1.42	−2.35	2.56	224.00
HLA-A*30:02	198	206	9	AMVVQRIKY	1.14	1.42	−2.37	2.56	233.00
HLA-A*30:01	193	201	9	GVRFIAMVV	0.96	0.12	−1.46	1.08	29.00
HLA-A*30:01	155	163	9	VKREFAVSK	1.18	0.26	−2.37	1.43	234.00
HLA-A*31:01	195	203	9	RFIAMVVQR	0.96	0.87	−0.78	1.84	6.00
HLA-A*32:01	157	165	9	REFAVSKAF	1.48	1.24	−2.19	2.72	155.00
HLA-B*15:01	157	165	9	REFAVSKAF	1.48	1.24	−1.76	2.72	57.00
HLA-B*18:01	157	165	9	REFAVSKAF	1.48	1.24	−1.08	2.72	12.00
HLA-B*40:01	157	165	9	REFAVSKAF	1.48	1.24	−1.38	2.72	24.00
HLA-B*40:02	157	165	9	REFAVSKAF	1.48	1.24	−1.20	2.72	16.00
HLA-B*44:02	157	165	9	REFAVSKAF	1.48	1.24	−1.54	2.72	35.00
HLA-B*44:03	157	165	9	REFAVSKAF	1.48	1.24	−1.30	2.72	20.00
HLA-A*68:02	180	188	9	EVSALAANI	1.02	0.20	−0.85	1.22	7.00
HLA-B*07:02	110	118	9	SPVFSILSL	1.48	0.30	−1.20	1.77	16.00
HLA-B*35:01	110	118	9	SPVFSILSL	1.48	0.30	−2.23	1.77	168.00
HLA-B*39:01	110	118	9	SPVFSILSL	1.48	0.30	−2.26	1.77	183.00
HLA-B*07:02	188	196	9	IPVYQGVRF	1.36	1.07	−2.14	2.43	139.00
HLA-B*35:01	188	196	9	IPVYQGVRF	1.36	1.07	−1.40	2.43	25.00
HLA-B*35:03	188	196	9	IPVYQGVRF	1.36	1.07	−2.17	2.43	148.00
HLA-B*53:01	188	196	9	IPVYQGVRF	1.36	1.07	−2.05	2.43	111.00
HLA-B*15:01	151	159	9	ALYGVKREF	1.42	1.18	−2.20	2.60	159.00
HLA-B*15:02	151	159	9	ALYGVKREF	1.42	1.18	−2.18	2.60	150.00
HLA-C*14:02	151	159	9	ALYGVKREF	1.42	1.18	−2.08	2.60	121.00
HLA-C*03:03	88	96	9	IASGCLDIL	1.47	0.47	−0.85	1.94	7.00
HLA-C*12:03	88	96	9	IASGCLDIL	1.47	0.47	−1.82	1.94	66.00
HLA-C*03:03	85	93	9	PASIASGCL	1.59	0.17	−1.61	1.76	41.00
HLA-C*12:03	189	197	9	PVYQGVRFI	1.31	0.17	−1.84	1.48	69.00
HLA-C*12:03	94	102	9	DILVGISSV	0.68	0.08	−2.36	0.76	227.00

TAP = Transporter of antigenic peptides, MHC = Major histocompatibility complex

DISCUSSION

Prediction of epitope and mapping theses on the protein surface is a vital step for epitope-based vaccine design. A number of ways were attempted in earlier studies but here tried to predict the epitopes more accurately by starting from the very basic step- like finding the hydrophilic regions of the proteins and ending by docking of epitopes to their respective receptors. To find out the hydrophobicity scores, which actually give us the index for hydrophilic regions, we used Eisenberg hydrophobicity and Wolfenden hydrophobicity scales. These scales put a positive score for the nonpolar residues and negative score for polar residues of a given protein. From these data, we

Supplementary material 3: HSP70-1 MHC 1 allele interaction

Allele	Start	End	Length	Sequence	Proteasome score	TAP score	MHC score	Processing score	MHC IC50 (nM)
HLA-A*02:06	21	29	9	FQHGKVEII	1.37	0.26	−1.81	1.62	65.00
HLA-B*39:01	21	29	9	FQHGKVEII	1.37	0.26	−1.95	1.62	89.00
HLA-C*06:02	21	29	9	FQHGKVEII	1.37	0.26	−2.10	1.62	125.00
HLA-C*07:01	21	29	9	FQHGKVEII	1.37	0.26	−2.18	1.62	153.00
HLA-C*12:03	21	29	9	FQHGKVEII	1.37	0.26	−1.28	1.62	19.00
HLA-A*11:01	120	128	9	SSMVLTKMK	0.97	0.27	−1.20	1.25	16.00
HLA-A*68:01	120	128	9	SSMVLTKMK	0.97	0.27	−2.30	1.25	198.00
HLA-A*11:01	179	187	9	AAIAYGLDK	0.72	0.39	−1.58	1.10	38.00
HLA-A*11:01	17	25	9	CVGVFQHGK	0.76	0.20	−1.71	0.95	51.00
HLA-A*68:01	17	25	9	CVGVFQHGK	0.76	0.20	−2.16	0.95	146.00
HLA-A*11:01	180	188	9	AIAYGLDKK	0.83	0.40	−1.79	1.22	61.00
HLA-A*30:01	276	284	9	SSKQASIEI	1.10	0.29	−1.43	1.40	27.00
HLA-C*15:02	276	284	9	SSKQASIEI	1.10	0.29	−1.91	1.40	82.00
HLA-A*68:02	367	375	9	EAVAYGAAV	0.95	0.04	−1.00	0.99	10.00
HLA-C*03:03	367	375	9	EAVAYGAAV	0.95	0.04	−1.61	0.99	41.00
HLA-C*12:03	367	375	9	EAVAYGAAV	0.95	0.04	−1.99	0.99	98.00
HLA-C*15:02	367	375	9	EAVAYGAAV	0.95	0.04	−2.33	0.99	216.00
HLA-A*68:02	66	74	9	TVFDAKRLI	1.05	0.39	−1.59	1.44	39.00
HLA-C*06:02	66	74	9	TVFDAKRLI	1.05	0.39	−1.85	1.44	71.00
HLA-C*07:01	66	74	9	TVFDAKRLI	1.05	0.39	−1.61	1.44	41.00
HLA-C*12:03	66	74	9	TVFDAKRLI	1.05	0.39	−1.60	1.44	40.00
HLA-A*68:02	283	291	9	EIDSLFEGI	1.06	0.13	−1.88	1.18	75.00
HLA-A*68:02	371	379	9	YGAAVQAAI	0.98	0.18	−1.94	1.16	88.00
HLA-C*03:03	371	379	9	YGAAVQAAI	0.98	0.18	−2.18	1.16	151.00
HLA-A*68:02	37	45	9	TTPSYVAFT	0.91	−0.27	−2.09	0.64	124.00
HLA-B*15:01	280	288	9	ASIEIDSLF	1.08	1.19	−2.27	2.27	188.00
HLA-B*58:01	280	288	9	ASIEIDSLF	1.08	1.19	−1.85	2.27	70.00
HLA-B*15:01	141	149	9	NVVVTVPAY	1.33	1.36	−2.37	2.69	235.00
HLA-B*35:01	141	149	9	NVVVTVPAY	1.33	1.36	−1.04	2.69	11.00
HLA-B*18:01	366	374	9	DEAVAYGAA	0.97	−0.48	−2.30	0.49	198.00
HLA-B*27:05	261	269	9	RRLRTACER	1.05	0.87	−1.45	1.92	28.00
HLA-B*58:01	142	150	9	VVVTVPAYF	1.43	1.28	−1.61	2.71	41.00
HLA-B*58:01	119	127	9	ISSMVLTKM	0.97	0.08	−1.72	1.05	53.00
HLA-B*58:01	82	90	9	GAQADMKHW	1.42	0.38	−1.86	1.80	73.00
HLA-C*03:03	372	380	9	GAAVQAAIL	1.41	0.35	−1.11	1.75	13.00
HLA-C*03:03	369	377	9	VAYGAAVQA	1.63	−0.03	−1.61	1.60	41.00
HLA-C*03:03	279	287	9	QASIEIDSL	1.46	0.49	−2.02	1.95	105.00
HLA-C*12:03	138	146	9	NCQNVVVTV	1.24	0.19	−2.13	1.43	134.00
HLA-C*14:02	20	28	9	VFQHGKVEI	1.27	0.30	−2.10	1.57	126.00
HLA-C*14:02	370	378	9	AYGAAVQAA	1.02	−0.10	−2.24	0.91	174.00

