

Synthetic construct of *Clostridium perfringens* epsilon-beta fusion protein gene

Reza Pilehchian Langroudi^{1&2}, Khosrow Aghaei Pour³, Mahdi Shamsara¹
and Saied Ali Ghorashi¹⁺

- ¹. Department of molecular genetics, National institute of genetic engineering and biotechnology, Tehran, Iran
². Department of anaerobic vaccine research & production, Razi serum and vaccine research institute, Karaj, Iran
³. Department of biotechnology, Razi serum and vaccine research institute, Karaj, Iran

Abstract. *Clostridium perfringens* one of the most important pathogens of humans and livestock is an anaerobic, spore-forming, heat-resistant bacterium which is subdivided to 5 groups (A to E) and produces 4 major and more than 12 minor toxins including epsilon and beta as major toxins. The aim of this study was to produce a new construction containing *Clostridium perfringens* type D epsilon toxin gene and type B beta toxin gene. A bioinformatics approach was used for ensuring that the designed epsilon-beta fusion gene is proper and would be cloned and expressed in the prokaryotic host cell. In silico study of predicted secondary and tertiary structures of the fusion protein construction, showed that a protein with two domains joined together by a small linker, will be produce. Sequence analysis showed that the designed fusion construction length is 1947 bp, which within the fragment consisting nucleotides 1 to 984 is complete epsilon prototoxin, 985 to 1020 (36 bp) is linker sequence which is optimized for *E.coli* and 1021 to 1947 (927 bp) is beta mature toxin. The synthetic construct sequence was submitted to GenBank (accession number JF833085). The present study clearly demonstrated that the designed epsilon-beta fusion gene could be clone and express in a suitable host cell.

Keywords: *Clostridium perfringens*, epsilon-beta, fusion, synthetic construct.

1. Introduction

The genus *Clostridium* is comprised of gram-positive, anaerobic, spore-forming bacilli. The natural habitat of these organisms is the environment and the intestinal tracts of humans and other animals. Clostridia are ubiquitous, and are commonly found in soil, dust, sewage, marine sediments, decaying vegetation, and mud (Sneath et al 1986). Despite the identification of approximately 100 species of *Clostridium*, only a small number have been recognized as relatively common etiologic agents of medical and veterinary importance. Nonetheless, some of these species are associated with very serious diseases. As virtually many of the species have been isolated from fecal samples of apparently healthy persons and animals, some of them may be transient, rather than permanent residents of the colonic flora. The majority of these organisms may be associated with serious and/or debilitating disease. In most cases, the pathogenicity of these organisms is related to the release of powerful exotoxins or highly destructive enzymes. Indeed, several species of the genus *Clostridium* produce toxins and other enzymes of great medical and veterinary

significance. (Hathewa 1990). *Clostridium perfringens* produces numerous toxins and is responsible for severe diseases including intestinal or foodborne diseases, hemorrhagic enteritis, enterotoxemia and gangrenes in human and animals. It is also a secondary pathogen in various diseases, such as necrotic enteritis (Sheedy et al 2004). Individual strains produce subsets of toxins and four of them iota (iA), alpha (cpa), beta (cpb) and epsilon toxins (etx) have been proposed to be used for classification of *Clostridium perfringens* into five isotypes A, B, C, D, and E, (McDonel et al 1986), (Petit et al 1999) where each type carries a different combination of the toxin genes. Alpha toxin (cpa) is found in *Clostridium perfringens* types A, B, C, D, and E, whereas the beta toxin (cpb) is found in types B and C. The epsilon toxin (etx) is found in types B, and D. The iota toxin (iA) is found only in type E. Epsilon-toxin is produced by strains of the bacterium that inhabit the gut of sheep and lambs, It is a potent toxin and a member of a group of highly toxic Clostridial toxins. Intoxication results in enterotoxaemia and neurological disorders. It is the third most potent Clostridial toxin after botulinium and tetanus neurotoxins, (McClane et al 2001).

