Designing and Structure Evaluation of Multi-Epitope Vaccine Against ETEC and EHEC: An In Silico Approach

Article *in* Protein and Peptide Letters · October 2015

DOI: 10.2174/0929866522666151026122116

CITATIONS
4

READS
288

6 authors, including:

Fatemeh Shojaei Jeshvaghani
National Institute of Genetic Engineering and Biotechnology
4 PUBLICATIONS **9** CITATIONS
SEE PROFILE

Jafar Amani
Baqiyatallah University of Medical Sciences, Tehran, Iran
214 PUBLICATIONS **1,344** CITATIONS
SEE PROFILE

Iman Rad
Stem cell technology research center
30 PUBLICATIONS **207** CITATIONS
SEE PROFILE

Mahya Jafari
National Institute of Genetic Engineering and Biotechnology
25 PUBLICATIONS **49** CITATIONS
SEE PROFILE

Send Orders for Reprints to reprints@benthamscience.ae

Protein & Peptide Letters, 2016, *23*, 000-000

Designing and Structure Evaluation of Multi-Epitope Vaccine Against ETEC and EHEC, an *In Silico* Approach

Fatemeh Shojaei jeshvaghani[1*], Ahmad Karimi Rahjerdi [1*], Jafar Amani[2], Iman Rad[3], Mahyat Jafari[1] and Ali Hatef Salmanian[1**]

[1]*Department of Plant Biotechnology. National Institute of Genetic Engineering and Biotechnology (NIGEB), Tehran, Iran;* [2]*Applied Microbiology Research Center, Baqiyatallah University of Medical Sciences, Tehran, Iran;* [3]*Molecular Cell Biology Department, Stem Cell Technology Research Center, Tehran, Iran*

Abstract: Diarrheal diseases represent a major health problem in developing countries. Several viruses and bacterial agents, such as Enterotoxigenic *Escherichia coli* (ETEC) and Enterohemorrhagic *Escherichia coli* (EHEC) are responsible for human enteric infections.

In humans, EHEC infections result in bloody or non-bloody diarrhea, which may be complicated by haemorrhagic colitis and haemolytic uraemic syndrome (HUS). Infection by ETEC is accompanied by a non inflammatory watery diarrhea. *E. coli* follows a common strategy of infection: colonization on a mucosal site, evasion of host defenses, multiplication, and host damage. Intimin, Stx, Lt and Cfa proteins are the virulence factors expressed by these strains. Antibiotic treatment is generally not recommended for most cases of diarrhea, since antibiotic usage may lead to antibiotic resistance in ETEC and may also change the intestinal flora. We hypothesized that the chimeric forms of these effectors as vaccine candidates would reduce the colonization of bacteria. This study is based on an *in silico* analysis of chimeric protein structure and its stability and solubility. The secondary and tertiary structures of selected domains were also predicted. Moreover, T and B cell epitopes were mapped. Protein structure Prediction showed that each domain of antigen was separated completely also stable for recombinant expression. We believe that this chimeric vaccine candidate is effective for prevention of bacteria caused diarrheal diseases.

Keywords: Bioinformatics analysis, chimeric vaccine, diarrheal diseases.

1. INTRODUCTION

Diarrheal diseases represent a major health problem in developing countries and also serious risk for travelers who visit these countries. It has been estimated nearly 2–4 billion episodes of diarrhea result in about 2.2 million deaths. The highest incidence for this disease is in children below the age of 5 years [1]. Multiple pathogens are responsible for this suffering and death, among these the diarrhea, *Escherichia coli* has been known as the most common bacterial cause of diarrhea.

Enterohemorrhagic *Escherichia coli* (EHEC, Shiga toxin-producing *E. coli* or STEC) is an *Escherichia coli* pathotype that causes diarrhea and also in some cases hemolytic–uremic syndrome (HUS). HUS is mostly characterized by thrombocytopenia, hemolytic anemia and renal failure [2, 3]. The most prominent serotype within the EHEC group is O157:H7 that frequently implicated in HUS [4-7]. EHEC strains are characterized by the ability to form attaching and effacing (A/E) lesions on the surface of epithelial cells in the gastrointestinal tract of mammals and are also known as the shiga toxins producers [8-10]. Shiga toxins (Stxs, also called Vero toxins) made and secreted by *E. coli* O157:H7, and is considered to be responsible for the development of HUS [11, 12]. Shiga toxins belong to the AB5 family of the toxins. The B subunit of this molecule would interact with globotriaosylceramides (Gb3s) which are found on the surface of human intestinal mucosa and kidney epithelial cells. The interaction between these two molecules results in the internalization of the toxin where the A subunit activation causes cell death [11, 13-15]. The Stx molecules are divided into two groups of Stx1 and Stx2 according to their antigenic and genetic differences. Nevertheless, Stx2 is known to be more interesting in vaccine exploration than Stx1, since it is produced by nearly all of the EHEC serotypes and also associated with HUS [12, 16, 17]. Recent *in vitro* studies demonstrated that Stx2 increases cell surface expression of nucleolin- a eukaryotic receptor for the *E. coli* O157:H7 adhesion molecule, “intimin” [11]. Intimin is amajor outer membrane protein of *E. coli* O157:H7 and is the first protein to be associated with the A/E activity [8]. This molecule is the product of the *E. coli* attach and efface gene (*eae*), which is located within the~43-kb pathogenicity island, called the locus of enterocyte effacement (LEE). The carboxy-terminal portion

*Address correspondence to this author at the Department of Plant Biotechnology, National Institute of Genetic Engineering and Biotechnology (NIGEB), Shahrak-e-Pajoohesh, 15th Km, Tehran -Karaj Highway, P.O. Box: 14965-161, Tehran, Iran; Tel: (+) 98 21 44580365;
Fax: (+) 98 21 44580395; E-mail: salman@nigeb.ac.ir

0929-8665/16 $58.00+.00

of intimin binds to the bacterium-encoded translocated intimin receptor (Tir) and the host cell receptor, nucleolin [17-21].

The other minor pathotype of *Escherichia coli* is Enterotoxigenic (ETEC), which at the first step of infection could adhere to intestinal epithelial cells via a heterogeneous group of surface proteins, termed as colonization factors (CFs) [13, 22]. More than 25 types of CFAs have been described. CFA/I is one of the most commonly antigenic types which has been identified in humans with diarrheas[3]. CFA/I is composed of two types of proteins. Thousands of repeats of CfaBs,as a major structural subunit forms a stalk that supports one or few CfaE minor tip subunits [3, 23-26].

Following the initial adhesion and colonization, ETEC strains cause diarrhea by producing plasmid-encoded heat-labile (LT) and/or heat-stable (ST) enterotoxins.LT,similarly to Shiga toxins is a member of AB5 family of toxins which are heterohexameric molecules consisting of five B subunits and a single A subunit. The B subunit binds irreversibly to GM1 ganglioside as receptors on the cell surface. Following to the toxin internalization, the A subunit phosphorylates the stimulatory guaninenucleotide binding protein. Then, it increases the levels of intracellular cyclic AMP which results in clinical manifestation of diarrhea [13, 15].

