Research Article

In silico analysis of chimeric recombinant immunogen against three diarrhea causing bacteria

Farzane Khaloie [1], Poune Porfarzam [1], Iraj Rasooli [1,2], Jafar Amani [3], Shahram Nazarian [4], Seyed Mousavi Latif[1]*

1. Department of Biology, Faculty of Basic science, Shahed University, Tehran, Iran
2. Molecular Microbiology research center, Shahed University, Tehran, Iran
3. Applied Microbiology Research Center, Baqiyatallah Medical Science University, Tehran, Iran
4. Department of Biology, Faculty of Basic science, Imam Hossain University, Tehran, Iran

Received 26 May 2013 Accepted 14 November 2013

Abstract

Shigella and *Escherichia* belong to the *Enterobacteriaceae* family which are the cause for most of the diarrheal cases in the world. *Shigella* can cause bacterial dysenteries and shigellosis. One of the most effective proteins for pathogenesis is invasion plasmid antigen C (IpaC). Other bacteria like Enterotoxogenic (ETEC), Enterohemorrhagic (EHEC), and *E.coli* can also cause diarrhea and produce intestinal disorders. Colonization factor antigen I (CFA/I), a critical virulence protein for these infections, has two subunits i.e. CfaB and CfaE. EHEC Attachment of bacteria is the main step of infection with intimin playing the key role in this function. This study was designed to elicit protection against the majority of diarrheal pathogens via development of polyvalent vaccine against *Shigella*, ETEC and EHEC. *In silico* techniques are as best tools to design new vaccines. For this purpose the immunogenic epitopes of CfaB, IpaC and Intimin were identified through bioinformatic tools and were then selected as major antigens to construct a chimeric protein (CII). The humoral and cellular immunities were analyzed bioinformatically. Prediction of allergens and mapping of IgE epitopes were carried out. The bioinformatic analysis showed each domain was folded separately in fusion structure. CII had many T and B cell epitopes in both linear and three-dimensional structures. This prediction of the chimeric construct had the potential to induce CD4+ and CD8+ immune responses against these pathogens. In addition CII could be accessible to surveillance by the immune system in mouse and human. In conclusion*, in silico* analysis showed that this chimeric protein can be used as a vaccine against *Shigella*, ETEC and EHEC simultaneously.

Keywords**:** Intimin; CfaB; IpaC; recombinant vaccine; chimeric protein

Introduction

Diarrheal diseases are among the most common causes of death in the world that can spread from person-to-person via the fecal-oral route. High rates of disease are found in daycare centers, hospitals and nursing homes. In addition, diarrhea is frequently reported in food-borne and water-borne outbreaks. *Escherichia coli, Shigella* and *Salmonella spp*. are the most important causing diarrhea in the world. The diseases attributed to *E. coli* include childhood and traveler's diarrhea (ETEC), bloody diarrhea and hemolytic uremic syndrome (HUS) (Bhatnagar et al., 1993). *Shigella* is a major source of bacterial diarrhea in developing countries. The most important diarrheogenic *Shigella* strains are *S. flexneri, S. sonnei, S. Boydii* and *S. dysenteriae* and untyped *Shigella* strains (Katouli et al., 1990). Diarrhea is the condition of having three or more loose or liquid bowel movements per day. The loss of fluids through diarrhea can cause dehydration and electrolyte imbalances. One of the most important bacteria producing severe diarrhea is *S. flexneri*, a gram-negative intracellular pathogen responsible for bacillary dysentery (shigellosis) (Terry et al., 2008b) which includes severe inflammation, fever, abdominal cramping, and ulceration of the colonic mucosa (Kueltzo et al., 2003; Nazarian et al., 2013). As estimated earlier, 164.7 million shigellosis occur per year, of which 1.1 million cases result in death, with a high rates of infant mortality in underdeveloped countries (Jennison and Verma, 2004). An initial important step in pathogenesis is the invasion of colonic epithelial cells that resulted in cytoskeletal rearrangements at the site of bacterial contact in the host cell.

*Corresponding author E-mail:
slmousavi@shahed.ac.ir

These localized changes lead to the formation of filopodia and trap the pathogen within a membrane-bound vacuole that is rapidly lysed, finally bacterium access the host cell cytoplasm (Picking et al., 2001). For the invasion of epithelial cell, *S. flexneri* requires encoding genes on a large virulence plasmid. The IpaA–D proteins play a central role in this process (Tran Van Nhieu et al., 1999). Upon host cell contact, IpaB and IpaC are embedded into the host cell membrane and generate a pore like transmembrane channel called the translocon. Effecter proteins can pass through the translocon into the host cell cytoplasm (Terry et al., 2008b; Tran Van Nhieu et al., 1999). Filopodial and lamellipodial extensions are induced when IpaC has access to the cell cytosol antibodies recognizing the C-terminal region of IpaC between residues 297 and 349. These data argue that the C-terminal domain of IpaC needs to interact with components present in the cell cytosol in order to promote actin polymerization (Tran Van Nhieu et al., 1999). The C-terminal tail residues 344-363 appears to contain a major effector function of IpaC required for cellular invasion (Terry et al., 2008b). In the vast majority of the cells, secretion of IpaC via the TTSS occurs at one pole of the *Shigella* cell during epithelial cell invasion. Destabilization of the cadherin complex by IpaC may modulate this transmission allowing bacterial enhanced multiplication within infected cells. Thus, IpaC might be a multifunctional protein that controls the invasion process in a finely tuned manner, and is, therefore, an important effector of *Shigella* pathogenesis (Barzu et al., 1997). (Picking et al., 2001).

