



Subtractive Proteomics and Immuno-informatics Approaches for Multi-peptide Vaccine Prediction Against *Klebsiella oxytoca* and Validation Through In Silico Expression

Qudsia Yousafi¹ · Humaira Amin² · Shabana Bibi³ · Rafea Rafi¹ · Muhammad S. Khan¹ · Hamza Ali¹ · Ashir Masroor⁴

Accepted: 8 September 2021 / Published online: 20 September 2021
© The Author(s), under exclusive licence to Springer Nature B.V. 2021

Abstract

Klebsiella oxytoca is a gram-negative bacterium. It is opportunistic in nature and causes hospital acquired infections. Subtractive proteomics and reverse vaccinology approaches were employed to screen out the best proteins for vaccine designing. Whole proteome of *K. oxytoca* strain ATCC 8724, consisting of 5483 proteins, was used for designing the vaccine. Total 1670 cytotoxic T lymphocyte (CTL) epitope were predicted through NetCTL while 1270 helper T lymphocyte (HTL) epitopes were predicted through IEDB server. The epitopes were screened for non-toxicity, allergenicity, antigenicity and water solubility. After epitope screening 300 CTL and 250 HTL epitopes were submitted to IFN- γ epitope server to predict their Interferon- γ induction response. The selected IFN- γ positive epitopes were tested for their binding affinity with MHCII-DRB1 by MHCpred. The 15 CTL and 13 HTL epitopes were joined by linkers AAY and GPGPG respectively in vaccine construct. Chain C of Pam3CSK4 (PDB ID; 2Z7X) was linked to the vaccine construct as an adjuvant. A 450aa long vaccine construct was submitted to I-TASSER server for 3D structure prediction. Thirteen Linear B cells were predicted by ABCPred server and 10 sets of discontinues epitopes for 3D vaccine structure were predicted by DiscoTope server. The modeled 3D vaccine construct was docked with human Toll-like receptor 2 (PDB ID: 6NIG) by PatchDock. The docked complexes were refined by FireDock. The selected docked complex showed five hydrogen bonds and one salt bridge. The vaccine sequence was reverse transcribed to get nucleotide sequence for In silico cloning. The reverse transcribed sequence strand was cloned in pET28a(+) expression vector. A clone containing 6586 bp was constructed including the 450 bp of query gene sequence.

Keywords Linear B cells epitopes · Disulfide engineering · Thermostability curve · Human TLR2 · pET28a (+)

Introduction

Klebsiella oxytoca is a non-motile, rod shaped gram-negative bacterium of the family Enterobacteriaceae (Gorkiewicz 2009). It is unscrupulous in nature and causes hospital

acquired infections in hospitalized patients, especially newborn infants (Podschun and Ullmann 1998). A large number of infants have been reported to die worldwide due to this infection (Reiss et al. 2000). After *Escherichia* sp. this genus includes second most infectious pathogens (Yinnon

✉ Qudsia Yousafi
qudsia@cuisahiwal.edu.pk

Humaira Amin
humaira.amin96@gmail.com

Shabana Bibi
dazzling.grin@gmail.com

Rafea Rafi
rafia.rafi000@gmail.com

Muhammad S. Khan
saad.khan@cuisahiwal.edu.pk

Hamza Ali
maikeehamza@gmail.com

Ashir Masroor
masroorakb@gmail.com

- 1 COMSATS University Islamabad, Sahiwal Campus, Sahiwal, Pakistan
- 2 COMSATS University Islamabad, Islamabad Campus, Islamabad, Pakistan
- 3 Yunnan Herbal Laboratory, College of Ecology and Environmental Sciences, Yunnan University, Kunming 650091, Yunnan, China
- 4 University of Agriculture Faisalabad, Sub Campus Burewala-Vehari, Burewala, Pakistan

et al. 1996) while *K. oxytoca* has been reported to be the second most frequent source for bacterial infection next to *Klebsiella pneumoniae* species (Singh et al. 2016). It is also found to be the causal organism of AAHC (antibiotic-associated hemorrhagic colitis) disease (Hoffmann et al. 2010). It uses many virulence factors to overcome the host innate immune system. The virulence factors are adhesions, siderophores, lipopolysaccharides (LPSs) on cell surface and capsular polysaccharides (CPLs). Although LPS have potential to activate infection pathway but K antigen of the capsule have been reported to be the main source of pathogenicity (Simoons-Smit et al. 1986; Podschun and Ullmann 1998). The bacterium attaches to host cell by adhesins as primary step for host cell infections. It attaches to intestinal, urogenital tracts and respiratory epithelial cells by Type 1 pili. It binds to the mannose-containing trisaccharides of the host glycoproteins via Type 4 pili (Podschun and Ullmann 1998).

Being unscrupulous in nature it causes infections in multiple fashions so that multiple drugs are used to treat the disease accordingly. This indiscriminate and heavy use of drugs leads to multiple drug resistance to commonly used antibiotics (Vali et al. 2015; Singh et al. 2016). Resistance development in different species of *Klebsiella* to different drugs like the oxyimino (carbapenemases), β -lactams (cef-tazidime and cefotaxime), penicillins (ampicillin, cephalosporinases, carbenicillin), monobactam and aztreonam have been reported from last so many decades (Decré et al. 2004; Wu et al. 2009). As this bacterium is not causal organism of a particular disease so, in spite of its multidrug resistance, very little attention have been given to the development of effective vaccine against this. Although it causes rare infectious disease but these can lead to life threatening complications in the patients especially in children. The development of an effective vaccine against this pathogen is need of this era.

Development of vaccine is a long tedious and time taking procedure which involves high cost of production even at initial screening steps at laboratory scale (André 2003). Some pathogens are very difficult or even impossible to handle in lab conditions so their vaccine development is very difficult even impossible (Tannock et al. 2020). The development of advanced technologies and improved bioinformatics techniques during recent past have given a breakthrough in many industries including medical technologies. The advanced approaches of bioinformatics can be used in vaccinology, an important branch of healthcare sciences. Reverse vaccinology (RV) is an innovative and smart in silico prediction technique for the identification of new protein-derived vaccine candidates to minimize the in vitro vaccinology limitation (Rappuoli 2000). A classic RV pipeline consists of screening the whole proteome- and choosing proteins that will become vaccine construct through computational methods. RV has been used against different kind of infectious diseases

(Delany et al. 2013; De la Fuente et al. 2016; Michalik et al. 2016; Mehla and Ramana 2017; Muruato et al. 2017) and the first successful application of this approach was achieved in *Neisseria meningitidis* serogroup B (Kelly & Rappuoli 2005). Conventional approaches, most of the time, failed in providing a universal and efficient vaccine for the most affected groups of patients. In contrast, reverse vaccinology has shown to be an efficient tool for the identification of protective vaccine candidates. Here, we have used subtractive proteomics and cheminformatics approaches for the screening the surface exposed proteins of *K. oxytoca* to design the effective epitopes for vaccine. These surface epitopes from surface exposed proteins can be used to design an effective and human safe vaccine against this bacterium.

Methods

Bacterial Proteome Acquisition and Subtraction

The proteome of *K. oxytoca* (strain ATCC8724) was retrieved from UniProt (UniProt ID: UP000007843). The whole proteome was passed through the pipeline of subtractive proteomics. The proteins, suitable for designing a potential vaccine against this pathogen infections, were shortlisted.

The proteome dataset may contain some similar or duplicate protein sequences which are unnecessary to carry forward. The proteome was screened through CD-HIT suite. The Paralogous or duplicates proteins arisen during evolution process were identified and then redundant protein entries were discarded at 0.8 cutoff values (Li & Godzik 2006).

The CD-HIT screened files of proteins were submitted to PSORTb (Yu et al. 2010) for the subcellular localization prediction. The proteins localized in four cell regions i.e., cytoplasmic membrane, outer membrane extracellular and periplasmic proteins, were selected. These membrane proteins were selected due to their important role in infection, attachment, and survival of pathogens (Nagpal et al. 2018; Ul-Ain et al. 2018).

The next requirement was to screen out virulent proteins. VFDB is a user-friendly database to find out the virulence factors in a variety of pathogenic bacteria (Chen et al. 2005). Protein BLAST (BLASTp), for similarity search with other infectious proteins, was used for screening in this database. Proteins showing score more than 100 and identity above 30% were selected. These virulent proteins screened by VFDB were kept in dataset to be analyzed further (Chen et al. 2005) and rest were discarded. The virulent proteins were submitted to an online antigenicity prediction server VaxiJen (Doytchinova and Flower 2007) and antigenicity was predicted at 0.4 threshold value (Barh et al. 2013; Hassan et al. 2016a,b). Antigenic virulent

proteins were subjected to further analysis. For the selection of small sized proteins we used ExPasy:ProtParam (Gasteiger et al. 2005). Proteins with molecular weight < 110 kDa were carried forward as potential candidates for vaccine construction. The shortlisted proteins for vaccine designs must be least identical with human protein or may be unique for bacteria. For making sure this condition similarity search of the proteins, with human (Human taxid: 9606), was done by NCBI BLASTp. The resultant proteins showing percent identity less than 35% or not found in human were selected for prediction of different type of epitopes.

MHCI Cytotoxic T Lymphocytes (CTL) Epitopes Prediction

Epitope is the portion of an antigen which is detected by immune system with the help of the B-lymphocytes (Chen et al. 2005). Many of the proteins of a pathogen can act as antigen and the portions of an antigen that attach with antigen binding site on either an antibody or a T-lymphocyte receptor sites are called epitopes which then help in producing cellular or humoral immune response.

NetCTL 1.2 server used to predict CTL epitopes in protein sequences at 0.75 thresholds level (Larsen et al. 2007). These predictions were done on the basis of (i) binding to MHC-I, (ii) proteasomal C-terminus cleavage score, (iii) transport efficiency of transporter associated with antigen processing (TAP).

Weight matrix was used for TAP score calculation while artificial neural network approach was deployed to calculate proteasomal C-terminal cleavage scores and peptide binding to MHC-I.

MHCII Helper T Lymphocyte (HTL) Epitopes Prediction

The prediction of MHCII HTL epitopes was done taking 27 human HLAs as a reference set, by Immune Epitope Database server (IEDB; Fleri et al. 2017). The server predicts the HTL epitopes based on their percentile rank or MHCII binding affinity which is expressed in terms of IC_{50} values which are inversely related to the binding affinity towards the MHC-II (Patronov & Doytchinova 2013). Ten epitopes from each protein were selected based on, the low percentile rank because the binding affinity to MHC-II, of predicted epitopes, is inversely proportional to the percentile rank (Paul et al. 2016).