Several epidemiological studies indicated that the treatment of either EHEC O157:H7 or ETEC infection using antibiotics may not be effective. The chance of HUS increases in case of EHEC infection and antibiotic resistance [2, 27, 28], hence, immunoprotection would be a rather efficient approach via the application of live attenuated or killed bacterial cell vaccines. This is mainly due to more complete protective response, generated by multi-component vaccine than a single one [29]. In the present study, we designed a novel multi subunit antigen that provides a suitable and safe vaccine candidate against two pathogenic *E.coli* strains of ETEC and EHEC. Using the synthetic biology, the major subunits from ETEC and EHEC virulence factors were constructed together to be expressed as a single chimeric protein. Here we describe the structure of the chimeric protein which was analyzed through an *In-silico* approach.

2. METHODS

2.1. Sequences of Antigenic Fragments, Applied Databases and Multiple Epitopic Chimer Design

Three antigenic fragments of Intimin,B subunit from LT and STX toxins and major subunit gene of CFA/I were selected for the chimer construction. Related sequences were retrieved from UniProtKB/Swiss-Prot data base (http://www.uniprot.org/help/uniprotkb).For the first fragment, a truncated form of stxB(lacking 18 amino acids from the N-terminal of the protein-signal peptide)- designated StxB. The 282 amino acids fragment from the carboxy terminus of Intimin which reported to be involved in binding to its receptor Tir[19], was selected. For the third fragment, complete sequence of CfaB was selected. The fourth fragment, LtB, was designated by removing 21 amino acids from the N-terminal of ltB. Gene order in the gene cassette was optimized from antigenicity perspective. In order to separate the different domains and find the best epitope exposing chimeric antigen, Linkers consisting of EAAAK repeats were designed. VaxiJen server(http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html) was used to predict the immunogenicity of the whole antigen and its subunit vaccine [30].

The synthetic sequence encoding the chimeric gene was optimized by the GenScript service (http://www.genscript.com/gene_synthesis.html), based on the codon usage table of Brassica napus L. Some parameters that are critical to the efficiency of gene expression including codon adaptation index (CAI), frequency of optimal codon (FOP) and GC content adjustment were considered for optimization. Finally the RNA secondary structure of the chimeric gene was analyzed by mfold(http://www.bioinfo.rpi.edu/applications/mfold), and RNAfold online servers (http://rna.tbi.univie.ac.at) [31, 32].

2.2. The Physico-Chemical Analysis of Chimer Protein Using Bioinformatics Tools

The physico-chemical parameters, (pI, MW, instability index and etc.) were computed using the Expasy's ProtParam(http://us.expasy.org/tools/ protparam.html)[33].

The protein secondary structure prediction was performed by GORIV (http://expasy.org/tools/gor4.html) and PSIPRED (http://bioinf.cs.ucl.ac.uk/psipred/) servers[34, 35].The I-TASSER server (http://zhanglab.ccmb.med.umich.edu/I-TASSER) was used to predict 3D structure of the chimeric protein[36]. Then, the Swiss-Pdb Viewer was used to visualize the modeled 3D structures.

The stereochemistry of structure validated using PROCHECK software (http://swissmodel.expasy.org/ workspace) and RAMPAGE server (http://mordred.bioc.cam.ac.uk/~ rapper/rampage.php) and demonstrated via Ramachandran plot.The energy minimization of modeled structure was calculated by Swiss-PDB Viewer software [37, 38].

2.3. Effect of Implicit Water on the Configuration of Chimer Protein

In order to have a stable configuration of chimeric protein, the geometry optimization launched using AMBER99 force field at 300 k temperature at a constant dielectric with default scale electrostatic and vander Waals factors. Molecular dynamic at the same condition was set up to 200 ps after geometry optimization to be acknowledged of dynamic conformation consistency. The total energy (ETOT) that is an overall estimation of the kinetic and potential energies, was then calculated at the constant 300 k temperature to filter the probable thermal noise.

2.4. Accessibility of Explicit Water Impact on the Linker Configuration of the Chimer Protein

The configuration of the chimeric protein is mostly defined either via change of length or torsion of linkers at different hydration levels. For this purpose, the number of explicit accessible waters that were used in molecular dynamic was set differentially. The number of waters that were proposed by the optimum box capacity was considered as high.

Subsequently, the different portions of 0.5, 0.25 and 0.125 of it were considered as moderate, low and very low hydrated linkers, respectively. This condition applied to all four linkers. All calculations were performed using the Hyperchem 8.0.7 portable software package. All linker configurative modifications in the presence of explicit water were studied via designing a fitted periodic box. Three linkers were solvated in explicit water differentially and their energetic indices are shown. All the simulation details are depicted in Fig. **1** and the differential solvation of linkers 1 to 3 are demonstrated in the panels b to c of Fig. **1**. Before any data was taken, the temperature was set at 300 k and the geometry optimization of linkers launched using AMBER99 force field of the molecular mechanic method. The negativity of calculated ETOT for all linkers with fourlevels of hydration (high, moderate, low and very low which are abbreviated as HH, MH, LH and VLH, respectively) was considered as a stability index and the results were compared. Molecular dynamic was set to 200 ps after geometry optimization. The simulation time step in each run was 10 picosecond and the relevant energies were stored every 50 ps. A typical 200 ps run took approximately 1 day on a two processor clustered 3 GHz computer.

2.5. Evaluation of Exposed Plausible Epitopes

Continuous B-cell epitopes were predicted by BcePred and IEDB (www.imtech.res.in/raghava), servers[39]. In the IEDB server, prediction is performed on the basis of Emini surface accessibility and Kolaskar-Tongaonkar antigenicity. Moreover, DiscoTope (http://tools.immuneepitope.org/stools/discotope/discotope and CBTOPE (http://www. im-

