Indian Journal of Biotechnology
Vol. 17, January 2018, pp 16-26

Computational analysis and gene cloning: design and preparation of a multi subunit vaccine consisting of EspA, Stx2B and Intimin antigens against enterohaemrrhagic *Escherichia coli*

Mostafa Bakhshi, Firouz Ebrahimi, Shahram Nazarian* and Jamil Zargan

Department of Biology, Faculty of Basic Science, Imam Hossein University, Tehran, Iran

Received 17 August 2015; revised 8 January 2016; accepted 13 January 2016

Escherichia coli O157:H7 is an intestinal pathogen that made diarrhoea, haemolytic uremia syndrome (HUS) and hemorrhagic colitis (HC) in patients. Roles of EspA and Intimin at the beginning of bacteria colonization in intestine are critical. Destruction of protein synthesis route with shiga toxins of *E. coli* O157:H7 is mediated through B-subunit of toxins. In this study, *in silico* approaches were performed to design a suitable construct from EspA, Intimin and Stx2B and a recombinant chimeric antigen was produced. Bioinformatics analyses such as physicochemical data, mRNA folding, 3D structures of chimera and various immunoinformatic data, such as linear and conformational B-cell epitopes, T-cell epitopes were reported according to authentic data base. The chimeric gene was prepared as synthetic construct after designing and cloning. The validation result showed that 83.9% residues lie in favoured or additional allowed region of the Ramachandran plot. Epitope prediction results proved very good distribution of conformational B-cell epitopes in the 3D structure of chimera. The identified T-cell epitopes are apt to bind MHC molecules. A good quantity of recombinant chimeric antigen was achieved in host cells. From *in silico* approach, an appropriate multi subunit vaccine candidate was designed and prepared for immunological examinations.

Keywords: Chimeric protein, bioinformatic, immunogenicity, EHEC, cloning, recombinant antigen

Introduction

Pathogenic strains of *Escherichia coli* cause millions of dead, especially between children each year. Five categories of *E. coli*, namely, enteropathogenic, enteroagregative, enterotoxigenic, enteroinvasive and Shiga toxin producing are responsible for intestinal and extra-intestinal diseases in humans[1]. Previous research were published about the prevalence of *E. coli* O157:H7 and its pathogenesis throughout the world. The pathogenicity of enterohemorrhagic *Escherichia coli* (EHEC) is very high because of colonization potency in human intestine and its toxins that lead to hemorrhagic diarrhea, hemorrhagic colitis (HC) and haemolytic uremic syndrome (HUS). The most important factors involved in bacterial colonization are Esp proteins family, Intimin and Tir proteins[4]. Some serotypes of *E. coli* produce one or more protein toxins that are closely related to Stx from *Shigella dysenteriae* serotype 1[5]. Stx1 or stx2 are the major toxins of EHEC caused diseases in human beings[6]. The stx types all have an AB_5 structure, in which a single A-subunit is associated with five B-subunits[7]. The A-subunit embodies the *N*-glycosidase catalytic activity; it acts by removing a specific adenine base from the 28S srRNA of the 60S ribosomal subunit within infected cells while B-subunits responsible for bacterial attachment to cell surface globotriaosylceramide (Gb3) receptor[7,8]. Type III secretion proteins and Tir/Intimin prepare the bacteria attachment to epithelial cells of intestine[9]. So, a good candidate for vaccination against this pathogen is the correct combination of different virulence factors which are involved in its pathogenicity.

Subunit vaccines are considerably safer, more specific with less adverse reactions, the ability to target the site where immunity is required, and large scale production of recombinant proteins by biotechnological revolution[10]. A more complete protective response is generated by multi-component vaccine than a single component. Recently, with use of bioinformatic approaches, we can design candidate protein vaccine antigens and predict the function and efficiency of that by high accuracy. Some studies are published about active and passive immunization with recombinant antigens against both colonization factors and Shiga

*Author for correspondence:
Tel: +9821-77104934
kpnazari@ihu.ac.ir

toxin from EHEC but some authors reported computational analysis for its construct[11]. In the present study, we designed a novel multi subunit antigen that provides an appropriate vaccine candidate against EHEC infection. So a chimeric protein containing the major subunit from EHEC virulence factors of EspA, Intimin and Stx2B was constructed together. Finally, the structure of the chimeric protein was analyzed through an *in silico* approach.

Materials and Methods

Sequences Analyses and Construct Design

Subunit genes with structural and functional properties in EspA, Intimin and Stx2B virulence factors are *espa*, *eae* and *stx2b*, respectively. Related sequences were obtained from available sequence databases, primarily from the National Centre for Biotechnology Information (NCBI) (www.ncbi.nlm.nih.gov). The sequences were retrieved in FASTA format for analysis. Multiple alignments were carried out using clustalw2 software of the European Bioinformatics Institute (EBI) website (www.ebi.ac.uk/Tools/sss/wublast/). The sequences were fused together by an appropriate linker. The final length of designed chimera by considering each subunit and linker sequences was 1671 bp. The *in silico* gene analysis and multi parameter gene optimization of the synthetic chimeric gene was performed using GenScript rare codon analysis tool (http://www.genscript.com/cgi-bin/tools/rare_codon_analysis), rare codon calculator tool (http://nihserver.mbi.ucla.edu/RACC/) and expasy translate tool (web.expasy.org/translate/). The chimeric gene was designed for cloning and expression in *E. coli*.

Prediction of RNA Secondary Structure

The messenger RNA secondary structure and different parameters of the chimeric gene was analyzed by Mfold web server (http://mfold.rna.albany.edu/?q=mfold)[12] and RNA predict a secondary structure server (http://rna.urmc.rochester.edu/RNAstructureWeb/index.html). RNA secondary structure was compared before and after gene optimization.

Bioinformatics Analyses of Chimeric Protein

Physico-Chemical Parameters

The physico-chemical parameters, theoretical isoelectric point (pI), molecular weight, total number of positive and negative residues, extinction coefficient, half-life, instability index, aliphatic index and grand average hydropathicity (GRAVY) of each subunits and different combinations of them were computed using the Expasy's ProtParam tool (web.expasy.org/protparam/). Chance of solubility of chimeric protein when overexpressed in *E. coli* was analyzed by recombinant protein solubility server (http://biotech.ou.edu/) and PROSI II server[13].