TAP = Transporter of antigenic peptides, MHC = Major histocompatibility complex

measured the hydrophilic regions of the selected proteins, which are supposed to be antigenic in nature and more exposed to the surface of the protein. The regions with maximum hydrophilic scores are analyzed as antigenic sites, because these regions are unstructured and solvent accessible which make it easier for antibodies to recognize the native proteins.[24] Wolfenden hydrophobicity plot puts an overall impression on hydrophilicity of a protein molecule. Numerical and graphical analysis showed that the proteins were hydrophilic in nature. According to Wolfenden hydrophobicity plot, glycine, leucine, and isoleucine are the most hydrophobic residues.

The local hydrophilic region of the protein which is typically more exposed to the surface is detected as the antigenic site and the corresponding amino acids of these sites are detected as the antigenic peptides. Hopp and Woods hydrophilicity scale and Kolaskar and Tangaonkar antigenecity scale were used to predict the antigenic peptides of the selected toxin proteins. Hopp and Woods hydrophobicity scale is actually a hydrophilicity scale in which window size 7 gives the ideal values for a protein hydrophilicity nature. Hopp and Woods scale assigns nonpolar residues with a negative value. Kolaskar and Tangaonkar antigenecity scale is the

Table 3: Prediction of MHC class II peptides of *Chironex fleckeri* venom by using artificial neural network method

Venom protein	Start	End	Sequence	Allele	IC50
Toxin-1	238	246	ILLDLYQLV	HLA-DPA1*01/DPB1*04:01	89.1
				HLA-DPA1*01:03/DPB1*02:01	41.5
				HLA-DPA1*02:01/DPB1*01:01	23.8
				HLA-DPA1*03:01/DPB1*04:02	22.6
				HLA-DPB1*03:01/DPB1*04:01	15.7
				HLA-DQA1*01:01/DQB1*05:01	246.2
				HLA-DRB1*03:01	9.7
				HLA-DRB1*04:05	185.2
				HLA-DRB1*13:02	194.9
				HLA-DRB1*15:01	114.1
				HLA-DRB3*01:01	13.2
				HLA-DRB4*01:01	73.6
Toxin-2	196	204	FIAMVVQRI	HLA-DPA1*01/DPB1*04:01	129.1
				HLA-DPA1*01:03/DPB1*02:01	192.3
				HLA-DPA1*02:01/DPB1*01:01	191.3
				HLA-DPA1*03:01/DPB1*04:02	29.2
				HLA-DPB1*03:01/DPB1*04:01	26.4
				HLA-DQA1*05:01/DQB1*02:01	200.7
				HLA-DRB1*01:01	5.4
				HLA-DRB1*04:01	26.7
				HLA-DRB1*04:05	44.6
				HLA-DRB1*07:01	12.7
				HLA-DRB1*09:01	11.4
				HLA-DRB1*11:01	27.9
				HLA-DRB1*13:02	106.9
				HLA-DRB1*15:01	23.2
				HLA-DRB4*01:01	23.5
HSP70-1	21	29	FQHGKVEII	HLA-DPA1*03:01/DPB1*04:02	136.7
				HLA-DQA1*05:01/DQB1*03:01	28.3
				HLA-DRB1*01:01	27.2
				HLA-DRB1*07:01	5.5
				HLA-DRB1*09:01	40

MHC = Major histocompatibility complex

simplest method for determining antigenic determinants. This method is based on the occurrence of amino acid residues in experimentally determined epitopes. In several experimental studies, it was found that the antigenic parts of a protein belong to the beta turn regions.[25,26] To validate the predicted hydrophilic and antigenic parts of the protein we analyzed Levitt and Deleage and Roux beta turn scale.

The AllerHunter score value is the probability that a particular sequence is a cross-reactive allergen. However, the threshold for prediction of cross-reactive allergen is adjusted such that a sequence is predicted as a cross-reactive allergen if its probability is >= 0.06. The probability threshold was determined during the fine-tuning of prediction model. AllerHunter has optimum prediction result at that particular threshold.

Supplementary material 4: Toxin 1 MHC II allele interaction

Start	End	Core	IC50	Allele
111	119	KFSPIFSIL	95.4	HLA-DPA1*01/DPB1*04:01
			85.1	HLA-DPA1*01:03/DPB1*02:01
			44.8	HLA-DPA1*02:01/DPB1*01:01
			112.2	HLA-DPA1*02:01/DPB1*05:01
			92.2	HLA-DPA1*03:01/DPB1*04:02
			173.5	HLA-DPB1*03:01/DPB1*04:01
193	201	IYQGVRFIA	8.1	HLA-DPA1*01/DPB1*04:01
			42.1	HLA-DPA1*02:01/DPB1*01:01
			241.4	HLA-DQA1*05:01/DQB1*03:01
			45.1	HLA-DRB1*11:01
194	202	YQGVRFIAM	8	HLA-DPA1*01/DPB1*04:01
			7.7	HLA-DPA1*01:03/DPB1*02:01
			157	HLA-DPA1*03:01/DPB1*04:02
			12.3	HLA-DPB1*03:01/DPB1*04:01
			182.4	HLA-DQA1*01:02/DQB1*06:02
			89.9	HLA-DQA1*05:01/DQB1*03:01
			147.3	HLA-DRB1*01:01
			134.8	HLA-DRB1*04:01
199	207	FIAMVVQRI	129.1	HLA-DPA1*01/DPB1*04:01
			192.3	HLA-DPA1*01:03/DPB1*02:01
			191.3	HLA-DPA1*02:01/DPB1*01:01
			29.2	HLA-DPA1*03:01/DPB1*04:02
			28.2	HLA-DPB1*03:01/DPB1*04:01
			200.6	HLA-DQA1*05:01/DQB1*02:01
			5.4	HLA-DRB1*01:01
			26.7	HLA-DRB1*04:01
			44.6	HLA-DRB1*04:05
			12.7	HLA-DRB1*07:01
			11.4	HLA-DRB1*09:01
			27.9	HLA-DRB1*11:01
			106.9	HLA-DRB1*13:02
			23.2	HLA-DRB1*15:01
			23.5	HLA-DRB4*01:01
227	235	LFTDLCSLR	161	HLA-DPA1*01/DPB1*04:01
			59.2	HLA-DPA1*02:01/DPB1*01:01
			22.5	HLA-DPA1*03:01/DPB1*04:02
			13	HLA-DRB1*03:01
231	239	LCSLRDLIL	190.6	HLA-DPA1*01/DPB1*04:01
			38.4	HLA-DRB1*01:01
			91.6	HLA-DRB1*04:05
			26.5	HLA-DRB1*07:01
238	246	ILLDLYQLV	89.1	HLA-DPA1*01/DPB1*04:01
			41.5	HLA-DPA1*01:03/DPB1*02:01
			23.8	HLA-DPA1*02:01/DPB1*01:01
			22.6	HLA-DPA1*03:01/DPB1*04:02
			15.7	HLA-DPB1*03:01/DPB1*04:01
			246.2	HLA-DQA1*01:01/DQB1*05:01
			9.7	HLA-DRB1*03:01
			185.2	HLA-DRB1*04:05
			194.9	HLA-DRB1*13:02
			114.1	HLA-DRB1*15:01
			13.2	HLA-DRB3*01:01
			73.6	HLA-DRB4*01:01

Contd...

