Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Human Immunology 80 (2019) 493–502

Contents lists available at ScienceDirect

Human Immunology

journal homepage: www.elsevier.com/locate/humimm

A multiepitopic theoretical fusion construct based on *in-silico* epitope screening of known vaccine candidates for protection against wide range of enterobacterial pathogens

Ashutosh Kumar, Kusum Harjai, Sanjay Chhibber*

Department of Microbiology, Panjab University, Chandigarh, India

ARTICLE INFO

Keywords:
In-silico
Epitope prediction
FyuA
Flagellin
IEDB

ABSTRACT

Enterobacterial pathogens that have acquired antibiotic resistance genes are a leading cause of community and hospital acquired infections. In such a situation vaccination is considered as a better option to prevent such infections. In the current study reverse vaccinology approach has been used to select peptides from already known immunogenic proteins to design a chimeric construct. We selected Yersiniabactin receptor of Escherichia coli UMN026 and Flagellin of Stenotrophomonas maltophila. B-cell linear epitopes were predicted using Bepipred prediction tool. Peptide binding with reference sets of 27 alleles of MHC class I and class II was also analyzed. The predicted peptides-MHC complexes were further validated using simulation dynamics. The in-silico construction of chimera was done by restriction mapping and codon optimization. Chimera was evaluated using the immunoinformatic approach as done for the selected proteins. From the 673 amino acids of FyuA protein, a region from 1 to 492 was selected for containing more linear epitopes and the processing scores obtained were significant for MHC class I and class II binding. Similarly, from Flagellin, a region between 60 and 328 amino acids was selected and the peptides present in the selected region showed lower percentile ranks for binding with MHC molecules. The simulation studies validated the predictions of peptide-MHC complexes. The selected gene fragments accommodating maximum part of these peptides were used to design a chimaeric construct of 2454 bp. From the immunoinformatic analysis, the chimera was found to be more immunogenic in terms of increased number of B-cell and T-cell epitopes along with increased coverage of global populations with allelic variability.

1. Introduction

Communicable diseases caused by members of *Enterobacteriaceae* family put a great burden on the society by affecting humans and their livestocks. These infections become quite severe when they are not controlled on time. Some of the examples include pneumonia, pyogenic liver abscess, pyelonephritis and septicemia [1]. The treatment has become difficult due to the emergence of antibiotic resistance among some of the pathogens. Due to the existing challenge of treating these infections with antibiotics, it is inevitable for the research community to look forward to prophylactic means for the prevention of these infections. A large number of vaccine candidates have been proposed by various researchers for specific infections. However, evaluation of individual vaccine candidates under *in-vivo* infection conditions is an enduring task. In the recent years, “Reverse Vaccinology” (RV) has come to play an important role in scrutinizing the vaccine candidates by *in-silico* analysis [2] thereby reducing the time required for ruling out ineffective candidates. This immunoinformatic approach is being frequently used by researchers to predict the epitopes on viruses. This has led to the finding of epitopes on nucleocapsid protein and ovarian tumour domain of Crimean–Congo hemorrhagic fever virus [3]. Another study investigated the variability among epitopes of Hepatitis C virus (HCV) identified in genotype 1 and also predicted the immunogenicity of their variants from other genotypes against South African human leukocyte antigen (HLA) backgrounds [4]. Epitopes of E1 protein isolated from HCV have also been identified using the similar approach

Abbreviations: MHC, Major histocompatibility complex; RV, reverse vaccinology; HLA, human leukocyte antigen; HCV, hepatitis c virus; MERS-CoV, middle east respirator syndrome coronavirus; OMPs, outer membrane proteins; IEDB, immune epitope database and analysis resource; NCBI, national center for biotechnology information; NEB, new england biolabs; DNA, deoxyribonucleic acid; PDB, protein data bank; USA, United States of America; BLASTp, basic local alignment tool for proteins

* Corresponding author.
E-mail address: sanjaychhibber@pu.ac.in (S. Chhibber).

https://doi.org/10.1016/j.humimm.2019.02.008
Received 29 October 2018; Received in revised form 23 January 2019; Accepted 11 February 2019
Available online 12 February 2019

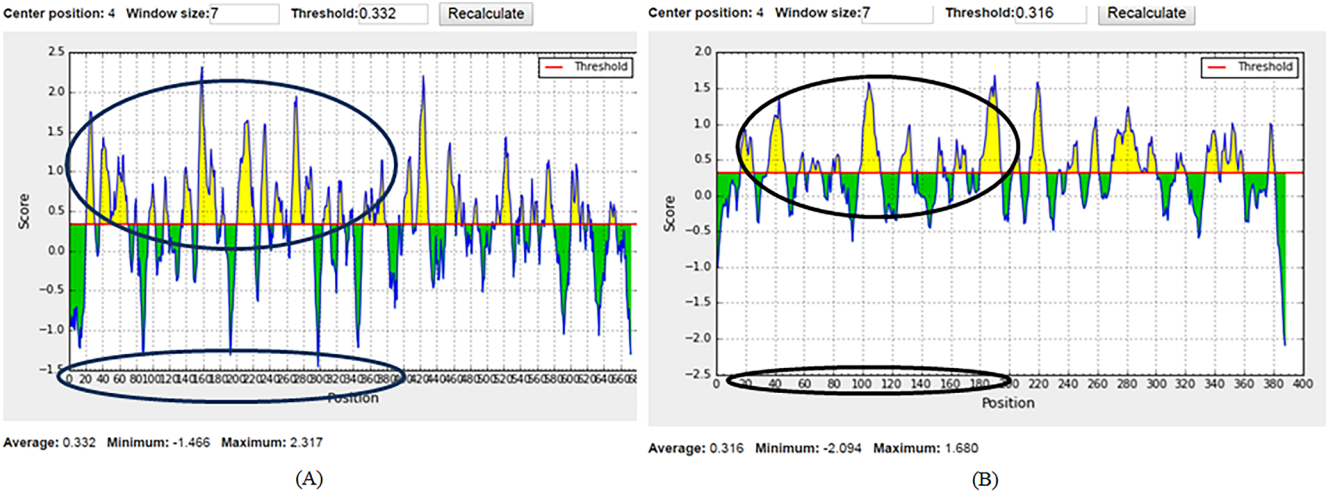

Fig 1. Linear epitope prediction from the protein sequences FyuA (A) and Flagellin (B) using Bepipred prediction portal of IEDB server. The yellow peaks show the peptide sequences that are potential epitopes whereas the green peaks show the peptides that are not epitopic in nature. The encircled area on the graphs shows the region of protein having higher frequency of epitopic peptides.

Table 1
. Significant linear peptides predicted from FyuA and Flagellin protein sequence.