Protein's Secondary Structure

The protein secondary structure prediction was performed by GOR secondary structure prediction (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_gor4.html)[14] and integrated methods[15]. The information about secondary structure, solvent accessibility, transmembrane helices, globular regions, coiled-coil regions, structural switch regions, B-values, disorder regions, intra-residue contacts, protein-protein and protein-DNA binding sites, sub-cellular localization, domain boundaries, beta-barrels, cysteine bonds, metal binding sites and disulphide bridges of chimeric antigen were adapted from Predict Protein servers (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_gor4.html/and https://www.predictprotein.org/).

Tertiary Structure Prediction

I-TASSER online software, the database of three-dimensional protein models (http://zhanglab.ccmb.med.umich.edu/I-TASSER/) calculated by comparative modeling such as SWISS-MODEL work space which was used to 3D structure prediction. The Swiss model is the automated modeling software, which develops the 3D structure model of unknown structure protein based on the sequence homology with the known structured protein[16]. The PyMOL software was used to visualize the modeled 3D structures.

Tertiary Structure Validation

To recognize the errors in the generated models, coordinates were supplied by uploading 3D structures in PDB format into ProSA web (https://prosa.services.came.sbg.ac.at/prosa.php), which is a valid software for this job[17]. The structure was validated to see the quality of the resulting stereochemistry of structure by Ramachandran plot in RAMPAGE web page (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php)[18]. Chiron minimizes the number of non-physical atomic interactions (clashes) in the given protein structure. Chiron has been benchmarked on high and low resolution crystal structures and homology models (http://troll.med.unc.edu/chiron/processManager.php). YASARA energy minimization server (http://www.yasara.org/minimizationserver.htm) also used for improving physical realism, stereochemistry, and side-chain accuracy in homology modeling[19].

Immunoinformatic Analyses of Chimeric Antigen

Antigenic Propensity

VaxiJen server (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html) was used to estimate the probability of antigenicity of the single and assembled forms of selected proteins. The predicted antigenic peptides server also recruited for this job. This program predicts those segments from within a protein sequence that are likely to be antigenic by eliciting an antibody response. In this software, antigenic peptides are determined using the method of Kolaskar and Tongaonkar[20]. Predictions are based on a table that reflects the occurrence of amino acid residues in experimentally known segmental epitopes.

B-Cell Epitopes

For prediction of B-cell epitopes, 5 different algorithms were used from immune epitope database (IEDB). The full-length protein sequence was subjected to BCPreds analysis tool (http://tools.immuneepitope.org/tools/bcell/iedb_input). Epitopes analysis by BCPreds is based on physico-chemical properties of amino acids such as hydrophilicity, flexibility/mobility, accessibility, polarity, exposed surface and turns. All predicted B-cell epitopes (20 mers) having a BCPreds cutoff score >1 were selected. DiscoTope 2.0 online software (http://www.cbs.dtu.dk/services/DiscoTope/) was used for predicting conformational B-cell epitopes from three dimensional protein structures[21]. Prediction of conformational B-cell epitope from primary sequence was done with web server CBTOPE (http://www.imtech.res.in/raghava/cbtope/)[22].

T-Cell Epitopes

For this purpose, prediction of MHC-I & MHC-II binding peptides were done by immune epitope database (IEDB). Also with CTLPred online software analysis (http://www.imtech.res.in/raghava/ctlpred/), among different haplotypes of major histocompatibility molecules (MHC) in mice, those of the BALB/c mice harboring H2d were selected. The MHC I binding predictions were made using the IEDB analysis resource consensus tool (http://tools.immuneepitope.org/mhci/) which combines the predictions from ANN aka NetMHC (3.4)[23,24], SMM[25] and comblib[26]. Quantitative matrix (QM), support vector machine (SVM) and artificial neural network (ANN) prediction approaches were used to recognizing T-cell epitopes crucially related to MHC class I[27].

Prediction of Allergens

IgE epitopes (allergens) of each subunit and whole chimeric antigen was predicted by Algpred (http://www.imtech.res.in/raghava/algpred/). This software allows to predict allergen using SVMc + IgE epitope + ARPs (allergen-representative peptides) BLAST + MAST[28]. AllerTOP 1.0 online software (http://www.pharmfac.net/allertop/) was also used. The principle of this method is based on auto cross covariance (ACC) transformation of protein sequences into uniform equal-length vectors.

Experimental Analysis

Construct Preparation and Cloning Confirmation

After all bioinformatics evaluations, chimeric gene sequence was ordered to Shine Gene Bio-Technologies Company (Shanghai, P. R. China) and prepared as a recombinant construct in pET28a (+) plasmid. Competent cells of *E. coli* BL21(DE3) were prepared by chemical method ($CaCl_2$) and recombinant construct was transferred to them by the heat shock method. Plasmid extraction was done by alkaline lysis of transgenic clones. Since chimeric gene was inserted between *Eco*RI and *Hin*dIII restriction sites, enzymatic digestion with both above enzymes was done to confirm the accurate cloning of chimeric gene in the plasmid.

Expression of Recombinant Protein in the Prokaryotic Host

The overnight culture of recombinant cells of *E. coli* BL21(DE3) was used to inoculate 5 ml of Luria–Bertani (LB) medium for recombinant protein expression. The culture was grown at 37°C up to an OD_{600} (OD, optical density) of 0.5 – 0.7. The protein expression was induced by the addition of 1 mM IPTG (isopropyl-β-d-thiogalactoside). Cells were harvested by centrifugation at 5,000 rpm for 5 min. Each pellet was resuspended in 50 µL of 1X SDS-PAGE sample buffer and was analyzed by 12% SDS-PAGE. The quantity of recombinant expression of ESI according to band intensity was calculated with Bio Doc Analysis software Ver.BDA-U-849.