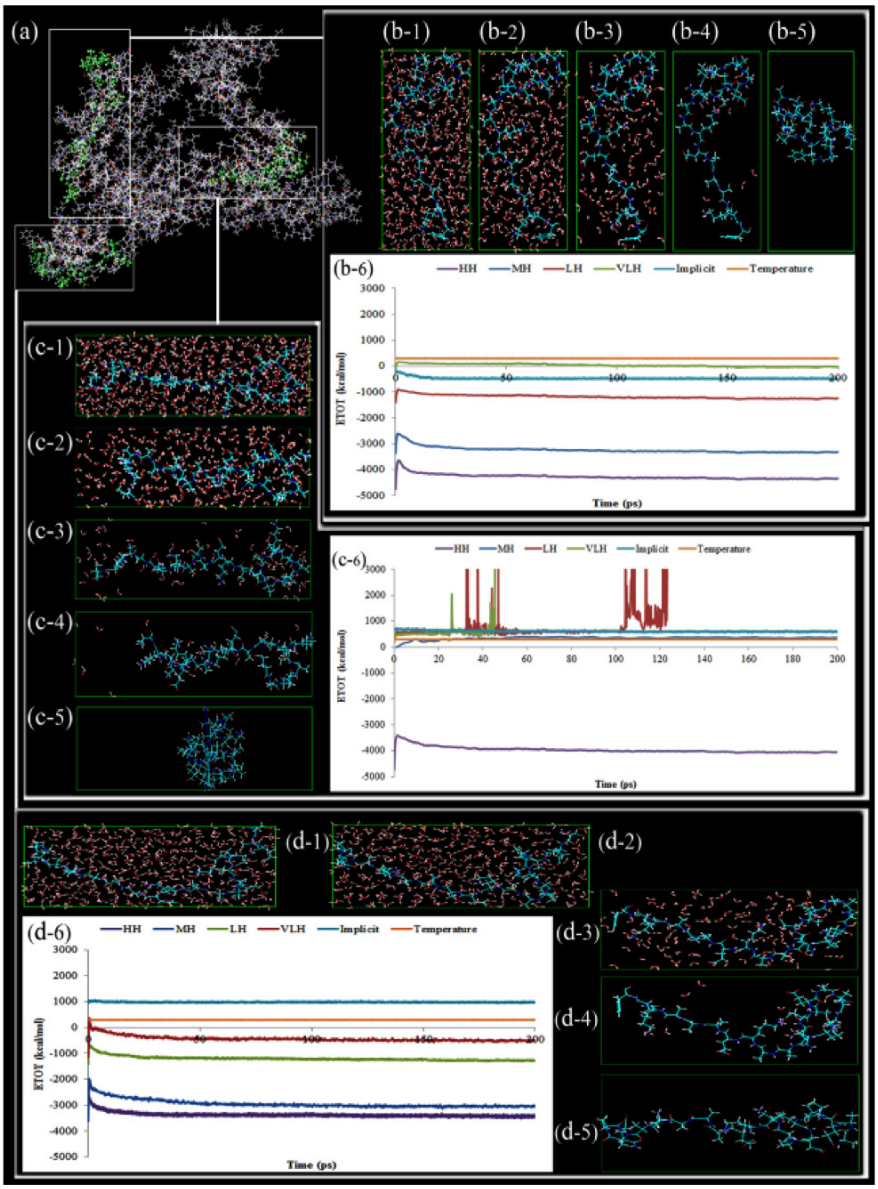

Figure 1. (**a**) All the three linkers are highlighted in the chimeric protein. Panels "b, c and d" represent the conformational status of linker1, 2 and 3, respectively. The 1-4 numbering of each panel demonstrate HH, MH, LH and VLH status of explicit solvation model. The 5th level of each panel demonstrates linkers' conformation in implicit solvation model. The 6th level of each panel depicts total energy of corresponding linkers' status during 200ps of molecular dynamic simulation.

tech.res.in/raghava/cbtope/submit.php)and SEPPA (http://lifecenter.sgst.cn/seppa/)servers were used to predict discontinuous epitopes using the 3D structures of the proteins in the protein data bank (PDB) format and amino acid composition [40-42]. Additionally,Ellipro server (http://tools.immuneepitope.org/tools/ElliPro/iedb_input) was used to predict both linear and discontinuous B-cell epitopes[43]. Propred-1 (http://www.imtech.res.in/raghava/propred/) was used for the prediction of peptides from the antigenic sequence binding to MHC class I [44]. IEDB (www.cbs.dtu.dk/services) server was used to identify common epitopes that bind to both MHC class molecules [45-49]. The prediction of epitopes for binding to MHC molecules, according to the distribution of MHC alleles in mouse (www.proimmune.com; http://biolinfo.org/mpid-t2/statistic.html),was performed.

The allergic properties of the construct were determined by using the Hybrid approach at AlegPred (http://www.imtech.res.in/raghava/algpred)[50]. The allergenicity was further analyzed by homology search in SDAP database (http://fermi.utmb.edu/) and an alignment-free method based on the main physicochemical properties of proteins in AllerTOP(http://fermi.utmb.edu/)for confirmation [51, 52].

The antigenic epitopes that were proposed via BcePred and SEPPA servers had enough length eligibility among other results. The antigenic epitope sequences that were proposed via these servers aligned to find a rather conserved sequential pattern. Then the aligned sequences were considered from the structural point of view to confirm the plausible conserved structural pattern.

3. RESULTS

3.1. Design and Construction of the Chimeric Gene

The chimeric gene consisting of Intimin, B subunits from LT and STX and major fragment of CFA/I, was examined from antigenicity perspective. Arrangements of the different fragments and the repeats of the linkers are shown in Figure **2**. A synthetic sequence encoding the whole chimeric protein was reverse translated into the nucleotide sequence and matched with the codon preference table of *Brassica napus L.*, to increase the target protein production in transgenic plant. Codon optimization increases the codon usage bias in B. napus by upgrading the CAI from 0.75 to 0.85. The overall GC content (which is considered as a measure value for transcriptional and translational efficiency) was improved from 68.84% to46.17% upon codon optimization. GC content and unfavorable peaks have been optimized to prolong the half-life of the mRNA. Within the synthetic construct all the splice sites, poly adenylation signal, destability elements and all the cis-acting sequences that may have a negative influence on the expression rate were removed. The *Xba*I and *Sac*I restriction sites for cloning in prokaryotic vectors were successfully introduced at the 3' and 5' sites of the sequence. In addition, The C-terminal of construct contained 6×His-tag for affinity purification. The result of "Mfold" server prediction for mRNA of optimized chimeric construct showed 38 structures. The comparative minimum free energy respect to the native mRNA showed that the ΔG of the best predicted structure for the optimized construct was -520.14 kcal/mol. These findings were in agreement with data generated by ''RNA structure '' web server.

3.2. Prediction of Secondary and Tertiary Structures of the Chimeric Protein

Physico-chemical characteristics of the synthetic construct which were computed by ''ProtParam'' show that some amino acids such as alanine (13.7%), serine (9.9%) and threonine (9.7%) were found to be more abundant in the protein than other residues. Extinction coefficient of SICL multiepitope peptide at 280 nm was 78645$M^{-1}cm^{-1}$. The structure of the chimeric protein was evaluated by GOR IV and the composition of predicted secondary structure for chimeric protein were 32.17% (helix/H), 24.89% (extended strand/E), and 42.94% (alpha helix/L). Three dimensional structures of SICL chimeric protein were successfully modeled using the I-TASSER server. The best model (model 1) was selected based on the C-score (-0.32) for estimating the quality of the predicted model. C-score is typically in the range of [−5 to 2], where a c-score of higher value signifies the model with a high confidence. In addition, the expected TM-score (template modeling) for this model was 0.67±0.13. The expected root mean square deviation (RMSD) for measuring the structural similarity between two structures)Protein with highly similar structure in PDB and SICL) was 8.7±4.5Å.