Enterotoxigenic *Escherichia coli* (ETEC), is another important etiological agent of travelers' diarrhea and a main cause of infantile death in developing countries (Luiz et al., 2008, Nazarian et al., 2013). ETEC pathogenesis depends on the ability to produce the heat-stable (ST) and/or heat-labile (LT) enterotoxins after attachment of bacteria to the intestinal epithelia with colonization factors CFs or CFAs (Luiz et al., 2008). Colonization factor antigen I (CFA/I) fimbriae contains a polymer consisting of 1,000 copies of the major pilin subunit CfaB, and one or a few copies of the tip-residing adhesive minor subunit CfaE (Li et al., 2009). It thus seems that CfaB subunit is a carbohydrate-binding protein which especially interacts with a number of carbohydrate sequences present in human small intestinal glycosphingolipids and glycoproteins. Carbohydrate-binding activity donates the attachment of CFA/I-fimbriated *E. coli* to host intestinal epithelium. Indeed, monoclonal antibodies directed against this protein could inhibit the binding of CFA/I-expressing cells to human intestinal cell (Jansson et al., 2006). Therefore, CfaB may be important targets for vaccine development.

Enterohemorrhagic *Escherichia coli* (EHEC) O157:H7, a human pathogen that causes bloody diarrhea and hemolytic uremic syndrome leading to kidney failure and even death. Cattle are the main reservoirs of EHEC O157:H7 strains and the associated diseases are brought by ingestion of undercooked or raw milk (Dean-Nystrom et al., 1998). EHEC belongs to a family of pathogens with ability to produce attaching and effacing (A/E) lesions characterized by degeneration of the epithelial intestinal microvilli. (DeVinney et al., 1999). The first gene for A/E activity is eae encoding intimin, an outer membrane adhesion protein essential for intimate bacterial attachment to eukaryotic host cells. Intimin is a 94-kDa outer membrane protein in enterohemorrhagic Escherichia coli (EHEC) (Tran Van Nhieu et al., 1999). Intimin is required for colonization of E. coli O157:H7 and mediates bacterial attachment to the plasma membrane of infected cells. eae is a part of a chromosomal island called the locus of enterocyte effacement (LEE). The LEE encodes a type III secretion system (TTSS) which includes a translocated intimin receptor (Tir) and three secreted proteins EspA, EspB, and EspD. These proteins are required for signal transduction in mammalian host cells and A/E lesion formation (Amani et al., 2010; Vlisidou et al., 2006). The Tir-Intimin interaction triggers actin cytoskeletal rearrangements, resulting in pedestal formation (Amani et al., 2009). Intimate attachment to the host cell leading to the formation of A/E lesions is an essential feature of EHEC pathogenesis. The *eaeA* gene plays a key role in the pathogenesis of the AE lesion therefore intimin is a critical factor for invasion (Shaikh et al., 2003).

Studies on the different intimins have shown that receptor-binding activity is localized to the C-terminal 280 amino acids (Int280). An experimental vaccination with the carboxy-terminal of intimin induced strong response of specific antibodies in serum and colostrums of pregnant swines (Law, 2000). In this work a new structural model containing three effective pathogenic factors with antigenic determinants of IpaC, Intimin and CfaB was designed. This chimeric gene was fused together by hydrophobic linkers and codon optimization for expression in *E. coli* was carried out. The chimeric protein structure and its ability to induce CD4+ and CD8+ immune responses against

these pathogens were predicted by *in silico* approaches.

Materials and Methods

Antigenic segment selection

cfab, ipac and eae sequences were obtained from GenBank. In order to identify the general and conserved antigenic fragments in these bacterial strain sequences, multiple sequence alignments were performed by Clustal W software (EBI, UK) at (http://www.ebi.ac.uk/Tools/clustalw2/ Thomopson et al., 1994).

Bioinformatic analysis of chimeric protein

For optimization of chimeric gene expression in *E. coli*, the Genscript Optimization Gene TM algorithm (www.genescript.com, piscataway. newjersy USA) was used. The secondary mRNA structure and mRNA stability was predicted by GeneBee (http://www.genebee.msu.su/services /rna2_reduced.html), mfold Web Server (Zuker, 2003) (http://mfold.rna.albany.edu/?q=mfold/RNA Folding-Form) and RNAfold web server.

Secondary and tertiary structure prediction

The secondary protein structure for each selected segment and the complete designed construct was predicted by PSIPRED (http://bioinf.cs.ucl.ac.uk /psipred/), SSpro8 (http://scratch.proteomics.ics.uci.edu/index.html) and GOR IV (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_gor4.html)(Garnie r et al., 1996).

Using ALPHAPRED (http://www.imtech.res.in /raghava/alphapred/index.html) (Wang et al., 2006) alpha turns residues were predicted in protein sequence. This method is based on the neural network training on PSI-BLAST generated position specific matrices and PSIPRED predicted secondary structure. Also different types of beta-turns such as Types I, II, IV, VIII were predicted by BetaTPred (http://www.imtech.res.in/raghava /betaturns/) (Kaur and Raghava, 2002).

For analyzing 2D structure stability, Ramachandran plot was drawn. For 3D protein structure predictions, several software based on homology modeling were used. They included SWISS-MODEL (http://swissmodel.expasy.org /workspace/)(Schwede et al., 2003), CPH models (http://www.cbs.dtu.dk/services/CPHmodels/) (Lund et al., 2002), phyre (http://www.sbg.bio.ic.ac.uk/~phyre/) (Kelley and Sternberg, 2009), FOLDpro and (PS)2 (http://ps2.life.nctu.edu.tw/) that combine PSI-BLAST, IMPALA, T-Coffee in both template selection and target-template alignment. The final three dimensional structures were built using the modeling package MODELLER (Eswar et al., 2007). Whole structures were then built by ab initio modeling using I-Tasser (http://zhanglab.ccmb.med.umich.edu/I-TASSER/) (Zhang, 2008). BHAGEERATH-H and Robetta software (http://robetta.bakerlab.org/index.html) \(Kim et al., 2004) combining homology and ab-intio modeling for Protein Tertiary Structure Prediction was used.