Results

Subunits Selection and Chimera Design

Entire EspA, belong to type III secretion proteins, 310 amino acids (aa) from C-terminal of Intimin and B subunit of Stx2 holotoxin were selected for the present study. A linker contained of 5 amino acids with rigid structure in order to find the best subunits separation and epitope exposing chimeric antigen was considered between segments[29]. Correct combination of subunits in chimera construction was choosen based on instability index and half-life of various combinations when overexpressed in different organisms (Table 1). Schematic diagram of protein domain structures with linker's sites depicted with

Table 1 — Instability and half-life of each subunit and whole chimera when overexpressed in different organisms

Protein	Instability index	Half-life (Mammalian reticulocytes, *in vitro*)	Half-life (Yeast, *in vivo*)	Half-life (*E.coli, in vivo*)
EspA	33.22	4.4 h	20 h	20 h
Stx2B	37.07	1.3 h	3 min	3 min
Intimin	23.15	1.3 h	3 min	3 min
EspA-Stx2B-Intimin	27.00	4.4 h	20 h	10 h
EspA-Intimin-Stx2B	27.58	4.4 h	20 h	10 h
Stx2b-EspA-Intimin	27.00	1.3 h	3 min	3 min
Stx2B-Intimin-EspA	26.97	1.3 h	3 min	3 min
Intimin-EspA-Stx2B	27.58	1.3 h	3 min	3 min
Intimin-Stx2B-EspA	27.58	1.3 h	3 min	3 min

DOG 2.0.1 software is shown in Figure 1. The codon adaptation index for the native fusion ESI gene was 0.67, while the optimized one had a codon adaptation index of 0.82. Both the wild type and the synthetic chimera were analyzed for their codon bias and GC content. GC content (important for transcription and translation processes) after optimization was improved from 41.67% to 47.79%. Percentage of codon having a frequency distribution of 91–100 in the native chimeric ESI gene was 44%, which was significantly improved to 51% in the optimized gene sequence. Negative *cis* elements in the native ESI gene sequence were reduced to one after codon optimization. From the RaCC results, after codon optimization 9, 3 and 2 codons respectively belong to Ilu, Leu and Arg changed according to *E. coli* codon preference. The *Eco*RI and *Hin*dIII restriction sites for cloning in pET family vectors were successfully introduced at the N and C-terminal of the sequence.

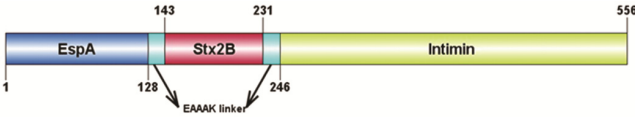

Fig. 1 — Schematic representation of chimera construct consists of EspA, Stx2B and Intimin proteins bound together by appropriate linkers.

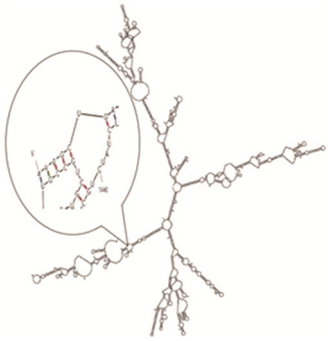

Fig. 2 — Prediction of RNA secondary structure of the ESI chimeric gene using Mfold algorithm. Predicted structure has no long stable hairpin and pseudoknot at 5′ site of mRNA. A: The one predicted folding for the sequence.

Messenger RNA Structure Prediction

Online servers combine four separate predictions and analysis algorithms: calculating a partition function, predicting a maximum free energy (MFE) structure, finding structures with maximum expected accuracy, and pseudoknot prediction. From this study, it was determined that the 5′ terminus of the gene (containing start codon) was folded in the similar pattern typical to the bacterial gene structures with the parameters of the percent sub optimality number, upper bound on the number of computed foldings, the maximum interior/bulge loop size, the maximum asymmetry of an interior/bulge loop equal to 5, 50, 30 and 30, respectively. All 40 structural elements obtained in this analysis revealed folding of the RNA construct. ΔG of the best predicted structure was -485.80 kcal/mol and the first nucleotides at 5′ inserted in a stem (Fig. 2), whereas in native mRNA, first nucleotides formed pseudoknot and the structure ΔG was -469.70 kcal/mol. The computed folding contained 3197 bp out of 103951 (3.1%) in the energy dot plot.

Physico-Chemical Parameters

The average molecular weight of ESI with 556 amino acids was 59.14 kDa. Isoelectric point (pI) was 8.65. Acidity of the protein was indicated by the pI value pIb7. Extinction coefficient of this chimera at 280 nm was 58580 M^{-1} cm^{-1}. The half-life was greater than 10 h when overexpressed in *E. coli*. Since the instability index of ESI was 27, Expasy's ProtParam classified this chimeric protein as stable. The grand average of hydropathicity (GRAVY) for ESI computed -0.279 which indicated that this protein

is soluble in water. From Pep server prediction, ESI has a 100% chance for soluble expression in the host cell.

Secondary and Tertiary Structures of ESI

Figure 3 shows the secondary structure prediction of the protein. Using the software Predict Protein and GOR IV, we obtained correctly predicted structural elements. Such prediction results lead to our confidence in the reliability of the ESI prediction. From the results, random coil, extended strand and alpha helix were structural contents of protein. Composition of the predicted secondary structure of chimeric protein was 0% (H), 41.01% (E), and 58.99% (L). The total amino acid residues are made up of 6 sheets, 7 hairpins, 1 psi loop, 1 beta bulges, 25 strands, 8 helices, 50 beta turns and 27 gamma turns. GOR analysis results showed three helix peaks located between positions 105–124, 267–289 and 434–453 corresponding to the linker fragments. Other sequence analyses and the prediction of protein structure and function such as low-complexity regions (SEG), regions lacking regular structure (NORS), transmembrane helices, and coiled-coil regions showed that the sequence does not contain long regions with any regular secondary structure. Analysis of the amino acid composition demonstrated three regions with a low sequence complexity. These regions have linker sequences.