3.3. Evaluation of Model Stability

The predicted structure was further validated for its reliability and structural quality based on the Ramachandran plot. Results of PROCHECK revealed that 77% of amino acid residues from modeled structure generated by I-TASSER were incorporated in the favored regions (A, B, and L) of the plot. Apart from that 20.7 % of residues were in allowed regions (a, b, l, and p) of the plot (Fig. **2b**). This results verified by the RAMPAGE server, which placed residues in core (82.2%), allowed (10.7%), and outlier (7.2%) regions. Further study on the chimer structure stability was performed via implicit and explicit solvation models also geometrically optimized the predicted structure (Fig. **2c**). In implicit model the dielectric constant is set to the one which will apply an extremely low hydration status in the solvation model. Hence, the implicit model would be rather low hydrated than the VLH level of the explicit model. The more the hydration level means the more diluted protein in the aqueous solvent.

3.4. Prediction of Immunogenic Epitopes

An antigen should be hydrophilic and could activate both the B-cell and T-cell mediated immunity for becoming an effective vaccine candidate. BcePred server based on some important properties of epitopes including hydrophilicity, antigenicity, flexibility, accessibility, polarity and exposed surfaces, was employed for prediction of continuous B-cell epitopes (Table **1a**). DiscoTope server was used for the prediction of discontinuous B-cell epitopes (Table **1b**). In addition, results indicated that amino acid residues were designed as a linker between different domains (amino acid residues 71-88, 370-392 and 539-556) did not have epitope proper-

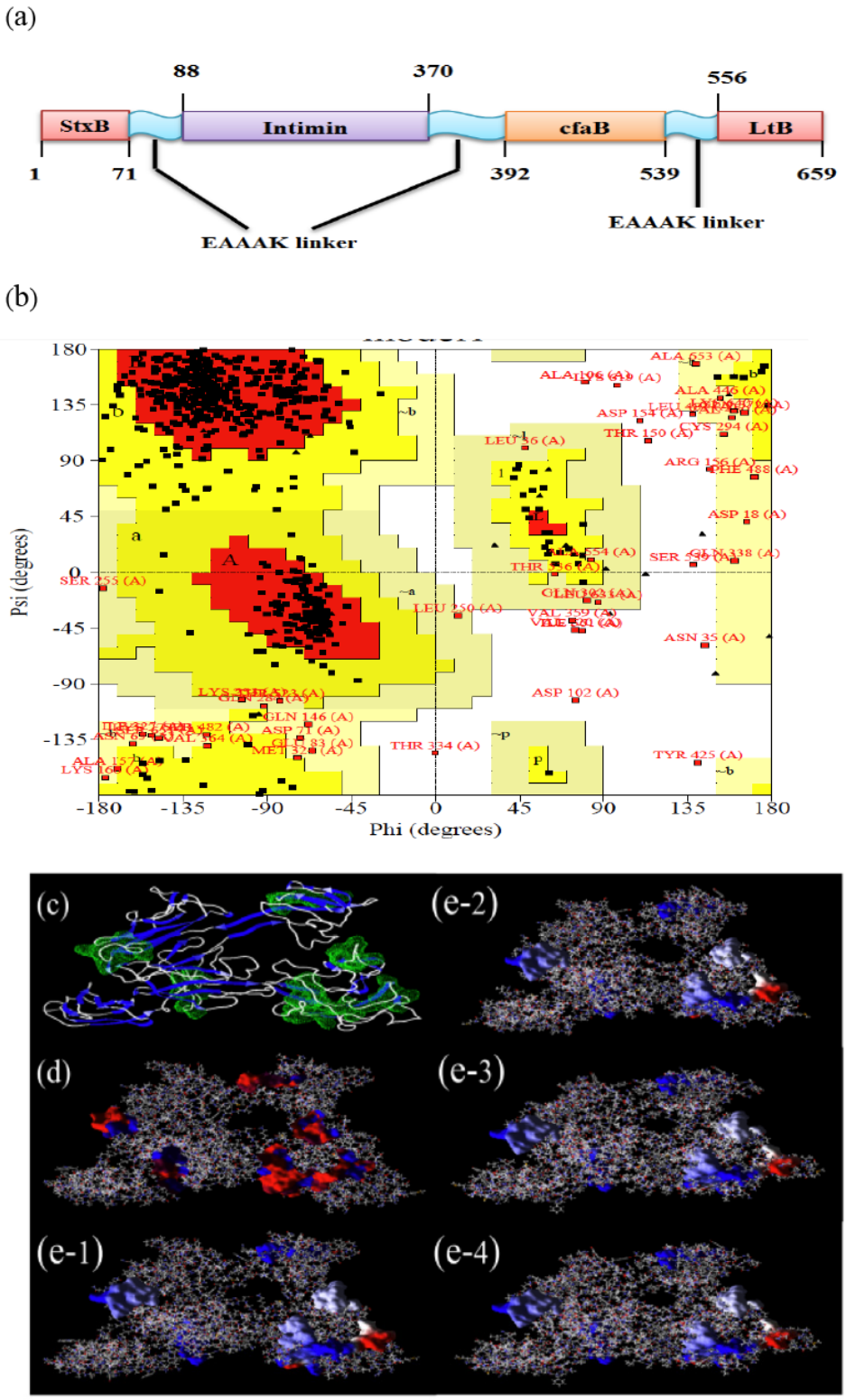

Figure 2. (**a**) Schematic model of the SICL chimeric construct consists of StxB (S), Intimin (I), CfaB(C) and LtB(L)attached together by linkers for eventual expression in plants. (**b**) Evaluation of model quality based on the Ramachandran plot. Percentage of the residue was 77% in the favored region, 20.7 % allowed and 2.3 % disallowed. (**c**) The secondary structure of chimer protein. The epitopes are marked via green dotted surfaces. (d) The antigenic epitopes are colored based on hydrophobicity. Hydrophilic and hydrophobic residues are colored in red and blue, respectively. (e-1) The antigenic epitopes are colored based on electrostatic distribution which red and blue colors correspond to negative and positive charge, respectively. (e-2 to e-4) Snapshots of 70, 140 and 200ps of implicit solvation model molecular dynamic are demonstrated, respectively. Position of electrostatic epitopes is shown in each snapshot.

Table 1. Prediction of linear B-cell epitopes in chimeric protein by different parameters based on BcePred software.

a) Prediction of linear epitopes by Ellipro server. Predicted peptides are ranked by score. The score ranges between 0.5 (threshold) to 1.