Supplementary material 4: Contd...

Start	End	Core	IC50	Allele
286	294	YLFLSYLYP	14	HLA-DPA1*01/DPB1*04:01
			12.4	HLA-DPA1*01:03/DPB1*02:01
			19.5	HLA-DPA1*03:01/DPB1*04:02
			19.1	HLA-DPB1*03:01/DPB1*04:01
			140.8	HLA-DRB1*01:01
			20	HLA-DRB1*04:04
223	231	TMLELFTDL	54.9	HLA-DPA1*01:03/DPB1*02:01
			56.7	HLA-DPA1*02:01/DPB1*01:01
234	242	LRDLILLDL	236.9	HLA-DPA1*01:03/DPB1*02:01
			67.7	HLA-DPA1*02:01/DPB1*01:01
			19.9	HLA-DPA1*03:01/DPB1*04:02
			42.6	HLA-DRB4*01:01
99	107	LVGISSVLK	181	HLA-DPA1*02:01/DPB1*01:01
			243.2	HLA-DRB1*01:01
			87.9	HLA-DRB1*07:01
			59.9	HLA-DRB5*01:01
114	122	PIFSILSLV	36.2	HLA-DPA1*02:01/DPB1*01:01
			59	HLA-DPA1*03:01/DPB1*04:02
			131.9	HLA-DPB1*03:01/DPB1*04:01
			158	HLA-DRB1*04:04
237	245	LILLDLYQL	37	HLA-DPA1*02:01/DPB1*01:01
			14.3	HLA-DPA1*03:01/DPB1*04:02
			72.8	HLA-DRB1*01:01
			75.7	HLA-DRB1*15:01
160	168	REYAVSKAF	204.6	HLA-DPA1*02:01/DPB1*05:01
			227.8	HLA-DRB1*07:01
201	209	AMVVQRIKY	36	HLA-DPA1*03:01/DPB1*04:02
224	232	MLELFTDLC	59.9	HLA-DQA1*01:01/DQB1*05:01
100	108	VGISSVLKD	219.2	HLA-DQA1*01:02/DQB1*06:02
181	189	PTEVSALAA	248.3	HLA-DQA1*01:02/DQB1*06:02
182	190	TEVSALAAN	62	HLA-DQA1*01:02/DQB1*06:02
			22.8	HLA-DQA1*05:01/DQB1*03:01
			170.4	HLA-DRB1*04:04
187	195	LAANVPIYQ	62.2	HLA-DQA1*01:02/DQB1*06:02
			182.7	HLA-DRB1*08:02
188	196	AANVPIYQG	75.9	HLA-DQA1*01:02/DQB1*06:02
197	205	VRFIAMVVQ	81	HLA-DQA1*01:02/DQB1*06:02
			202.7	HLA-DRB4*01:01
69	77	VMGAIGSLS	6	HLA-DQA1*05:01/DQB1*03:01
			9.3	HLA-DRB1*01:01
			219.2	HLA-DRB1*04:01
			204	HLA-DRB1*08:02
			211.5	HLA-DRB5*01:01
72	80	AIGSLSTAV	12	HLA-DQA1*05:01/DQB1*03:01
			148.1	HLA-DRB1*08:02
			211.8	HLA-DRB1*09:01
73	81	IGSLSTAVG	170.7	HLA-DQA1*05:01/DQB1*03:01
			16.2	HLA-DRB1*01:01
			152.9	HLA-DRB1*04:01
			161.1	HLA-DRB1*08:02
98	106	ILVGISSVL	70	HLA-DQA1*05:01/DQB1*03:01
			21	HLA-DRB1*01:01
			176.7	HLA-DRB1*09:01
			110.1	HLA-DRB1*15:01

Contd...

Supplementary material 4: Contd...

Start	End	Core	IC50	Allele
155	163	LYGVKREYA	246.9	HLA-DQA1*05:01/DQB1*03:01
			44.1	HLA-DRB1*11:01
162	170	YAVSKAFLD	139.7	HLA-DQA1*05:01/DQB1*03:01
			12	HLA-DRB1*01:01
			35.5	HLA-DRB1*04:05
			40.5	HLA-DRB1*07:01
			119.8	HLA-DRB1*09:01
			179.4	HLA-DRB1*11:01
			14.7	HLA-DRB5*01:01
184	192	VSALAANVP	36.2	HLA-DQA1*05:01/DQB1*03:01
			22.9	HLA-DRB1*01:01
			205.5	HLA-DRB1*04:04
186	194	ALAANVPIY	142.2	HLA-DQA1*05:01/DQB1*03:01
96	104	LDILVGISS	12.6	HLA-DRB1*01:01
			173.5	HLA-DRB1*04:04
			182.7	HLA-DRB1*11:01
112	120	FSPIFSILS	10.9	HLA-DRB1*01:01
			143.5	HLA-DRB1*04:01
			114.5	HLA-DRB1*04:05
			33.3	HLA-DRB1*07:01
			101.5	HLA-DRB5*01:01
116	124	FSILSLVVG	24.9	HLA-DRB1*01:01
			77.5	HLA-DRB1*04:04
156	164	YGVKREYAV	94.9	HLA-DRB1*01:01
196	204	GVRFIAMVV	143.6	HLA-DRB1*01:01
			224	HLA-DRB1*09:01
			37.2	HLA-DRB1*15:01
202	210	MVVQRIKYI	152.1	HLA-DRB1*01:01
			165.8	HLA-DRB1*08:02
			38.5	HLA-DRB1*15:01
228	236	FTDLCSLRD	26.2	HLA-DRB1*01:01
			42.6	HLA-DRB1*04:01
			100.3	HLA-DRB1*04:04
			49.4	HLA-DRB1*04:05
			239.2	HLA-DRB1*07:01
242	250	LYQLVATPG	4.9	HLA-DRB1*01:01
			34.9	HLA-DRB1*04:01
			22.5	HLA-DRB1*04:05
			81.9	HLA-DRB1*07:01
			40.9	HLA-DRB1*09:01
233	241	SLRDLILLD	43.3	HLA-DRB1*03:01
			86.6	HLA-DRB1*04:05
70	78	MGAIGSLST	68.5	HLA-DRB1*04:04
117	125	SILSLVVGL	63.6	HLA-DRB1*04:04
198	206	RFIAMVVQR	74.4	HLA-DRB1*04:04
240	248	LDLYQLVAT	34.5	HLA-DRB1*04:04
115	123	IFSILSLVV	24.2	HLA-DRB1*07:01
			89.5	HLA-DRB1*11:01
			119.4	HLA-DRB1*15:01
185	193	SALAANVPI	30	HLA-DRB1*07:01
			75.8	HLA-DRB1*13:02
102	110	ISSVLKDFA	28.2	HLA-DRB1*11:01
203	211	VVQRIKYIK	57.1	HLA-DRB1*11:01
			15.3	HLA-DRB5*01:01

Contd...