Epitope Screening

Allergenicity Antigenicity, Toxicity and Water Solubility Prediction for Epitopes

Allergenicity prediction with better precision was achieved by the AllgPred server (Christensen et al. 2003). The

multiple algorithms approach was employed by the server. The accuracy of the server is nearly 85% for the hybrid algorithm at threshold (– 0.4). Allows to predict allergens based on SVM modules using amino acid composition and similarity with known allergenic epitopes. It does prediction of allergens based on similarity of known epitope with any region of protein.

VaxiJen server (<http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html/>) was employed to predict antigenicity of the vaccine sequence. This server predicts the antigenic score wholly based on the query amino acid sequence and the target strain was selected as bacteria. 0.4 was the threshold value selected for it. VaxiJen precision for antigenicity is quite better, 70–89%. The antigenic scores for the vaccine sequence at 0.4 thresholds for designating the vaccine. Prediction of toxicity of peptides by using computational methods not only saves resources, but also helps in designing better least toxic therapeutic peptides by retaining the functionalities undisturbed (Gupta et al. 2013).

The list of the non-allergen and antigenic epitopes was imported in ToxinPred online server as batch submission for their toxicity prediction. The non-toxic epitopes were selected and rest were discarded.

INNOVAGEN peptide property calculator tool PepCalc was used to calculate the water solubility of peptides. Water soluble peptides were shortlisted.

Interferon- γ Epitope Prediction

The IFN- γ cytokine is a very important factor induced in response of pathogen attack. It activates macrophages as an immune response in the cell. It enhances the response of MHCII to antigens. A potential candidate of vaccine is a peptide not only be epitope for B and T cell but also be capable of inducing the appropriate immune cells to generate desired immunity response (Shtrichman and Samuel 2001). The vaccine candidate required to induce IFN- γ to destroy the infection agents in the cell (Flynn et al. 1993; Reljic 2007). The IFN- γ epitope server (Dhanda et al. 2013) was used to identify IFN- γ epitopes. The IFN- γ epitopes in CTL and HTL epitopes was predicted by IFN- γ epitope server by using support vector machine (SVM) hybrid approach. The epitopes with value $IC_{50} < 100$ with positive results were selected.

Prediction of CTL Epitope-HLA-DRB1 complex Binding Affinity

The peptides which do not bind to any of an individual's allelic MHC variants cannot be antigenic within a cellular context (Hattotuwigama et al. 2004). The human version of

MHC complex is human leukocytes antigen complex (HLA). The 9-mer CTL epitopes were docked with HLA-DRB1*101 allele through MHCpred (Guan et al. 2003). MHCpred is an online available (<http://www.jenner.ac.uk/MHCpred>) used to predict binding affinity of CTL epitope-HLA-DRB1 complex in terms of IC50 values. The 15-mer HTL epitopes were predicted through IEDB on the basis of seven HLA-DRB alleles (Wang et al. 2010) that's why they were not docked by MHCpred server.

Vaccine Construct

MHC class I and II binding epitopes were screened based on their high binding affinity and non-allergenic nature for vaccine construction. AAY linkers are used to join the CTL epitopes and GPGPG linkers for HTL epitopes (Hasan et al. 2019). These linkers are helpful for differentiation and improvement of epitope presentation (Saadi et al. 2017; Dorosti et al. 2019). The adjuvant protein sequence SKKKK, of TLR2 agonist Pam3CSK4 Chain C, was attached in the start of vaccine construct sequence. This adjuvant enhances the immunogenic property of the designed vaccine. The adjuvant was joined to CTL epitope by another linker sequence EAAAK at N-terminus of vaccine construct (Arai et al. 2001).

Antigenicity, Allergenicity and Thermostability Test of the Vaccine Construct

The antigenicity and non-allergenicity of the vaccine construct was predicted by VaxiJen v 2.0 and AllerTOP v 2.0 respectively. Thermal stability of the vaccine construct was measured by an online available server SCooP (prediction of the stability curve of proteins). It measures the thermal stability in terms of T_m (the protein melting point) which shows the temperature at which half of the folds can become unfolded and denatured. It also measures the values of change in free energy (ΔG), standard enthalpy of folding (ΔH_m) and standard heat capacity of folding (ΔC_p) of protein.

Tertiary Structure Prediction, Refinement and Validation

The vaccine construct sequence was subjected to NCBI BLASTp but no suitable template was found for homology based modeling so it was submitted to an online available server I-TASSER for ab initio based structure prediction. The quality of the predicted structure was improved by using ModRefiner.

B-Cell Epitope Prediction

Receptors present in the B-lymphocytes surface recognize and binds B cell epitopes. ABCpred server was used

for linear B cell epitopes prediction. The accuracy of the ABCpred server is 75% (0.75 specificities and 0.49 sensitivity) (Saha and Raghava 2006). The amino acid sequence of vaccine construct was submitted for 16-mer B cell linear epitope prediction. Discontinuous B-cell epitopes were predicted by an online server Discotope2.0 (Kringelum et al. 2012) by submitting refined 3D structure of vaccine construct. Discotope2.0 conformational B cell epitopes prediction is based on amino acids composition ratio between residues of epitope and non-epitope. The server specificity is 0.75, whereas the sensitivity is 0.47, at 3.7 threshold value.

Cysteine Mutation to Improve Disulfide Bonding

Cysteine is a sulfur containing amino acid. Two cysteine molecules combine to form cysteine by forming disulfide bond which improves protein stability (Liu et al. 2016). An online freely available server Disulfide by Design 2 v 2.13 (DbD2) was used to enhance disulfide bonding in vaccine construct. The server can predicts the suitable pairs of amino acids to be mutated with cysteine for disulfide bond formation. A list of potential amino acid pairs was generated and suitable pairs were selected keeping in view the resultant bond energy and B-factor (Craig & Dombkowski 2013).

Molecular Docking with Human Toll Like Receptor 2 (TLR-2)

The three dimensional structure of human TLR2 (PDB ID: 6NIG) was downloaded from RCSB Protein Data Bank (PDB). The structure was refined by using Chimera and chain A was selected for docking purpose. Vaccine construct was docked with Human TLR2 by online available server PatchDock. The resulting solutions were submitted to FireDock for refinement. The best refined complex model with lowest binding energy was selected. The interactions in docked complex were visualized by LIGPLOT: DIMPLOT (Laskowski & Swindells 2011).

Reverse Translation and Expression Vector Cloning of Vaccine Construct

The vaccine construct was prepared and optimized to be incorporated in expression vector for expression. The amino acid sequence of vaccine was reverse translated and optimized by JCAT server (JAVA Codon Adaptation Tool) server. GC content and codon adaptation (CAI) index was calculated for optimized sequence to ensure its suitability to be carried forward in cloning vector. The DNA sequence obtained from reverse translation was modified by conjugation with XhoI and NdeI restriction site at the N-terminal and C-terminal sites respectively through SnapGene software. The modified DNA sequence was inserted in pET28a (+) vector at multiple

cloning site (MCS) between the XhoI (158) and NdeI (1450) by using SnapGene restriction cloning module.

Results

Proteome Subtraction

The whole proteome of *K. oxytoca* strain ATCC 8724 (UniProt ID: UUP000007843), containing 5483 proteins was retrieved from the UniProt database. The whole proteome was passed through the subtractive proteomics pipeline.

CD-HIT refined the proteome and 5470 non-paralogous proteins at the 80% threshold were carried on for further analysis. Prediction of subcellular localization was the next step executed by submitting proteins to PSORTb. The proteins from four cellular location i.e., periplasm (188), extracellular (54), outer membrane (108) and cytoplasmic membrane (1398) were selected and carried forward. These total 1748 screened proteins were submitted to ExPASy ProtParam to get small sized proteins. The proteins with less than 110 kDa were selected for vaccine construction (Hassan et al. 2016a,b) and number of proteins reduced to 197. NCBI BLASTp was used for taking off the non-human homologous proteins keeping “Human taxid: 9606” as the reference organism. The proteins showing more than 35% identity with human proteins were discarded. The number of qualified proteins reduced to 172.

Virulent proteins interact with different host pathways and play important role in pathogenicity that's why are considered best for vaccine construction (Baseer et al. 2017). BLASTp search of VFDB (virulence factor database) was performed

for virulent proteins identification (Ul Ain et al. 2018). Proteins showing identity and bit score greater than 30% and 100 respectively, were considered virulent and were further evaluated for antigenicity. Total 156 virulent proteins were submitted to VaxiJen v 2.0 server (Doytchinova & Flower 2007) for antigenicity score prediction at 0.4 threshold value (Naz et al. 2015; Hassan et al. 2016a,b). Total 127 proteins with VaxiJen score greater than 0.4 were selected and all other were discarded. Antigenic virulent proteins were subjected to further analysis. The number of protein reduced from 5483 to only 127 and these virulent and antigenic proteins were selected for CTLs and HTL epitope prediction and further analysis (Fig. 1).

Cytotoxic T Lymphocytes (CTL) and Helper T Lymphocyte (HTL) Epitopes Prediction

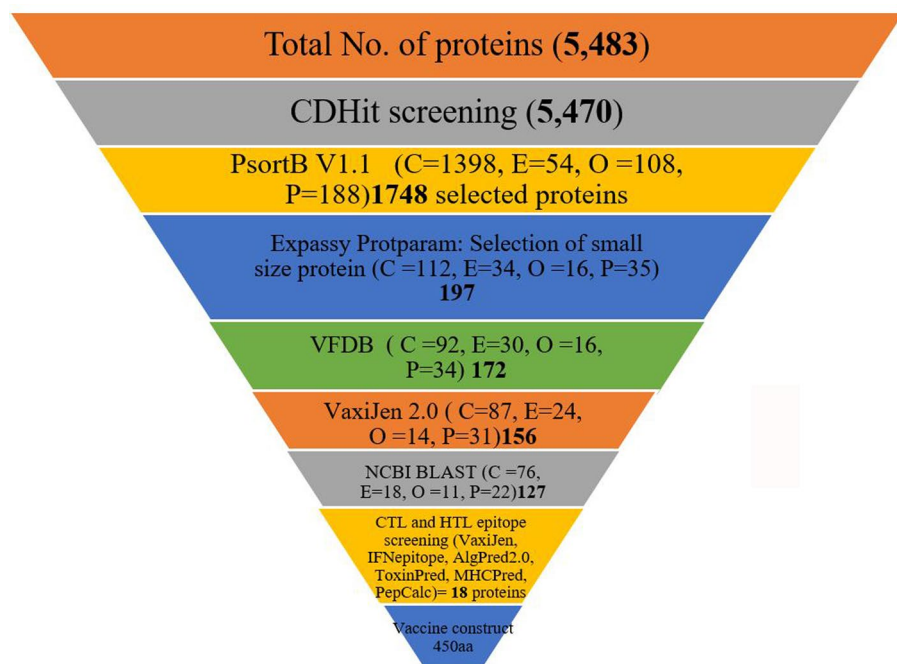
CTLs (MHC-I binding) 1670 (9-mer) epitopes were predicted by NetCTL 1.2 server. The Immune Epitope Database (IEDB) server predicted 1270 HTL (MHC-II binding), 15-mer, epitopes against a reference set of seven human leukocyte antigen (HLAs) alleles.

