BIOSCIENCES BIOTECHNOLOGY RESEARCH ASIA, September 2015. Vol. **12**(Spl. Edn. 2), p. 725-730

Designing, Optimizing, and Structure Prediction of Chimeric Protein CTB-IpaD for Expression in *E.coli*

Seyed Mohammad Gheibi Hayat[1], Azam Sadeghinia[2]*, Babak Elyasifar[3], Ghasem Abbaszadeh-Goudarzi[4], Zahra Adeli[5] and Somaieh Salimi[6]

[1]Young Researchers and Elite Club, Andimeshk Branch, Islamic Azad University, Andimeshk, Iran.
[2]Department of Cardiology, Faculty of Medicine, Dezful University of Medical Sciences, Dezful, Iran.
[3]Department of Drug Biotechnology, Faculty of Pharmacy,
Tabriz University of Medical Sciences, Tabriz, Iran.
[4]Department of Medical Biotechnology, Faculty of Advanced Technologies in Medicine, Tehran University of medical Sciences, Tehran, Iran.
[5]Department of Microbiology, Islamic Azad University, Damghan Branch, Damghan, Iran
[6]Department of Marine Biology, Faculty of Marine Science,
Chabahar Marine University, Chabahar, Iran.

doi: http://dx.doi.org/10.13005/bbra/2253

(Received: 01 April 2015; accepted: 23 June 2015)

The most common cause of diarrhea is Shigella and no vaccine has been found to provide protection against it. IpaD play an important role in invasion, infection, and pathogenesis caused by Shigella. This protein has been one of the most important protein for shigella vaccine candidate. This study aims to link the gene sequence IpaD and CTB with In-silico analysis. Nucleotide sequences were obtained from NCBI database and optimized. Sequences were fused together by a flexible linker in order to find the best epitope exposing chimeric antigen. After Protein half-life and instability index were determined, and then the prediction of the secondary structure and the three-dimensional structure was analyzed and also immunodominant linear B-cell epitopes were also settled at the final step. Codons were changed with respect to codon bias of *E.coli* and the GC content was changed to an optimal level. Inappropriate structures of RNA were removed. However, these changes led to a prolonged half-life and increased mRNA expression of the recombinant protein. Sequences of the chimeric gene was optimized based on codon usage pattern, In-silico analysis represented that sequence can have a high expression in *E.coli* and designed to enhance the expression of proteins which contain rare codons at high frequency. (EAAAK) 4 hydrophobic linkers could prevent the domain interactions of the chimeric protein. Chimeric protein can be a good candidate in immunogenicity study to produce an effective vaccine.

Key words: Shigella dysenteriae, IpaD protein, B subunit cholera toxin, Bioinformatics.

Shigella and *Escherichia are members* of the family *Enterobacteriaceae* and are genetically similar[1]. The name of Shigella that was first isolated in 1894, derived from the Japanese bacteriologist Kiyoshi shiga[2]. This bacterium is an intracellular pathogen and as few as ten bacteria of the *Shigella* in water or food can cause the symptom to occur[3].

The natural host of this organism is the human and mode of transmission of infection is fecal-oral route through infected food and water. Diarrhea with abdominal pain, generalized muscle aches and pains are the most common early symptoms. Shigellosis appear 24-48 hours after

* To whom all correspondence should be addressed.
Mob.:+989168456907;
E-mail: gheibi65@yahoo.com

eating contaminated food, and the enterotoxin by the effect on intestinal epithelial cells causes loss of water and electrolytes[4].

Each year, more than 164 million infections and 1.1 million deaths due to infection have been world-widely reported and 69% of the victims are children under 5 years of age[5]. According to several studies, prevalence rate have been reported different. According to studies in Tehran, the prevalence rate of shigellosis were 1% and 2% and this percentage was 5.2% in another study over 3 years in Qazvin[6-8].

The mechanism of shigella entry into epithelial cells is dependent on a number of bacterial surface proteins. Important proteins are encoded by invasive plasmid. These proteins include IpaA, B, C, D and H protein. IpaB, C and D proteins proposed as *Shigella* vaccine candidate[9]. IpaD protein play an important role during invasion of the bacterium, and causes IpaB and IpaC proteins to come more quickly to the surface of the of the cell for pathogenesis and invasion[10-11].

IpaD protein is one of the most important virulence factors of Shigella species. IpaD With molecular weight of 37 kDa, by using of IpaB preceded the onset of the invasion which existed on the surface of bacteria. IpaD has dumbbell-shaped structure, is connected to the bacterial type III secretion system by the C-terminal region and IpaD directly interact with the environment by the N-terminal region[12-13].