Results of tertiary structure of the fusion protein construction using I-TASSER showed a protein with three main domains joined together with tight linkers (Fig. 4). Potential errors of 3D models of protein structures were considered with ProSA tool. The z-score of the input structure (-4.55) was within the range of scores typically found in native proteins of similar size (Fig. 5). The confidence score (C-score) for estimating the quality of predicted models by I-TASSER was -1.70. It 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 template modeling's score (TM-score) for this model was 0.51 ± 0.15. The expected root mean squared deviation (RMSD) was 11.6 ± 4.5Å.

Evaluation of Model Stability

The Ramachandran plot analysis revealed that 465 amino acid residues (83.9%) from modeled structure generated by I-TASSER were incorporated in the

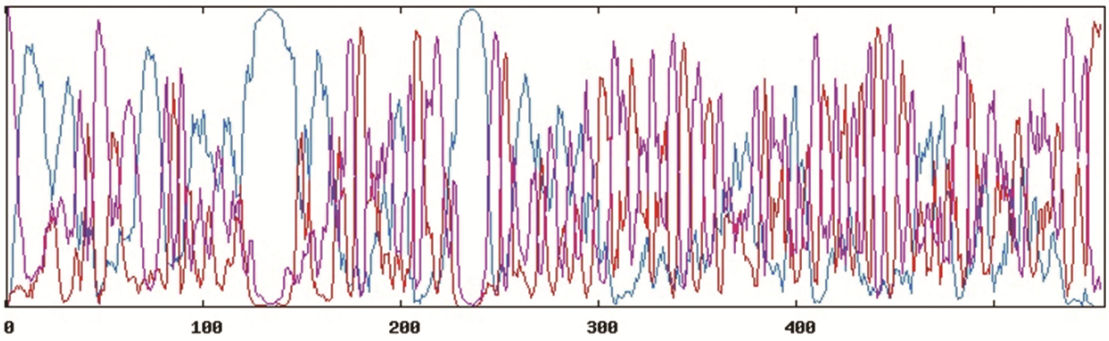

Fig. 3 — Graphical representation for secondary structure prediction of ESI. Extended strand: purple, Coil: red, Helix: blue. Two helix peaks are corresponded to the linker fragments.

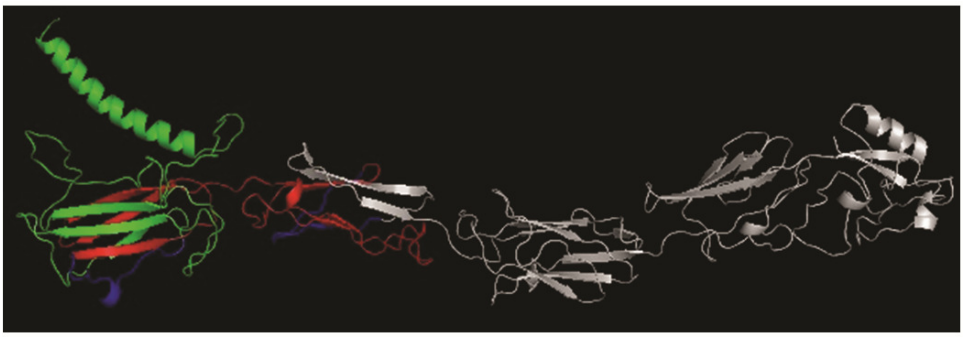

Fig. 4 — Tertiary structure of chimeric protein by I-TASSER software. Three main domains are separated by rigid linkers (blue color). EspA, Stx2B and Intimin are shown by green, red and gray colors, respectively.

favored regions (A, B, and L) of the plot (Fig. 6). Apart from that 48 amino acid residues were in allowed regions (a, b, l, and p) of the plot, the number of residues in outlier region was 41.

B-Cell, T-Cell and Allergenic Site Prediction

B-cell and T-cell mediated immunity can be produced from a good antigen which is vaccine candidate. Best epitopes were selected based on Parker, Karplus, Emini , Pellequer and Kolaskar methods as mentioned in BCPred data base. From the IEDB server, linear B-cell epitopes by different methods were predicted (Table 2). Furthermore, the conformational epitopes for B cells were predicted by the CBTOPE (Table 3). DiscoTope results proved very good distribution of conformational B-cell epitopes in the 3D structure of the chimera (Table 4). For screening T-cell

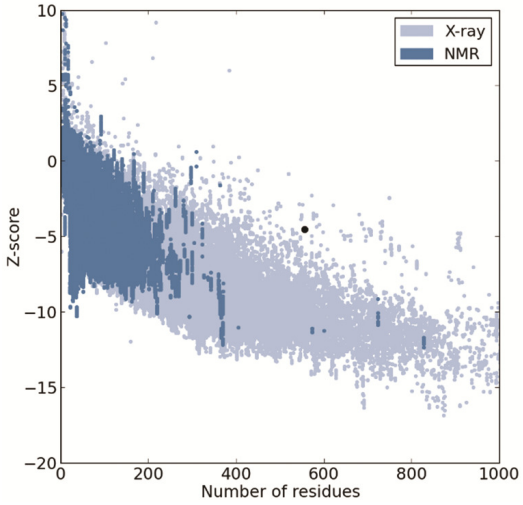

Fig. 5 — ProSA-web z-score chimeric protein plot. The z-score indicates overall model quality. ProSA-web z-scores of all protein chains in PDB determined by X-ray crystallography (light blue) or NMR spectroscopy (dark blue) with respect to their length. The plot shows results with a z-score ≤ 10. The z-score of ESI is highlighted as a large dot. The value is in the range of native conformations.