Supplementary material 4: Contd...

Start	End	Core	IC50	Allele
191	199	VPIYQGVRF	43.7	HLA-DRB1*15:01
285	293	TYLFLSYLY	21	HLA-DRB1*15:01
113	121	SPIFSILSL	247.2	HLA-DRB4*01:01
200	208	IAMVVQRIK	15.5	HLA-DRB4*01:01
			13.4	HLA-DRB5*01:01
235	243	RDLILLDLY	80.3	HLA-DRB4*01:01
158	166	VKREYAVSK	11.8	HLA-DRB5*01:01

Supplementary material 5: Toxin 2 MHC II allele interaction

Start	End	Core	IC50	Allele
108	116	KFSPVFSIL	95.7	HLA-DPA1*01/DPB1*04:01
			83.7	HLA-DPA1*01:03/DPB1*02:01
			44.9	HLA-DPA1*02:01/DPB1*01:01
			183.6	HLA-DPA1*02:01/DPB1*05:01
			78.4	HLA-DPA1*03:01/DPB1*04:02
			212.6	HLA-DPB1*03:01/DPB1*04:01
190	198	VYQGVRFIA	18.6	HLA-DPA1*01/DPB1*04:01
			143	HLA-DPA1*01:03/DPB1*02:01
			51.5	HLA-DPA1*02:01/DPB1*01:01
			237.3	HLA-DQA1*05:01/DQB1*02:01
			123.7	HLA-DRB1*11:01
191	199	YQGVRFIAM	8.1	HLA-DPA1*01/DPB1*04:01
			8	HLA-DPA1*01:03/DPB1*02:01
			158.5	HLA-DPA1*03:01/DPB1*04:02
			12	HLA-DPB1*03:01/DPB1*04:01
			189.9	HLA-DQA1*01:02/DQB1*06:02
			96.5	HLA-DQA1*05:01/DQB1*02:01
			148.9	HLA-DRB1*01:01
			135.4	HLA-DRB1*04:01
196	204	FIAMVVQRI	129.1	HLA-DPA1*01/DPB1*04:01
			192.3	HLA-DPA1*01:03/DPB1*02:01
			191.3	HLA-DPA1*02:01/DPB1*01:01
			29.2	HLA-DPA1*03:01/DPB1*04:02
			26.4	HLA-DPB1*03:01/DPB1*04:01
			200.7	HLA-DQA1*05:01/DQB1*02:01
			5.4	HLA-DRB1*01:01
			26.7	HLA-DRB1*04:01
			44.6	HLA-DRB1*04:05
			12.7	HLA-DRB1*07:01
			11.4	HLA-DRB1*09:01
			27.9	HLA-DRB1*11:01
			106.9	HLA-DRB1*13:02
			23.2	HLA-DRB1*15:01
			23.5	HLA-DRB4*01:01
283	291	FLFLSYLYP	14.7	HLA-DPA1*01/DPB1*04:01
			6.7	HLA-DPA1*01:03/DPB1*02:01
			29	HLA-DPA1*03:01/DPB1*04:02
			8.6	HLA-DPB1*03:01/DPB1*04:01
			154.3	HLA-DRB1*01:01
			23.4	HLA-DRB1*04:04
159	167	FAVSKAFLD	128.4	HLA-DPA1*01:03/DPB1*02:01

Contd...

Supplementary material 5: Contd...

Start	End	Core	IC50	Allele
			191.3	HLA-DPA1*02:01/DPB1*01:01
			122.4	HLA-DPA1*02:01/DPB1*05:01
			227.9	HLA-DQA1*01:02/DQB1*06:02
			108.1	HLA-DQA1*05:01/DQB1*02:01
			11.5	HLA-DRB1*01:01
			26.2	HLA-DRB1*04:05
			33	HLA-DRB1*07:01
			98.8	HLA-DRB1*09:01
			192.1	HLA-DRB1*11:01
			17.9	HLA-DRB5*01:01
96	104	LVGISSVLK	181	HLA-DPA1*02:01/DPB1*01:01
			238.8	HLA-DPA1*03:01/DPB1*04:02
			243.2	HLA-DRB1*01:01
			87.9	HLA-DRB1*07:01
			59.9	HLA-DRB5*01:01
111	119	PVFSILSLV	55.8	HLA-DPA1*02:01/DPB1*01:01
			68.5	HLA-DPA1*03:01/DPB1*04:02
			130.6	HLA-DPB1*03:01/DPB1*04:01
			200.6	HLA-DRB1*04:04
158	166	EFAVSKAFL	246.5	HLA-DPA1*02:01/DPB1*01:01
			174.8	HLA-DQA1*01:02/DQB1*06:02
224	232	LFTDLCSIR	117	HLA-DPA1*02:01/DPB1*01:01
			36.4	HLA-DPA1*03:01/DPB1*04:02
			21.4	HLA-DRB1*03:01
228	236	LCSIRDLIL	184.6	HLA-DPA1*02:01/DPB1*01:01
			85.2	HLA-DRB1*01:01
			16.6	HLA-DRB1*07:01
			238.9	HLA-DRB4*01:01
231	239	IRDLILLDL	74.6	HLA-DPA1*02:01/DPB1*01:01
			29.7	HLA-DPA1*03:01/DPB1*04:02
			217.5	HLA-DRB1*01:01
			20.9	HLA-DRB4*01:01
234	242	LILLDLHQL	32.8	HLA-DPA1*02:01/DPB1*01:01
			11.3	HLA-DPA1*03:01/DPB1*04:02
			32.9	HLA-DRB1*01:01
			81.1	HLA-DRB1*04:05
			163.8	HLA-DRB1*15:01
102	110	VLKDFAKFS	77.8	HLA-DPA1*02:01/DPB1*05:01
			212.4	HLA-DRB1*01:01
			38.6	HLA-DRB1*03:01
			187.1	HLA-DRB1*11:01
			68.7	HLA-DRB3*01:01
106	114	FAKFSPVFS	137.9	HLA-DPA1*02:01/DPB1*05:01
			98.5	HLA-DQA1*05:01/DQB1*02:01
			13.5	HLA-DRB1*01:01
			173.6	HLA-DRB1*04:04
			156.1	HLA-DRB1*07:01
			143.2	HLA-DRB1*09:01
			81.5	HLA-DRB1*11:01
			177.1	HLA-DRB5*01:01
157	165	REFAVSKAF	114.8	HLA-DPA1*02:01/DPB1*05:01
			226.2	HLA-DRB1*07:01
112	120	VFSILSLVV	100.9	HLA-DPA1*03:01/DPB1*04:02
			154.5	HLA-DRB1*11:01
			192.6	HLA-DRB1*15:01

Contd...

Supplementary material 5: Contd...