Prediction of immunogenic epitopes

B-cell epitopes: The linear B-cell sequences were obtained from ABCpred (Saha and Raghava, 2008) using Recurrent Neural Network, Bcepred (Saha and Raghava, 2004) base on using physico-chemical properties and BepiPred (http://www.cbs.dtu.dk/services/BepiPred/) (Ole and Morten, 2006) using a combination of a hidden Markov model and a propensity scale method. The residues were analyzed to predict discontinuous B cell epitopes using DiscoTope 1.2 Server (http://www.cbs.dtu.dk/services/DiscoTope/) (Haste Andersen et al., 2006).

MHC Binding Peptide

Prediction of MHC Class-I Binding Peptide was obtained from ProPred-I (http://www.imtech.res.in /raghava /propred1/) (Singh and Raghava, 2001) and nHLAPred (Zhang et al., 2009) and NetMHC (http://www.cbs.dtu.dk/services/NetMHC/) (Buus et al., 2003). NetMHC 3.2 server predicts binding of peptides to a number of different HLA alleles using weight matrices. Prediction of MHC Class-II Binding Peptide was performed by HLA-DR4Pred HLA-DR4Pred (Bhasin and Raghava, 2004b). The HLA-DR4Pred is a Support Vector Machine (SVM) and artificial neural networks (ANNs) based HLA-DRB1*0401(MHC class II alleles) binding peptides prediction method. But MHC molecules are highly polymorphic and MHC haplotype antigens of BALB/c mice are H-2Kd, H-2Dd, H-2Ld for MHC I and IAd, and I-Ed for MHC II. NetMHC 3.2 and ProPred-I was performed to predict H-2Kd, H-2Dd and H-2Ld.

The Rankpep (http://imed.med.ucm.es/Tools /rankpep.html) (Reche et al., 2002; Reche et al., 2004) is software for prediction of binding peptides to both Class I and Class II MHC molecules.

T-cell epitopes: CTLPred (http://www.imtech.res.in/raghava/ctlpred/) (Bhasin and Raghava, 2004a) is a direct method for prediction of CTL epitopes crucially related to MHC class I using Quantitative Matrix (QM), SVM and ANN in subunit vaccine design.

IgE epitopes: Prediction of allergens and mapping of IgE epitopes were obtained from AlgPred(http://www.imtech.res.in/raghava/algpred/) (Saha and Raghava, 2006). Algpred allows to predict allergen using SVMc + IgE epitope + ARPs (allergen-representative peptides) BLAST + MAST.

Protease and Hydropathicity prediction

Prediction of proteasome and immune-proteasome cleavage sites in CII obtained from Pcleavage (http://www.imtech.res.in/raghava/pcleavage/)(Bhasin and Raghava, 2005) using SVM and NetChop 3.1 Server (http://www.cbs.dtu.dk/services/NetChop/)(Saxov، et al., 2003) using ANN. Hydropathicity of CII was predicted by IEDB (http://tools.immuneepitope.org/tools/bcell/iedb_input)(Zhang et al., 2008) and ProtScale (http://web.expasy.org/protscale/) (Gasteiger et al., 2005).

Antigenic propensity and solvent accessibility

The probability of antigenicity of the construct was estimated by vaxijen (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html)(Doytchinova and Flower, 2007). Antigenicity Prediction in single and assembled forms of selected segments was performed at http://www.pbcpeptide.com/Feedback.htm by Kolaskar & Tongaonkar Antigenicity and ANTIGENpro (http://scratch.proteomics.ics.uci.edu/). Prediction of solvent accessibility in CII was performed by Recombinant Protein Solubility Prediction (http://biotech.ou.edu/#r)(Davis et al., 1999; Wilkinson and Harrison, 1991), NetSurfP ver. 1.1, SARpred(Garg et al., 2005) using multiple sequence alignment and secondary structure and IEDB (http://tools.immuneepitope.org/tools/bcell/iedb_input).

Prediction of subcellular localization: Subcellular localization of CII in gram-negative bacteria was predicted by CELLO (http://e093.life.nctu.edu.tw/index.html) and PSLpred (http://www.imtech.res.in/raghava/) (Bhasin et al., 2005). PSLpred is a hybrid approach-based method integrating PSI-BLAST and three SVM modules.

Results

Antigenic segment selection

Entire cfaB protein (Bouzari et al., 2010, Nazarian et al., 2012), 282 amino acids from C-terminal of intimin (Amani et al.)and 64 amino acids in C-terminal of ipaC (Terry et al., 2008a) were selected. EAAAK (Arai et al., 2001) linker was used between each segment(Figure1).

Figure1: Schematic model which shows the construction of whole of CfaB, Intimin 282 and Ipac 64, bound together by the linkers.

Bioinformatic analysis of chimeric protein

The CII mRNA structure (Figure 2) had a free energy of -361.3 Kcal/mol. The analysis of the sequence encoding the optimized chimeric gene and the wild type is shown in Figure3. As a result of reduction of overall GC content to 51.56%, the mRNA stability and ribosomal binding was optimized for transcription by changing the stem loop. The codon adaptation index (CAI) was increased. Restriction sites interfering with cloning, instability elements, and all the *cis*-acting sites were removed. The necessary restriction sites of *Eco*RI and *Hin*dIII were at the ends of the sequence for subsequent cloning.