Many studies have shown that using a specific antibody against IpaD can prevent the transfer of bacteria into the host cell. Because of this reason, IpaD has become a major antigen in the vaccine[14-16].

Today, CTB is known as a strong immuno adjuvant which is related to mucosal immunity[17-18], because of being an affective translocator for the systemic antibody secretion and mucosal antibody secretion for the conjugated antibody, it has been chemically and genetically altered. Dendritic cells activated by CTB has the ability to increase primary histocompatibility complex and induce the secretion of secondary molecules such as CD80 and CD86 on the surface of dendritic cells at the same time and also causes the secretion of proinflammatory cytokines[19-21]. For this reason, the chimeric protein consisted of CTB and IpaD, first CTB as an adjuvant causes to increases the immunogenicity against of IpaD, and second it can create immunity against both infectivity agents. The aim of this study is to construct a gene that consist of IpaD and CTB and subsequent bioinformatics feature.

MATERIALS AND METHODS

Sequences, databases and construct design

IpaD and *CTB* with accession number of YP_406165 and AAC34728 were obtained from National Centre for Biotechnology Information (http://www.ncbi.nlm.nih.gov), and saved as FASTA format. Signal peptides and sequences removed that are not an urgent necessity for immunological response.

It is required to maintain the protein structure and the prevention of domain interactions. The sequences fused together by (**EAAAK**) **4** hydrophobic linkers in order to find the best epitope, exposing chimeric antigen. In order to express in E. coli, codon usage optimization was done by using bioinformatics software, Optimum Gene TM Algorithm.

The nucleotide sequences were translated by use of the site (http://web.expasy.org/translate), and the antigenicity of the proteins is identified by the VaxiJen server (http://www.darrenflower.info/VaxiJen), this server used an independent alignment method to predict physical and chemical properties of proteins. The combination of these genes constructed based on highest immunogenicity scores.

The physico-chemical parameters of proteins

For the analysis of the physico-chemical parameters of proteins, sequences were analyzed via ProtParam (http://web.expasy.org/protparam/) from ExPASy, (Swiss Institute of Bioinformatics), in order to gain information about protein properties such as molecular weight, amino acid composition, the number of positively and negatively charged amino acids, the biochemical characteristics of protein, half-life of a protein, and finally protein instability index(PII).

Prediction of RNA secondary structure

The secondary structure of the mRNA was evaluated by RNAalifold program after codon optimization. The program used multiple sequence alignment as input information for the analysis of patterns keep changing, and a scoring matrix is

made for the minimum free energy and synchronous information changes. The dynamic programming is used to select a structure has a minimal energy for a complete set of RNA sequences alignment.

Prediction of protein secondary structure

Protein secondary structure prediction was performed by SOPMA. The server uses a self-optimization method (SOPM) in order to improve the prediction.

Prediction of protein tertiary structure

Protein tertiary structure prediction was performed using I-TASSER server online (http://zhanglab.ccmb.med.umich.edu/I-TASSER/) and the work was done on the basis of using combination of threading and Ab-initio technique.

Prediction of antigenic B-cell epitopes

The server ABCpred (http://www.imtech.res.in/raghava/abcpred) was used to identify and predict B cell epitopes of chimeric protein sequences. In this server, prediction was performed using recurrent neural networks (RNN).

RESULTS

Designing and optimizing

The optimization was performed in two steps; involving optimization of codon bias and increasing the GC content, these result in enhancing the CAI from 22 to 100%, that showing a significant increase in codon usage as a function of the level of gene expression and It is obvious that an improvement in codon usage frequency will improve protein expression levels in *E. coli* (Fig. 1)

The results of the Blast-X sequence analysis confirmed that gene sequence optimization result no changes in chimeric protein sequence.

Physicochemical properties

Biochemical characterization of the protein was analyzed by using the program ProtParam. The calculated molecular weight and isoelectric point of the target protein were respectively 25,647 Daltons and 6.69 with molecular formula C1126H1818N312O356S7. The instability index calculated in the In-vitro, and proteins with instability index less than 40 are stable proteins. The instability index of the protein was computed as 36.71 and stable, due to the presence of a conserved amino acid (threonine), in vivo half-life of protein in *E. coli* was measured more than 10 hours. The aliphatic index defined as the relative volume of a protein occupied by aliphatic side chains and is a measure of the thermo-stability of the protein and aliphatic index for the desired protein was 85.41, and also antigenicity prediction by Vaxijen to determine antigenicity of the protein and the value was 0.5011.