Start	End	Core	IC50	Allele
198	206	AMVVQRIKY	36	HLA-DPA1*03:01/DPB1*04:02
235	243	ILLDLHQLI	204.6	HLA-DPA1*03:01/DPB1*04:02
			90.3	HLA-DPB1*03:01/DPB1*04:01
			7.9	HLA-DRB1*03:01
			173.6	HLA-DRB1*11:01
			83.2	HLA-DRB1*13:02
			157.6	HLA-DRB1*15:01
			7.1	HLA-DRB3*01:01
			21.3	HLA-DRB4*01:01
221	229	MLELFTDLC	57.8	HLA-DQA1*01:01/DQB1*05:01
90	98	SGCLDILVG	239.1	HLA-DQA1*01:02/DQB1*06:02
97	105	VGISSVLKD	219.2	HLA-DQA1*01:02/DQB1*06:02
179	187	TEVSALAAN	63.7	HLA-DQA1*01:02/DQB1*06:02
			26.4	HLA-DQA1*05:01/DQB1*02:01
			133.6	HLA-DRB1*04:04
184	192	LAANIPVYQ	64.7	HLA-DQA1*01:02/DQB1*06:02
			218.1	HLA-DRB1*01:01
			110.4	HLA-DRB1*08:02
			232.8	HLA-DRB1*09:01
			135.9	HLA-DRB1*13:02
194	202	VRFIAMVVQ	81	HLA-DQA1*01:02/DQB1*06:02
			222.4	HLA-DRB4*01:01
66	74	VMGAIGSLG	11	HLA-DQA1*05:01/DQB1*02:01
			12	HLA-DRB1*01:01
95	103	ILVGISSVL	70	HLA-DQA1*05:01/DQB1*02:01
			21	HLA-DRB1*01:01
			176.7	HLA-DRB1*09:01
			110.1	HLA-DRB1*15:01
181	189	VSALAANIP	48.9	HLA-DQA1*05:01/DQB1*02:01
			21.8	HLA-DRB1*01:01
			134	HLA-DRB1*04:04
183	191	ALAANIPVY	39.6	HLA-DQA1*05:01/DQB1*02:01
93	101	LDILVGISS	12.6	HLA-DRB1*01:01
			173.5	HLA-DRB1*04:04
			182.7	HLA-DRB1*11:01
109	117	FSPVFSILS	12.9	HLA-DRB1*01:01
			187.3	HLA-DRB1*04:01
			121.9	HLA-DRB1*04:05
			19.5	HLA-DRB1*07:01
			57.7	HLA-DRB1*11:01
			133.2	HLA-DRB5*01:01
113	121	FSILSLVVG	29.7	HLA-DRB1*01:01
			77.5	HLA-DRB1*04:04
155	163	VKREFAVSK	51.7	HLA-DRB1*01:01
			216.5	HLA-DRB1*04:01
			12.3	HLA-DRB5*01:01
193	201	GVRFIAMVV	146.2	HLA-DRB1*01:01
			85.6	HLA-DRB1*15:01
199	207	MVVQRIKYI	152.1	HLA-DRB1*01:01
			165.8	HLA-DRB1*08:02
			38.5	HLA-DRB1*15:01
225	233	FTDLCSIRD	22.4	HLA-DRB1*01:01
			60.9	HLA-DRB1*04:01
			135.4	HLA-DRB1*04:04
			56.5	HLA-DRB1*04:05

Contd...

Supplementary material 5: Contd...

Start	End	Core	IC50	Allele
239	247	LHQLIATPG	5.7	HLA-DRB1*01:01
			72.6	HLA-DRB1*04:01
			57.6	HLA-DRB1*04:04
			28.8	HLA-DRB1*04:05
			145.7	HLA-DRB1*07:01
			192.6	HLA-DRB1*09:01
			169.9	HLA-DRB1*11:01
			32.3	HLA-DRB4*01:01
230	238	SIRDLILLD	44.8	HLA-DRB1*03:01
			79.6	HLA-DRB1*04:05
67	75	MGAIGSLGT	122	HLA-DRB1*04:04
114	122	SILSLVVGL	63.6	HLA-DRB1*04:04
195	203	RFIAMVVQR	74.4	HLA-DRB1*04:04
237	245	LDLHQLIAT	197.6	HLA-DRB1*04:04
107	115	AKFSPVFSI	30.8	HLA-DRB1*07:01
180	188	EVSALAANI	241.9	HLA-DRB1*07:01
182	190	SALAANIPV	72.3	HLA-DRB1*07:01
			86.1	HLA-DRB1*13:02
99	107	ISSVLKDFA	28.2	HLA-DRB1*11:01
			227.6	HLA-DRB4*01:01
152	160	LYGVKREFA	53.7	HLA-DRB1*11:01
200	208	VVQRIKYIK	57.1	HLA-DRB1*11:01
			15.3	HLA-DRB5*01:01
233	241	DLILLDLHQ	208.4	HLA-DRB1*11:01
105	113	DFAKFSPVF	59.5	HLA-DRB1*15:01
188	196	IPVYQGVRF	30.5	HLA-DRB1*15:01
197	205	IAMVVQRIK	15.5	HLA-DRB4*01:01
			13.4	HLA-DRB5*01:01
232	240	RDLILLDLH	24.2	HLA-DRB4*01:01

Supplementary material 6: HSP70-1 MHC II allele interaction

Start	End	Core	IC50	Allele
282	290	IEIDSLFEG	85	HLA-DPA1*01:03/DPB1*02:01
283	291	EIDSLFEGI	130.5	HLA-DPA1*01:03/DPB1*02:01
205	213	FDVSILTIE	233	HLA-DPA1*02:01/DPB1*01:01
			71.1	HLA-DQA1*03:01/DQB1*03:02
			174.6	HLA-DRB1*04:05
280	288	ASIEIDSLF	171.7	HLA-DPA1*02:01/DPB1*01:01
			145.8	HLA-DQA1*03:01/DQB1*03:02
			53.1	HLA-DQA1*05:01/DQB1*02:01
21	29	FQHGKVEII	136.7	HLA-DPA1*03:01/DPB1*04:02
			28.3	HLA-DQA1*05:01/DQB1*03:01
			27.2	HLA-DRB1*01:01
			5.5	HLA-DRB1*07:01
			40	HLA-DRB1*09:01
119	127	ISSMVLTKM	173.4	HLA-DPA1*03:01/DPB1*04:02
			194	HLA-DQA1*01:02/DQB1*06:02
			177.3	HLA-DRB1*01:01
			77.8	HLA-DRB1*04:01
			122.1	HLA-DRB1*09:01
			186.7	HLA-DRB4*01:01
343	351	IPKIQQLLS	208.1	HLA-DPA1*03:01/DPB1*04:02
			205.3	HLA-DRB1*04:04
			105.4	HLA-DRB1*08:02

Contd...

Supplementary material 6: Contd...