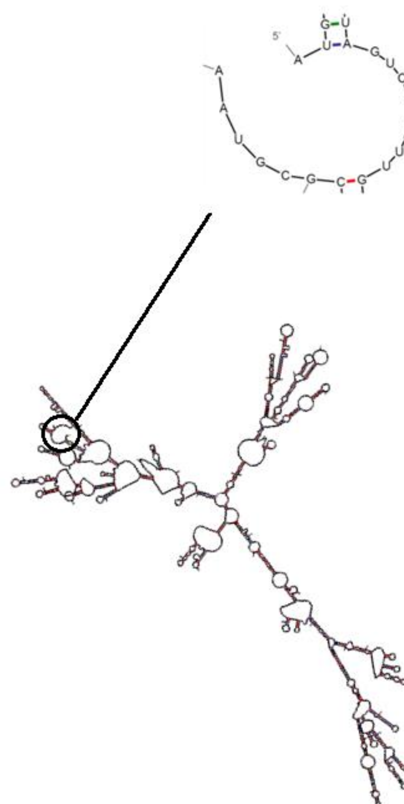

Figure 2: Analyzed of CII mRNA stability and first nucleotide position in this structure.

Secondary and tertiary structure prediction

The secondary structure, alpha and beta turn achieved by several online programs is given in Figure 4. No difference was seen comparing 2D structure of each selected segment and the complete designed construct. Secondary structure stability determined by Ramachandran plot is shown in

Figure 5. Tertiary structures predicted with several softwares are shown in Figure 6 and the linkers in this structure are highlighted in Figure7.

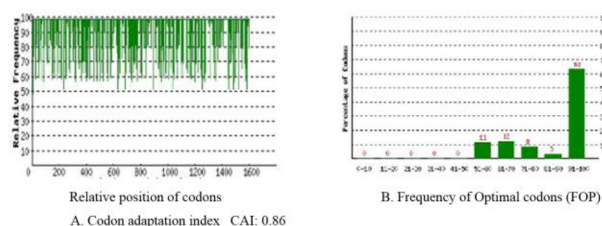

Figure 3: The sequence of the chimeric gene (CII) was optimized by changing some factors to increase gene expression. A) Codon adaptation index (CAI), B) Frequency of Optimal codons (FOP).

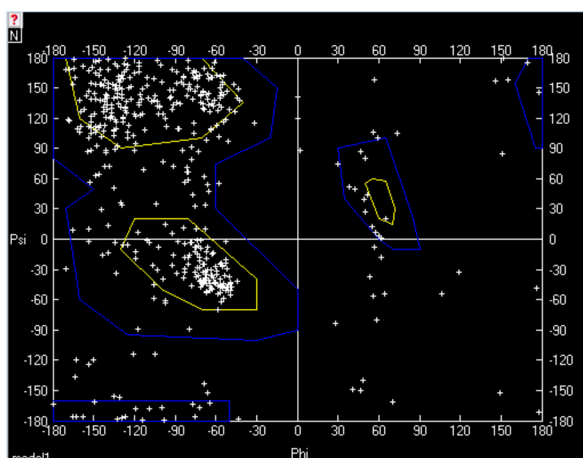

Figure 4: Evaluation of CII stability based on a Ramachandran plot.

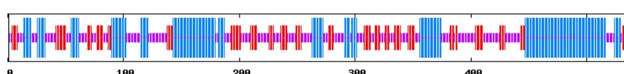

Figure 5: Analysis of CII protein secondary structure.

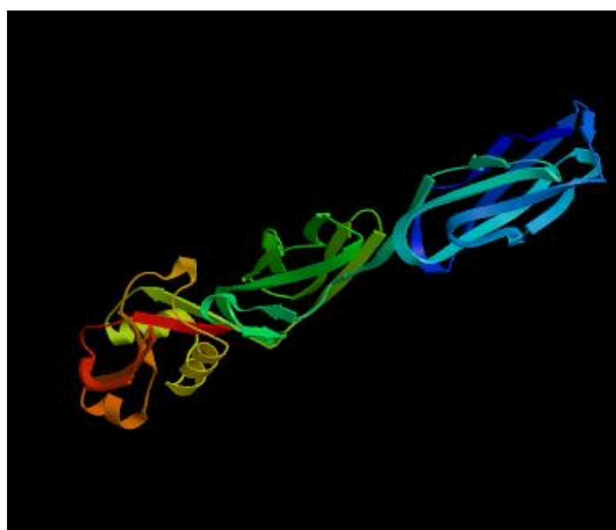

Figure 6: *Ab initio* modeling for prediction of the 3D structure of CII. The result was viewed by Rasmol software.

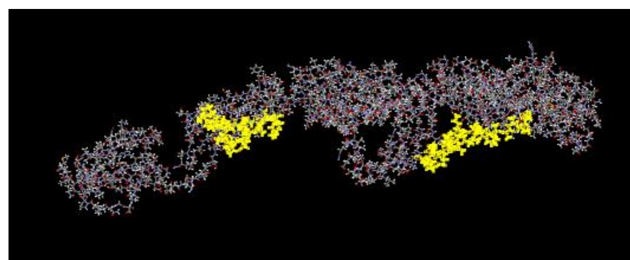

Figure 7: Analysis of the linkers' position in 3d structure of CII.

Prediction of immunogenic epitopes

B-cell epitopes: Bcepred utilized seven different physico-chemical scales including hydrophilicity, flexibility, mobility, accessibility, polarity, exposed surface, turns, and antigencity for prediction existing Continuous B-cell epitopes (Table 1). ABCpred is based on recurrent neural network, which ranked according to their score obtained and above the chosen threshold value (Table 2).