No.	Start Position	End Position	Peptide	Number of Residues	Score
1	594	640	VIITFKSGATFQVEVPGSQHIDSQKKAIERMKDTLRI-TYLTETKIDK	47	0.843
2	1	14	MADCAKGKIEFSKY	14	0.769
3	211	276	NIWLQYGQFKLKASGGDGTYSWYSENTSIATVDASG-KVTLNGKGSVVIKATSGDKQTVSYTIKAPS	66	0.759
4	17	91	NDTFTVKVAGKEYWTSRWNLQPLLQSAQLTGMT-VTIKSSTCESGSGFAEVQFNNDAEAAAKEAAAKEAA-AKAFDQ	75	0.751
5	504	527	SSSQELVISAAPKTAGTAPTAGNY	24	0.751
6	654	659	AISMEN	6	0.722
7	317	343	KYSHYSSMNSITAWIKQTSSEQRSGVS	27	0.698
8	349	383	ITQNPLPGVNVNTPNVYAVCVEAEAAAKEAAAKEA	35	0.682
9	531	546	VSLVMTLGSAEAAAKE	16	0.612
10	310	314	DSWGA	5	0.541
11	293	297	ICKNL	5	0.538
12	566	573	SEYRNTQI	8	0.530
13	183	194	ATEVTFFDELKI	12	0.517

b) Prediction of conformational epitopes by DiscoTope server (The final scores are calculated by combining the propensity scores of residues in spatial proximity and the contact numbers,Threshold = -3.700)

Position	Epitope	DiscoTope score
123	N	-3.094
128-9	NN	-2.692,-3.652
180	E	-3.449
194	I	-3.409
215	Q	-3.199
228	G	-3.675
246	G	-3.371
263-5	GDK	-2.339, -2.296, -1.827
334-7	TSSE	-3.443, -2.254, -2.707, -1.690
339	R	-2.802
524-6	AGN	-3.573, -3.199, -3.480
607	E	-3.240
609	P	-3.538
611-3	SQH	-3.462, -1.886, -2.660
623-4	RM	-3.446, -3.693

ties. The conformational epitopes for B cells were predicted by CBTOPE and SEPPA servers. Furthermore, in parallel, Ellipro server was used for predicting both linear and discontinuous epitopes. For screening the T-cell epitopes, Propred-I

and IEDB servers were used to identify common T-cell epitopes (Table **2a**, **b**). In order to make reliable predictions, IEDB tools were also employed for predicting the affinity of both B and T-cell epitopes. Furthermore, the allergenicity of the sequence was predicted using AlgPred tool. Based on different allergenicity prediction approaches in AlgPred tool, this protein was not detected as a potential allergen. Finally, search for allergens showed no significant similarities between any region in SDAP and AllerTOP allergen library.

3.5. Comparative Study of Implicit and Explicit Solvation Models

The ETOT of molecular dynamic studies reveals that all of the linkers show the higher stability at the HH level which is highly correlated with the conformational consistency (Fig. **1**). This status is especially detected in linker 1 and 2 (Fig. **1** panels **b** and **c**). Due to gradual increment of ETOT positivity in molecular dynamic study of linker 1 and also appearance of molecular dynamic instability at MH, LH, VLH and implicit grades of linker 2 (Fig. **1 c-6**), it is concluded that the high susceptibility of these linkers to hydration degree causes conformational changes in implicitly solvated chimer protein (Fig. **2 e-1** to **e-4**). However, conformational stability of the chimer protein at the vicinity of linker 3 which connects domain 3 to 4, is guaranteed due to imperviousness of linker 3 to hydration degree (Fig. **1** panel d). By this regard, linker1, 2 and 3 are structurally stable at concentrations less than 29, 33 and 35mM, respectively. Since the linker1needs the highest hydration degree, the least concentration that linker1is stable would be named as the "limiting structural stability". This dilution rate would eventually mask the least stable concentration of other linkers.

3.6. Characterization of Main Exposed Epitopes

Studying potential epitopes that are concluded from BcePred and SEPPA servers proposes seven rather conserved regions that are introduced as main epitopes. There are at least two epitopic regions that preserve beta sheet as their secondary structure (Fig. **2c**). The electrostatic and hydrophobicity distribution of the epitopic residues reveals that they consist of positive hydrophilic residues (Fig. **2d**, **e**). This point would guarantee the fact that the main epitopes would be never buried in the stable conformations of the chimer protein, even in destabilizing conditions like implicit solvation model (Fig. **2 e-1** to **e-4**).

4. DISCUSSION

Treatment of ETEC and EHEC infections remains problematic, since conventional use of antibiotics enhances pathogenesis [16]. Generally, it is proved that vaccination is the most effective means to prevent infectious diseases. Due to the large number of pathogens and the requirement to induce immunity that is effective in the digestive tract (especially in the gut), the development of vaccines against enteric pathogens represents a serious challenge. Development of immunity would be expedited if the specific epitopes that are

Table 2. Prediction of T-cell epitopes.

a) Prediction of MHC I epitopes by ProPred-I server.

Epitopes	MHC alleles
11-19, 31-39, 66-74, 105-114, 119-132, 138-146, 151-159, 192-209, 290-298, 312-331, 340-351, 356-364, 401-410, 415-423, 439-447,473-481, 497-509, 527-537, 556-564, 642-650,	MHC-Db
24-32, 132-140, 153-161, 184-192, 213-221, 231-239, 282-293, 297-305, 316-327, 340-349, 360-368, 409-417, 430-438, 456-470, 490-498, 501-509, 525-533, 556-587	MHC-Kb

b) Prediction of MHC II epitopes by IEDB server. (IEDB recommended prediction method is Consensus approach, combining NN-align, SMM-align)

Allele	Start	End	Peptide	Method used	Percentile rank
H2-IAb	491	505	AALGYSASGVNGVSS	Consensus (smm/nn)	1.45
H2-IAb	490	504	AAALGYSASGVNGVS	Consensus (smm/nn)	1.51
H2-IAb	489	503	EAAALGYSASGVNGV	Consensus (smm/nn)	1.68
H2-IAb	492	506	ALGYSASGVNGVSSS	Consensus (smm/nn)	1.68
H2-IAb	510	524	VISAAPKTAGTAPTA	Consensus (smm/nn)	1.82
H2-IAd	433	447	ESYRVMTQVHTNDAT	Consensus (smm/nn)	1.69
H2-IAd	434	448	SYRVMTQVHTNDATK	Consensus (smm/nn)	1.70
H2-IAd	432	446	FESYRVMTQVHTNDA	Consensus (smm/nn)	1.88
H2-IAd	444	458	NDATKKVIVKLADTP	Consensus (smm/nn)	1.96

responsible for triggering protective immune responses are defined precisely[3, 23]. Chimeric proteins carrying epitopes from different parts of ETEC and EHEC virulence factors present either increased immunogenicity of the recombinant antigen and the possibility of eliciting abroad cellular or humoral immune response. We designed a novel multi subunit antigen that provides a suitable and safe vaccine candidate against ETEC and EHEC infections. The mechanism that guarantees achievement to this aim is governed via the proper folding and exposure of epitopes in chimeric vaccine, which in turn relies on the appropriate linker conformation. Here, the hierarchical dependence of mechanisms is considered thoroughly via designing different assessments. The applicability of designed chimeric gene construct evaluated through prediction of secondary and tertiary protein structures and prediction of the immunogenic epitopes.

Intimin ,B subunit from LT and STX toxins and major subunit gene in CFA/I were used for *in silico* design of a chimeric subunit vaccine, candidate against two important enteric pathogens ,ETEC and EHEC. In order to separate different domains of the chimeric protein several repeats of EAAAK linkers, which expected to form a monomeric hydrophobic α-helix were used. Multi-domain proteins which are composed of two or more fused genes can provide multiple functional properties resulting from each of their parts. Furthermore, it has also been reported that fusion proteins spatially separated by appropriate linker peptides might be more effective than independent domains [53, 54]. It has been shown that the completed linker sequence $A(EAAAK)_n A$(n = 3-5)which provides a stabilized $Glu^{-}Lys^{+}$ salt bridge, shows eighty percent of helicity. It would indicate the effective separation ofthe domains of the fusion protein[53]. Our successful experience of using four EAAAK repeat sequences in chimeric gene has demonstrated that it could lead to logically acceptable results [20].