Peptides predicted from FyuA protein sequence				
Peak. No	Start	End	Peptide	Length
1	21	30	NAQTSQQDES	10
2	37	70	SKQSSRSASANNVSSTVVSAPELSDAGVTASDKL	34
7	137	147	QGTLYGKSAQG	11
8	154	176	TQQPDSTPRGYIEGGVSSRDSYR	23
10	201	222	DDGDMINPATGSDDLGGTRASI	22
11	229	238	LAPDDQPWEM	10
12	245	255	ECTRATQDAYV	11
14	266	291	SISDGSPDPYMRRCTDSQTLSGKYTT	26
15	304	312	QQHYSRTFP	9
19	352	364	NTREKLNSAYDMP	13
20	366	379	MPYLSSTGYTTAET	14
23	401	411	RFSHDKSSTQY	11
24	417	431	GNPFGDQGKSNDDQV	15
25	452	463	AQGYKPSGYNIV	12
31	512	530	PVGMQTLSNAGKADATGVE	19
38	571	581	RYGAGSSVNGV	11
39	602	612	YFDGDNQLRQG	11
Peptides predicted from Flagellin protein sequence				
Peak. No	Start	End	Peptide	Length
1	17	25	NLSTSGSSL	9
2	35	47	GSRINSAKDDAAG	13
4	65	71	IRNANDG	7
7	98	112	QASNATNSASDRKAL	15
8	126	133	VAKQSDFN	8
12	163	170	IDAKANAL	8
14	180	194	IAAPTLTGTPATTAD	15
16	214	225	VENTDAAGAGKA	12
18	240	246	LAEADAT	7
20	270	289	TGGVTVAGNTTVPARQYADK	20
24	334	342	QTSSENLSA	9
25	345	355	SRIKDTDFAKE	11

[5]. Effective immunogens of MERS-CoV have been discovered through immunoinformatics-driven genome-wide screening strategy [6]. Goodswen et al. [7] have also used this technique for designing protein based vaccines against eukaryotic pathogens. Reverse vaccinology has already been used against other bacterial pathogens like Group B *streptococcus*, where genomic analysis has led to the development of a vaccine composed of four proteins giving protection against all serotypes [8]. Another *In-silico* study has found a protein Bam A of *Acinetobacter baumanii* to be a potential immunogen [9]. RV has also been implied to predict the potential vaccine candidates from the proteome of *Burkholderia pseudomallei* [10]. The outer membrane proteins (OMPs) of these Gram-negative bacteria are usually considered as potent vaccine candidates as they are exposed to the host immune defenses [11]. These OMPs are not always conserved in different genus of bacteria but their lies a probability of presence of some conserved peptide sequence in these OMPs. In this study we have taken into account the proteins which have proved to be potential vaccine candidates on the basis of *in-vivo* research work on animal models. Yersiniabactin receptor FyuA is highly conserved protein prevalent in various members of *Enterobacteriaceae*. As per the reports, FyuA mediates the uptake of ferric-yersiniabactin [12,13] confirming its role in the virulence of bacteria, which makes it an important vaccine candidate. Moreover researchers have found it as a potential vaccine candidate against pyelonephritis in a murine model of urinary tract infection [14]. It has also been found to be protective in murine model of pneumonia caused by *K. pneumoniae* 43816 in our laboratory (manuscript under communication). The Flagellin protein is another potential vaccine candidate which has been included in this study. Various studies have established the role of flagellin in inducing a systemic inflammatory response via intraperitoneal and intravenous administration [15,16] and a local inflammatory response with intraintestinal administration [17,18]. In our lab, flagellin of *Stenotrophomonas maltophilia* has been shown to induce non-specific immune response which protected mice against subsequent bacterial challenge [19]. In the current study both of these vaccine candidates have been analyzed using IEDB server to design a novel *in-silico* vaccine construct harbouring the properties of both these proteins.

2. Methodology

2.1. Protein sequence retrieval

Gene sequence of FyuA accession no. NC_011751.1 and Flagellin NC_010943.1 were taken from NCBI database [20]. The obtained gene sequences were translated into the protein sequences using Expasy-Translate tool (Swiss Institute of Bioinformatics).

2.2. B-cell linear epitope prediction

Protein sequences of Yersiniabactin receptor of *Escherichia coli* UMN026 and Flagellin of *Stenotrophomonas maltophilia* were analyzed for the presence of linear epitopes using Bepipred portal of IEDB server. Both proteins were also analyzed on other algorithms, Parker

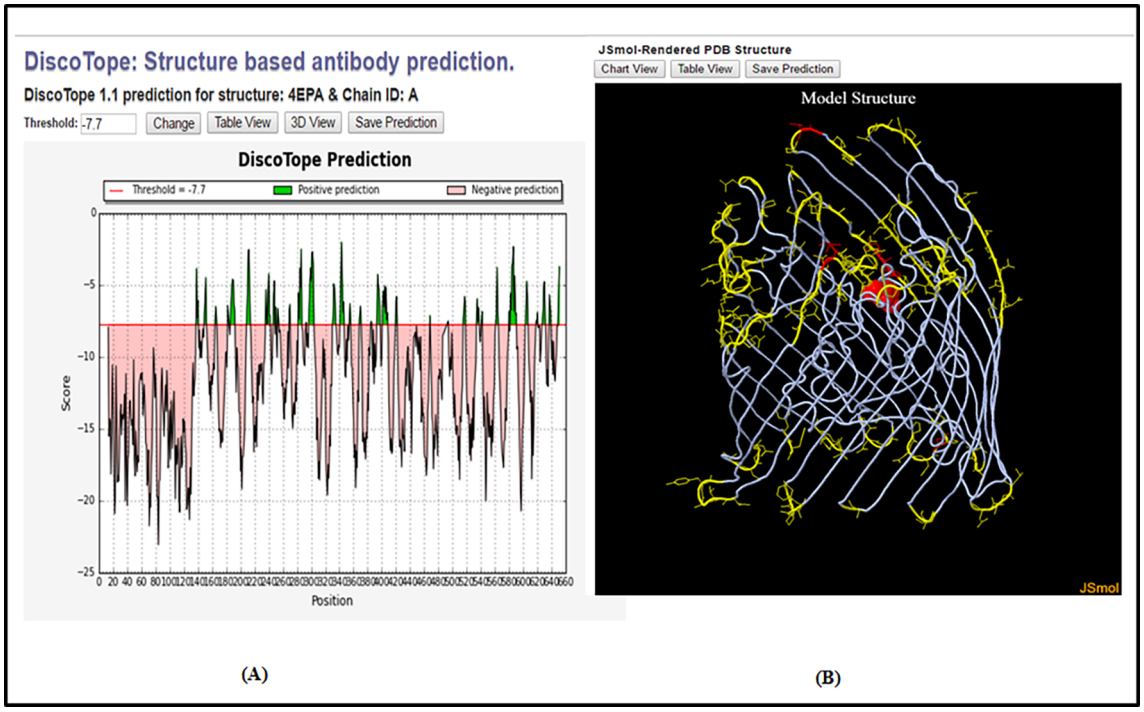

Fig 2. Presence of discontinuous or conformational epitopes for FyuA protein using Discotope tool on IEDB server. (A) The result window shows the presence of discontinuous epitopes in the form of green peaks. (B) The 3D structure shows the presence of conformational epitopes in yellow colour on various parts of protein. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