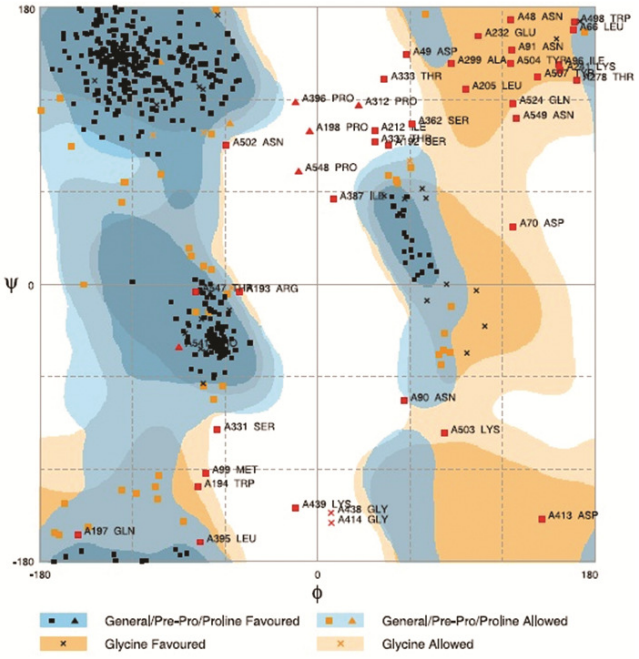

Fig. 6 — Evaluation of ESI stability based on a Ramachandran plot.

Table 2 — Linear B-cell epitopes prediction. 4 top scores of peptide from each methods of prediction are shown

Peptide	Residue	Position	Score	Method of predication	Peptide	Residue	Position	Score	Method of predication
SGG**D**GTY	**D**	410	1.381	Chou &Fasman Beta-Turn	VYA**V**CVE	V	553	1.234	Kolaskar&Tongaonkar antigenicity
GGD**G**TYS	**G**	411	1.381	Chou &Fasman Beta-Turn	NVY**A**VCV	A	552	1.223	Kolaskar&Tongaonkar antigenicity
NGQ**P**VNN	**P**	309	1.320	Chou &Fasman Beta-Turn	AVL**F**ALV	F	152	1.212	Kolaskar&Tongaonkar antigenicity
KAS**G**GDG	**G**	408	1.320	Chou &Fasman Beta-Turn	VLF**A**LVS	A	153	1.205	Kolaskar&Tongaonkar antigenicity
KY**N**END	**N**	175	5.277	Emini Surface Accessibility	QSS**T**DKN	T	31	6.700	Parker hydrophilicity
ND**P**RND	**P**	50	4.918	Emini Surface Accessibility	TSG**D**KQT	D	450	6.329	Parker hydrophilicity
SK**Y**NEN	**Y**	174	4.235	Emini Surface Accessibility	KQT**S**SEQ	S	521	6.243	Parker hydrophilicity
YN**E**NDT	**E**	176	3.808	Emini Surface Accessibility	SSE**Q**RSG	Q	524	6.171	Parker hydrophilicity
KQ**T**SSEQ	**S**	521	1.135	Karplus& Schulz Flexibility	TSS**E**QRS	E	523	1.127	Karplus & Schulz flexibility
QTS**S**EQR	**S**	522	1.134	Karplus& Schulz Flexibility	SSE**Q**RSG	Q	524	1.121	Karplus& Schulz flexibility

Table 3— Conformational B-cell epitopes predicted by CBTOPE server from amino acid sequence of ESI. Above scale 4 can be considered as epitope residue

Amino acid	Position	Probability scale	Amino acid	Position	Probability scale
ADMNE	1-5	4	VIFF	272-275	4
ANLV	17-20	4	Q	277	4
A	37	4	IKADKTTAV	284-293	4
QDVIDYIND	41-49	4	Y	302	4
IS	54-55	4	VKVMKNGQPVNNQSVTFSTNFGMFNGKSQTQA	304-335	4
T	57	4	T	336	5
IR	59-60	4	TG	337-338	4
LQTV	71-74	4	AT	343-344	4
K	81	4	AGK	352-354	4
VV	87-88	4	VS	361-362	4
N	89	5	F	374	4
NSQLEI	91-96	4	VD	384-385	4
LGK	126-128	4	G	387	4
E	170	4	N	390	4
KY	173-174	5	N	397	4
N	175	4	QYGQ	401-404	4
ENDT	176-179	5	KLKASGG	406-412	4
F	180	4	D	413	5
T	181	4	GTY	414-416	4
V	182	5	W	418	4
V	184	5	S	424	4
EYWTS	188-192	4	S	431	4
R	183	5	VTLNG	434-437	4
WNL	184-186	4	SVVIKA	441-446	4
L	200	4	M	477	4
T	209	4	IYD	494-496	4
VTI	210-212	5	SIT	512-514	4
ST	215-216	4	G	527	4
ES	218-219	4	T	531	4
G	220	5	YNL	532-534	5
KK	246-247	4	E	556	4

epitopes, IEDB analysis resource consensus tool for MHC I with H2Dd allele and MHC II with H-2-ID allele were used and 10 best percentile rank peptide were selected and then analyzed with Vaxijen. Propred-I (47 MHC Class-I alleles), Propred (51 MHC Class-II alleles), and MHCPred were used to identify common T-cell epitopes that share B-cell epitope sequence, which can interact with both MHC class I & II with the highest number (Table 5). Also a pictorial representation of chimera antigen with common epitopes is shown in Figure 7. According to different allergenicity prediction approaches in AlgPred and AllerTOP tools, this protein was not detected as a potential allergen. Finally, search for potential allergens in ESI sequence showed any similarities between any region and SDAP allergen library.

Confirmation of Chimeric Gene Cloning

Enzymatic digestion with restriction endonucleases according to MCS of pET28a (+) plasmid was done on recombinant construct consist of chimeric gene and then the product of digestion and intact plasmid were electrophoresed on 1% agarose gel (Fig. 8).