Epitope Screening

Allergenicity Antigenicity, Toxicity and Water Solubility Prediction for Epitopes

Allergenicity scores were predicted through AlgPred 2.0 server. The results showed 1322 CTL and 1107 HTL

Fig. 1 Subtractive proteomic schema for vaccine construction against *Klebsiella oxytoca*



epitopes which can activate the immune response in the cell (Sharma et al. 2020). VaxiJen filtered out 1109 CTL and 968 HTL epitopes. Total 675 CTL and 548 HTL nontoxic epitopes were predicted by ToxinPred server. PepCalc tool was used to predict water soluble peptides. Total 300 CTL and 250 HTL epitopes were screened out from water solubility filter.

Epitope Interferon- γ Induction Prediction

The vaccine candidate 12 HTL epitopes were selected based on IFN- γ induction (positive score) and MHCII binding affinity (Table 1). Fifty six IFN- γ positive CTL epitopes were further analyzed to qualify to be part of vaccine construct.

Prediction of CTL Epitope-HLA-DRB1 Complex Binding Affinity

The IC₅₀ values are the most widely quoted binding affinity measures (Ruppert et al. 1994) and it is inversely proportional to the nominal binding affinity (Guan et al. 2003). If the IC₅₀ value is above 5000 then the peptide cannot bind to the MHC molecules so, the peptides with lower IC₅₀ values (< 100) were selected (Table 2).

Vaccine Construct

Fifteen CTL (MHCI class) and 13 HTL (MHCII class) epitopes were selected, based on their high binding affinity and non-allergenic nature, for vaccine construction (Fig. 2).

Antigenicity, Allergenicity and Thermostability Test of the Vaccine Construct

The vaccine construct was predicted as non-allergenic and antigenic in nature. SCooP (prediction of the stability curve of proteins) showed $T_m = 61.5$ °C, $\Delta Gr = -7.9$ kcal/mol, $\Delta C_p = -5.19$ kcal/(mol K) and $\Delta H_m = -170.5$ kcal/mol (Fig. 3). Which reflect the stability of protein.

Tertiary Structure Prediction, Refinement and Validation of Vaccine Construct

The vaccine construct sequence was subjected to NCBI BLASTp but no suitable template was found for homology based modeling so it was submitted to an online available server I-TASSER for ab initio based structure prediction. Refinement of the predicted model was done by using ModRefiner (Fig. 4).

Prediction of B (Lymphocytes) Cell Epitopes

The B-cell (lymphocytes) epitopes identification is very important for the vaccine development research (Shirai et al. 2014). Eight (16-mer) linear B-cell epitopes were predicted by ABCpredserver at 0.85 threshold level (Table 3). B-cell epitopes are mostly present at discontinuous locations within the sequence far apart from each other but brought together during folding in 3D structure (Kringelum et al. 2013). Thirty four discontinuous epitopes were predicted by DiscoTope server at threshold of -3.7 (Table 4).

Cysteine Mutation to Improve Disulfide Bonding

The analysis provides option of 61 pairs of amino acids which can be potentially mutated to cysteine along with estimated bond energy χ^3 and β factors as a result of mutation. Residues pairs with lower bond energy < 1.5 were mutated to cysteine 184Val-218Val, 156Ala-163Asp, 420Ala-424Asp, 43Asp-264Leu and 179Ala-221Ile (Table 5). The χ^3 of selected pairs ranged from -93.27 to 101.17 and β factor 10.42 to 23.92 . The disulfide bonds appeared as yellow cylinders in the structure (Fig. 5). The mutated residues are encircled in the sequence (Fig. 6). The thermal stability of the vaccine can be enhanced by increasing disulfide bindings without affecting its efficacy (Bouvier et al. 2003). Our current study proves this finding. The mutated sequence 3D vaccine structure submitted to SCooP resulted in 1 point rise in T_m (Fig. 7). This shows that the thermostability of vaccine can be enhanced by disulfide engineering.

Molecular Docking of Vaccine Construct with Human TLR-2

Human toll like receptor 2 (TLR-2) is widely used as vaccine target receptor (Basto and Leitão 2014) it is a protein that in humans is encoded by the TLR2 gene (Rock et al. 1998). It is a membrane protein, a receptor, which is produced on the surface of certain cells and detects foreign invaders and transmits the appropriate signals to immune cells (Borrello et al. 2011). The 3D model of human TLR2 (PDBID: 6NIG) was downloaded from RCSB Protein Data Bank (PDB). The refund chain A is shown in Fig. 8. Vaccine construct was docked with Human TLR2 by online available server PatchDock. The vaccine-HTLR2 docked models (solutions) were received via email as top 20 models with their respective docking score. The top 10 solutions (vaccine-HTLR2 complexes) were refined using FireDock. The solution/complex showing the highest refinement rank (solution 6) with lowest binding affinity of -85.71 kcal/mol, was selected. Five hydrogen bonds with bond distance less than 3 Å were formed between vaccine and HTLR2. A salt bridge was found between Arg375 (HTL2) and Glu141

Table 1 HTL epitopes selected for vaccine construct against *Klebsiella oxytoca*

Protein ID	Alleles	Start	End	Method	Peptide sequence	Percentile rank	Adjusted rank	Allergenicity score	Antigenicity	IFN epitope score
A0A0H3HBY6	HLA-DRB5*01:01	3	17	Consensus (simm/nm/ sturniolo)	EKKVVMRIRALHKRF	0.21	0.21	-0.4281	0.57	0.030451
	HLA-DRB5*01:01	6	20	Consensus (simm/nm/ sturniolo)	VVMRIRALHKRFGAQ	0.58	0.58	-0.598	1.24	0.013655
	HLA-DRB5*01:01	2	16	Consensus (simm/nm/ sturniolo)	SEKKVVMRIRALHKR	0.21	0.21	-0.5404	0.46	0.188259
A0A0H3HCH0	HLA-DRB3*02:02	142	156	NetMHCIIpan	QQRVAIARALAMKPS	3.6	3.6	-4.5988	0.73	0.336966
A0A0H3HAB9	HLA-DRB3*01:01	117	131	Consensus (comb.lib./ simm/nm)	HVKGM- KRDEAKAIAM	0.71	0.71	-0.5438	2.1	0.206067
	HLA-DRB3*01:01	115	129	Consensus (comb.lib./ simm/nm)	LLHVKGMRDEAKAI	0.72	0.72	-0.4656	0.75	0.055908
	HLA-DRB3*01:01	118	132	Consensus (comb.lib./ simm/nm)	VKGM- KRDEAKAIAMQ	0.72	0.72	-0.5340	0.49	0.170946
	HLA-DRB3*01:01	116	130	Consensus (comb.lib./ simm/nm)	LHVKGMRDEAKAIA	0.74	0.74	-0.5529	0.63	0.105924
A0A0H3HGA0	HLA-DRB5*01:01	85	99	Consensus (simm/nm/ sturniolo)	TVSRFMLLRPGTHKR	0.88	0.88	-0.7711	0.42	0.602909
	HLA-DRB1*03:01	15	29	Consensus (simm/nm/ sturniolo)	GQRRVLSDSLTKP	1.1	1.1	-9.901	0.53	0.3684091
A0A0H3HDE1	HLA-DRB1*03:01	151	165	Consensus (simm/nm/ sturniolo)	HGWSAVNVDDR- WLFR	0.24	0.24	-0.5072	0.58	0.937815
A0A0H3HT729	HLA-DRB5*01:01	214	228	Consensus (simm/nm/ sturniolo)	RSVTRAFNHMAA- GVK	0.16	0.16	-0.697	1.43	0.509832

Table 2 CTL epitopes selected for vaccine construct against *Klebsiella oxytoca*

Cellular location	Protein ID	Peptide sequence	VaxiJen score	IFN epitope score	Water solubility	MHCPred	AllerTOP v. 2.0	ToxinPred
Periplasmic	A0A0H3HK70	TYDKLAKKY	0.5864	0.045	Good	1.53		
	A0A0H3H8B0	MQEQRASAY	0.5668	0.213	Good	6.32	Non-allergenic	Non toxic
	A0A0H3HFM6	LAEGLRADY	0.4985	0.378	Good	13.40	Non-allergenic	Non toxic
	A0A0H3HAJ2	QSDAEKSGV	1.6633	0.486	Good	18.49	Non-allergenic	Non toxic
Outer membrane	A0A0H3H611	DSSSSDSVY	1.4043	0.521	Good	19.19	Non-allergenic	Non toxic
		SSNHSADIY	0.7153	0.085	Good	23.99	Non-allergenic	Non toxic
		FSDIDLSDV	0.5651	0.089	Good	3.33	Non-allergenic	Non toxic
	A0A0H3H6X7	NSDDTSYAR	0.9284	0.211	Good	5.30	Non-allergenic	Non toxic
	A0A0H3H9M6	RTDSRQLDS	2.1329	0.006	Good	6.17	Non-allergenic	Non toxic
Extracellular	A0A0H3H7X1	SEEPITFEY	0.6935	0.416	Good	15.63	Non-allergenic	Non toxic
	A0A0H3HDA8	HLEQIELRY	1.8453	0.191	Good	6.95	Non-allergenic	Non toxic
Cytoplasmic membrane	A0A0H3HF34	QSDEVAAAA	0.9697	0.348	Good	5.40	Non-allergenic	Non toxic
	A0A0H3HFF1	HIELRALS	1.5293	0.667	Good	0.89	Non-allergenic	Non toxic
		SADSASALM	0.5090	0.977	Good	7.80	Non-allergenic	Non toxic
	A0A0H3HCS4	ATSRARQPY	0.4767	0.059	Good	41.78	Non-allergenic	Non toxic

>Vaccine construct 15 MHC I, 13 MHC II

SKKKKEAAKHAYADRSEYAAAYLAEGLRADYAAAYDSSSSDSVYAAAYHSATAAENVAAYTTLDLGVRY
 AAYHLETLELRYAAAYKTNQRSLGYAAAYMSAPGGASYAAAYRLKQDNVIYAAAYQSDEVAAAAAYRSDH
 RPSELAAAYGMAHKADVYAAAYETTISDSRYAAAYTVDALNSEKAAAYISVRQAVSYGPGPGEKKVVMRIRAL
 HKRF GPGPGKVVMRIRALHKRFGA GPGPGQQRVAIARALAMKPS GPGPGLSWKMTKPLRLVKEI GPGPG
 WQGVETDLPQLRALD GPGPGMHNSVMRLTISNKKI GPGPGVTHDLQLAKRMSRQL GPGPGVTHDLQL
 AKRMSRQ GPGPGVTHDLQLAKRMSRQLE GPGPGTVSRFMLLRPGTHKR GPGPGGQRRVLSDISLTLKPGPG
 PGHGWSAVNVDDRWLFR GPGPGRSVTRAFNHMAAGVK

Fig. 2 Amino acid sequence of vaccine construct for *Klebsiella oxytoca*

(Vaccine) (Fig. 9). Three dimensional interaction complex was visualized by BIOVADiscovery Studio (Fig. 10).