Table1: Epitope predicted CII chimeric protein by physico and chemical properties based on Bcepred

Prediction parameters	Epitope positions and segments
Hydrophilic	48-59, 110-117, 147-164, 180-186, 222-238, 242-250, 252-262, 282-288, 302-311, 321-329, 339-349, 412-424, 450-466, 482-505, 515-527.
Flexibility	32-41, 50-56, 107-116, 219-234, 237-246, 299-307, 327-334, 336-345, 409-423, 479-499, 513-525.
Accessibility	34-45, 48-59, 62-71, 92-98, 153-164, 167-175, 177-186, 189-197, 199-211, 221-236, 270-279, 296-302, 309-318, 339-351, 360-369, 394-403,411-421, 450-466, 472-509, 515-527.
Turns	49-56, 205-212.
Exposed surface	52-58, 271-279, 360-366, 412-418.
Polarity	39-45, 92-98, 153-164, 177-184, 269-279, 450-466, 472-482, 485-493.
Antigen propensity	56-63, 67-83, 114-121, 136-146, 194-200, 332-339, 372-380, 383-390, 433-439, 505-512.

Table 2: The predicted B cell epitopes by ABCpred obtained by trained recurrent neural network.

Rank	Sequence	Start position	Score
1	YNLITQNPLPGVNVNT	426	0.95
2	SYTIKAPSYMIKVDKQ	349	0.94
3	AWIKQTSSEQRSGVSS	409	0.93
4	DGTYSWYSENTSIATV	307	0.90
5	SKTFESYRVMTQVHTN	38	0.88
6	VKLADTPQLTDVLNST	61	0.87

Table 3: The result of prediction for MHCI antigens of BALB/c mice

MHC-Db	10-18, 24-32, 61-73, 110-118, 185-193, 199-213, 229-237, 264-280, 312-320, 350-358, 401-409, 420-428, 436-447, 471-480
MHC-Dd	18-30, 65-73, 85-94, 107-118, 137-145, 189-197, 203-212, 231-242, 273-312, 333-350, 353-361, 377-385, 420-428, 431-439, 466-474, 503-513
MHC-Kb	18-26, 39-47, 65-79, 99-107, 110-118, 134-142, 212-220, 232-241, 264-272, 293-301, 311-319, 362-373, 377-385, 396-407, 420-429, 440-448, 498-514
MHC-Kd	10-18, 24-41, 110-118, 135-146, 168-179, 186-194, 295-303, 311-320, 365-377, 397-408, 420-434, 472-480, 505-517
MHC-Kk	2-16, 52-61, 64-81, 97-105, 112-120, 148-166, 168-194, 231-239, 259-278, 284-292, 311-338, 351-359, 365-373, 380-388, 399-411, 415-429, 449-485, 499-517, 524-532.
MHC-Ld	10-21, 26-50, 66-74, 81-97, 104-112, 136-144, 205-221, 243-251, 291-301, 378-386, 390-398, 401-410, 414-422, 432-449, 481-489.

MHC Binding Peptide: Prediction for MHC I antigens of BALB/c mice is shown in Table 3. RANKPEP denoted binding peptides to Class II MHC molecules in mice (Table 4).

T-cell epitopes: T-cell epitope was directly predicted by CTLpred (Table 5).

IgE epitopes: Algpred predicted allergens based on similarity of known epitopes with any region of CII. The protein sequence does not contain experimentally proven IgE epitope.

Table 4: Prediction of MHC II antigens in mice by RANKPEP

I-Ek	407-415, 194-202, 187-195, 462-470, 457-465, 452,460, 160-168, 155-163, 150-158, 359-36
I-Ab	169-177, 125-133, 246-254, 365-373, 434-442, 104-112
I-Ad	483-491, 88-96, 403-411, 146-154, 466-474, 482-490, 246-254, 445-453, 208-216
I-Ed	87-95
I-Ag7	189-197, 391-399, 448-456, 458-466, 453-461, 156-164, 151-159, 131-139, 464-472

Table 5: Prediction of T-cell epitope by CTL pred

Peptide Rank	Start Position	Sequence	Score	Prediction
1	360	KVDKQAYYA	0.990	Epitope
2	439	VNTPNVYAV	0.990	Epitope
3	524	TASQIAGNI	0.990	Epitope

Table 6: Predicted peptide with high accessibility by Emini Surface Accessibility Prediction

No.	Start Position	End Position	Peptide	Peptide Length
1	35	46	SPASKTFESYRV	12
2	92	97	TTAKEF	6
3	168	173	KFDQTK	6
4	179	184	IKADKT	6
5	203	208	NGQPVN	6
6	223	235	GKSQTQATTGNDG	13
7	272	277	LKIDNK	6
8	308	316	GTYSWYSEN	9
9	341	348	TSGDKQTV	8
10	361	366	VDKQAY	6
11	395	403	ANKYSHYSS	9
12	412	420	KQTSSEQRS	9
13	484	492	ASSKQAEEA	9
14	494	503	QVSKEASQAT	10
15	517	524	INQSKNST	8

Antigenic propensity and solvent accessibility

For building a construct with higher antigencity, different states of chimer was checked. This model

had the highest score (0.7086) in vaxijen. CII has a 79.2 percent chance of insolubility when over expressed in *E. coli*. The average of surface accessibility was 1.00 and the maximum and minimum of CII surface accessibility were 3.763 and 0.081 respectively. Surface accessibility graph was predicted by Emini Surface Accessibility Prediction (Figure 8A). Table 6 shows the predicted peptides with high accessibility.