Start	End	Core	IC50	Allele
			7.7	HLA-DRB4*01:01
120	128	SSMVLTKMK	126.9	HLA-DQA1*01:02/DQB1*06:02
			119.7	HLA-DRB1*11:01
			69.7	HLA-DRB5*01:01
137	145	KNCQNVVVT	141.2	HLA-DQA1*01:02/DQB1*06:02
138	146	NCQNVVVTV	135.8	HLA-DQA1*01:02/DQB1*06:02
140	148	QNVVVTVPA	198.5	HLA-DQA1*01:02/DQB1*06:02
277	285	SKQASIEID	144.3	HLA-DQA1*01:02/DQB1*06:02
278	286	KQASIEIDS	249.4	HLA-DQA1*01:02/DQB1*06:02
370	378	AYGAAVQAA	17.6	HLA-DQA1*01:02/DQB1*06:02
			96.2	HLA-DQA1*03:01/DQB1*03:02
			113.2	HLA-DQA1*04:01/DQB1*04:02
			3	HLA-DQA1*05:01/DQB1*03:01
178	186	AAAIAYGLD	129.1	HLA-DQA1*03:01/DQB1*03:02
368	376	AVAYGAAVQ	89.8	HLA-DQA1*03:01/DQB1*03:02
			16.7	HLA-DQA1*05:01/DQB1*03:01
176	184	PTAAAIAYG	169.6	HLA-DQA1*04:01/DQB1*04:02
371	379	YGAAVQAAI	249.7	HLA-DQA1*05:01/DQB1*02:01
			10.6	HLA-DQA1*05:01/DQB1*03:01
			14.7	HLA-DRB1*01:01
			235	HLA-DRB1*04:05
			108.5	HLA-DRB1*07:01
			10.5	HLA-DRB1*09:01
			189.5	HLA-DRB5*01:01
15	23	YSCVGVFQH	209.8	HLA-DQA1*05:01/DQB1*03:01
37	45	TTPSYVAFT	240.9	HLA-DQA1*05:01/DQB1*03:01
216	224	IFEVKATAG	218.4	HLA-DQA1*05:01/DQB1*03:01
			87.6	HLA-DRB1*01:01
			157.8	HLA-DRB1*07:01
366	374	DEAVAYGAA	152.8	HLA-DQA1*05:01/DQB1*03:01
367	375	EAVAYGAAV	48.7	HLA-DQA1*05:01/DQB1*03:01
369	377	VAYGAAVQA	3.5	HLA-DQA1*05:01/DQB1*03:01
			13.5	HLA-DRB1*01:01
			185.5	HLA-DRB1*08:02
			106.8	HLA-DRB1*09:01
372	380	GAAVQAAIL	55.5	HLA-DQA1*05:01/DQB1*03:01
373	381	AAVQAAILQ	108.4	HLA-DQA1*05:01/DQB1*03:01
374	382	AVQAAILQG	203	HLA-DQA1*05:01/DQB1*03:01
65	73	NTVFDAKRL	107.5	HLA-DRB1*01:01
260	268	VRRLRTACE	21.4	HLA-DRB1*01:01
			36.1	HLA-DRB1*04:01
			207.1	HLA-DRB1*04:04
			37.2	HLA-DRB1*04:05
			113.3	HLA-DRB1*08:02
			116.2	HLA-DRB5*01:01
66	74	TVFDAKRLI	211.6	HLA-DRB1*03:01
141	149	NVVVTVPAY	122.1	HLA-DRB1*04:04
344	352	PKIQQLLSD	54.9	HLA-DRB1*04:04
142	150	VVVTVPAYF	133.4	HLA-DRB1*08:02
			84.4	HLA-DRB1*09:01
			119	HLA-DRB1*13:02
			25.4	HLA-DRB1*15:01
67	75	VFDAKRLIG	28.1	HLA-DRB1*11:01
18	26	VGVFQHGKV	200	HLA-DRB1*15:01
17	25	CVGVFQHGK	37.9	HLA-DRB5*01:01
179	187	AAIAYGLDK	149.6	HLA-DRB5*01:01

Table 4: Binding site coordinates for protein-ligand docking between MHC molecules and peptides prepared by autodock tools

MHC molecule PDB ID	Axis	Center box	Size
1A1O	X	3.12	44
	Y	26.631	22
	Z	19.201	16
1DUZ	X	5.499	22
	Y	18.462	22
	Z	8.057	38
1JHT	X	20.706	26
	Y	37.098	36
	Z	72.438	20
1AQD	X	12.939	34
	Y	24.708	18
	Z	43.286	22
1DLH	X	4.301	40
	Y	74.656	14
	Z	19.422	22
1H15	X	95.81	22
	Y	-5.497	16
	Z	16.03	36

*PDB ID = Protein data bank identity, MHC = Major histocompatibility complex

Table 5: Docking simulation results prepared by autodock vina

Epitope sequence/ligand	MHC	Receptor PDB ID	Affinity (Kcal/mol)	Dist. from RMSD l.b.	Best mode RMSD u.b.
Toxin-1	MHC I	1A1O	−6.6	0.0	0.0
		1DUZ	−6.0	0.0	0.0
		1JHT	−6.4	0.0	0.0
	MHC II	1AQD	−6.1	0.0	0.0
		1DLH	−6.1	0.0	0.0
		1H15	−6.6	0.0	0.0
Toxin-2	MHC I	1A1O	−6.8	0.0	0.0
		1DUZ	−6.2	0.0	0.0
		1JHT	−6.5	0.0	0.0
	MHC II	1AQD	−6.2	0.0	0.0
		1DLH	−5.2	0.0	0.0
		1H15	−6.2	0.0	0.0
HSP70-1	MHC I	1A1O	−7.0	0.0	0.0
		1DUZ	−5.7	0.0	0.0
		1JHT	−6.4	0.0	0.0
	MHC II	1AQD	−6.4	0.0	0.0
		1DLH	−5.3	0.0	0.0
		1H15	−6.7	0.0	0.0

*PDB ID = Protein data bank identity, MHC = Major histocompatibility complex, RMSD = Root mean square deviation

The FAO and WHO evaluation scheme is a guideline by the FAO and WHO for sequence-based alelrgenicity prediction. This guideline clearly states that a sequence can be a potentially allergenic if it either has an approximated identity of at least 6 contiguous amino acids or >35% sequence identity over a window of 80 amino acid chains when compared to known allergens.[27] So, if a vaccine