Reverse Translation and Expression Vector Cloning of Vaccine Construct

Reverse translation of the primary sequence of vaccine construct was done through JCAT (Grote et al. 2005) by selection *Escherichia coli* as reference organism. The human expression system is different from that of *E. coli* that's why codon adaptation index (CAI) was calculated to ensure high accuracy of translated sequence (Khatoun et al. 2017). CAI is a measure of the synonymous DNA/RNA codon usage bias and reflects the resemblance in the synonymous codon usage of query gene set and that of a reference set (Sharp and Li 1987; Carbone et al. 2003). Vaccine protein was reverse transcribed and codon adaptation index of the adapted codons was found to be 0.991

indicating the higher proportion of most abundant similar codons. The frequency of G and C base pair is called GC content (Carbone et al. 2003). The GC content calculated for the optimized codons was 55.45%. Computational cloning, by using the SnapGene software, was done to check the expression of inserted gene, The optimized codon sequence, of vaccine, obtained by reverse translation, was inserted into pET28a (+) vector plasmid at multiple cloning site (MCS) between XhoI and NdeI restriction sites (Sikarwar et al. 2013; Hasan et al. 2019; Khan et al. 2020). Multiple cloning sites are a feature that allows for the insertion of foreign DNA without disrupting the rest of the plasmid which makes it extremely useful in biotechnology, bioengineering, and molecular genetics (Clark 2005). The pET vector system is a powerful and widely used system for expressing recombinant proteins in *E. coli* (Studier et al. 1990). The gene of interest is initially cloned into the pET vector in a bacteria host that lacks the T7 RNA polymerase gene. This

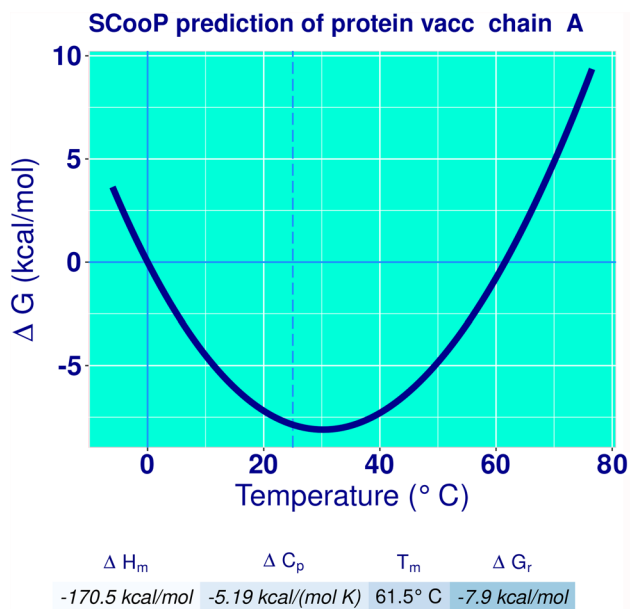


Fig. 3 Thermostability curve of *Klebsiella oxytoca* vaccine protein

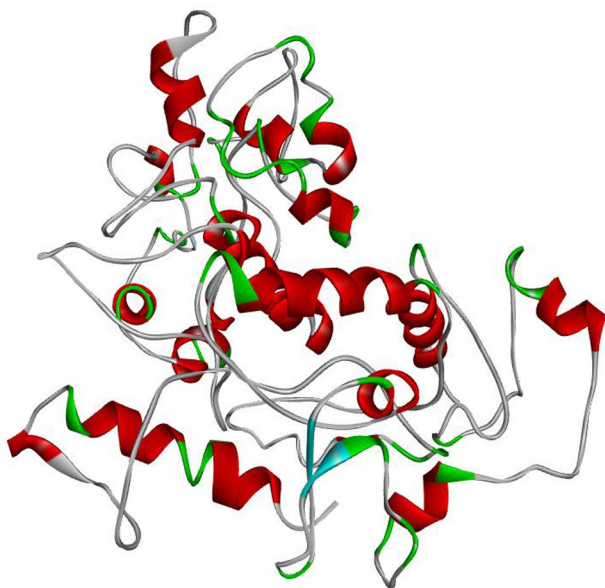


Fig. 4 Predicted 3D structure of *Klebsiella oxytoca* vaccine construct

eliminates plasmid instability due to expression of proteins of interest that may be harmful to host cells (Dubendorf and Studier 1991). The construct codons lacked restriction sites for XhoI and NdeI which were added manually, ensuring its compatibility with vector restriction site, for the cloning purpose (Fig. 11). A clone containing 6586 bp was constructed including the 450 bp of query gene sequence. The query sequence was shown as red region in pET28a (+) vector sequence (Fig. 12).

Table 3 Predicted B cell linear epitopes in vaccine construct for *Klebsiella oxytoca*

Rank	Sequence	Start position	Score
1	GPGGQRRVLSDISLTL	393	0.93
2	RALAMKPSGPGPLSW	243	0.92
3	SRQGGPGTHDLQLAK	348	0.91
3	SDSRYAAYTVDALNSE	162	0.91
3	YGMAHKADVYAAYETT	145	0.91
4	YHLETLELRYAAYKTN	73	0.89
5	PGPGMHNSVMRLTISN	292	0.88
5	PGPGEKKVVMRIRALH	192	0.88
6	PGPGPGHGWSAVNVDD	410	0.87
7	QGVETDLPQLRALDGP	277	0.86
7	QAVSYGPGPGEKKVVM	186	0.86
8	GTHKRGPGGQRRVL	386	0.85
8	VVMRIRALHKRFGAGP	217	0.85

Discussion

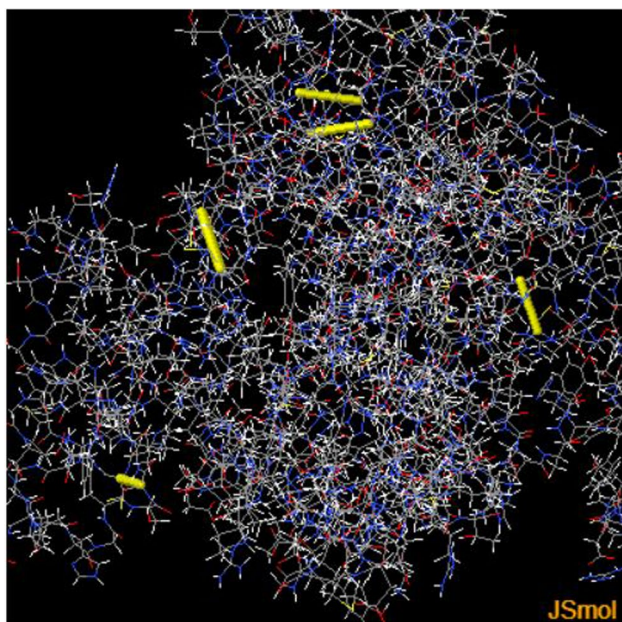
The modern developments in Bioinformatics techniques had opened diversified ways to develop protection against human pathogens (Khan et al. 2019). The use of a vaccine against a particular pathogen is an efficient way to protect against infections. The classical vaccine development process against infectious microorganisms is successfully working in medical science worldwide (Serruto & Rappuoli 2006; Heldens, et al. 2008). However vaccine development by these traditional approaches have becoming more tedious laborious and expensive, which is not suitable in current era of advancement in science (Khan et al. 2019). Computational based methods are employed efficiently to overcome these drawbacks of conventional vaccine development. These computational approaches are rapid, inexpensive, and reliable with particular results (Saha and Raghava 2006). With the passage of time computational strategy for vaccines development of has become lucrative for scientist because of its safety, simplicity and specificity (Sette & Rappuoli 2010). So far, no approved vaccine for the *K. oxytoca* has been produced; an effective vaccine is required to efficiently manage the clinical borne infections by this pathogen. In silico multiepitope vaccine designing have many benefits compared to conventional vaccines (Li et al. 2014). The vaccine designed by using this method allows the immune system of the host to target specifically appropriate antigenic epitopes, thus avoiding non-protective reactions, the autoimmune response from the host and pathogen evasion of the immune system (Wiederstein and Sippl 2007; Li et al. 2014). We used a subtractive proteomics pipeline to identify antigenic proteins for designing vaccine against *K. oxytoca*. CD-HIT tools were used to refine the proteome refined the proteome and resulted in 5470 non-paralogous

Table 4 Predicted discontinuous epitopes in 3D structure of vaccine construct for *Klebsiella oxytoca*

Nos	Residues	Contact numbers	DiscoTope score
1	GLU123, GLU130, GLY238	0	− 1.906, − 3.442, − 0.980
2	SER66, HIS67, THR329, SER66,	1	1.392, 0.663, − 2.111, 0.667
3	GLY65, ARG70, ASN124, ASN176, VAL330, ARG70, ASN124	3	1.621, − 1.187, − 2.030, − 3.616, − 3.040, − 1.044, − 1.73
4	LEU69, GLU173 VAL330	4	− 2.100, − 2.387, − 3.220
5	GLU71, TYR121, PRO236	5	− 1.514, − 3.301, 0.379
6	GLY65, GLU173	6	1.094, − 3.147
7	LYS72, GLU180, THR219, LYS285	7	− 3.456, − 3.119, − 0.831, − 3.436
8	MET68,	8	− 0.689
9	SER73, GLU183, MET68, LYS285	9	− 2.647, − 2.797, − 1.266, − 3.673
10	ALA127, ASN217	10	− 2.900, − 1.729

Table 5 Mutated amino acid pairs in vaccine construct of *Klebsiella oxytoca*

Amino acid				Bond		
Seq#	AA	Seq#	AA	χ^3	Energy (kcal/mol)	Sum B-factors
184	Val	218	Val	− 86.21	0.07	10.42
156	Ala	163	Asp	− 83.27	0.72	12.78
420	Ala	424	Asp	94.41	0.99	14.29
43	Asp	264	Leu	96.71	1.07	11.12
179	Ala	221	Ile	− 76.81	1.07	11.90

**Fig. 5** Disulfide engineered structure of *Klebsiella oxytoca* vaccine construct

proteins at the 80% threshold which were carried on for further analysis (Hassan et al. 2016a,b). The proteins of four location i.e., periplasm, extracellular, outer membrane and

cytoplasmic membrane were selected by PSORTb. The proteins from these four locations were selected on the basis of their importance in vaccine designing (Nagpal et al. 2018) and due to their important role in infection, attachment, and survival of pathogens (Ul Ain et al. 2018). Smaller sized proteins are good candidates for vaccine design (Hassan et al. 2016a,b) so we selected protein with weight less than 110 kDa.