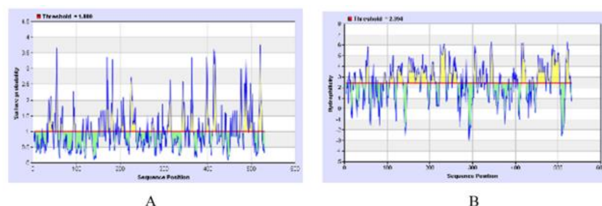

Figure 8: Analizing of CII surface accessibility(A) and hydrophibisity(B) by IEDB

Protease and Hydropathicity prediction

Pcleavage did not find the residues with proteasome cleavage site in antigenic sequence. Hydropathicity graph was drawn by Parker Hydrophilicity Prediction (Figure 8B). The minimum and maximum Hydropathicity ranked between -3.043 and 6.326 with an average rank of 2.394.

Discussion

An effective vaccine against ETEC in adult travelers is licensed in few countries, and the development of a new ETEC vaccine is sought (Svennerholm and Tobias, 2008). In dysentery caused by *Shigella*, treatment with appropriate antibiotics shortens the duration of the symptoms and reduces the duration of host excretion of the pathogen. Sulfonamides, ampicillin, and nalidixic acid were used to be the first-line therapies. Gradually, *Shigella* became resistant to each of these antibiotics. Ciprofloxacin is the first choice of antibiotics recommended by the World Health Organization for treating shigellosis and currently, increasing resistance to ciprofloxacin has been documented (Ahs et al., 2010). Antibiotic use in a case of EHEC O157:H7 may place the patient at greater risk of HUS (Dundas et al., 2001) and could increase the amount of toxin produced by strains of Shiga toxigenic group of *Escherichia coli* (STEC) bacteria (Kimmitt et al., 2000). The abilities of Shigella to enter epithelial cells require Ipa proteins which play prominent roles in the infection process (Picking et al., 2001). Intimin is encoded by the *E. coli* attaching and effacing (eae) gene required for intimate adhesion to epithelial cells and cytoskeletal reorganization (Tran Van Nhieu et al., 1999). CfaB is a major subunit in CFA/I fimbriae polymer of ETEC bacteria (Jaumouille et al., 2008) (Terry et al., 2008b). CfaB is an important target to prevent ETEC invasion. In this work we designed a construct containing epitopic segments of CfaB, Intimin and IpaC. A subunit vaccine containing these epitopes could be an effective way of simultaneous protection aganist shigella, EHEC and ETEC. Antigen index of different orientations of the epitopes in the chimeric protein was analyzed. The bacterial threshold of 0.4 renders this protein as a suitable antigen. The structure shown in Figure 1 was the most optimum choice. The codon usage of chimeric gene adapted to the codon usage of highly expressed genes in *E. coli.* GC content, CpG dinucleotides content, Cryptic splicing sites, Repeat sequences, Premature PolyA sites, Negative CpG islands, Internal chi sites and ribosomal binding sites, RNA instability motif and Restriction sites that may interfere with cloning were considered. In the mRNA structure, the position of start codon stands in the loop (Figure 2) that helps better ribosome binding and subsequent start of translation. Because of low ΔG (-361.3 Kcal/mol.), the mRNA is expected to be highly stable. This phenomenon is further supported by Ramachandran plot. When the secondary structure of chimeric protein was compared with 2D structure of single proteins, no significant changes were noted. Each domain in 3D protein structure maintained their conformations in fusion structure. The presence of helix-forming peptide linker caused reduced interference between the fragments. Finally localization prediction in gram negative bacteria shows CII as a cytoplasmic protein.

For immunological study, the humoral and cellular immunity was analyzed. CII had many B cell epitopes in both linear and three-dimensional structure. B cell response is critical, especially in production of IgA and IgG. This prediction of the chimeric construct had the potential to induce CD4+ and CD8+immune responses against these pathogens in the BALB/c model.

The CII peptide fragments are carried to the surface of the presenting cell on MHC proteins, which present the fragments to helper T lymphocytes. This stimulates B cells to make antibodies, macrophages to destroy any intracellular pathogen multiplying and cytotoxic T cells to kill infected target cells. The prediction of allergenic proteins is becoming very important due to use of modified proteins in therapeutics. This prediction study also revealed that the chimeric construct was not allergen. In conclusion, as this chimeric structure has a potential to induce humoral and cellular immunity response, CII could be a candidate

subunit vaccine against EHEC, ETEC and shigella.

Acknowledgements

The authors wish to thank Shahed University for their financial assistance to conduct the present study.

References

1- Ahs J. W., Tao W., Lfgren J. and Forsberg B. C. (2010) Diarrheal Diseases in Low-and Middle-Income Countries: Incidence, Prevention and Management. Open Infectious Diseases Journal 4:113-124.

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

3- Amani J., Salmanian A. H., Rafati S. and Mousavi S. L. Immunogenic properties of chimeric protein from espA, eae and tir genes of Escherichia coli O157: H7. Vaccine 28:6923-6929.

4- Amani J., Salmanian A. H., Rafati S. and Mousavi S. L. (2010) Immunogenic properties of chimeric protein from espA, eae and tir genes of Escherichia coli O157:H7. Vaccine 28:6923-6929.

5- Arai R., Ueda H., Kitayama A., Kamiya N. and Nagamune T. (2001) Design of the linkers which effectively separate domains of a bifunctional fusion protein. Protein engineering 14:529.

6- Barzu S., Benjelloun-Touimi Z., Phalipon A., Sansonetti P. and Parsot C. (1997) Functional analysis of the Shigella flexneri IpaC invasin by insertional mutagenesis. Infect Immun 65:1599-1605.