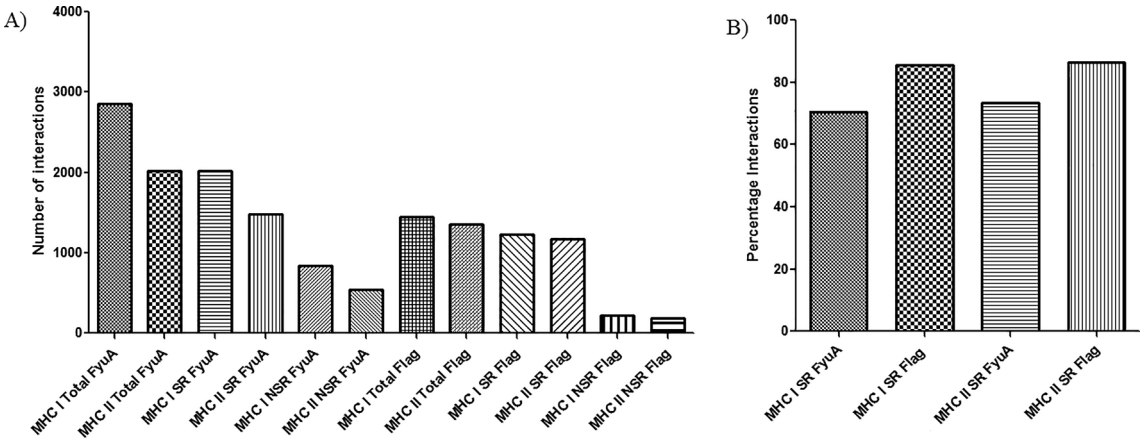

Fig 3. Comparative number (A) and percentage (B) of peptide interactions from the selected (SR) and non-selected regions (NSR) of FyuA and Flagellin respectively.

hydrophilicity prediction, Emini Surface Accessibility Prediction [21]. Antigenicity prediction was done according to Kolaskar and Tongaonkar scale. Prediction of Beta-Turns was done by Chou and Fasman scale on IEDB server.

2.3. Conformational B-cell epitope prediction

Conformational epitopes on yersiniabactin receptor were predicted using Discotope portal of IEDB server by analyzing solvent-accessibility.

2.4. MHC class I and II binding peptide prediction

In human population Major Histocompatibility complexes (MHC-I and MHC-II) are encoded by human leukocyte antigen (HLA) alleles and are required for the presentation of antigen to T cells. The peptides which could bind to MHC I molecules were predicted by MHC class I binding peptide prediction portal on IEDB server using the consensus method. Interactions were evaluated in terms of percentile ranks. Similarly, peptides binding to MHC II molecules were predicted by MHC class II binding peptide prediction portal on IEDB server using the consensus method.

2.5. Docking

3D structures of alleles were retrieved from RCSB PDB database [22]. Predicted peptide sequences and 3D structure of MHC class I and Class II alleles were submitted to CABS Dock server for docking and simulation studies [23]. Secondary structures of peptides were generated from PSIPRED [24]. The simulation time was set to 50 cycles. The results were clustered according to the distance between the residues of the peptide and MHC molecules.

Table 2
. Predicted MHC I binding peptides from the selected region of FyuA and Flagellin.

Peptides from FyuA						
Allele	start	end	length	peptide	Method	Percentile rank
HLA-A*30:02	340	349	10	RTVDMVFGLY	Consensus (ann/smm)	0.06
HLA-B*35:01	366	374	9	MPYLSSTGY	Consensus (ann/comblib_sidney2008/smm)	0.1
HLA-A*26:01	375	383	9	TTAETLAAY	Consensus (ann/smm)	0.11
HLA-A*68:01	85	93	9	MLFSTISLR	Consensus (ann/smm)	0.11
HLA-A*01:01	267	275	9	ISDGSPDPY	Consensus (ann/smm)	0.12
HLA-A*01:01	341	349	9	TVDMVFGLY	Consensus (ann/smm)	0.12
HLA-A*30:02	446	455	10	RVYTRVAQGY	Consensus (ann/smm)	0.13
HLA-A*68:01	341	350	10	TVDMVFGLYR	Consensus (ann/smm)	0.14
HLA-B*53:01	311	320	10	FPSGSLIVNM	Consensus (ann/smm)	0.14
HLA-B*51:01	70	79	10	LPRVLPGLNI	Consensus (ann/smm)	0.14
Peptides from Flagellin						
Allele	start	end	length	Peptide	Method	Percentile rank
HLA-A*02:06	112	120	9	LQAEVTQLV	Consensus (ann/smm)	0.11
HLA-A*68:02	373	382	10	ETAELTRTQI	Consensus (ann/smm)	0.14
HLA-A*68:01	44	53	10	DAAGLAISER	Consensus (ann/smm)	0.17
HLA-A*68:02	233	241	9	MVAAINAKI	Consensus (ann/comblib_sidney2008/smm)	0.2
HLA-B*44:03	114	123	10	AEVTQLVSEI	Consensus (ann/smm)	0.23
HLA-A*68:02	233	242	10	MVAAINAKIG	Consensus (ann/smm)	0.24
HLA-A*02:01	173	182	10	AMFDSVSFTI	Consensus (ann/smm)	0.25
HLA-A*30:01	91	99	9	RVRELSVQA	Consensus (ann/comblib_sidney2008/smm)	0.3
HLA-A*02:03	173	182	10	AMFDSVSFTI	Consensus (ann/smm)	0.3
HLA-B*44:02	375	383	9	AELTRTQIL	Consensus (ann/smm)	0.3

Table 3
. Predicted MHC II binding peptides from the selected region of FyuA and Flagellin.

Peptides from FyuA					
Allele	Start	End	Peptide	Method used	Percentile rank
HLA-DRB3*01:01	435	449	LSAGYMLTDDWRVYT	Consensus (comb.lib./smm/nn)	0.01
HLA-DRB3*01:01	436	450	SAGYMLTDDWRVYTR	Consensus (comb.lib./smm/nn)	0.01
HLA-DRB3*01:01	437	451	AGYMLTDDWRVYTRV	Consensus (comb.lib./smm/nn)	0.01
HLA-DRB3*01:01	438	452	GYMLTDDWRVYTRVA	Consensus (comb.lib./smm/nn)	0.01
HLA-DRB3*01:01	434	448	QLSAGYMLTDDWRVY	Consensus (comb.lib./smm/nn)	0.02
HLA-DRB3*01:01	439	453	YMLTDDWRVYTRVAQ	Consensus (comb.lib./smm/nn)	0.03
HLA-DRB1*09:01	1	15	MKMTRLYPLALGGLL	Consensus (comb.lib./smm/nn)	0.08
HLA-DRB1*08:02	13	27	GLLLPAIANAQTSQQ	Consensus (smm/nn/sturniolo)	0.08
HLA-DRB1*08:02	14	28	LLLPAIANAQTSQQD	Consensus (smm/nn/sturniolo)	0.1
HLA-DRB3*01:01	433	447	GQLSAGYMLTDDWRV	Consensus (comb.lib./smm/nn)	0.15
Peptides from Flagelin					
Allele	Start	End	Peptide	Method used	Percentile rank
HLA-DQA1*01:02/DQB1*06:02	380	394	TQILQQAGTAMLAQA	Consensus (comb.lib./smm/nn)	0.01
HLA-DQA1*01:02/DQB1*06:02	381	395	QILQQAGTAMLAQAN	Consensus (comb.lib./smm/nn)	0.01
HLA-DQA1*01:02/DQB1*06:02	382	396	ILQQAGTAMLAQANQ	Consensus (comb.lib./smm/nn)	0.01
HLA-DQA1*01:02/DQB1*06:02	383	397	LQQAGTAMLAQANQV	Consensus (comb.lib./smm/nn)	0.01
HLA-DQA1*01:02/DQB1*06:02	384	398	QQAGTAMLAQANQVP	Consensus (comb.lib./smm/nn)	0.01
HLA-DQA1*01:02/DQB1*06:02	379	393	RTQILQQAGTAMLAQ	Consensus (comb.lib./smm/nn)	0.03
HLA-DQA1*04:01/DQB1*04:02	239	253	AKIGETGVLAEVDAA	Consensus (comb.lib./smm/nn)	0.1
HLA-DRB4*01:01	377	391	LTRTQILQQAGTAML	Consensus (comb.lib./smm/nn)	0.19
HLA-DRB1*09:01	175	189	FDSVSFTIAAPADAT	Consensus (comb.lib./smm/nn)	0.21
HLA-DRB1*09:01	176	190	DSVSFTIAAPADATT	Consensus (comb.lib./smm/nn)	0.21