Analysis of Recombinant Expression of Chimeric Antigen

Because of compatibility of the protein expression system of *E. coli* BL21(DE3) and pET plasmids, IPTG at the final concentration of 1 mM was used to induce recombinant expression of chimeric antigen in transgenic host cells and result of this examination was analyzed on 12% SDS-PAGE. By considering the length of gene which is coded chimera residues and N-terminal fusion from pET28 (a) construct, we expected that recombinant protein must had a

Table 4 — Conformational B-cell epitopes from full length proteins using DiscoTope server

Amino acid	Position	Contact number	DiscoTope result
ADMNEASKAST	1-11	7-4-0-0-2-3-3-5-5-6-5	-2.244,-1.418,-0.916,-0.439,-.759,-2.261,-2.105, -2.48, 5-3.191, -3.573,-3.646
KN	33-34	2-2	-2.689,-2.053
K	36	1	-3.154
D	45	5	-1.845
INDPRN	47-52	2-1-9-0-12-5	-1.845,0.437,3.164,2.518 2.581, -1.219,-2.807
S	63	0	-3.344
N	195	11	-3.542
STPG	249-252	2-7-18-9	-2.914-3.111,-3.698 ,-3.661
NG	309-310	15-8	-2.491,-1.671
P	312	11	-3.337
DG	340-341	2-3	-1.372,-2.279
K	383	9	-3.273
G	388	2	-1.539
N	390	3	-2.043
R	392	0	-1.275
E	394	9	-3.367
GG	411-412	2-9	-3.357,-3.525
G	499	12	-3.668
N	502	10	-3.641
QRSG	524-527	3-1-2-5	-3.693,-1.282,-3.193,-1.833

Table 5 —Common epitopes from protein that can produce both the B- and T-cell mediated immunity are represented along with their various parameters

Predicted epitopes	IC50 value	VaxiJen score	Number of MHC Class I binding alleles	Number of MHC Class II binding alleles	Total number of MHC binding alleles
QDVIDYIND	3303.70	-1.7027	11	0	11
IKADKTTAV	426.58	0.8932	14	11	25
KLKASGG	1690.44	1.2027	3	4	7
KQTSSEQ	2162.72	2.3121	6	9	15
KYNENDTF	543.25	-0.8699	10	0	10
PNVYAVCVE	3819.44	1.2251	12	0	12
FMAVLFALV	18.79	1.1534	12	17	29

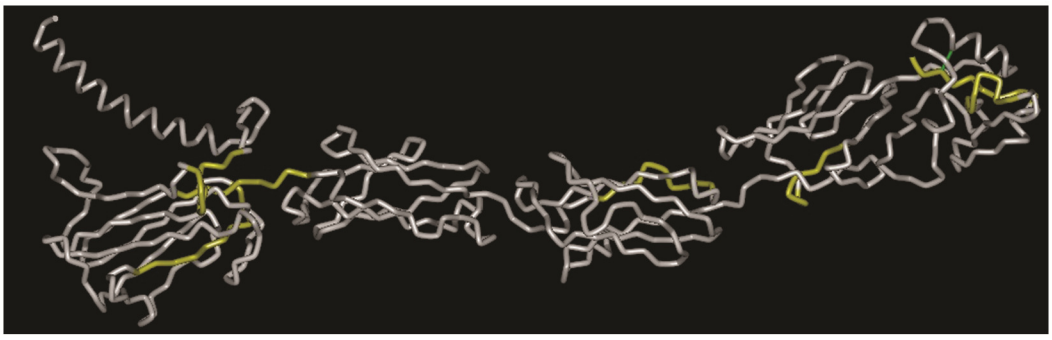

Fig. 7 — Representation of common B & T cell epitopes on chimera tertiary structure (yellow parts).

molecular weight ≅ 63 kDa. According to Figure 9, the band of ESI antigen in induced cells is clear in comparison with non-induced cells. From the results of concentration assay of total cell proteins by Bradford reagent and percentage of band intensity belong to recombinant chimera, the quantity of ESI was calculated equal to 1.9 g/1 medium.

Discussion

Vaccine candidate molecules must be safe and immunogenic and capable to induce protective immunity against a special or wide range pathogen(s). The aim of this study was to design a chimeric protein as a vaccine candidate with *in silico* approach which carries epitopes from different virulence factors,

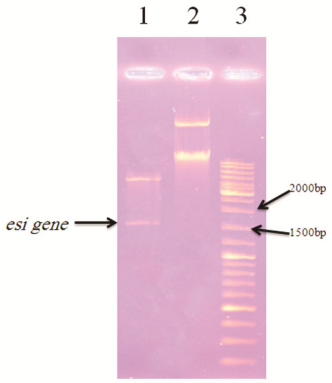

Fig. 8 — Gene cloning analysis on 1% agarose: Lane 1: Digested pET28a-ESI with *Eco*RI and *Hin*dIII, Lane 2: Intact pET28a-ESI construct, Lane 3: DNA molecular marker (100 bp DNA size marker ThermoSIENTIFIC. #SM1173). The band of *esi* gene is clear on gel near to 1500 bp band of DNA ladder.

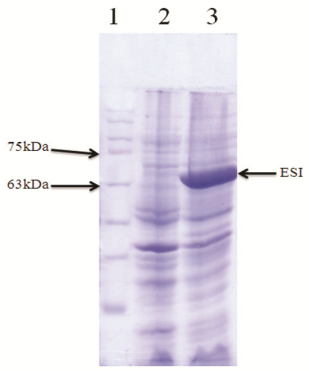

Fig. 9 — Recombinant expression analysis of ESI antigen on 12% SDS-PAGE: Lane 1: Protein molecular marker (prestained protein ladder (10-250 kDa) Sinaclon. PR911641), Lane 2: Non-induced lysed cells (negative control), Lane 3: Induced lysed cells. The recombinant expression band belong to ESI chimera in induced cells (≅ 63 kDa) is clear in comparison with non-induced cells.