Virulence is one of the vaccine's vital characteristics. Virulent proteins can potentially evolve the infection pathways as compared to non-virulent proteins (Naz et al. 2015). We selected virulent protein from VFDB generated results. Human homologs proteins were removed, to avoid autoimmune response, followed by the removal of paralogs and proteins that are not essential for the survival of the pathogen (Solank et al. 2021). We removed all the protein having more than 30% identity with human protein. The final set of screened proteins subjected to immunoinformatics tools for designing an effective vaccine. MHCs, also called human leukocyte alleles (HLA) are among the most polymorphic proteins in higher vertebrates, with more than 6000 class I and class II MHC molecules listed in IMGT/HLA. The selections of these alleles are the prerequisite for vaccine development. MHC class I and MHC class II binding epitopes were predicted at the first step of the immunoinformatics pipeline. The

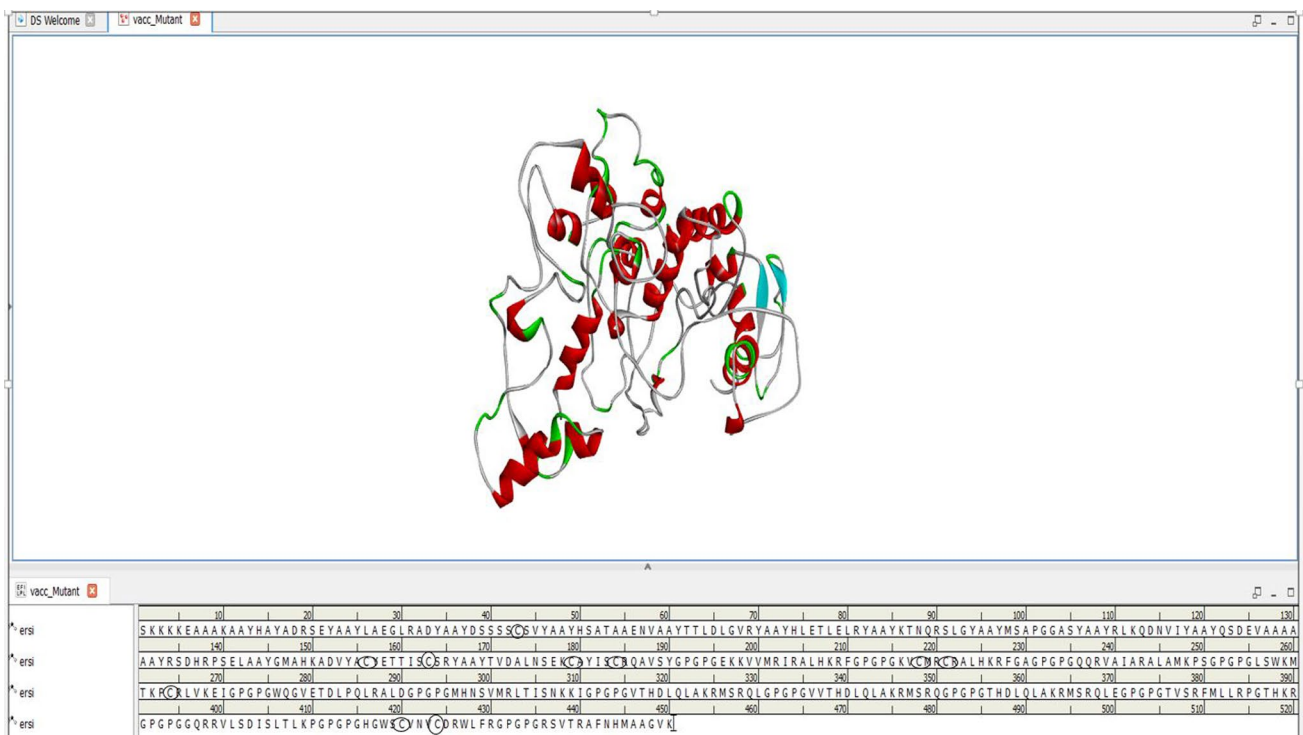


Fig. 6 Amino acid sequence of *Klebsiella oxytoca* vaccine construct with cysteine mutations

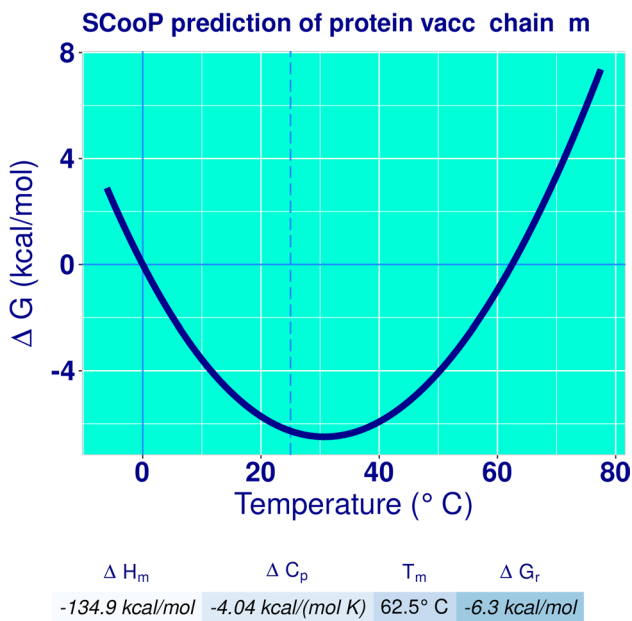


Fig. 7 Thermostability curve of *Klebsiella oxytoca* vaccine protein after disulfide engineering

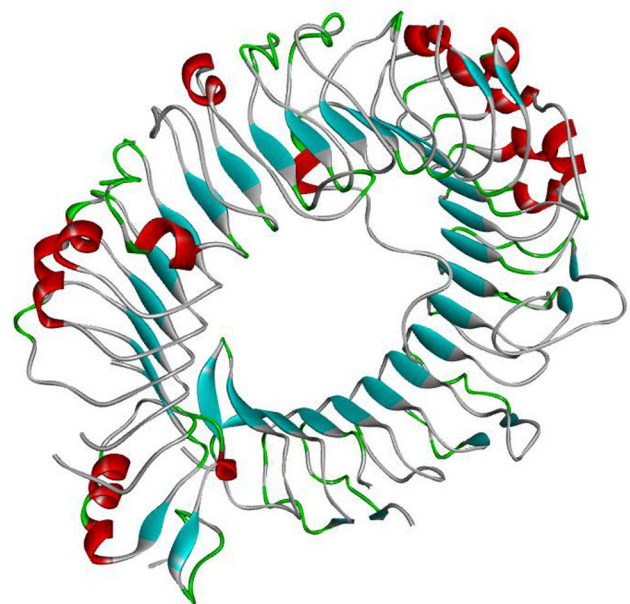


Fig. 8 Chain A of human toll like receptor 2 (PDB ID: 6NIG)

immune system has T cells that readily identify MHC class I and MHC class II epitopes. MHC molecules are present on antigenic cell surface recognized by the T cell receptors (Chen et al. 2005). Class I MHCs protein are present

on almost all the nucleated body cells (Sanchez-Trincado et al. 2017). MHC class I represents the CTL, peptides that have been through the cytosolic pathway are represented by MHC-I, these peptides are epitopes, either antigenic portion of pathogen or endogenous proteins (Wieczorek

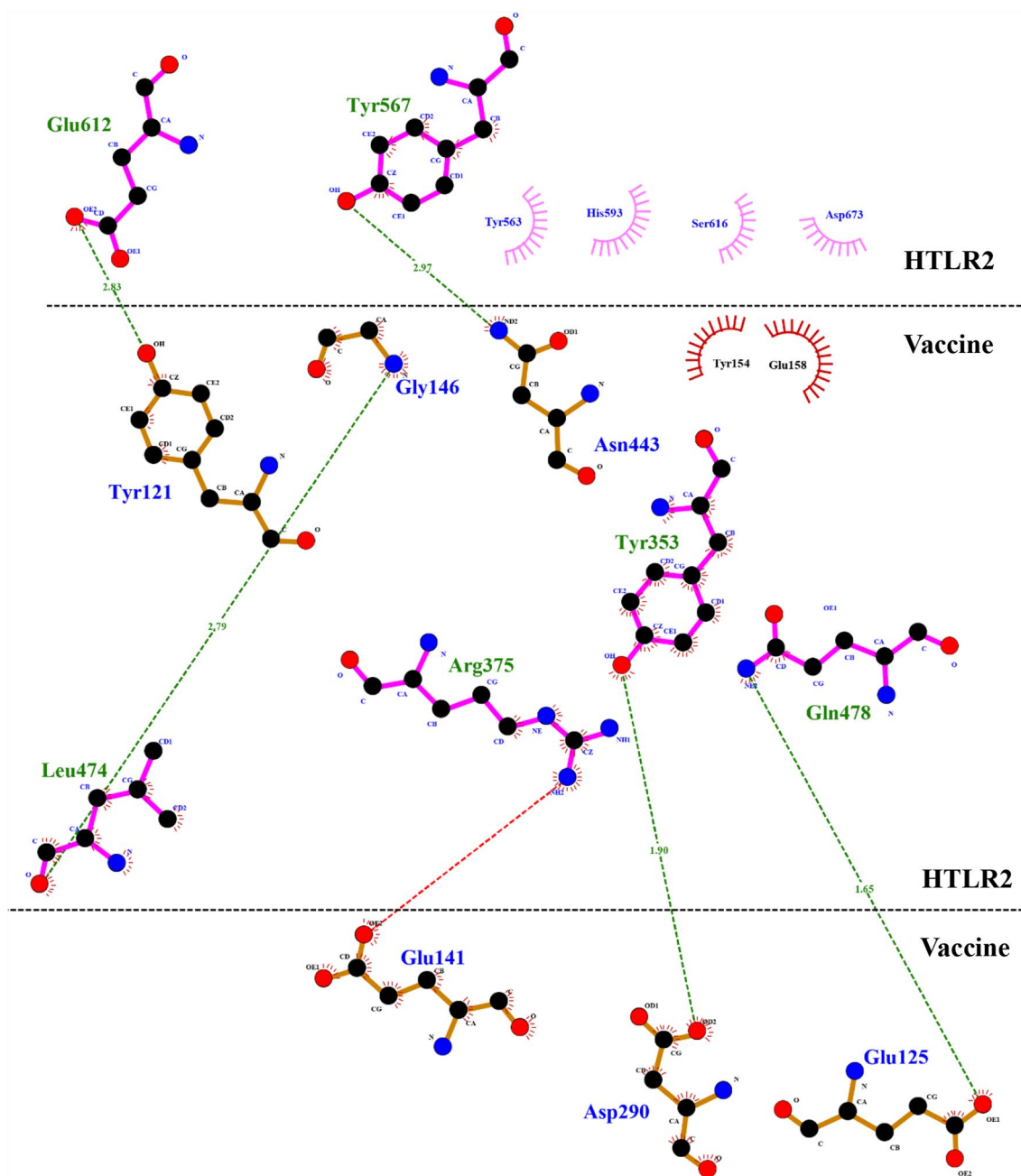


Fig. 9 Two dimensional interactions in docked complex of human TLR2 and *Klebsiella oxytoca* vaccine construct