Table 1. Results of sequence optimization

Type	Codon adaptation index	Effective number of codon	%GC	%AT
Query	0.22	43	35	65
Optimized	1	21	47	53

Table 2. Linear B cell epitopes prediction of chimeric protein by ABCpred server

Start position	Epitope	Score
33	GKREMAIITFKNGATF	0.92
7	DLCAEYHNTQIHTLND	0.91
194	YQMISHRELWDKIAKS	0.91
64	AIERMKDTLRIAYLTE	0.9
27	YTESLAGKREMAIITF	0.83
21	NDKIFSYTESLAGKRE	0.82
110	AAKEAAAKEAAAKRTT	0.82
116	AKEAAAKRTTNQALKK	0.81
207	AKSINNINEQYLKVYE	0.8

Prediction of RNA secondary structure

Prediction of RNA secondary structure of 5' end in mRNA was revealed after optimization in chimeric gene (fig. 2). The minimum energy for this structure was calculated to be -421 kcal/mol.

Secondary structure prediction of chimeric protein

Chimeric protein secondary structure was predicted by using SOPMA program showed that the chimeric protein has 67.5% a-helix, 9.5% extended strand, 4.3% β-turn, and 18.6% random coil.

3D structure prediction of chimeric protein

Three-dimensional (3D) structure of

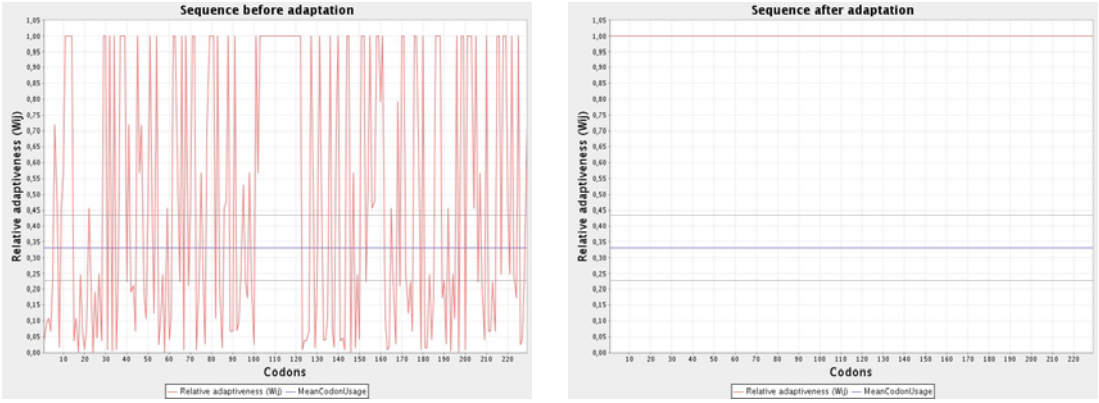

Fig. 1. Codon Adaptation Index (CAI) for chimeric gene (*CTB-IpaD*), before and after optimization

Fig. 2. Prediction of 5' end mRNA, after optimization of chimeric gene

chimeric proteins were predicted by the I-TASSER web tool. The results of this analysis are shown in Fig. 4.

Linear B cell epitope prediction

The presence of B cell epitopes are often the determining factor for the immunogenicity of a protein. Linear B cells epitopes in the chimeric protein were predicted by ABCpred server and the result was depicted in Table 2.

```
         10        20        30        40        50        60        70
          |         |         |         |         |         |         |
TPQNITDLCAEYHNTQIHTLNDKIFSYTESLAGKREMAIITFKNGATFQVEVPGSQHIDSQKKAIERMKD
ccccchhhhhhhhcccchhhhhhhhhhhhhhhhhhtchheeeeeettcceeeeecccccchhhhhhhhhhhhh
TLRIAYLTEAKVEKLCVWNNKTPHAIAAISMAEAAAKEAAAKEAAAKEAAAKRTTNQALKKDLSQKTLTK
hhhhhhhhhhhhhheeeeccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcch
TSLEEIALHSSQISMDVNKSAQLLDILSKKEYPINKDARELLHSAPEEAELDGYQMISHRELWDKIAKSI
hhhhhhhhccccceeecccchhhhhhhhhttccccchhhhhhhhccttthhhhttceeecchhhhhhhhhhhh
NNINEQYLKVYEHAVSSYT
hhhhhhhhhhhhhhhhhhh
```

Fig. 3. Secondary structure prediction of chimeric protein (h: α helix, e: Extended strand t: β turn and c: Random coil)