In this study, GenScript gene synthesis service was utilized to optimize a variety of parameters that are critical to the efficiency of gene expression. These include the codon adaptation index(CAI), which is a measurement of the relative adaptiveness of the codon usage bias for a particular DNA or RNA sequence [55] and FOP, which is used to show the optimization level of synonymous codon choice in each gene during the translation process [56]. The optimized gene sequence had a CAI of 0.85 and average GC content of 46.17%, indicating that the synthetic gene sequence could be theoretically expressed in the host well. In addition, negative elements and repetitive sequences that may form unfavorable secondary structures on mRNA were eliminated.

In order to analyze the physico-chemical parameters of the chimeric protein,pI value of the SICL was calculated as 7.93. High aliphatic index of SICL indicates that the protein may be stable for a wide range of temperature. On the basis of instability index, Expasy's ProtParam classifies the SICL protein as stable (Instability index<40). GRAVY of the chimeric construct was -0.253, which indicates hydrophilicity and better interaction of the construct with the surrounding water molecules. For an antigen molecule, the more interaction with water means the more communication with immune system elements, specifically antibodies.

The chimer protein consists of four domains that are connected together via three linkers. Since the folding of domains occurs independently, the chimer protein stability would be affected significantly by linkers' conformation. The hydration degree of linkers defines its conformation which consequently affects chimer stability. The geometry optimization of chimer, prepares a perceptive stable conformation including linkers stable conformation. The more deviation from stable conformation would be considered as increased instability. The increased instability is observed at higher concentrations of solvated chimer which is also acknowledged as lower dilution. Considering the structural limiting dilution rate of linkers and their corresponding molecular weight, we could roughly estimate the "limiting structural stability" for the chimer protein which is almost 29mMSince most of epitopic residues are positively hydrophilic, in explicit solvation model the pattern of epitope exposure would be always superficial.

The linkers' conformational changes would affect chimer conformation and stability in implicit solvation model. However, the inherent property of the epitope residues would always keep majority of them at the surface, even if there are conformational changes in the other regions of the chimer protein.

One of the softwares used for predicting RNA secondary structure was Mfold. Advantages of Mfold are that it employs a theoretically tract able DP algorithm which can find the structure with the minimum ΔG within its thermodynamic model and its ability to predict true positive base pairs. The 5' terminus of the gene had no Stem-Loop structure, which impacts the ribosomal binding capacity and the stability of mRNA. The data showed the mRNA was stable enough for efficient translation in the new host. The study of protein secondary structure plays an important role in the prediction of protein tertiary structure with the ab-initio method or protein fold recognition by providing additional constraints [57].The predicted model had acceptable stability (-6162.187 kJ/mol). Assessment of the accuracy and reliability of experimental and theoretical models of protein structures is necessary. For the evaluation of the predicted models, both RMSD and TM-scores were used. Expected TM-score of 0.67±0.13, where the TM-score of higher than 0.5 indicates a model of correct topology, validates the accuracy of the model. Our chimeric structure showed desirable protein stability based on Ramachandran plot predictions. In Ramachandran plot analysis, almost a negligible 2.3% of the residues were found to be in the outliner region that could probably be due to the presence of chimeric junctions.

To be effective, a vaccine candidate should be able to induce strong Bcell and T cell responses. For this reason, the ability to map T cell and Bcell epitopes is important for optimal vaccine design and development. Bcepred software was used to determine the continuous B cell epitope based on single characters including hydrophilicity, antigenicity, flexibility, accessibility, polarity and exposed surface (Table 2). These parameters have been correlated with the location of continuous epitopes. The continuous B cell epitopes predicted by Bcepred and IEDB servers, were similar and this results were confirmed with predicted epitopes by Ellipro server. The conformational epitopes derived from CBTOPE,

SEPPA and DiscoTope servers were almost similar. Some epitopes were recognized by single software only. Among these results, the similarity between predicted epitopes by CBTOPE and DiscoTope servers, was rather higher than others, which seems to be more reliable. By considering the distribution of MHC alleles in mouse, potential epitopes in chimeric protein for binding to MHC molecules, were predicted. The results of prediction showed that the "Intimin" fragment of chimeric construct has more putative epitopes. Based on important parameters, namely; Emini surface accessibility [58], Kolaskar-Tongaonkar antigenicity [59] we found thatmost regions of the synthetic construct have good design criteria which confirmed our strategy for selecting accurate epitopes of each antigen. Therefore, this chimeric protein has epitopes likely to induce both the B-cell and T-cell mediated immune responses. Because of the high current incidence of allergenicity, with serious outcomes in many cases, the accurate prediction of allergenicity for new proteins is crucial. Allergenicity is not straightforward to predict using alignment-based methods, since it is a phenomenon mediated by specific IgE antibodies requiring the presence of conformational B-cell epitopes in allergens. Based on Hybrid approach in AlgePred tool, that improves the sensivity as well as the specificity of the allergen prediction method and the main physicochemical properties of proteins in AllerTOP, there is no significant similarity between any region and allergen library. Thus the chimeric protein is not allergen. Although different antigenic parts of construct are selected to be immunogenic, and are not toxigenic.

Stx is an N-glycosidase that inhibits the protein synthesis by depurination of a critical ribosomalresidue, important for protein elongation. It has also demonstrated *in vitro* that Stx2 increases cell surface expression of nucleolin, a eukaryotic receptor for the E. coliO157:H7 adhesion intimin. Epidemiological data suggests that Stx2 may play a more critical role in the development and severity of HUS. LTB also plays adjuvant role, enhancing the immunogenicity of chimeric antigen. *E. coli* can trigger a series strategy of infection: colonization of a mucosal site, evasion of host defenses, multiplication, and host damage. In our study, we targeted the adhesion factors required for this strain to reduce the colonization as well as the enterotoxins to avoid creating further damage in the host. At low concentrations of solvated chimer and at the expense of highly diluted chimer, we would expect a stable conformation with proper folding that is concurrent with main antigenic epitopes exposure. This fact supports the high solubility of chimer protein which prevents aggregation of the vaccine. On the other hand, the effectiveness of the chimer protein is identified at low concentrations. It would propose that the highly diluted chimer must be administrated as proper dose which would drastically decline the chance of immunological side effect outbreak.

CONFLICT OF INTEREST

Authors declare that they have no conflicts of interest.

ACKNOWLEDGEMENTS

This work was supported by National Institute of Genetic Engineering and Biotechnology.

REFERENCES

[1] Svennerholm, A.-M.; Steele, D. Progress in enteric vaccine development. Best Practice & Research Clinical Gastroenterology. **2004**, 18, 421-445.