et al. 2017). MHC class II represents HTLs, peptides that have been through the endocytic pathways, the peptides are antigenic portions of the pathogen surface (Wosen et al. 2018). The oligopeptides from the pathogen protein bind to MHC II protein and these bond pathogenic proteins are recognized by T-cells (Jones et al. 2006). Appropriate number of the HTL epitopes results in the induced pathogen specific immune response triggered by T cell. Class II MHCs are expressed on antigen presenting cells (APCs) such as B cells, macrophages and dendritic cells (Nielsen

et al. 2003). In the current study we predicted 1670 CTL and 1270 HTL epitopes.

The final vaccine construct consisted of a good number of B and T cell epitopes. Thermostability of a vaccine is one of the most important property of a successful vaccine (Yadav, et al 2021a) and thermostability was checked by SCooP. High thermal stability of a protein can be judged by values of standard enthalpy of folding measured at the melting temperature intervals (ΔH_m), in combination with T_m values (Pucci & Rooman 2016).

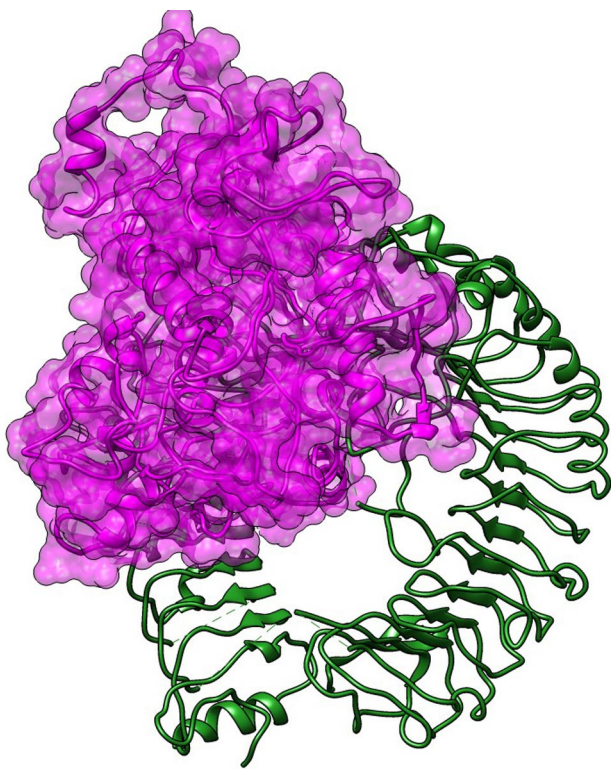


Fig. 10 Three dimensional structure of docked complex of human TLR2 and *Klebsiella oxytoca* vaccine construct (vaccine construct: purple and solid, HTLR2: green)

To attain thermal stability, proteins need large negative values of ΔH that make the value of ΔG more negative, moreover a less negative value of ΔC_p increases the T_m values and in turn the thermal stability. The large T_m values with large negative ΔH and ΔG values give evidence for the observed thermal stability of the enzymes standard heat capacity of folding (ΔC_p), and free energy change (ΔG) correlated with the protein folding with respect to temperature (Yadav et al. 2021a). In silico studies have proved high melting temperature as an indicator of thermostability (Khairudin and Mazlan 2013). The values of $T_m = 61.5$ °C, $\Delta G_r = -7.9$ kcal/mol, $\Delta C_p = -5.19$ kcal/(mol K) and $\Delta H_m = -170.5$ kcal/mol shows stability of

vaccine protein. The vaccine was thermostable, non-allergen, nontoxic and antigenic in nature which reflects the standards of a good vaccine (Liu and Irvine 2015). The tertiary structure provides information such as a normal function of the protein, dynamics, interaction with other proteins and ligands (Mittag et al. 2010). The suitable templates for homology modeling of vaccine construct were not found so, online available server I-TASSER was used for tertiary structure prediction by using multiple templates (Zhang 2008). Five models were generated for query sequence and good quality structure with high C score was selected. C-score is a confidence score for quality estimation of I-TASSER generated models. It is calculated on the basis of significance of threading template alignments and the convergence parameters of the structure assembly simulations. C-score is generally in the range between -5 and 2 and high value reflects the high confidence level on predicted structure (Yang et al. 2015). The PDB files of vaccine construct and HTLR2 were submitted to PatchDock server for protein–protein docking. PatchDock is a geometry-based molecular docking algorithm (Duhovny et al. 2002). The results received via email in the form of 20 complex solution PDB files. Top 10 complex solution were selected for refinement. This algorithm is aimed at finding docking transformations that yield good molecular shape complementarity (Schneidman-Duhovny et al. 2005) a transformation file was also generated along with the complex PDB files. The PDB file of vaccine and HTLR2 and transformation file for top ten complexes were submitted to FireDock for refinement. The methods used by this server include optimization of side-chain conformations and rigid-body orientation and allows a high-throughput refinement (Mashiach et al. 2008). The protocol of FireDock thoroughly tested (Andrusier et al. 2007) on protein–protein docking benchmarks (Mintseris et al. 2005) which makes the results more reliable. JCat server was used for the optimization of codons, in order to increase the expression of vaccine protein in the *E. coli* system. The vaccine was reverse translated and CAI and GC content were calculated by this server. CAI is a measure of the synonymous DNA/RNA codon usage bias and reflects the resemblance in the synonymous codon usage of query gene

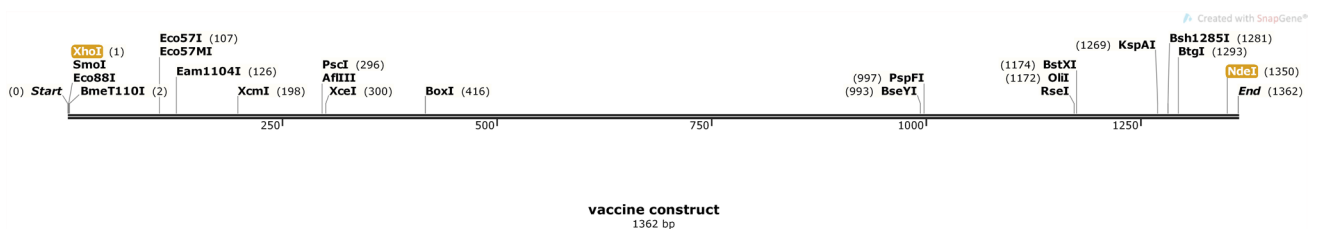


Fig. 11 The engineered *Klebsiella oxytoca* vaccine construct map for cloning

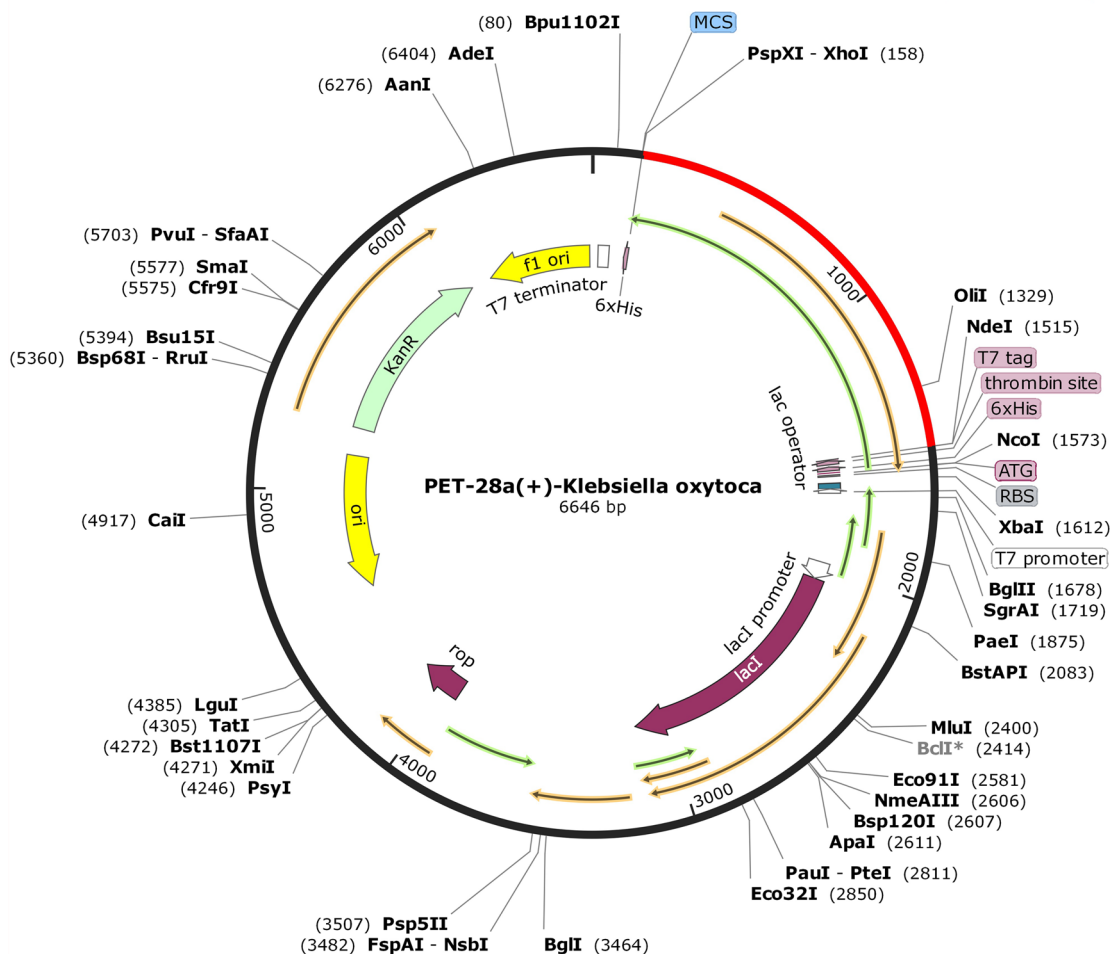


Fig. 12 The cloned pET28a (+) vector showing *Klebsiella oxytoca* vaccine construct as red region

set and that of a reference set (Sharp & Li 1987; Carbone et al. 2003). The CAI was 0.991 and GC content 55.4% have been reported for our designed vaccine. These values (CAI and GC content) support the fact that vaccine will be highly expressed in *E. coli* expression system (Khan et al. 2019). Vaccines are generally tested on model organisms before getting approval for human use. However, most of the time, the vaccine may show promising results in model organisms but can't show same successful results in human being. This might be because of the complicated human immune system. The in silico designed vaccines are computationally designed strictly by using human proteins and immune system which make it specific and effective.