7- Bhasin M., Garg A. and Raghava G. (2005) PSLpred: prediction of subcellular localization of bacterial proteins. Bioinformatics 21:2522-2524.

8- Bhasin M. and Raghava G. (2004a) Prediction of CTL epitopes using QM, SVM and ANN techniques. Vaccine 22:3195-3204.

9- Bhasin M. and Raghava G. (2004b) SVM based method for predicting HLA-DRB1* 0401 binding peptides in an antigen sequence. Bioinformatics 20:421-423.

10- Bhasin M. and Raghava G. (2005) Pcleavage: an SVM based method for prediction of constitutive proteasome and immunoproteasome cleavage sites in antigenic sequences. Nucleic acids research 33:W202-W207.

11- Bhatnagar S., Bhan M. K., Sommerfelt H., Sazawal S., Kumar R. and Saini S. (1993) Enteroaggregative Escherichia coli may be a new pathogen causing acute and persistent diarrhea. Scand J Infect Dis 25:579-583.

12- Bouzari S., Shahrokhi N., Dashti A., Janani A. R. and Jafari A. (2010) Construction and evaluation of chimeric heat-labile toxin B subunit and N-terminal (1–75) fragment of colonization factor antigen I gene of enterotoxigenic Escherichia coli. Annals of microbiology 60:255-261.

13- Buus S., Lauemí¸ller S., Worning P., Kesmir C., Frimurer T., Corbet S., Fomsgaard A., Hilden J., Holm A. and Brunak S. (2003) Sensitive quantitative predictions of peptide-MHC binding by a 'Query by Committee' artificial neural network approach. Tissue antigens 62:378-384.

14- Davis G. D., Elisee C., Newham D. M. and Harrison R. G. (1999) New fusion protein systems designed to give soluble expression in Escherichia coli. Biotechnol Bioeng 65:382-388.

15- Dean-Nystrom E. A., Bosworth B. T., Moon H. W. and O'Brien A. D. (1998) Escherichia coli O157:H7 requires intimin for enteropathogenicity in calves. Infect Immun 66:4560-4563.

16- DeVinney R., Stein M., Reinscheid D., Abe A., Ruschkowski S. and Finlay B. B. (1999) Enterohemorrhagic Escherichia coli O157:H7 produces Tir, which is translocated to the host cell membrane but is not tyrosine phosphorylated. Infect Immun 67:2389-2398.

17- Doytchinova I. and Flower D. (2007) VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. Bmc Bioinformatics 8:4.

18- Dundas S., Todd W. T., Stewart A. I., Murdoch P. S., Chaudhuri A. K. and Hutchinson S. J. (2001) The central Scotland Escherichia coli O157:H7 outbreak: risk factors for the hemolytic uremic syndrome and death among hospitalized patients. Clin Infect Dis 33:923-931.

19- Eswar N., Webb B., Marti-Renom M. A., Madhusudhan M., Eramian D., Shen M. Y., Pieper U. and Sali A. (2007) Comparative protein structure modeling using Modeller. Curr Protoc Protein Sci 50:1-2.9.

20- Garg A., Kaur H. and Raghava G. (2005) Real value prediction of solvent accessibility in proteins using multiple sequence alignment and secondary structure. Proteins: Structure, Function, and Bioinformatics 61:318-324.

21- Garnier J., Gibrat J., Robson B. and Doolittle R. (1996) GOR secondary structure prediction method version IV. Methods Enzymol 266:540-553.

22- Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M. R., Appel R. D. and Bairoch A. (2005) Protein identification and analysis tools on the ExPASy server. The

proteomics protocols handbook:571-607.

23- Haste Andersen P., Nielsen M. and Lund O. (2006) Prediction of residues in discontinuous B-cell epitopes using protein 3D structures. Protein Science 15:2558-2567.

24- Jansson L., Tobias J., Lebens M., Svennerholm A. M. and Teneberg S. (2006) The major subunit, CfaB, of colonization factor antigen i from enterotoxigenic Escherichia coli is a glycosphingolipid binding protein. Infect Immun 74:3488-3497.

25- Jaumouille V., Francetic O., Sansonetti P. J. and Tran Van Nhieu G. (2008) Cytoplasmic targeting of IpaC to the bacterial pole directs polar type III secretion in Shigella. EMBO J 27:447-457.

26- Jennison A. V. and Verma N. K. (2004) Shigella flexneri infection: pathogenesis and vaccine development. FEMS Microbiol Rev 28:43-58.

27- Katouli M., Jaafari A., Farhoudi-Moghaddam A. A. and Ketabi G. R. (1990) Aetiological studies of diarrhoeal diseases in infants and young children in Iran. J Trop Med Hyg 93:22-27.

28- Kaur H. and Raghava G. (2002) BetaTPred: prediction of Î²-TURNS in a protein using statistical algorithms. Bioinformatics 18:498.

29- Kelley L. A. and Sternberg M. J. E. (2009) Protein structure prediction on the Web: a case study using the Phyre server. Nature protocols 4:363-371.

30- Kim D. E., Chivian D. and Baker D. (2004) Protein structure prediction and analysis using the Robetta server. Nucleic acids research 32:W526-W531.

31- Kimmitt P. T., Harwood C. R. and Barer M. R. (2000) Toxin gene expression by shiga toxin-producing Escherichia coli: the role of antibiotics and the bacterial SOS response. Emerg Infect Dis 6:458-465.

32- Kueltzo L. A., Osiecki J., Barker J., Picking W. L., Ersoy B., Picking W. D. and Middaugh C. R. (2003) Structure-function analysis of invasion plasmid antigen C (IpaC) from Shigella flexneri. J Biol Chem 278:2792-2798.