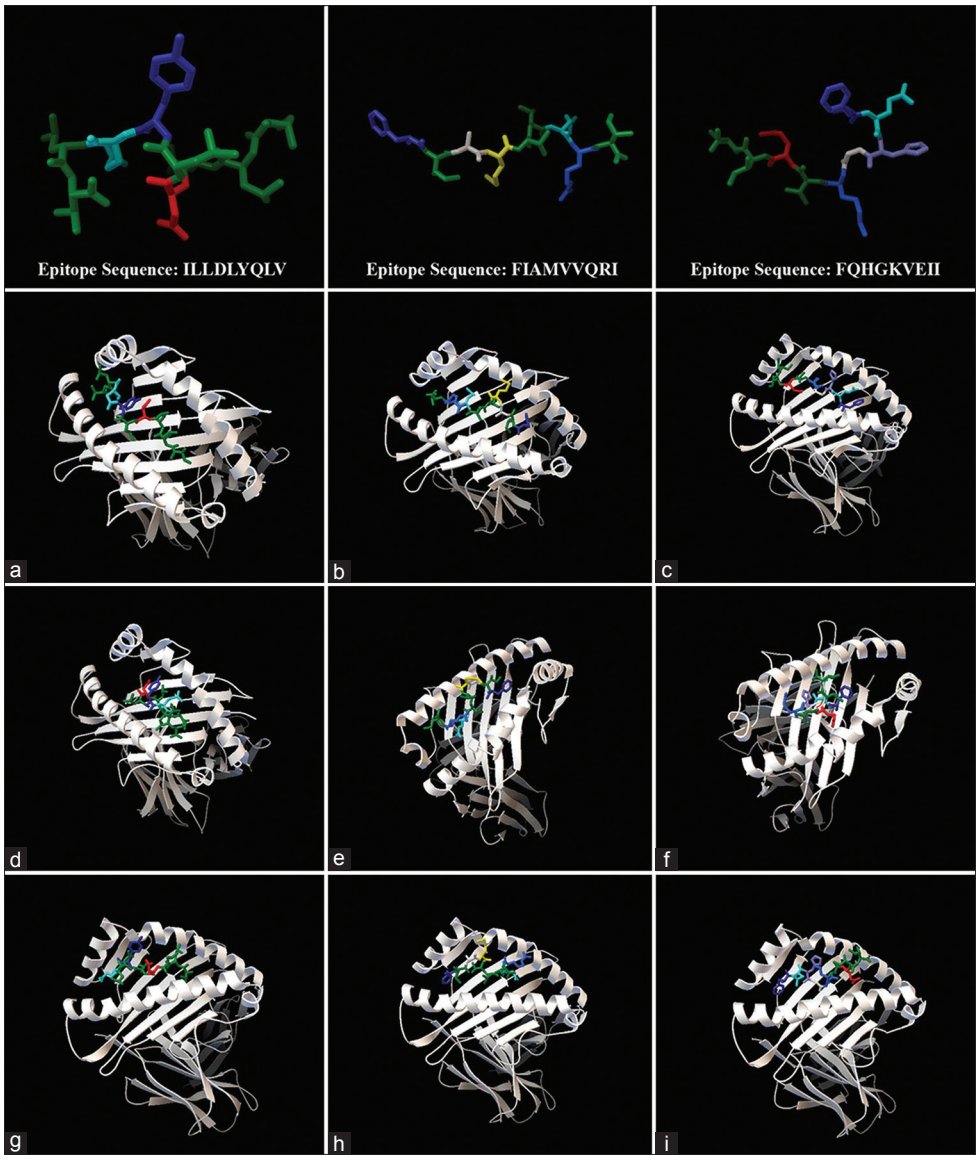

Figure 3: Visualization of best docking results for predicted peptides with MHC class I receptors by using Autodock Tools. (a-c) represents docking images with 1A1O, (d-f) represents docking images with 1DUZ, (g-i) represents docking images with 1JHT

was developed by using the venom peptides, it will not create allergic reactions.

MHC class II molecules are highly polymorphic in nature, and this polymorphism exclusively corresponds with a few differences along the peptide-binding groove in antigenic fragments.[28] The binding between antigenic peptides (epitopes) and the MHC molecule is a crucial step in the cellular immune response. For the prediction of MHC binding molecules in both cases (MHC I and II) artificial neural network method was used. For T cell class I epitope prediction, the neural network method was designed by combining sparse encoding, blosum encoding, and input derived from hidden markov models.[21] In this study, for MHC class II peptide prediction, we used artificial neural network-based method NN-align which was evaluated by 14 human MHC class II alleles.[23]

In this study, we tried to minimize the predicted promiscuous epitopes and pinpoint the efficient epitope sequences that have the greatest chance for eliciting cell-mediated immunity in human body against box jellyfish venom. As it is a concern that the prediction-based epitope design might not work in reality, we docked the predicted MHC peptides with HLA molecules to find out whether or not the vaccine designed by using the predicted epitopes will elicit sufficient immunological response *in vivo*. Lower energy scores represent better binding between receptor and ligand.[29]

More interestingly, sequence similarity search between toxin-1 and toxin-2 by using CLUSTALW multiple sequence alignment web server [Figure 5] showed that the epitope that was predicted for toxin-2 (FIAMVVQRI) would also elicit immune response against toxin-1. Again,

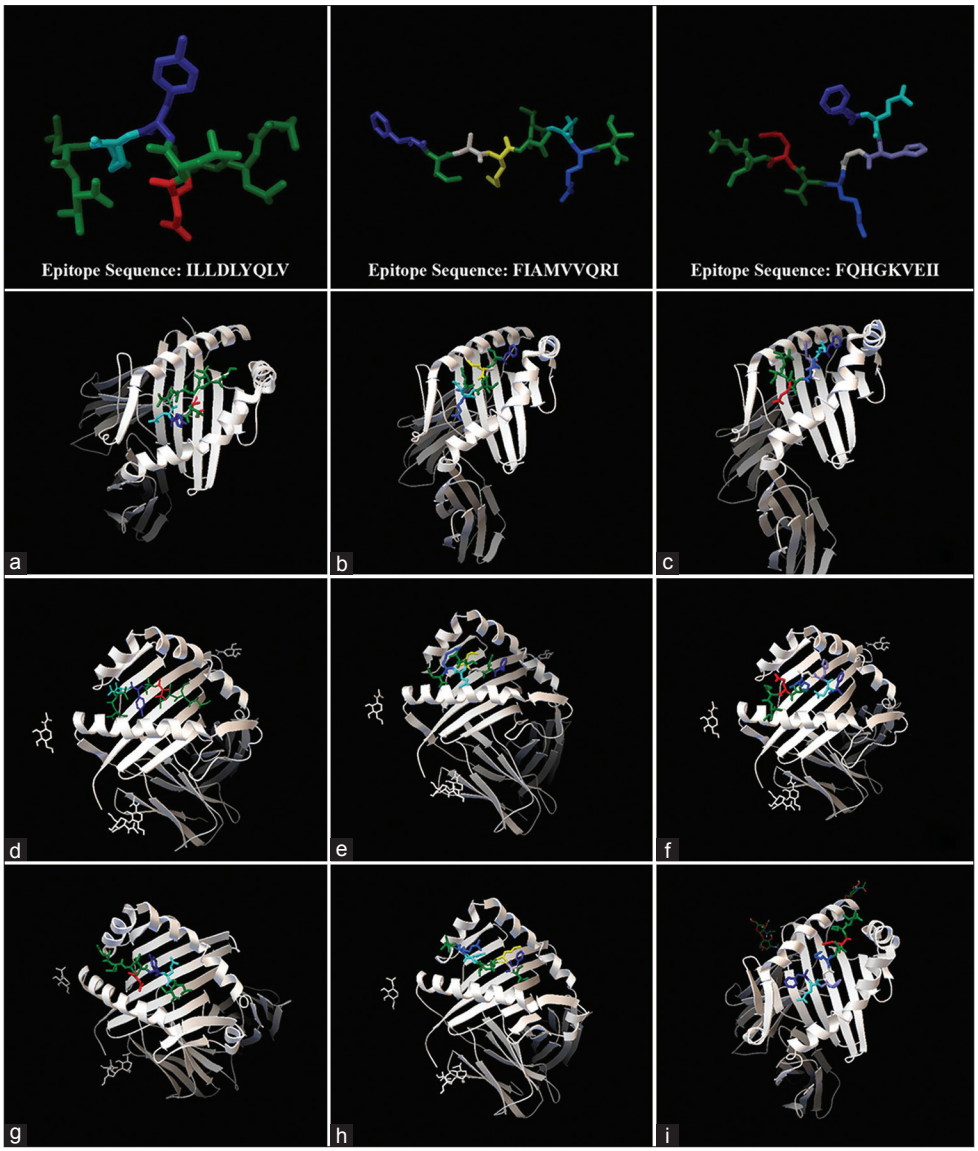

Figure 4: Visualization of best docking results for predicted peptides with MHC class II receptors by using Autodock tools. (a-c) represents docking images with 1AQD (d-f) represents docking images with 1DLH, (g-i) represents docking images with 1H15

we analyzed the B-cell epitope prediction method by Kolaskar and Tangaonkar method [Table 6] and found out that the selected epitopes fall inside the sequences predicted for B-cell immunity. So, we summarize that designing of a trivalent vaccine by using these three epitopes may elicit both humoral and cell-mediated immunity against box jellyfish venom.