Conclusion

We presented a study about the prediction of an effective, non-toxic and non-allergenic vaccine against of *K. oxytoca* strain ATCC 8724. A smart and comprehensive approach

of subtractive proteomics was employed to screen out the potential 28 peptides (CTL and HTL epitopes) sequences out of 5483 protein. Epitopes were joined with appropriate linkers to design 450 aa long vaccine construct. The predicted 3D structure of vaccine construct was further tested for stability and B cell epitopes. It showed good binding affinity with human Toll like receptor protein. The vaccine construct was reverse transcribed and confirmed its expression by inserting in pET28a (+) vector. This predicted vaccine construct showed stability up to 61.5 °C temperature. This computational investigation can be used as baseline study for vaccine designating in wet labs. The preliminary study result can save time and resources in vaccine production experiments.

Author Contributions QY: Conceived idea, designed methodology, helped in analysis and Manuscript writing. HA: Helped in data interpretation and write-up. SB: Helped in result interpretation drafting. RR: Execution of methodology (partly). MSK: Execution of methodology

(partly). HA: drafting and analysis (partly). AM: Helped in manuscript writing and analysis.

Funding No funding was obtained for this research work.

Data Availability The data that support the findings of this study are available from the corresponding author upon reasonable request.

Declarations

Conflict of interest The authors declared no conflict of interest.

References

- Andrusier N, Nussinov R, Wolfson HJ (2007) FireDock: fast interaction refinement in molecular docking. *Proteins* 69:139–159
- André FE (2003) Vaccinology: past achievements, present roadblocks and future promises. *Vaccine* 21(7–8):593–595
- Arai R, Ueda H, Kitayama A, Kamiya N, Nagamune T (2001) Design of the linkers which effectively separate domains of a bifunctional fusion protein. *Protein Eng* 14(8):529–532
- Barh D, Barve N, Gupta K, Chandra S, Jain N, Tiwari S, Azevedo V (2013) Exoproteome and secretome derived broad spectrum novel drug and vaccine candidates in *Vibrio cholerae* targeted by Piper betel derived compounds. *PLoS ONE* 8(1):e52773
- Baseer S, Ahmad S, Ranaghan KE, Azam SS (2017) Towards a peptide-based vaccine against *Shigella sonnei*: a subtractive reverse vaccinology based approach. *Biologicals* 50:87–99
- Basto AP, Leitão A (2014) Targeting TLR2 for vaccine development. *J Immunol Res*. <https://doi.org/10.1155/2014/619410>
- Borrello S, Nicolo C, Delogu G, Pandolfi F, Ria F (2011) TLR2: a crossroads between infections and autoimmunity? *Int J Immunopathol Pharmacol* 24(3):549–556
- Bouvier A, Chapline J, Boerner R, Jeyarajah S, Cook S, Acharya PS, Shepard SR (2003) Identifying and modulating disulfide formation in the biopharmaceutical production of a recombinant protein vaccine candidate. *J Biotechnol* 103(3):257–271
- Carbone A, Zinovyev A, Képes F (2003) Codon adaptation index as a measure of dominating codon bias. *Bioinformatics* 19(16):2005–2015
- Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y, Jin Q (2005) VFDB: a reference database for bacterial virulence factors. *Nucleic Acids Res* 33(Suppl_1):D325–D328
- Christensen JK, Lamberth K, Nielsen M, Lundegaard C, Worning P, Lauemøller SL, Lund O (2003) Selecting informative data for developing peptide-MHC binding predictors using a query by committee approach. *Neural Comput* 15(12):2931–2942
- Clark DP (2005) *Molecular biology*. Academic, p 611. ISBN 0-12-175551-7
- Craig DB, Dombkowski AA (2013) Disulfide by Design 2.0: a web-based tool for disulfide engineering in proteins. *BMC Bioinform* 14(1):1–7
- Decré D, Burghoffer B, Gautier V, Petit JC, Arlet G (2004) Outbreak of multi-resistant *Klebsiella oxytoca* involving strains with extended-spectrum β -lactamases and strains with extended-spectrum activity of the chromosomal β -lactamase. *J Antimicrob Chemother* 54(5):881–888
- Delany I, Rappuoli R, Seib KL (2013) Vaccines, reverse vaccinology, and bacterial pathogenesis. *Cold Spring Harb Perspect Med* 3:a012476
- Dhanda SK, Vir P, Raghava GP (2013) Designing of interferon-gamma inducing MHC class-II binders. *Biol Direct* 8:30
- Dorosti H, Eslami M, Negahdaripour M, Ghoshoon MB, Gholami A, Heidari R, Ghasemi Y (2019) Vaccinomics approach for developing multi-epitope peptide pneumococcal vaccine. *J Biomol Struct Dyn* 37(13):3524–3535
- Doytchinova IA, Flower DR (2007) VaxiJen: a server for prediction of protective antigens, tumor antigens and subunit vaccines. *BMC Bioinform* 8(1):1–7
- Dubendorf JW, Studier FW (1991) Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor. *J Mol Biol* 219(1):45–59
- Duhovny D, Nussinov R, Wolfson HJ (2002) Efficient unbound docking of rigid molecules. In: Guigo R, Gusfield D (eds) *Proceedings of the fourth international workshop on algorithms in bioinformatics, 17–21 September 2002, Rome, Italy*. Springer, pp 185–200
- Fleri W, Paul S, Dhanda SK, Mahajan S, Xu X, Peters B, Sette A (2017) The immune epitope database and analysis resource in epitope discovery and synthetic vaccine design. *Front Immunol* 8:278
- Flynn JL, Chan J, Triebold KJ, Dalton DK, Stewart TA, Bloom BR (1993) An essential role for interferon gamma in resistance to *Mycobacterium tuberculosis* infection. *J Exp Med* 178:2249–2254
- De la Fuente J, Kopacek P, Lew-Tabor A, Maritz-Olivier C (2016) Strategies for new and improved vaccines against ticks and tick-borne diseases. *Parasite Immunol* 38:754–769
- Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD, Bairoch A (2005) Protein identification and analysis tools on the ExPASy server. In: *The proteomics protocols handbook*. Humana Press, Totowa, pp 571–607
- Gorkiewicz G (2009) Nosocomial and antibiotic-associated diarrhoea caused by organisms other than *Clostridium difficile*. *Int J Antimicrob Agents* 33:S37–S41
- Grote A, Hiller K, Scheer M, Münch R, Nörtemann B, Hempel DC, Jahn D (2005) JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Res* 33(Suppl_2):W526–W531
- Guan P, Doytchinova IA, Zygouri C, Flower DR (2003) MHCPreD: a server for quantitative prediction of peptide-MHC binding. *Nucleic Acids Res* 31:3621–3624
- Gupta S, Kapoor P, Chaudhary R, Gautam A, Kumar R, Raghava GP, Open Source Drug Discovery Consortium (2013) In silico approach for predicting toxicity of peptides and proteins. *PLoS ONE* 8(9):e73957
- Hasan M, Ghosh PP, Azim KF, Mukta S, Abir RA, Nahar J, Khan MMH (2019) Reverse vaccinology approach to design a novel multi-epitope subunit vaccine against avian influenza A (H7N9) virus. *Microb Pathog* 130:19–37
- Hassan A, Naz A, Obaid A, Paracha RZ, Naz K, Awan FM, Ali A (2016a) Pangenome and immuno-proteomics analysis of *Acinetobacter baumannii* strains revealed the core peptide vaccine targets. *BMC Genomics* 17(1):1–25
- Hassan A, Naz A, Obaid A, Paracha RZ, Naz K, Awan FM, Muhammad SA, Janjua HA, Ahmad J, Ali A (2016b) Pangenome and immuno-proteomics analysis of *Acinetobacter baumannii* strains revealed the core peptide vaccine targets. *BMC Genomics* 17(1):732
- Hattotuwigama CK, Guan P, Doytchinova IA, Zygouri C, Flower DR (2004) Quantitative online prediction of peptide binding to the major histocompatibility complex. *J Mol Graph Model* 22:195–207
- Heldens JGM, Patel JR, Chanter N, Ten Thij GJ, Gravendijck M, Schijns VEJC, Schetters TP (2008) Veterinary vaccine development from an industrial perspective. *Vet J* 178(1):7–20
- Hoffmann KM, Deutschmann A, Weitzer C, Joainig M, Zechner E, Högenauer C, Hauer AC (2010) Antibiotic-associated hemorrhagic colitis caused by cytotoxin-producing *Klebsiella oxytoca*. *Pediatrics* 125(4):e960–e963