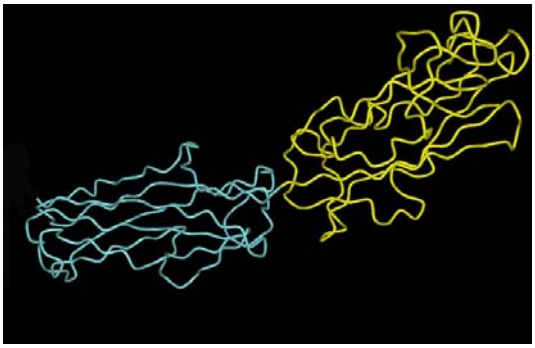

Fig. 4. Prediction of 3D structure of chimeric protein (helical model)

DISCUSSION

Despite the great efforts made in the past, an effective vaccine is not available for Shigella disease. Children in progress communities being the highest risk group for the disease. It previously reported that Shigella proteins such as IpaB and IpaD are extremely immunogenic and effective against Shigellosis[22].

Taking this into consideration, in order to identify IpaD epitopes, dominant available IpaD epitopes depend on the the function of this area.

As if its activity becomes inactive, the ability of Shigella invasion will be suppressed. The results show that IpaD and specially the N-terminal of the protein are the important factors for the entry of Shigella into host cells. Altogether, the IpaD N-terminal region is vaccine candidate against Shigellosis.

In many studies, the immuno-adjuvant properties of CTB from bacterium Vibrio cholerae have been found to boost the immune response[19-20]. Recently to design vaccine candidate, chimeric proteins has been the focus of many research. A chimeric protein comprises subunits, a flexible linker, and sequences with adjuvant properties can improve immunogenic properties of recombinant proteins. These two subunits of the protein connected by appropriate linker .The hydrophobic residues of the linkers (EAAAK) 4 mediate electrostatic interactions between the glutamic acid and lysine and thereby forming the salt bridge with respect to each other, the chemical nature of neighboring residues and the positioning of the charged groups of the residues form a stable helix structure and prevents domain interaction of chimeric protein[23]. In the study by Chen et al, the number of linker repeat are between 1 and 5 and it has been shown by analysis that repeat 4 or 5 are particularly critical, and can prevents domain interaction of protein[24].

Since IpaD and CTB gene sequences are rich in rare codons, and the bases of adenine and thymine, the expression of genes are not likely to be in BL21 (DE3), and Rosetta (DE3) that is a strain derivative of BL21 and designed to enhance the expression of proteins which contain rare codons at high frequency. For this purpose, sequences of the chimeric gene was optimized based on codon usage pattern for high expression in E. coli. In order to investigate the stability of the RNA structure, pseudo loop structure and ΔG were evaluated for each RNA structure. Also the prediction of secondary structures and three-dimensional structure were evaluated. Finally, for the prediction of linear epitopes were measured using the protein sequence as input.

CONCLUSION

With the Bioinformatics tools, a good construct was designed that expected to be largely expressed in E. coli host. The protein derive from this construct could be a candidate vaccine against Shigelosis.

ACKNOWLEDGMENTS

We are grateful to Dr. Monireh Gharibeh niazi for critical editing and evaluation of the manuscript. This study was financially supported by Young Researchers and Elite Club. The author have no conflict of interest to declare.

REFERENCES

1. Pupo GM, Lan R, Reeves PR. Multiple independent origins of Shigella clones of Escherichia coli and convergent evolution of many of their characteristics. *Proceedings of the National Academy of Sciences*. 2000; **97**(19): 10567-10572.
2. Trofa AF, Ueno-Olsen H, Oiwa R, Yoshikawa M. Dr. Kiyoshi Shiga: discoverer of the dysentery bacillus. *Clinical infectious diseases*. 1999; **29**(5): 1303-1306.
3. Niyogi SK. Shigellosis. *Journal of microbiology* (Seoul, Korea). 2005; **43**(2):133-143.
4. Sansonetti PJ. III. Shigellosis: from symptoms to molecular pathogenesis. *American Journal of Physiology-Gastrointestinal and Liver Physiology*. 2001; **280**(3): G319-G323.
5. Kotloff K, Winickoff J, Ivanoff B, Clemens JD, Swerdlow D, Sansonetti P, *et al*. Global burden of Shigella infections: implications for vaccine development and implementation of control strategies. *Bulletin of the World Health Organization*. 1999; **77**(8): 651-666.
6. Pourakbari B, Mamishi S, Mashoori N, Mahboobi N, Ashtiani MH, Afsharpaiman S, et al. Frequency and antimicrobial susceptibility of Shigella species isolated in Children Medical Center Hospital, Tehran, Iran, 2001-2006. *Brazilian Journal of Infectious Diseases*. 2010; **14**(2):153-157.
7. Mahyar A, Mahyar S. 896 Antimicrobial Resistance Patterns of Shigella Isolates in *Iranian Children. Pediatric Research.* 2010;**68**:449-458.
8. Jafari F, Hamidian M, Rezadehbashi M, Doyle M, Salmanzadeh-ahrabi S, Derakhshan F, et al. 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**(3):e56.
9. Barry EM, Pasetti MF, Sztein MB, Fasano A, Kotloff KL, Levine MM. Progress and pitfalls

in Shigella vaccine research. *Nature Reviews Gastroenterology and Hepatology.* 2013; **10**(4): 245-255.