2.6. Gene sequence analysis of both genes for common site of restriction enzyme

Both gene sequences were analyzed on NEB cutter for mapping restriction sites. The enzymes which could cut the DNA at only single site were identified and the one that was common in both sequences was selected.

2.7. Strategy for construction of chimeric construct by ligation of fragments from FyuA and Flagellin

The chimeric construct could be constructed after ligation of gene fragments excised from FyuA and Flagellin. Briefly, both the genes could be digested with AatII to produce linear sticky ended fragments of 7.4 and 6.5 kb respectively. AatII will digest FyuA at position 1475 bp and Flagellin at position 188 bp. These fragments could then be digested with *XhoI* to generate sticky ended fragments of 6.9 kb and 0.5 kb from FyuA and 5.5 kb and 0.9 kb from Flagellin. The 0.9 kb fragment from Flagellin and 6.9 kb fragment from FyuA could be ligated with T4 DNA ligase followed by transformation, screening and sequencing.

2.8. ORF analysis of chimeric construct AKSC2

Open reading frame for the sequence of chimeric gene was analyzed on Expasy-Translate tool (Swiss institute of Bioinformatics).

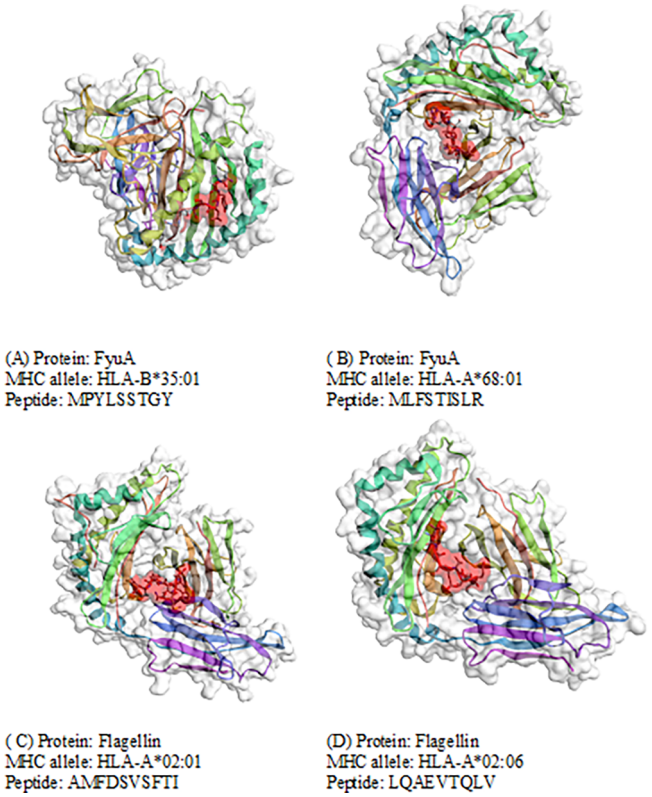

Fig 4. Peptide-MHC complexes obtained after 3D simulations using CABS server.

2.9. Dimensional modeling and stability analysis

The model structure of the chimera AKSC2 was generated using Modeller 9.20 [25]. The structure was energy minimized and validated using Molprobity [26] server for its stereophysical characteristics.

2.10. Linear epitope analysis of AKSC2

Protein sequence of AKSC2 was analyzed using Bepipred portal of IEDB server. The obtained linear epitopes were analyzed for changes in comparison to the epitopes predicted from the individual proteins. The epitopes were also analyzed for the presence of overlapping regions between individual proteins.

2.11. MHC class I and class II binding peptide prediction for AKSC2

Protein sequence of AKSC2 was analyzed to observe the peptides which could bind to MHC class I molecules. These peptides were predicted using MHC class I binding peptide prediction portal on IEDB server. Similarly, peptides binding to MHC II molecules were predicted using MHC class II binding peptide prediction portal on IEDB server. Interactions were evaluated in terms of percentile ranks.

3. Results and discussion

3.1. B-cell linear epitope prediction

Amino acid sequences of both the proteins were analyzed for the presence of linear epitopes using the Bepipred prediction tool of IEDB server. For FyuA a threshold score of 0.332 for a window size of 7 amino acids was generated in the portal. Results in Fig. 1(A) show that a large number of linear epitopes are present in FyuA and most of the prominent epitopes were in the region from amino acid position 20 to 520. The maximum score of 2.317 was obtained for the amino acids near position 160. Similarly, linear epitopes were predicted for Flagellin with the Bepipred generated threshold value of 0.316 for a window size of 7 amino acids. Results in Fig. 1(B) show a large number of linear epitopes on Flagellin and most of the prominent epitopes were in the region from amino acid position 20 to 380. The maximum value of 1.680 was obtained for the amino acids near position110. The peptides that were significant linear epitopes are shown here in Table 1. Results of Bepipred prediction as shown in Fig. 1 revealed the presence of a wide range of linear epitopes on both FyuA and Flagellin proteins. These results were further verified using other parameters like hydrophilicity, surface accessibility and antigenicity. Similar results were obtained from all the above predictions as both the proteins were found to possess large number of linear epitopes (Supplementary information SI.1.). The region accommodating the maximum number of epitopes was then selected to be taken further for theoretical construction of chimaera. Further, evaluation of other parameters was done by keeping the selected regions into consideration. Since, there lies a possibility that the chimeric protein may not take up the proper folding when over expressed due to physical constraint of large size. Hence, it may be expressed as inclusion bodies when subjected to over expression. Misfolding of protein would certainly not affect the T- cell dependent response since the generation of T-cell response depends on the processing and presentation of peptides on MHC molecules but it may affect the antibody response to conformational epitopes. However, Results in Table 1show some large linear epitopes of length upto 34, 23, 22 and 26 amino acids in the peak number 2, 8, 10 and 14 respectively in FyuA protein. Whereas, most of the epitopes in Flagellin were not large and the maximum length of a large epitope was 20 amino acids as shown by peak number 20 in Table 1. Therefore, the presence of linear epitopes can be considered as a very significant feature of the vaccine candidate protein as it could help in the generation of antibody response even if the protein is administered in denatured form.