- Jones EY, Fugger L, Strominger JL, Siebold C (2006) MHC class II proteins and disease: a structural perspective. *Nat Rev Immunol* 6(4):271–282
- Kelly DF, Rappuoli R (2005) Reverse vaccinology and vaccines for serogroup B *Neisseria meningitidis*. In: Hot topics in infection and immunity in children II. Springer, Boston, pp 217–223
- Khairudin NBA, Mazlan NSF (2013) Molecular docking study of beta-glucosidase with cellobiose, cellotetraose and cellotriose. *Bioinformatics* 9:813–815
- Khan S, Khan A, Rehman AU, Ahmad I, Ullah S, Khan AA, Ali SS, Gul S, Wei DQ (2019) Immunoinformatics and structural vaccinology driven prediction of multi-epitope vaccine against Mayaro virus and validation through in silico expression. *Infect Genet Evol J Mol Epidemiol Evol Genet Infect Dis* 73:390–400
- Khan S, Ali SS, Zaheer I, Saleem S et al (2020) Proteome-wide mapping and reverse vaccinology-based B and T cell multi-epitope subunit vaccine designing for immune response reinforcement against *Porphyromonas gingivalis*. *J Biomol Struct Dyn* 1:15
- Khatoun N, Pandey RK, Prajapati VK (2017) Exploring *Leishmania* secretory proteins to design B and T cell multi-epitope subunit vaccine using immunoinformatics approach. *Sci Rep* 7(1):1–12
- Kringelum JV, Lundegaard C, Lund O, Nielsen M (2012) Reliable B cell epitope predictions: impacts of method development and improved benchmarking. *PLoS Comput Biol* 8(12):e1002829
- Kringelum JV, Nielsen M, Padkjaer SB, Lund O (2013) Structural analysis of B-cell epitopes in antibody: protein complexes. *Mol Immunol* 53:24–34
- Larsen MV, Lundegaard C, Lamberth K, Buus S, Lund O, Nielsen M (2007) Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC Bioinform* 8(1):1–12
- Laskowski RA, Swindells MB (2011) LigPlot+: multiple ligand–protein interaction diagrams for drug discovery. *J Chem Inf Model* 51(10):2778–2786
- Li W, Godzik A (2006) CD-Hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22(13):1658–1659
- Li W, Joshi M, Singhanian S, Ramsey K, Murthy A (2014) Peptide vaccine: progress and challenges. *Vaccines* 2(3):515–536
- Liu H, Irvine DJ (2015) Guiding principles in the design of molecular bioconjugates for vaccine applications. *Bioconjug Chem* 26(5):791–801
- Liu T, Wang Y, Luo X, Li J, Reed SA, Xiao H, Young S, Schultz PG (2016) Enhancing protein stability with extended disulfide bonds. *Proc Natl Acad Sci USA* 113(21):5910–5915
- Mashiach E, Schneidman-Duhovny D, Andrusier N, Nussinov R, Wolfson HJ (2008) FireDock: a web server for fast interaction refinement in molecular docking. *Nucleic Acids Res* 36(Suppl_2):W229–W232
- Mehla K, Ramana J (2017) Surface proteome mining for identification of potential vaccine candidates against *Campylobacter jejuni*: an in silico approach. *Funct Integr Genomics* 17(1):27–37
- Michalik MM, Djahanshiri B, Leo JC, Linke D (2016) Reverse vaccinology: the pathway from genomes and epitope predictions to tailored recombinant vaccines. *Methods Mol Biol* 1403:87–106
- Mintseris J, Wiehe K, Pierce B, Anderson R, Chen R, Janin J, Weng Z (2005) Protein–protein docking benchmark 2.0: an update. *Proteins* 60:214–216
- Mittag T, Kay LE, Forman-Kay JD (2010) Protein dynamics and conformational disorder in molecular recognition. *J Mol Recognit Interdiscip J* 23(2):105–116
- Muruato LA, Tapia D, Hatcher CL, Kalita M, Brett PJ, Gregory AE et al (2017) The use of reverse vaccinology in the design and construction of nano-glycoconjugate vaccines against *Burkholderia pseudomallei*. *Clin Vaccine Immunol* 24(11):e00206–e217
- Nagpal G, Usmani SS, Raghava GPS (2018) A web resource for designing subunit vaccine against major pathogenic species of bacteria. *Front Immunol* 9:2280
- Naz A, Awan FM, Obaid A, Muhammad SA, Paracha RZ, Ahmad J, Ali A (2015) Identification of putative vaccine candidates against *Helicobacter pylori* exploiting exoproteome and secretome: a reverse vaccinology based approach. *Infect Genet Evol* 32:280–291
- Nielsen M, Lundegaard C, Worning P, Lauemøller SL, Lamberth K, Buus S, Brunak S, Lund O (2003) Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. *Protein Sci* 12(5):1007–1017. <https://doi.org/10.1110/ps.0239403>
- Patronov A, Doytchinova I (2013) T-cell epitope vaccine design by immunoinformatics. *Open Biol* 3(1):120139
- Paul S, Sidney J, Sette A, Peters B (2016) TepiTool: a pipeline for computational prediction of T cell epitope candidates. *Curr Protoc Immunol* 114(1):18–19
- Podschun R, Ullmann U (1998) *Klebsiella* spp. as nosocomial pathogens: epidemiology, taxonomy, typing methods, and pathogenicity factors. *Clin Microbiol Rev* 11(4):589–603
- Pucci F, Rooman M (2016) Towards an accurate prediction of the thermal stability of homologous proteins. *J Biomol Struct Dyn* 34(5):1132–1142
- Rappuoli R (2000) Reverse vaccinology. *Curr Opin Microbiol* 3(5):445–450
- Reiss I, Borkhardt A, Füssle R, Sziegoleit A, Gortner L (2000) Disinfectant contaminated with *Klebsiella oxytoca* as a source of sepsis in babies. *Lancet* 356(9226):310
- Reljic R (2007) IFN-gamma therapy of tuberculosis and related infections. *Interferon Cytokine Res* 27:353–364
- Rock FL, Hardiman G, Timans JC, Kastelein RA, Bazan JF (1998) A family of human receptors structurally related to *Drosophila* Toll. *Proc Natl Acad Sci USA* 95(2):588–593
- Ruppert J, Sidney J, Celis E, Kubo RT, Grey HM, Sette A (1994) Prominent role of secondary anchor residues in peptide binding to HLA-A2.1 molecules. *Cell* 74:929–934
- Saadi M, Karkhah A, Nouri HR (2017) Development of a multi-epitope peptide vaccine inducing robust T cell responses against brucellosis using immunoinformatics based approaches. *Infect Genet Evol* 51:227–234
- Saha S, Raghava GPS (2006) Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins Struct Funct Bioinform* 65(1):40–48
- Sanchez-Trincado JL, Gomez-Perosanz M, Reche PA (2017) Fundamentals and methods for T- and B-cell epitope prediction. *J Immunol Res*. <https://doi.org/10.1155/2017/2680160>
- Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ (2005) PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res* 33(Suppl_2):W363–W367
- Serruto D, Rappuoli R (2006) Post-genomic vaccine development. *FEBS Lett* 580(12):2985–2992
- Sette A, Rappuoli R (2010) Reverse vaccinology: developing vaccines in the era of genomics. *Immunity* 33(4):530–541
- Sharma N, Patiyal S, Dhali A, Pande A, Arora C, Raghava GP (2020) AlgPred 2.0: an improved method for predicting allergenic proteins and mapping of IgE epitopes. *Brief Bioinform*. <https://doi.org/10.1093/bib/bbaa294>
- Sharp PM, Li WH (1987) The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acid Res* 15:1281–1295
- Shirai H, Prades C, Vita R, Marcatili P, Popovic B, Xu J, Overington JP, Hirayama K, Soga S, Tsunoyama K, Clark D, Lefranc M-P, Ikeda K (2014) Antibody informatics for drug discovery. *Biochim Biophys Acta* 1844(11):2002–2015
- Shtrichman R, Samuel CE (2001) The role of gamma interferon in antimicrobial immunity. *Curr Opin Microbiol* 4:251–259
- Sikarwar J, Kaushik S, Sinha M, Kaur P, Sharma S, Singh TP (2013) Cloning, expression, and purification of nucleoside diphosphate kinase from *Acinetobacter baumannii*. *Enzyme Res*. <https://doi.org/10.1155/2013/597028>

- Simoons-Smit AM, Verweij-van Vught AMJJ, MacLaren DM (1986) The role of K antigens as virulence factors in *Klebsiella*. *J Med Microbiol* 21(2):133–137
- Singh L, Cariappa MP, Kaur M (2016) *Klebsiella oxytoca*: an emerging pathogen? *Med J Armed Forces India* 72:S59–S61
- Solanki V, Sharma S, Tiwari V (2021) Subtractive proteomics and reverse vaccinology strategies for designing a multiepitope vaccine targeting membrane proteins of *Klebsiella pneumoniae*. *Int J Pept Res Ther* 27:1177–1195
- Studier FW, Rosenberg AH, Dunn JJ, Dubendorff JW (1990) Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol* 185:60–89
- Tannock GA, Kim H, Xue L (2020) Why are vaccines against many human viral diseases still unavailable; an historic perspective? *J Med Virol* 92(2):129–138
- Ul Ain Q, Ahmad S, Azam SS (2018) Subtractive proteomics and immunoinformatics revealed novel B-cell derived T-cell epitopes against *Yersinia enterocolitica*: an etiological agent of Yersiniosis. *Microb Pathog* 125:336–348
- Vali L, Dashti AA, El-Shazly S, Jadaon MM (2015) *Klebsiella oxytoca* with reduced sensitivity to chlorhexidine isolated from a diabetic foot ulcer. *Int J Infect Dis* 34:112–116
- Wang P, Sidney J, Kim Y, Sette A, Lund O, Nielsen M, Peters B (2010) Peptide binding predictions for HLA DR, DP and DQ molecules. *BMC Bioinform* 11:568
- Wieczorek M, Abualrous ET, Sticht J, Álvaro-Benito M, Stolzenberg S, Noé F, Freund C (2017) Major histocompatibility complex (MHC) class I and MHC class II proteins: conformational plasticity in antigen presentation. *Front Immunol* 8:292
- Wiederstein M, Sippl MJ (2007) ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res* 35(Web Server Issue):W407–W410
- Wosen JE, Mukhopadhyay D, Macaubas C, Mellins ED (2018) Epithelial MHC class II expression and its role in antigen presentation in the gastrointestinal and respiratory tracts. *Front Immunol* 9:2144
- Wu KM, Li LH, Yan JJ, Tsao N, Liao TL, Tsai HC et al (2009) Genome sequencing and comparative analysis of *Klebsiella pneumoniae* NTUH-K2044, a strain causing liver abscess and meningitis. *J Bacteriol* 191(14):4492–4501
- Yadav PD, Ella R, Kumar S, Patil DR, Mohandas S, Shete AM, Bhargava B (2021a) Immunogenicity and protective efficacy of inactivated SARS-CoV-2 vaccine candidate, BBV152 in rhesus macaques. *Nat Commun* 12(1):1–11
- Yadav S, Pandey AK, Dubey SK (2021b) Molecular modeling, docking and simulation dynamics of β -glucosidase reveals high-efficiency, thermo-stable, glucose tolerant enzyme in *Paenibacillus lautus* BHU3 strain. *Int J Biol Macromol* 168:371–382
- Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y (2015) The I-TASSER Suite: protein structure and function prediction. *Nat Methods* 12:8
- Yinnon AM, Butnaru A, Raveh D, Jerassy Z, Rudensky B (1996) *Klebsiella bacteraemia*: community versus nosocomial infection. *QJM Int J Med* 89(12):933–942
- Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, Brinkman FS (2010) PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* 26(13):1608–1615
- Zhang Y (2008) I-TASSER server for protein 3D structure prediction. *BMC Bioinform* 9:40

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