10. Wilson BA, Salyers AA, Whitt DD, Winkler ME. Bacterial pathogenesis: a molecular approach: *American Society for Microbiology* (ASM); 2011.
11. Roehrich AD, Guillossou E, Blocker AJ, Martinez Argudo I. Shigella IpaD has a dual role: signal transduction from the type III secretion system needle tip and intracellular secretion regulation. *Molecular microbiology.* 2013; **87**(3): 690-706.
12. Espina M, Olive AJ, Kenjale R, Moore DS, Ausar SF, Kaminski RW, et al. IpaD localizes to the tip of the type III secretion system needle of Shigella flexneri. *Infection and immunity*. 2006; **74**(8): 4391-4400.
13. De Geyter C, Wattiez R, Sansonetti P, Falmagne P, Ruysschaert JM, Parsot C, et al. Characterization of the interaction of IpaB and IpaD, proteins required for entry of Shigella flexneri into epithelial cells, with a lipid membrane. *European Journal of Biochemistry.* 2000; **267**(18): 5769-5776.
14. Robbins JB, Schneerson R, Kubler-Kielb J, Keith JM, Trollfors B, Vinogradov E, et al. Toward a new vaccine for pertussis. *Proceedings of the National Academy of Sciences*. 2014; **111**(9): 3213-3216.
15. Camacho AI, Irache JM, Gamazo C. Recent progress towards development of a Shigella vaccine. 2013.
16. Allaoui A, Sansonetti P, Sani M, Botteaux A, Parsot C. Shigella Ipad protein and its use as a vaccine against Shigella infection. *Google Patents;* 2007.
17. Odumosu O, Nicholas D, Yano H, Langridge W. AB toxins: a paradigm switch from deadly to desirable. *Toxins*. 2010; **2**(7):1612-1645.
18. Arêas APM, Oliveira MLS, Miyaji EN, Leite LCC, Araújo Aires K, Dias WO, et al. Expression and characterization of cholera toxin B pneumococcal surface adhesin A fusion protein in Escherichia coli: ability of CTB-PsaA to induce humoral immune response in mice. *Biochemical and biophysical research communications*. 2004; **321**(1): 192-196.
19. Sharma MK, Singh NK, Jani D, Sisodia R, Thungapathra M, Gautam J, et al. Expression of toxin co-regulated pilus subunit A (TCPA) of Vibrio cholerae and its immunogenic epitopes fused to cholera toxin B subunit in transgenic tomato (Solanum lycopersicum). *Plant cell reports.* 2008; **27**(2): 307-318.
20. Ruhlman T, Ahangari R, Devine A, Samsam M, Daniell H. Expression of cholera toxin B–proinsulin fusion protein in lettuce and tobacco chloroplasts–oral administration protects against development of insulitis in non obese diabetic mice. *Plant biotechnology journal*. 2007;**5**(4):495-510.
21. Rosales Mendoza S, Alpuche Solís ÁG, Soria Guerra RE, Moreno Fierros L, Martínez González L, Herrera Díaz A, et al. Expression of an Escherichia coli antigenic fusion protein comprising the heat labile toxin B subunit and the heat stable toxin, and its assembly as a functional oligomer in transplastomic tobacco plants. *The Plant Journal.* 2009;**57**(1):45-54.
22. Martinez-Becerra FJ, Kissmann JM, Diaz-McNair J, Choudhari SP, Quick AM, Mellado-Sanchez G, et al. Broadly Protective Shigella Vaccine Based on Type III Secretion Apparatus Proteins. *Infection and Immunity*. 2012; **80**(3): 1222-1231.
23. Arai R, Ueda H, Kitayama A, Kamiya N, Nagamune T. Design of the linkers which effectively separate domains of a bifunctional fusion protein. *Protein engineering*. 2001; **14**(8): 529-532.
24. Chen X, Bai Y, Zaro JL, Shen W-C. Design of an in vivo cleavable disulfide linker in recombinant fusion proteins. *BioTechniques.* 2010; **49**(1): 513.