linkers, and adjuvant sequences offer not only increased immunogenicity of the recombinant antigen, but also the possibility to excite a broad cellular or humoral immune response. Three putative antigens were considered in a recombinant construct. Receptor binding subunit of Stx2 (Stx2B) and 2 major antigens involving in bacterial adherence (EspA & Intimin) present on the most prevalent EHEC pathogens were used for *in silico* designing of a chimeric subunit vaccine against EHEC. Stx2B is a nontoxic subunit of Stx2 holotoxin that plays an important role in EHEC virulence and pathogenesis and possesses adjuvant properties[10]. Since linkers can play an important role in displaying specific epitopes in the total structure of the chimeric protein, selection of an appropriate linker was done precisely. Besides appropriate amino acid composition, the overall folding of the linker must be considered. In order to separate different domains of our chimeric protein, linkers consisting of 3 EAAAK repeats were designed. There is some report that the salt bridge Glu^{-}-Lys^{+} between repeated Ala can stabilize helix formation[30]. Chimeric proteins with a helical linker assumed a more elongated conformation compared to those with a flexible linker[31]. Our successful experience of using four repeated EAAAK sequences in chimeric gene has shown that it could lead to logically acceptable results[32]. The results from instability index and half-life of various combinations of subunits were considered for selection of a good subunit arrangement. The lowest instability index and the highest half-life for a combination is suitable for a chimeric antigen. So, EspA-Stx2B-Intimin arrangement was chosen for construct sequence. Codon optimization was performed by improving the overall GC content of the gene, codon adaptation index and codon frequency distribution and reducing negative elements that may form unfavorable secondary structures on mRNA. The optimized gene sequence had a codon adaptation index of 0.82, indicating that the optimized gene sequence could be expressed well in *E. coli*. The software used for prediction of RNA secondary structure was Mfold. Advantages of Mfold are that it employs a theoretically tractable DP algorithm which can find the minimum ΔG structure within its thermodynamic model and high ability to predict true positive base pairs. Prediction data for mRNA structure showed that the mRNA was stable for efficient translation in the 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[33]. GOR method was used in studying the secondary structure because this method takes into account not only the probability of each amino acid having a particular secondary structure, but also the conditional probability of the amino acid assuming each structure given the contributions of its neighbours. The approach is both more sensitive and more accurate than that of Chou and Fasman[34] because amino acid structural propensities are only strong for a small number of amino acids such as proline and glycine. The theoretical pI value of protein (pI = 8.65) showed a little alkaline nature of the protein. Extinction coefficient of ESI at 280 nm was high, indicating the presence of high concentration of Cys, Trp and Tyr. Expasy's ProtParam classifies the ESI protein as stable (instability index = 27). High aliphatic index of

ESI (76.76) indicates that the protein may be stable for a wide range of temperature. From the secondary structure prediction results it was relieved that no disulphide bond involved in the final structure. It means that ESI can fold simply in prokaryotic organisms as expression hosts. Functional properties of a protein sequence are one of the problems in proteomic. This task is usually facilitated by accurate three-dimensional structure of the protein in question and can be done with comparative and *ab initio* protein structure prediction methods. Both methods were used for predicting three-dimensional structure of the chimeric protein. EspA and Intimin but not Stx2b monomer have templates in the Protein Data Bank (PDB) library on the basis of which the 3D modeled structure of the protein was generated by Swiss model software. Our result showed that *ab initio*I-TASSER software could predict the 3D structure of chimeric ESI protein. Also, we understood that the linkers separate domains from each other in a good manner and each subunit has the same fold as PDB library.

Assessment of the accuracy and reliability of experimental and theoretical models of protein structures is necessary. RMSD and TM-score were used for the evaluation of the predicted model. TM-score and RMSD are known standards for measuring structural similarity between two structures which are usually used to measure the accuracy of structural modeling when the native structure is known. In case where the native structure is not known, it becomes necessary to predict the quality of the modeling prediction, i.e. what is the distance between the predicted model and the native structures? To answer this question, we tried to predict the TM-score and RMSD of the predicted models relative to the native structures based on the C-score. The best RMSD value was the result of superimposing our model on a template which consisted of 556 amino acids; 100% of total protein residues. Expected TM-score of 0.51 ± 0.15 validates the accuracy of the ESI model (TM-score >0.5 indicates a model of correct topology). The z-score indicates overall model quality and measures the deviation of the total energy of the structure with respect to an energy distribution derived from random conformations. The ProSA web results indicate that ESI has features which are the characteristics of native structures. Our chimeric structure showed desirable protein stability based on Ramachandran plot predictions. In Ramachandran plot analysis, nearly, a negligible 7.4% of the residues were found to be in outlier region that could probably be due to the presence of chimeric junctions.

A vaccine candidate should be able to induce strong B cell and T cell responses. For this reason, the ability to introduce T cell and B cell epitopes is important for vaccine design and development. BCpreds was used to predict linear B-cell epitopes from each protein. The methods employed for prediction of linear epitopes led to almost identical results with some minor differences. The recognition of conformational epitopes in antibody-antigen interaction is a necessary for the logical design of novel vaccines[35]. Some epitopes were recognized by single software only. DiscoTope uses 3D structure which seems to be more reliable. Immune epitope database (IEDB) used for the purpose of prediction of T-cell epitopes and binding affinity of MHC molecules. The present study identified potential T-epitopes derived from antigenic B-cell epitopes of chimeric proteins. Selected T-epitopes are antigenic with the potential to interact with H2D alleles. Therefore, this chimeric protein has epitopes likely to induce both the B-cell and T-cell mediated immune responses. Finally, non-allergenic region was recognized in in chimera sequence. Enzymatic digestion of recombinant construct showed that *esi* chimeric gene was inserted in desired sites (*Eco*RI and *Hind*III) and analyze of recombinant expression of ESI showed that this antigen could be produced in good quantity in *E. coli* BL21(DE3) as host cells.

Our data indicate that epitopes of the chimeric protein ESI, designed from colonization factors and toxin subunit of EHEC could induce both B-cell and T-cell mediated immune responses. Stx2B also plays an adjuvant role, enhancing the immunogenicity of chimeric antigen. The possibility of successful production of a chimeric antigen that is composed of the important virulence factors of *E. coli* O157:H7 was prepared. It also emphasizes the capacity of immunoinformatics in assimilation of immunological data with bioinformatics to design broadly effective vaccine candidates.

Acknowledgment

We wish to thank Imam Hossein University, Iran for the sanction of grants to conduct the present study. The authors also declare no conflict of interests.