[2] Babiuk, S.; Asper, D.J.; Rogan, D.; Mutwiri, G.K.; Potter, A.A. Subcutaneous and intranasal immunization with type III secreted proteins can prevent colonization and shedding of Escherichia coli O157: H7 in mice. Microb. Pathog., **2008**, 45, 7-11.

[3] Girard, M.P.; Steele, D.; Chaignat, C.-L.; Kieny, M.P. A review of vaccine research and development: human enteric infections. Vaccine. **2006**, 24, 2732-2750.

[4] Phillips, A.; Navabpour, S.; Hicks, S.; Dougan, G.; Wallis, T.; Frankel, G. Enterohaemorrhagic Escherichia coliO157: H7 target Peyer's patches in humans and cause attaching/effacing lesions in both human and bovine intestine. Gut. **2000**, 47, 377-381.

[5] Noris, M.; Remuzzi, G. Hemolytic uremic syndrome. J. Am. Soc. Nephrol., **2005**, 16, 1035-1050.

[6] Scheiring, J.; Andreoli, S.P.; Zimmerhackl, L.B. Treatment and outcome of Shiga-toxin-associated hemolytic uremic syndrome (HUS). Pediatr. Nephrol., **2008**, 23, 1749-1760.

[7] Al-Charrakh, A.; Al-Muhana, A. Prevalence of Verotoxin-producing Escherichia coli (VTEC) in a survey of dairy cattle in Najaf, Iraq. Iranian journal of microbiology. **2010**, 2, 130.

[8] Bonyadian, M.; Momtaz, H.; Rahimi, E.; Habibian, R.; Yazdani, A.; Zamani, M. Identification & characterization of Shiga toxin-producing Escherichia coli isolates from patients with diarrhoea in Iran. **2010**.

[9] Cheng, Y.; Feng, Y.; Luo, P.; Gu, J.; Yu, S.; Zhang, W.-j.; Liu, Y.-q.; Wang, Q.-x.; Zou, Q.-m.; Mao, X.-h. Fusion expression and immunogenicity of EHEC EspA-Stx2Al protein: implications for the vaccine development. The Journal of Microbiology. **2009**, 47, 498-505.

[10] Kühne, S.A.; Hawes, W.S.; La Ragione, R.M.; Woodward, M.J.; Whitelam, G.C.; Gough, K.C. Isolation of recombinant antibodies against EspA and intimin of Escherichia coli O157: H7. Journal of clinical microbiology. **2004**, 42, 2966-2976.

[11] Mohawk, K.L.; Melton-Celsa, A.R.; Robinson, C.M.; O'Brien, A.D. Neutralizing antibodies to Shiga toxin type 2 (Stx2) reduce colonization of mice by Stx2-expressing Escherichia coli O157: H7. Vaccine. **2010**, 28, 4777-4785.

[12] Gao, X.; Cai, K.; Shi, J.; Liu, H.; Hou, X.; Tu, W.; Xiao, L.; Wang, Q.; Wang, H. Immunogenicity of a novel Stx2B–Stx1B fusion protein in a mice model of Enterohemorrhagic Escherichia coli O157: H7 infection. Vaccine. **2009**, 27, 2070-2076.

[13] Jafari, A.; Aslani, M.M.; Bouzari, S. Escherichia coli: a brief review of diarrheagenic pathotypes and their role in diarrheal diseases in Iran. Iranian journal of microbiology. **2012**, 4, 102.

[14] Schuller, S. Shiga toxin interaction with human intestinal epithelium. Toxins (Basel). **2011**, 3, 626-39.

[15] Odumosu, O.; Nicholas, D.; Yano, H.; Langridge, W. AB toxins: a paradigm switch from deadly to desirable. Toxins. **2010**, 2, 1612-1645.

[16] Gu, J.; Liu, Y.; Yu, S.; Wang, H.; Wang, Q.; Yi, Y.; Zhu, F.; Yu, X.J.; Zou, Q.; Mao, X. Enterohemorrhagic Escherichia coli trivalent recombinant vaccine containing EspA, intimin and Stx2 induces strong humoral immune response and confers protection in mice. Microbes Infect., **2009**, 11, 835-41.

[17] Judge, N.A.; Mason, H.S.; O'Brien, A.D. Plant cell-based intimin vaccine given orally to mice primed with intimin reduces time of Escherichia coli O157: H7 shedding in feces. Infect. Immun., **2004**, 72, 168-175.

[18] Mohawk, K.L.; O'Brien, A.D. Mouse models of Escherichia coli O157: H7 infection and shiga toxin injection. BioMed Research International. **2011**, 2011.

[19] Amani, J.; Mousavi, S.L.; Rafati, S.; Salmanian, A.H. In silico analysis of chimeric espA, eae and tir fragments of Escherichia coli O157:H7 for oral immunogenic applications. Theor. Biol. Med. Model., **2009**, 6, 28.

[20] Amani, J.; Salmanian, A.H.; Rafati, S.; Mousavi, S.L. Immunogenic properties of chimeric protein from espA, eae and tir genes of Escherichia coli O157:H7. Vaccine. **2010**, 28, 6923-9.

[21] Amani, J.; Mousavi, S.L.; Rafati, S.; Salmanian, A.H. Immunogenicity of a plant-derived edible chimeric EspA, Intimin and Tir of Escherichia coli O157: H7 in mice. Plant Sci. (Amsterdam, Neth.). **2011**, 180, 620-627.

[22] Steffen, R.; Castelli, F.; Dieter Nothdurft, H.; Rombo, L.; Jane Zuckerman, N. Vaccination against Enterotoxigenic Escherichia coli, a Cause of Travelers—Diarrhea. J. Travel Med., **2005**, 12, 102-107.

[23] Baker, K.; Levine, M.; Morison, J.; Phillips, A.; Barry, E. CfaE tip mutations in enterotoxigenic Escherichia coli CFA/I fimbriae define critical human intestinal binding sites. Cell. Microbiol., **2009**, 11, 742-754.

[24] Fratamico, P.M.; Smith, J.L. Escherichia coli infections. Foodborne infections and intoxications. **2006**, 205-208.

[25] Turner, A.K.; Stephens, J.C.; Beavis, J.C.; Greenwood, J.; Gewert, C.; Randall, R.; Freeman, D.; Darsley, M.J. Generation and characterization of a live attenuated enterotoxigenic Escherichia coli combination vaccine expressing six colonization factors and heat-labile toxin subunit B. Clin. Vaccine Immunol., **2011**, 18, 2128-2135.

[26] Tchesnokova, V.; McVeigh, A.L.; Kidd, B.; Yakovenko, O.; Thomas, W.E.; Sokurenko, E.V.; Savarino, S.J. Shear-enhanced binding of intestinal colonization factor antigen I of enterotoxigenic Escherichia coli. Mol. Microbiol., **2010**, 76, 489-502.