3.2. Conformational B-cell epitope prediction

Conformational epitopes of FyuA were predicted using Discotope tool on IEDB server. Results in Fig. 2(A) as depicted by the green peaks is the region from amino acid position 200 to 400 that possess maximum conformational epitopes. These epitopes (yellow) were also shown on the 3-D structural image created by J-mol-PDB Fig. 2(B). Since most of the conformational epitopes are present on the exposed surface, these epitopes may become a target of antibodies for effectively neutralizing the bacteria during infection. Both the protein could prove to be good vaccine candidates if they are able to generate both B cell as well T cell responses. However, both the proteins qualified for the generation of B cell response therefore predictions were made for their ability to generate T-cell responses. This was done by predicting MHC class I and II binding peptides.

3.3. MHC class I binding peptide prediction

MHC Class I binding peptides were predicted using the MHC Class I binding prediction tool on IEDB server. Peptides with a percentile rank below 10 were considered significant. Results in Fig. 3(A) show that 2858 significant interactions were obtained for FyuA protein and among these, 2017 interactions were from the selected region of amino acid position 1 to 490. Similarly, for Flagellin protein results in Fig. 3(A) show about 1443 significant interactions and among these, 1226 interactions were from the selected region of amino acid position 61 to 389 (Supplementary information SI.2.). Therefore the results in Fig. 3(B) show that 70% of the significant interactions were from the selected region of FyuA and 85% of significant interactions were from the selected region of Flagellin Fig. 3(B). The top 10 peptides from the selected region having lowest percentile rank were shortlisted and are shown in Table 2. Data in Table 2 show that both of these proteins possessed significant number of peptidic regions that could bind to

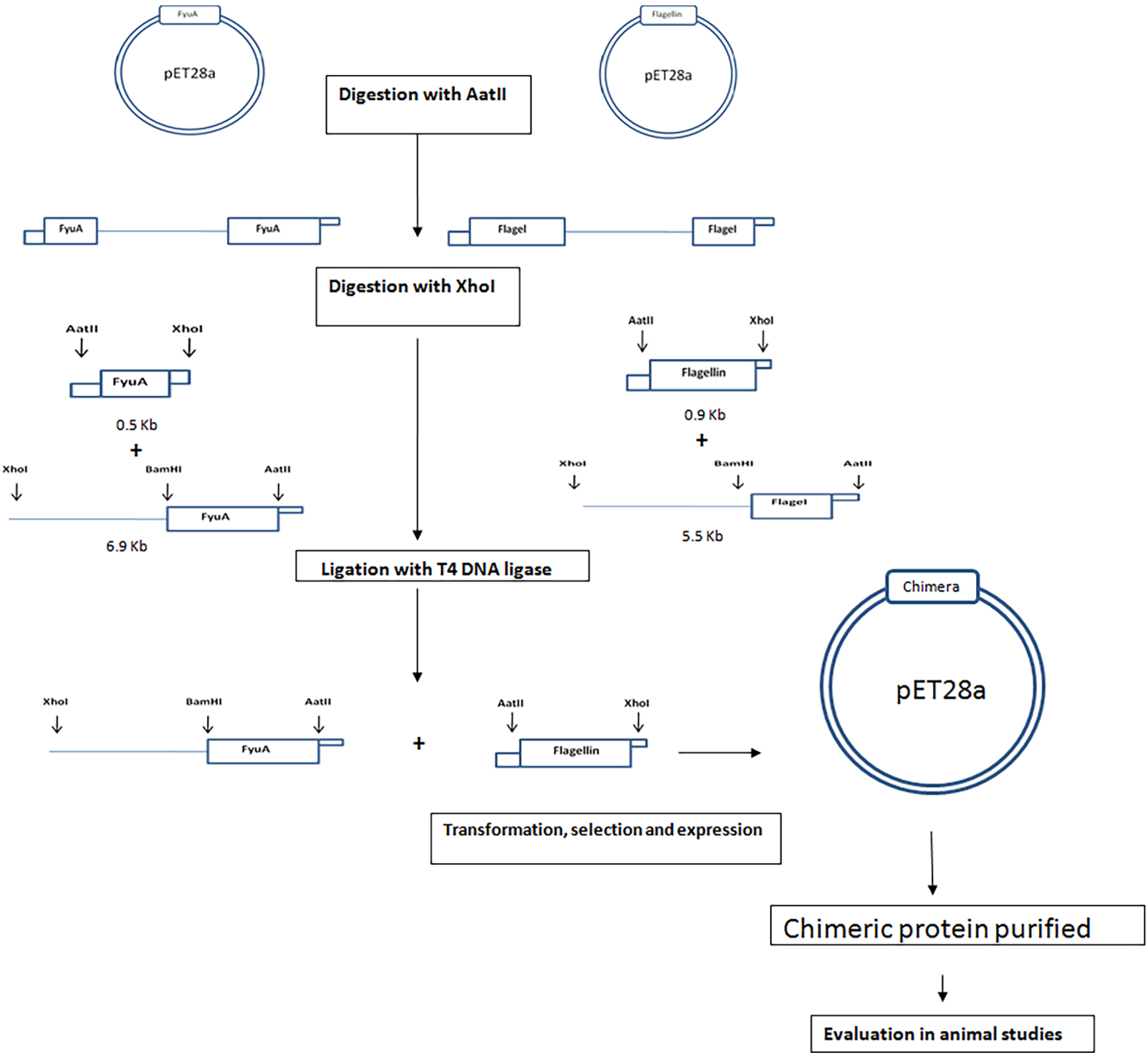

Fig 5. Graphical representation of the process for the construction of chimaera AKSC2.

MHC class I molecules and presented to $CD8^+$ T-cells. Population having the alleles shown in table 2 were determined using the allelefrequencies.net and the allele HLA-A*30:02 and HLA-A*53:01 are found in South African population, HLA-A*35:01 is present in USA, Mexico and Oman. HLA-A*26:01 and HLA-A*68:01 are found among Asians and Europeans. HLA-A*01:01 is found in Ireland, USA and Uganda. It is also seen that the peptides of both these proteins bind with different alleles which are present in different parts of world. This difference in the binding will turn out to be very beneficial when AKSC2 will be used as vaccine as more the number of binding alleles more would be the coverage of human population.

3.4. MHC class II binding peptide prediction

Peptides binding to MHC class II molecules were predicted using MHC class II binding peptide prediction tool on IEDB server. For MHC class II binding, IEDB recommended method and a reference set of 27 alleles was used. Results in Fig. 3(A) show that 2018 significant interactions were obtained for FyuA protein. Out of which 1479 interactions with the reference set of alleles were from the selected region. Similarly, for Flagellin protein results in Fig. 3(A) show that 1353 significant peptide allele interactions were predicted by the server and among these, 1168 interactions were from the selected region (Supplementary information SI.2.). Also the results in Fig. 3(B) show that 73% of the significant interactions were from the selected region of FyuA and 86% of significant interactions were from the selected region of Flgellin Fig. 3(B). Table 3 depict the top 10 peptides from the selected region having lowest percentile rank for FyuA for Flagellin respectively. Alleles HLA-DRB3*01:01, HLA-DRB1*09:01 and HLA-DRB1*08:02 are present largely in Asian and Russian populations (allelefrequencies.net). Allele HLA-DQA1*01:02/DQB1*06:02 is found in Asian, African, Israel and French populations and HLA-DQA1*04:01/DQB1*04:02 is present in German, African and Asian populations.