References

1 Nazarian S, Gargari S L, Rasooli I, Alerasol M, Bagheri S *et al*, Prevalent phenotypic and genotypic profile of

enterotoxigenic *Escherichia coli* among Iranian children, *Jpn J Infect Dis,* 67 (2014) 78-85.
2 Karch H, Tarr P I & Bielaszewska M, Enterohaemorrhagic *Escherichia coli* in human medicine, *Int J Med Microbiol,* 295 (2005) 405-418.
3 Larrie-Bagha S M, Rasooli I, Mousavi-Gargari S L, Rasooli Z & Nazarian S, Passive immunization by recombinant ferric enterobactin protein (FepA) from *Escherichia coli* O157, *Iran J Microbiol*, 5 (2013) 113-119.
4 Khaloie F, Porfarzam P, Rasooli I, Amani J, Nazarian S *et al*, *In silico* analysis of chimeric recombinant immunogen against three diarrhea causing bacteria, *J Cell Mol Res,* 5 (2013) 65-74.
5 O'Loughlin E V & Robins-Browne R M, Effect of Shiga toxin and Shiga-like toxins on eukaryotic cells, *Microbes Infect*, 3 (2001) 493-507.
6 Law D, Virulence factors of *Escherichia coli* O157 and other Shiga toxin-producing *E. coli*, *J Appl Microbiol*, 88 (2000) 729-745.
7 Fraser M E, Fujinaga M, Cherney M M, Melton-Celsa A R, Twiddy EM *et al*, Structure of Shiga toxin type 2 (Stx2) from *Escherichia coli* O157:H7, *J Biol Chem*, 25 (2004) 27511-27517.
8 Torgersen M L, Engedal N, Bergan J & Sandvig K, The intracellular journey of Shiga toxins, *Open Toxinology J,* 3 (2010) 3-12.
9 Moxley R A, *Escherichia coli* O157:H7: an update on intestinal colonization and virulence mechanisms, *Animal Health Res Rev,* 5 (2004) 15-33.
10 Nazarian S, Mousavi Gargari S L, Rasooli I, Amani J, Bagheri S *et al*, An *in silico* chimeric multi subunit vaccine targeting virulence factors of enterotoxigenic *Escherichia coli* (ETEC) with its bacterial inbuilt adjuvant, *J Microbiol Methods,* 90 (2014) 36-45.
11 Garcia-Angulo V A, Kalita A & Torres A G, Advances in the development of enterohemorrhagic *Escherichia coli* vaccines using murine models of infection, *Vaccine*, 11 (2013) 3229-3235.
12 Zuker M, Mfold web server for nucleic acid folding and hybridization prediction, *Nucleic Acids Res*, 31 (2013) 3406-3415.
13 Smialowski P, Martin-Galiano A J, Mikolajka A, Girschick T, Holak T A *et al*, Protein solubility: sequence based prediction and experimental verification, *Bioinformatics,*1 (2007) 2536-2542.
14 Garnier J, Gibrat J F & Robson B, GOR method for predicting protein secondary structure from amino acid sequence, *Methods Enzymol,* 266 (1996) 540-553.
15 Rost B, Yachdav G & Liu J, The Predict Protein server, *Nucleic Acids Res*, 1 (2004) 321-326.
16 Yang J, Yan R, Roy A, Xu D, Poisson J *et al*, The I-TASSER Suite: protein structure and function prediction, *Nat Methods,*12 (2015) 7-8.
17 Wiederstein M & Sippl M J, ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins, *Nucleic Acids Res,* 35 (2007) 407-410.
18 Lovell SC, Davis IW, Arendall WB, de Bakker PI, Word JM *et al*, Structure validation by Calpha geometry: phi, psi and Cbeta deviation, *Proteins*, 15 (2003) 437-450.
19 Krieger E, Joo K, Lee J, Lee J, Raman S *et al*, Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8, *Proteins,* 77 (2009) 114-122.
20 Kolaskar A S & Tongaonkar P C, A semi-empirical method for prediction of antigenic determinants on protein antigens, *FEBS Lett,*10 (1990) 172-174.
21 Kringelum J V, Lundegaard C, Lund O & Nielsen M, Reliable B cell epitope predictions: impacts of method development and improved benchmarking, *PLoS Comput Biol,* 8 (2012) 1-10.
22 Ansari H R & Raghava G P, Identification of conformational B-cell epitopes in an antigen from its primary sequence, *Immunome Res*, 6 (2010) 1-9.
23 Nielsen M, Lundegaard C, Worning P, Lauemøller S L, Lamberth K *et al*, Reliable prediction of T-cell epitopes using neural networks with novel sequence representations, *Protein Sci,*12 (2003) 1007-1017.
24 Lundegaard C, Lamberth K, Harndahl M, Buus S, Lund O *et al*, 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,* 1 (2008) 509-512.
25 Peters B & Sette A, Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method, *BMC Bioinformatics,* 31 (2005) 1-9.
26 Sidney J, Assarsson E, Moore C, Ngo S, Pinilla C *et al*, Quantitative peptide binding motifs for 19 human and mouse MHC class I molecules derived using positional scanning combinatorial peptide, *Immunome Res*, 25 (2008) 1-14.
27 Bhasin M & Raghava G P S, Prediction of CTL epitopes using QM, SVM and ANN techniques, *Vaccine*, 22 (2004) 3195-3201.
28 Saha S & Raghava G P, AlgPred: prediction of allergenic proteins and mapping of IgE epitopes, *Nucleic Acids Res*, 1 (2006) 202-209.
29 Chen X, Zaro J L & Shen W C, Fusion protein linkers: property, design and functionality, *Adv Drug Deliv Rev,* 65 (2013) 1357-1369.
30 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,*14 (2001) 529-532.
31 Arai R, Wriggers W, Nishikawa Y, Nagamune T & Fujisawa T, Conformations of variably linked chimeric proteins evaluated by synchrotron X-ray small angle scattering, *Proteins,* 57 (2004) 829-838.
32 Bagheri S, Mousavi Gargari S L, Rasooli I, Nazarian S & Alerasol M, A CssA, CssB and LTB chimeric protein induces protection against enterotoxigenic *Escherichia coli*, *Braz J Infect Dis*, 18 (2014) 308-314.
33 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*, 30 (2011) 1367-1382.
34 Chou P Y & Fasman G D, Prediction of the secondary structure of proteins from their amino acid sequence, *Adv Enzymol*, 47 (1987) 45-148.
35 Negi S S & Braun W, Automated detection of conformational epitopes using Phage display peptide sequences, *Bioinform Biol Insights,* 1 (2009) 71-81.