[27] Liu, L.; Zeng, H.; Luo, P.; Wu, J.; Chen, H.; Shi, Y.; Zhang, W.; Mao, X.; Xiao, B.; Zou, Q. Cloning a Truncated Fragment (stx2a1) of the Shiga-Like Toxin 2A1 Subunit of EHEC O157: H7: Candidate Immunogen for a Subunit Vaccine. Mol. Biotechnol., **2009**, 43, 8-14.

[28] Jafari, F.; Hamidian, M.; Rezadehbashi, M.; Doyle, M.; Salmanzadeh-ahrabi, S.; Derakhshan, F.; Zali, M.R. Prevalence and antimicrobial resistance of diarrheagenic Escherichia coli and Shigella species associated with acute diarrhea in Tehran, Iran. The Canadian Journal of Infectious Diseases & Medical Microbiology. **2009**, 20, e56.

[29] Nazarian, S.; Gargari, S.L.M.; Rasooli, I.; Amani, J.; Bagheri, S.; Alerasool, M. An in silico chimeric multi subunit vaccine targeting virulence factors of enterotoxigenic Escherichia coli (ETEC) with its bacterial inbuilt adjuvant. J. Microbiol. Methods. **2012**, 90, 36-45.

[30] Doytchinova, I.A.; Flower, D.R. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinformatics. **2007**, 8, 4.

[31] Gruber, A.R.; Lorenz, R.; Bernhart, S.H.; Neuböck, R.; Hofacker, I.L. The vienna RNA websuite. Nucleic Acids Res., **2008**, 36, W70-W74.

[32] Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res., **2003**, 31, 3406-3415.

[33] Gasteiger, E.; Hoogland, C.; Gattiker, A.; Wilkins, M.R.; Appel, R.D.; Bairoch, A., *Protein identification and analysis tools on the ExPASy server*, in *The proteomics protocols handbook*. 2005, Springer. p. 571-607.

[34] Garnier, J.; Gibrat, J.; Robson, B.; Doolittle, R. GOR secondary structure prediction method version IV. Methods Enzymol., **1996**, 266, 540-553.

[35] Buchan, D.W.; Minneci, F.; Nugent, T.C.; Bryson, K.; Jones, D.T. Scalable web services for the PSIPRED Protein Analysis Workbench. Nucleic Acids Res., **2013**, 41, W349-W357.

[36] Zhang, Y. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics. **2008**, 9, 40.

[37] Laskowski, R.A.; MacArthur, M.W.; Moss, D.S.; Thornton, J.M. PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr., **1993**, 26, 283-291.

[38] Lovell, S.C.; Davis, I.W.; Arendall, W.B.; de Bakker, P.I.; Word, J.M.; Prisant, M.G.; Richardson, J.S.; Richardson, D.C. Structure validation by Cα geometry: ϕ, ψ and Cβ deviation. Proteins: Struct., Funct., Bioinf., **2003**, 50, 437-450.

[39] Saha, S.; Raghava, G., *BcePred: Prediction of continuous B-cell epitopes in antigenic sequences using physico-chemical properties*, in *Artificial Immune Systems*. 2004, Springer. p. 197-204.

[40] Sun, J.; Wu, D.; Xu, T.; Wang, X.; Xu, X.; Tao, L.; Li, Y.; Cao, Z.-W. SEPPA: a computational server for spatial epitope prediction of protein antigens. Nucleic Acids Res., **2009**, 37, W612-W616.

[41] Kringelum, J.V.; Lundegaard, C.; Lund, O.; Nielsen, M. Reliable B cell epitope predictions: impacts of method development and improved benchmarking. PLoS Comput. Biol., **2012**, 8, e1002829.

[42] Ansari, H.R.; Raghava, G. Identification of conformational B-cell Epitopes in an antigen from its primary sequence. Immunome Res. **2010**, 6, 1-9.

[43] Ponomarenko, J.; Bui, H.-H.; Li, W.; Fusseder, N.; Bourne, P.E.; Sette, A.; Peters, B. ElliPro: a new structure-based tool for the prediction of antibody epitopes. BMC Bioinformatics. **2008**, 9, 514.

[44] Singh, H.; Raghava, G. ProPred1: prediction of promiscuous MHC Class-I binding sites. Bioinformatics. **2003**, 19, 1009-1014.

[45] Wang, P.; Sidney, J.; Kim, Y.; Sette, A.; Lund, O.; Nielsen, M.; Peters, B. Peptide binding predictions for HLA DR, DP and DQ molecules. BMC Bioinformatics. **2010**, 11, 568.

[46] Nielsen, M.; Lundegaard, C.; Worning, P.; Lauemøller, S.L.; Lamberth, K.; Buus, S.; Brunak, S.; Lund, O. Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. Protein Sci., **2003**, 12, 1007-1017.

[47] Lundegaard, C.; Lamberth, K.; Harndahl, M.; Buus, S.; Lund, O.; Nielsen, M. NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8–11. Nucleic Acids Res., **2008**, 36, W509-W512.

[48] Peters, B.; Sette, A. Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. BMC Bioinformatics. **2005**, 6, 132.

[49] Sidney, J.; Assarsson, E.; Moore, C.; Ngo, S.; Pinilla, C.; Sette, A.; Peters, B. Quantitative peptide binding motifs for 19 human and mouse MHC class I molecules derived using positional scanning combinatorial peptide libraries. Immunome research. **2008**, 4, 2.

[50] Saha, S.; Raghava, G. AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. Nucleic Acids Res., **2006**, 34, W202-W209.

[51] Ivanciuc, O.; Schein, C.H.; Braun, W. SDAP: database and computational tools for allergenic proteins. Nucleic Acids Res., **2003**, 31, 359-362.

[52] Dimitrov, I.; Flower, D.R.; Doytchinova, I. AllerTOP-a server for in silico prediction of allergens. BMC Bioinformatics. **2013**, 14, S4.

[53] Arai, R.; Ueda, H.; Kitayama, A.; Kamiya, N.; Nagamune, T. Design of the linkers which effectively separate domains of a bifunctional fusion protein. Protein Eng., **2001**, 14, 529-532.

[54] Chen, X.; Zaro, J.L.; Shen, W.-C. Fusion protein linkers: property, design and functionality. Adv. Drug Delivery. Rev., **2013**, 65, 1357-1369.

[55] Sharp, P.M.; Li, W.-H. The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res., **1987**, 15, 1281-1295.

[56] Ikemura, T. Correlation between the abundance of Escherichia coli transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the E. coli translational system. J. Mol. Biol., **1981**, 151, 389-409.

[57] Soria-Guerra, R.E.; Moreno-Fierros, L.; Rosales-Mendoza, S. Two decades of plant-based candidate vaccines: a review of the chimeric protein approaches. Plant Cell Rep., **2011**, 30, 1367-1382.

[58] Emini, E.A.; Hughes, J.V.; Perlow, D.; Boger, J. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. J. Virol., **1985**, 55, 836-839.

[59] Kolaskar, A.; Tongaonkar, P.C. A semi-empirical method for prediction of antigenic determinants on protein antigens. FEBS Lett., **1990**, 276, 172-174.

Received: September 1, 2015 **Revised: April 16, 2015** **Accepted: April 20, 2015**