3.5. Docking and simulations

Results obtained after simulated 3-Dimensional docking of predicted peptides with the predicted MHC molecules were presented in

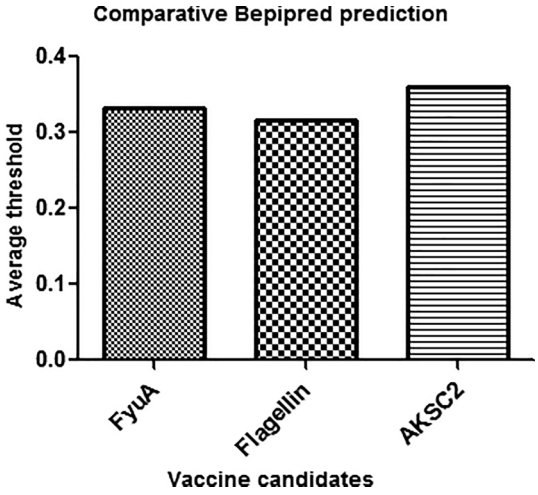

Fig 6. Comparative threshold of peptide prediction in comparison to all three vaccine candidates.

Tables 5 and 6. The data shows the average RMSD values of the each peptide-MHC complex. The average RMSD values between 3 and 5 are considered of medium accuracy whereas the values below 3 are considered as highly accurate [27]. Along with the RMSD values the distance between the interacting amino acid residues of MHC molecules and that of the interacting peptide was also analyzed. The results in Table 5and Fig. 4 show that the MHC class I and class II binding peptide predictions made using the IEDB server, were also found to be significantly accurate using the 3-D simulations for docking. This could be interpreted as the values of average RMSD were below 5 in the simulated complexes. The interacting residues shown in the table were lying at a distance of < 3 Å.

3.6. Proposed strategy for construction of chimaeric construct by ligation of fragments from FyuA and Flagellin

These results of all the *in-silico* predictions helped us in choosing the regions from both the proteins, so that they could be combined to produce a single chimaeric construct. We therefore propose a strategy which could lead to the formation of the chimaeric construct. For designing the chimaeric gene construct, cloning strategy was adopted using genetic engineering tool for restriction digestion. Gene sequences of both the proteins were analyzed on NEB cutter tool of New England Biolabs. Restriction enzyme AatII (Supplementary information SI.1.) was found to be present in both the protein sequences and cleaves FyuA at position 1475 and Flagellin at position 188 and generates sticky ends. Fig. 5shows the graphical representation of the cloning strategy.

Table 4
. Overlapping peptides binding with MHC class I and class II molecules.

MHC Class I					
Allele	Start	End	Peptide	Method used	Percentile rank
HLA-A*68:02	487	496	ETADVAIRNA	Consensus (ann/smm)	0.47
HLA-A*68:02	487	495	ETADVAIRN	Consensus (ann/comblib_sidney2008/smm)	0.92
HLA-A*24:02	485	493	RYETADVAI	Consensus (ann/smm)	1.1
HLA-A*26:01	487	496	ETADVAIRNA	Consensus (ann/smm)	1.55
HLA-A*23:01	485	493	RYETADVAI	Consensus (ann/smm)	1.7
HLA-A*30:01	483	491	GTRYETADV	Consensus (ann/comblib_sidney2008/smm)	2.4
HLA-B*40:01	486	494	YETADVAIR	Consensus (ann/smm)	2.4
HLA-A*68:01	486	494	YETADVAIR	Consensus (ann/smm)	3.05
HLA-A*68:01	487	495	ETADVAIRN	Consensus (ann/smm)	3.4
HLA-A*24:02	485	494	RYETADVAIR	Consensus (ann/smm)	3.5
HLA-A*31:01	485	494	RYETADVAIR	Consensus (ann/smm)	3.55
HLA-B*40:01	486	495	YETADVAIRN	Consensus (ann/smm)	4.25
HLA-A*32:01	485	493	RYETADVAI	Consensus (ann/comblib_sidney2008/smm)	4.3
HLA-A*26:01	487	495	ETADVAIRN	Consensus (ann/smm)	4.95
HLA-B*44:03	486	494	YETADVAIR	Consensus (ann/smm)	5.2
HLA-B*58:01	488	497	TADVAIRNAN	Consensus (ann/smm)	5.3
HLA-A*33:01	485	494	RYETADVAIR	Consensus (ann/smm)	5.45
HLA-A*01:01	487	496	ETADVAIRNA	Consensus (ann/smm)	5.95
HLA-B*44:02	486	494	YETADVAIR	Consensus (ann/smm)	6.35
HLA-A*33:01	486	494	YETADVAIR	Consensus (ann/smm)	6.9
HLA-B*51:01	484	493	TRYETADVAI	Consensus (ann/smm)	7.5
HLA-B*40:01	485	493	RYETADVAI	Consensus (ann/smm)	7.6
HLA-A*23:01	485	494	RYETADVAIR	Consensus (ann/smm)	7.9
HLA-A*02:03	483	491	GTRYETADV	Consensus (ann/smm)	8.8
HLA-A*30:01	485	493	RYETADVAI	Consensus (ann/comblib_sidney2008/smm)	9.9
MHC Class II					
Allele	Start	End	Peptide	Method used	Percentile rank
HLA-DRB1*08:02	480	494	YELGTRYETADVAIR	Consensus (smm/nn/sturniolo)	6.69
HLA-DRB5*01:01	480	494	YELGTRYETADVAIR	Consensus (smm/nn/sturniolo)	7.03
HLA-DRB5*01:01	481	495	ELGTRYETADVAIRN	Consensus (smm/nn/sturniolo)	7.03
HLA-DRB5*01:01	482	496	LGTRYETADVAIRNA	Consensus (smm/nn/sturniolo)	7.03
HLA-DRB5*01:01	483	497	GTRYETADVAIRNAN	Consensus (smm/nn/sturniolo)	7.03
HLA-DRB5*01:01	484	498	TRYETADVAIRNAND	Consensus (smm/nn/sturniolo)	7.03
HLA-DRB3*01:01	479	493	NYELGTRYETADVAI	Consensus (comb.lib./smm/nn)	7.99
HLA-DRB3*01:01	480	494	YELGTRYETADVAIR	Consensus (comb.lib./smm/nn)	7.99
HLA-DRB3*01:01	481	495	ELGTRYETADVAIRN	Consensus (comb.lib./smm/nn)	7.99
HLA-DRB3*01:01	482	496	LGTRYETADVAIRNA	Consensus (comb.lib./smm/nn)	7.99
HLA-DRB3*01:01	483	497	GTRYETADVAIRNAN	Consensus (comb.lib./smm/nn)	7.99
HLA-DRB3*01:01	484	498	TRYETADVAIRNAND	Consensus (comb.lib./smm/nn)	7.99
HLA-DRB3*01:01	485	499	RYETADVAIRNANDG	Consensus (comb.lib./smm/nn)	7.99

Table 5
. Structural clustering between MHC molecules and predicted peptides from FyuA and Flagellin.