33- Law D. (2000) Virulence factors of Escherichia coli O157 and other Shiga toxin-producing E. coli. J Appl Microbiol 88:729-745.

34- Li Y. F., Poole S., Nishio K., Jang K., Rasulova F., McVeigh A., Savarino S. J., Xia D. and Bullitt E. (2009) Structure of CFA/I fimbriae from enterotoxigenic Escherichia coli. Proc Natl Acad Sci U S A 106:10793-10798.

35- Luiz W. B., Cavalcante R. C., Paccez J. D., Souza R. D., Sbrogio-Almeida M. E., Ferreira R. C. and Ferreira L. C. (2008) Boosting systemic and secreted antibody responses in mice orally immunized with recombinant Bacillus subtilis strains following parenteral priming with a DNA vaccine encoding the enterotoxigenic Escherichia coli (ETEC) CFA/I fimbriae B subunit. Vaccine 26:3998-4005.

36- Lund O., Nielsen M., Lundegaard C. and Worning P. 2002. CPHmodels 2.0: X3M a computer program to extract 3D models. Vol. 102. 2002.

37- Nazarian S., Mousavi S. L., Rasooli I., Hasannia S. and Pirooznia N. (2013) A PLGA-encapsulated chimeric protein protects against adherence and toxicity of enterotoxigenic Escherichia coli. Microbiol Res.

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

39- Ole L. and Morten N. (2006) Improved method for predicting linear B-cell epitopes. Immunome Research 2.

40- Picking W. L., Coye L., Osiecki J. C., Barnoski Serfis A., Schaper E. and Picking W. D. (2001) Identification of functional regions within invasion plasmid antigen C (IpaC) of Shigella flexneri. Mol Microbiol 39:100-111.

41- Reche P. A., Glutting J. P. and Reinherz E. L. (2002) Prediction of MHC class I binding peptides using profile motifs. Human immunology 63:701-709.

42- Reche P. A., Glutting J. P., Zhang H. and Reinherz E. L. (2004) Enhancement to the RANKPEP resource for the prediction of peptide binding to MHC molecules using profiles. Immunogenetics 56:405-419.

43- Saha S. and Raghava G. (2004) BcePred: Prediction of continuous B-cell epitopes in antigenic sequences using physico-chemical properties. Artificial Immune Systems:197-204.

44- Saha S. and Raghava G. (2006) AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. Nucleic acids research 34:W202-W209.

45- Saha S. and Raghava G. (2008) ABCPred benchmarking datasets. 2006a.

46- Saxova'P., Buus S., Brunak S. and Keلmir C. (2003) Predicting proteasomal cleavage sites: a comparison of available methods. International immunology 15:781.

47- Schwede T., Kopp J., Guex N. and Peitsch M. C. (2003) SWISS-MODEL: an automated protein homology-modeling server. Nucleic acids research 31:3381-3385.

48- Shaikh N., Terajima J. and Watanabe H. (2003) IpaC of Shigella binds to the C-terminal domain of beta-catenin. Microb Pathog 35:107-117.

49- Singh H. and Raghava G. (2001) ProPred: prediction of HLA-DR binding sites. Bioinformatics 17:1236-1237.

50- Svennerholm A. M. and Tobias J. (2008) Vaccines against enterotoxigenic Escherichia coli. Expert Rev Vaccines 7:795-804.

51- Terry C. M., Picking W. L., Birket S. E., Flentie K., Hoffman B. M., Barker J. R. and Picking W. D. (2008a) The C-terminus of IpaC is required for effector activities related to Shigella invasion of host cells. Microbial pathogenesis 45:282-289.

52- Terry C. M., Picking W. L., Birket S. E., Flentie K., Hoffman B. M., Barker J. R. and Picking W. D. (2008b) The C-terminus of IpaC is required for effector activities related to Shigella invasion of host cells. Microb Pathog 45:282-289.

53- Thomopson J., Higgins D. G. and Gibson T. (1994) ClustalW. Nucleic Acids Res 22:4673-4680.

54- Tran Van Nhieu G., Caron E., Hall A. and Sansonetti P. J. (1999) IpaC induces actin polymerization and filopodia formation during Shigella entry into epithelial cells. EMBO J 18:3249-3262.

55- Vlisidou I., Dziva F., La Ragione R. M., Best A., Garmendia J., Hawes P., Monaghan P., Cawthraw S. A., Frankel G., Woodward M. J. and Stevens M. P. (2006) Role of intimin-tir interactions and the tir-cytoskeleton coupling protein in the colonization of calves and lambs by Escherichia coli O157:H7. Infect Immun 74:758-764.

56- Wang Y., Xue Z. and Xu J. (2006) Better prediction of the location of α-turns in proteins with support vector machine. Proteins: Structure, Function, and Bioinformatics 65:49-54.

57- Wilkinson D. L. and Harrison R. G. (1991) Predicting the solubility of recombinant proteins in Escherichia coli. Biotechnology (N Y) 9:443-448.

58- Zhang P., Brusic V. and Basford K. (2009) A hybrid model for prediction of peptide binding to MHC molecules. Advances in Neuro-Information Processing 529-536.

59- Zhang Q., Wang P., Kim Y., Haste-Andersen P., Beaver J., Bourne P. E., Bui H. H., Buus S., Frankild S. and Greenbaum J. (2008) Immune epitope database analysis resource (IEDB-AR). Nucleic acids research 36:W513-W518.

60- Zhang Y. (2008) I-TASSER server for protein 3D structure prediction. Bmc Bioinformatics 9:40.

61- Zuker M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucleic acids research 31:3406-3415.