CONCLUSION

All of these computational approaches demonstrate the importance of *C. fleckeri* venom proteins as valuable immunodiagnostic tool for initial research methodologies with a view to future disease diagnosis and drug design against this fatal venom. The findings of this study yet need to be validated in future by experimental procedures; however, the given information and approaches in this study will be more blissful for researchers to investigate novel human therapeutics-like design of subunit and synthetic peptide vaccine from this world's most venomous marine creature *C. fleckri*. This superficial concept can be implemented to design synthetic and subunit peptide vaccine against these lethal venom proteins that may save thousand lives especially in Australia where it poses a major problem.

ACKNOWLEDGMENT

Authors are thankful to Mr. Najmul Islam Chowdhury (Southern University, Bangladesh) for his technical support to run the docking simulations.

Figure 5: Multiple sequence alignment between toxin-1 and toxin-2 by using CLUSTALW web server [* (star mark) designates similarity between two proteins]

Table 6: B cell epitope regions of *Chironex fleckeri* venom proteins predicted by Kolaskar and Tangaonkar antigenecity scale

Venom	Start	End	Peptide	Peptide length
Toxin-1	224	250	MLELFTDLCSLRDLILLDLYQLVATPG	17
Toxin-2	178	208	PTEVSALAANIPVYQGVRFIAMVVQRIKYIK	31
HSP70-1	14	29	TYSCVGVFQHGKVEII	16

REFERENCES

1. Marine-medic.com. Chironex fleckeri-The north Australian box-jellyfish (2000). Available from: http://www.marine-medic.com.au/pages/biology/biologyBreakup/jellyfishChironex.pdf [Last accessed on 2013 Jul 22].
2. Brinkman DL. The molecular and biochemical characterisation of venom proteins from the box jellyfish, Chironex fleckeri. [PhD thesis], Queensland, Australia: James Cook Univ; 2008. [Available Online at: http://eprints.jcu.edu.au/5682/].
3. Beadnell CE, Rider TA, Williamson JA, Fenner PJ. Management of a major box jellyfish (Chironex fleckeri) sting. Lessons from the first minutes and hours. Med J Aust 1992;156:655-8.
4. Northern Territory Government of Australia. Centre for Disease Control-Chironex fleckeri (Box jellyfish) (2012). Available from: http://www.health.nt.gov.au/library/scripts/objectifyMedia.aspx?file=pdf/26/02.pdf and siteID=1 and str_title=Box%20Jellyfish.pdf [Last accessed on 2013 Jul 22].
5. Brinkman DL, Burnell JN. Biochemical and molecular characterisation of cubozoan protein toxins. Toxicon 2009;54:1162-73.
6. Brinkman DL, Aziz A, Loukas A, Potriquet J, Seymour J, Mulvenna J. Venom proteome of the box jellyfish Chironex fleckeri. PLoS One 2012;7:e47866.
7. Tsan MF, Gao B. Heat shock proteins and immune system. J Leukoc Biol 2009;85:905-10.
8. Pagetta A, Tramentozzi E, Frigo G, Finotti P. Heat shock protein-derived peptides as a potential immuno-regulatory vaccine in allergy. "In silico" analysis. Immunome Res 2012;8:27.
9. Wagstaff SC, Laing GD, Theakston RD, Papaspyridis C, Harrison RA. Bioinformatics and multiepitope DNA immunization to design rational snake antivenom. PLoS Med 2006;3:e184.
10. Kolaskar AS, Tongaonkar PC. A semi-empirical method for prediction of antigenic determinants on protein antigens. FEBS Lett 1990;276:172-4.
11. Hopp TP, Woods KR. Prediction of protein antigenic determinants from amino acid sequences. Proc Natl Acad Sci U S A 1981;78:3824-8.
12. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, *et al*. Protein identification and analysis tools on the ExPASy Server. In: Walker JM, editor. The Proteomics Protocols Handbook. New York: Humana Press; 2005. p. 571-607.
13. Garnier J, Osguthorpe DJ, Robson B. Analysis of the accuracy and

implications of simple methods for predicting the secondary structure of globular proteins. J Mol Biol 1978;120:97-120.
14. Robson B, Garnier J. Protein structure prediction. Nature 1993;361:506.
15. King RD, Sternberg MJ. Identification and application of the concepts important for accurate and reliable protein secondary structure prediction. Protein Sci 1996;5:2298-310.
16. Wolfenden R, Andersson L, Cullis PM, Southgate CC. Affinities of aminoacid side-chains for solvent water. Biochemistry 1981;20:849-55.
17. Eisenberg D, Schwarz E, Komaromy M, Wall R. Analysis of membrane and surface protein sequences with the hydrophobic moment plot. J Mol Biol 1984;179:125-42.
18. Eisenberg D, Weiss RM, Terwilliger TC. The hydrophobic moment detects periodicity in protein hydrophobicity. Proc Natl Acad Sci U S A 1984;81:140-4.
19. Levitt M. Conformational preferences of amino acids in globular proteins. Biochemistry 1978;17:4277-85.
20. Trott O, Olson AJ. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem 2010;31:455-61.
21. Nielsen M, Lundegaard C, Worning P, Lauemøller SL, Lamberth K, Buus S, *et al*. Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. Protein Sci 2003;12:1007-17.
22. Lundegaard C, Lamberth K, Harndahl M, Buus S, Lund O, Nielsen M. 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 2008;36:W509-12.
23. Nielsen M, Lund O. NN-align. An artificial neural network-based alignment algorithm for MHC class II peptide binding prediction. BMC Bioinformatics 2009;10:296.
24. Reche PA, Glutting JP, Zhang H, Reinherz EL. Enhancement to the RANKPEP resource for the prediction of peptide binding to MHC molecules using profiles. Immunogenetics 2004;56:405-19.
25. Rini JM, Schulze-Gahmen U, Wilson IA. Structural evidence for induced fit as a mechanism for antibody-antigen recognition. Science 1992;255:959-65.
26. Hinds MG, Welsh JH, Brennand DM, Fisher J, Glennie MJ, Richards NG, *et al*. Synthesis, conformational properties, and antibody recognition of peptides containing beta-turn mimetics based on alpha-alkylproline derivatives. J Med Chem 1991;34:1777-89.
27. Muh HC, Tong JC, Tammi MT. AllerHunter: A SVM-Pairwise system for assessment of allergenicity and allergic cross-reactivity in proteins. PLoS One 2009;4:e5861.
28. Consogno G, Manici S, Facchinetti V, Bachi A, Hammer J, Conti-Fine BM, *et al*. Identification of immunodominant regions among promiscuous HLA-DR-restricted CD4+T-cell epitopes on the tumor antigen MAGE-3. Blood 2003;101:1038-44.
29. Alam MJ, Ashraf KU, Gupta SD, Emon MA. Computational approach for the prediction of potential MHC binding peptides and epitope mapping in order to develop sero-diagnostic immunogen against potato virus Y. Int J Comput Bioinfo Silico Model 2013;2:186-98.

How to cite this article: Alam M, Muhammad Ashraf KU. Prediction of an epitope-based computational vaccine strategy for gaining concurrent immunization against the venom proteins of Australian Box Jellyfish. Toxicol Int 2013;20:235-53.
Source of Support: Nil. **Conflict of Interest:** None declared.