Allele	Peptide	RMSD	Interacting residues		Distance (A°)
FyuA			Receptor	Peptide	
HLA-A*68:01	MLFSTISLR	4.156	THR A 233	PHE 3	2.537
			THR A 233	SER 4	2.697
			TYR B 63	SER 4	2.731
			TYR B 63	THR 5	2.733
			TYR B 26	SER 4	2.769
HLA-B*35:01	MPYLSSTGY	3.970	TYR A 7	THR 7	2.670
			THR A 143	MET 1	2.715
			ASN A 63	TYR 9	2.737
			TYR A 159	GLY 8	2.779
			TYR A 59	TYR 9	2.786
HLA-DRB1*08:02	LLLPAIANAQTSQQD	2.371	ASN A 62	ALA 7	2.73
			TYR B 78	GLN13	2.735
			VAL B 11	PRO 4	2.74
			ASN B 82	GLN 14	2.742
			ASN A 69	LEU 3	2.745
HLA-DRB3*01:01	LSAGYMLTDDWRVYT	4.023	THR 83	THR 15	2.631
			THR A 80	TYR 14	2.705
			TYR A 13	TRP 11	2.729
			ASP A 66	TRP 11	2.754
			ILE A 82	THR 15	2.808
Flagellin					
HLA-A*02:01	AMFDSVSFTI	1.418	SER A 4	ILE 10	2.675
			ARG A6	PHE 8	2.758
			SER B 57	ASP 4	2.853
			ASP A 30	SER 7	2.863
			PRO A 235	SER 5	2.868
HLA-A*02:06	LQAEVTQLV	0.421	ASP A 30	THR 6	2.653
			GLN A 32	VAL 9	2.738
			TYR A 113	LEU 1	2.744
			LEU A 126	LEU 1	2.763
			GLU A 212	GLU 4	2.796
HLA-DRB4*01:01	LTRTQILQQAGTAML	2.236	ASN A 69	LEU 1	2.532
			ASN A 69	ARG 3	2.653
			LEU B 67	GLN 5	2.667
			PHE A 54	MET 14	2.722
			VAL B 85	LEU 15	2.735

3.7. ORF analysis of chimeric construct

In-silico created chimeric construct was found to be consisting of 2454 base pairs (Supplementary information SI.1.) which was further translated using Expasy/Translate tool and a protein of 817 amino acids was formed to give a molecular weight of approximately 90 KDa (Supplementary information SI.1.). This ruled out the presence of any stop codon within the whole sequence. The main objective of combining large gene fragments from both the proteins was considered so as to provide the cellular protein processing machinery with enough of proteasome cleavage sites. The possibility of a single peptide to act as a vaccine is usually a rare chance. Hence a protein can be processed to in different ways to create epitopes whose presentation to T-cells may add to protection during pathogenesis.

3.8. Dimensional modeling and stability analysis

Analysis of Ramachandran plot of the chimera AKSC2 generated using the Molprobity server suggested that 94.2% residues lied in the favoured and allowed region while only 5.8% residues were in the outlier region (supplementary information SI.3.). The obtained values suggest that the model is structurally stable.

3.9. Linear epitope analysis of AKSC2

Further analysis of AKSC2 showed an increase in the average threshold value in linear epitope prediction this probably resulted by combining the two proteins (Fig. 6). There were also the peptides in AKSC2 that were epitopic and lied in the region joining both the proteins. This was another advantage of this chimaera as these overlapping peptides could increase the population coverage of vaccine.

3.10. Overlapping epitopes in AKSC2

Table 4 show the peptides that are able to bind to MHC class I and Class II and are present in the region joining the two proteins. These peptides are able to bind to different alleles.

Results in Table 7 show the similarity of overlapping peptides which were generated as a result of fusion of the two proteins. The peptide analyzed using BLASTp showed its identity with peptides present in microorganisms other than the source of peptide. This also gave a hope that the chimaeric vaccine candidate may confer protection upon global populations from the infectious diseases caused by a wide range of pathogens.

4. Conclusion

From the *in-silico* analysis it is concluded that reverse vaccinology can be used to create novel chimeric constructs from the already known vaccine candidates to make them more effective and to confer protection among diverse populations against a wide range of enterobacterial pathogens.

5. Funding information

For this research work no specific grant from any funding agency was provided. Mr. Ashutosh Kumar was granted financial support in the form of Senior Research Fellowship from Indian Council of Medical

Table 6
. Structural clustering between the predicted alleles and peptides of overlapping region adjoining the two proteins.

Allele	Peptide	RMSD	Interacting residues		Distance (A°)
			Receptor	Peptide	
HLA-A*68:02	ETADVAIRNA	0.574	PHE A 33	VAL 5	2.732
			TYR A 99	GLU 1	2.737
			PHE A 33	ALA 6	2.761
			ASN A 63	ASP 4	2.785
			ASN A 174	ASN 9	2.816
HLA-A*68:02	ETADVAIRN	1.287	TYR A 7	ARG 8	2.703
			THR A 73	ASP 4	2.806
			ASP A 77	GLU 1	2.832
			HIS A 114	ALA 3	2.834
			GLN A 70	ALA 6	2.847
HLA-A*68:02	YETADVAIR	0.643	VAL A 34	ARG 9	2.734
			VAL A 152	GLU 2	2.739
			ASN A 66	VAL 6	2.742
			ARG A 97	TYR 1	2.768
			THR A 73	THR 3	2.771
HLA-B*58:01	TADVAIRNAN	1.931	THR A 73	ASN 8	2.575
			TYR 9	ASP 3	2.65
			THR A 143	ASN 10	2.669
			SER A 116	ARG 7	2.726
			ARG A 97	ALA 5	2.732
HLA-DRB3*01:01	NYELGTRYETADVAI	3.0601	THR A 113	TYR 2	2.673
			GLN A 9	GLU 9	2.707
			PHE A 24	VAL 13	2.735
			ILE A 7	VAL 13	2.738
			GLU A 11	THR 6	2.740
HLA-DRB3*01:01	YELGTRYETADVAIR	1.747	ASN A 62	THR 5	2.704
			ASP B 57	LEU 3	2.744
			TYR B 78	ARG 6	2.867
			THR B 77	ARG 6	3.008
			GLY A 58	TYR 7	3.011
HLA-DRB5*01:01	ELGTRYETADVAIRN	4.839	VAL A 91	ALA 12	2.565
			GLU B 4	LEU 2	2.740
			LYS B12	GLU 1	2.744
			VAL A 91	VAL 11	2.745
			ASP A 27	TYR 6	2.766

Table 7
. BLAST analysis of overlapping peptides.

Description	Peptide	Max score	Total score	Query cover	E value	Ident	Accession
hypothetical protein [Microbacterium humi]	**ETADVAIRNA**	29.5	29.5	100%	35	91%	WP_091187671.1
ribosome recycling factor [Mycoplasma agalactiae]	**ETADVAIRNA**	28.6	28.6	100%	70	90%	WP_004024227.1
DUF4192 family protein [Mycobacterium avium]	**ETADVAIRNA**	27.8	27.8	90%	139	89%	WP_084051761.1
hypothetical protein [Burkholderia sp. AU31652]	**ETADVAIRNA**	27.4	27.4	80%	198	100%	WP_089433309.1
hypothetical protein [Acinetobacter sp. WCHA45]	**ETADVAIRNA**	26.9	26.9	90%	277	89%	WP_111313911.1

Research, New Delhi, India. Fellowship award letter No. 67/13/2018-IMM/BMS.

Conflict of interest

None to declare.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.humimm.2019.02.008.

References

[1] Bush and Perez. Klebsiella, Enterobacter, and Serratia Infections. Retrieved from https://www.msdmanuals.com/professional/infectious-diseases/gram negativebacilli/klebsiella-,-enterobacter-,-and-serratia-infections, 2018 (accessed 11.09.18).
[2] S. Vivona, J.L. Gardy, S. Ramachandran, F.S. Brinkman, G.P.S. Raghava, D.R. Flower, F. Filippini, Computer-aided biotechnology: from immuno-informatics to reverse vaccinology, Trends Biotechnol. 26 (2008) 190–200.
[3] P. Srinivasan, S.P. Kumar, M. Karthikeyan, J. Jeyakanthan, Y.T. Jasrai, Y, H.A. Pandya, et al., Epitope-based immunoinformatics and molecular docking studies of nucleocapsid protein and ovarian tumor domain of Crimean-Congo hemorrhagic fever virus, Front. Genet. 2 (2011) 72.
[4] N. Prabdial-Sing, A.J. Puren, S.M. Bowyer, Sequence-based in silico analysis of well studied hepatitis C virus epitopes and their variants in other genotypes (particularly genotype 5a) against South African human leukocyte antigen backgrounds, BMC Immunol. 13 (2012) 67.
[5] S. Idrees, U.A. Ashfaq, Structural analysis and epitope prediction of HCV E1 protein isolated in Pakistan: an *in-silico* approach, Virology J. 10 (2013) 113.
[6] J. Shi, J. Zhang, S. Li, J. Sun, Y. Teng, M. Wu, et al., Epitope-based vaccine target screening against highly pathogenic MERS-CoV: an in silico approach applied to emerging infectious diseases, PloS One 10 (2015) e0144475.
[7] S.J. Goodswen, P.J. Kennedy, J.T. Ellis, Enhancing in silico protein-based vaccine discovery for eukaryotic pathogens using predicted peptide-MHC binding and peptide conservation scores, PloS One 9 (2014) e115745.
[8] D. Maione, I. Margarit, C.D. Rinaudo, V. Masignani, M. Mora, M. Scarselli, Identification of a universal Group B *streptococcus* vaccine by multiple genome screen, Science 309 (2005) 148–150.
[9] R. Singh, N. Garg, N. Capalash, R. Kumar, M. Kumar, P. Sharma, In silico analysis of *Acinetobacter baumannii* outer membrane protein BamA as a potential immunogen, Int. J. Pure. Appl. Sci. Technol. 21 (2014) 32–39.
[10] R. Singh, *In-silico* analysis of *Burkholderia pseudomallei* proteome to predict potential vaccine candidate proteins, Int. J. Pharm. Bio. Sci. 6 (2015) 957–962.
[11] A. Wieser, E. Romann, G. Magistro, C. Hoffmann, D. Nörenberg, K. Weinert, et al., A

multiepitope subunit vaccine conveys protection against extraintestinal pathogenic *Escherichia coli* in mice, Infect. Immun. 78 (2010) 3432–3442.
[12] A. Rakin, E. Saken, D. Harmsen, J. Heesemann, The pesticin receptor of *Yersinia enterocolitica*: a novel virulence factor with dual function, Mol. Microbiol. 13 (1994) 253–263.
[13] M.C. Miller, J.D. Fetherston, C.L. Pickett, A.G. Bobrov, R.H. Weaver, E. DeMoll, et al., Reduced synthesis of the Ybt siderophore or production of aberrant Ybt-like molecules activates transcription of yersiniabactin genes in *Yersinia pestis*, Microbiology 156 (2010) 2226–2238.
[14] A.R. Brumbaugh, S.N. Smith, H.L. Mobley, Immunization with the yersiniabactin receptor, FyuA, protects against pyelonephritis in a murine model of urinary tract infection, Infect. Immun. (2013) IAI-00470.
[15] F. Hayashi, K.D. Smith, A. Ozinsky, T.R. Hawn, C.Y. Eugene, D.R. Goodlett, et al., The innate immune response to bacterial flagellin is mediated by Toll-like receptor 5, Nature 410 (2001) 1099.
[16] T. Eaves-Pyles, K. Murthy, L. Liaudet, L. Virág, G. Ross, F.G. Soriano, et al., Flagellin, a novel mediator of *Salmonella*-induced epithelial activation and systemic inflammation: IκBα degradation, induction of nitric oxide synthase, induction of proinflammatory mediators, and cardiovascular dysfunction, J. Immunol. 166 (2001) 1248–1260.
[17] C. Szabó, Role of flagellin in the pathogenesis of shock and acute respiratory distress syndrome: therapeutic opportunities, Crit. Care Med. 31 (2003) S39–S45.
[18] C.J. Sanders, Y. Yu, D.A. Moore, I.R. Williams, A.T. Gewirtz, Humoral immune response to flagellin requires T cells and activation of innate immunity, J. Immunol. 177 (2006) 2810–2818.
[19] A.K. Zgair, S. Chhibber, *Stenotrophomonas maltophilia* flagellin induces a compartmentalized innate immune response in mouse lung, J. Med. Microbiol. 59 (2010) 913–919.
[20] D.A. Benson, I. Karsch-Mizrachi, D.J. Lipman, J. Ostell, D.L. Wheeler, GenBank, Nucl. Acids Res. 33 (2005) D34–D38.
[21] E.A. Emini, J.V. Hughes, D. Perlow, J. Boger, Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide, J. Virol. 55 (1985) 836–839.
[22] H.M. Berman, J. Westbrook, Z. Feng, G. Gillilan, T.N. Bhat, H. Weissi, et al., The protein data bank, Nucl. Acids Res. 28 (2000) 235–242 https://www.rcsb.org/.
[23] M. Kurcinski, M. Jamroz, M. Blaszczyk, A. Kolinski, S. Kmiecik, CABS-dock web server for the flexible docking of peptides to proteins without prior knowledge of the binding site, Nucl. Acids Res. 43 (2015) W419–W424 http://212.87.3.12/CABSdock/job/579d06c3d309924/.
[24] D.T. Jones, Protein secondary structure prediction based on position-specific scoring matrices, J. Mol. Biol. 292 (1999) 195–202 http://bioinf.cs.ucl.ac.uk/psipred/.
[25] A. Sali, T.L. Blundell, Comparative protein modelling by satisfaction of spatial restraints, J. Mol. Biol. 234 (1993) 779–815.
[26] V.B. Chen, W.B. Arendall, J.J. Headd, D.A. Keedy, R.M. Immormino, G.J. Kapral, et al., MolProbity: all-atom structure validation for macromolecular crystallography, Acta Crystallogr. D. 66 (2009) 12–21.
[27] M. Blaszczyk, M. Kurcinski, M. Kouza, L. Wieteska, A. Debinski, A. Kolinski, et al., Modeling of protein–peptide interactions using the CABS-dock web server for binding site search and flexible docking, Methods 93 (2016) 72–83.