Epsilon toxin is normally produced as a relatively inactive prototoxin and activates by proteolytic hydrolysis and conformational changes to cytotoxic, causing rapidly fatal enterotoxemia in sheep and goats, (Habeeb et al 1973), (Brown et al 197). The structure of the prototoxin had been solved by X-ray analysis. This structure suggests that it is a pore forming toxin that changes conformation between the soluble and transmembrane states. Using this form by X-ray diffraction and electron microscopy, a heptameric form of the protein is reported that is probably involved in pore formation in the host cell membrane. The main biological activity of epsilon toxin is the production of oedema in various organs. Epsilon toxin is known as one of the *Clostridium/Bacillus mosquitocidal* toxin ETX /MTX2, Family members.

Beta toxin causes enterotoxaemia and necrotic entrities in lambs, piglets and calves, (Hogh et al 1976). Beta-toxin gene sequencing has revealed similarity (17%-29%) to alpha-toxin, gamma toxin and leukocidin from *Staphylococcus aureus*. Beta toxin is known as one of the combined Leucocidin/ASH4 hemolysin, Domain and also *staphylococcal* bi-component toxin family members which are beta channel forming cytolysins, (Hunter et al 1993).

In the present study a new synthetic structural model containing *Clostridium perfringens* epsilon toxin and beta toxin genes was designed. A bioinformatics approach was used for in silico analysis of the structure of designed chimeric fusion protein. Fusion construction was inserted into pJET1.2blunt cloning vector and E.coli strain Top10 was transformed by the recombinant cloning vector.

2. Material and methods

2.1 Fusion gene construction

The classical end to end gene fusion technique has widely been used for monitoring gene expression, biological screening and purification of recombinant proteins (Doi et al 1999). In this study, end to end gene fusion was considered for design of epsilon-beta fusion gene construction and primers were designed for this purpose. The complete nucleotide sequences of *etx* and *cpb* were retrieved from the GenBank. The first data used was GenBank HQ179546, accession number, which is the sequence of the *Clostridium perfringens* epsilon toxin gene, as the representative sequence of this group of toxin genes. The same steps performed on GenBank HQ179547, accession number which is the sequence of the *Clostridium perfringens* beta toxin gene, as the representative sequence of this group of toxin genes. Nucleotide sequences were used to design the chimeric fusion gene. This construction is consisting of *etxD* and *cpbB* linked together by the linker fragment. *NdeI* restriction site and it's flanking region at the 3' end of epsilon toxin gene and *XhoI* restriction site and it's flanking region for the 5' end of beta toxin gene also was designed.

2.2 Fusion protein structural prediction

Patterns and profiles of the fusion chimeric protein were predicted by Interpro online software Physico-chemical parameters of the protein sequence were identified by ProtParam. Secondary structure predicted using PROMOTIF program, (Hutchinson et al 1996). Three set of analysis prepared using epsilon toxin peptide (328 amino acids), beta toxin peptide (309 amino acids) and the fusion chimeric peptide (649 amino acids). GOR was used for showing the location of corresponded linker fragment, (Garnier et al 1996). SignalP Prediction was used for locating the signal peptide cleavage sites based on neural networks (NN)

trained on Gram-positive bacteria and HMMTOP was carried out for prediction of the transmembrane helices location in the fusion protein. Tertiary structure of the fusion protein was predicted using online software. Several models were uploaded to the servers including Swiss model and I-TASSER, (Zhang 2008, 2009, Zhang et al 2010). Uploaded models were including complete epsilon toxin, epsilon prototoxin peptide, epsilon activated toxin, complete beta toxin and beta mature toxin. Fusion construction model was also uploaded including complete epsilon toxin - linker - beta mature toxin, containing 649 amino acids. At the end protein docking was done for these data confirmation.

3. Results

3.1 Fusion gene construction

Analyzing nucleotide sequences of epsilon toxin gene showed that this gene codes for a protein containing of 328 amino acids. A fragment of 32 amino acids at the N-terminal is a signal peptide and the next 296 amino acids fragment is an unactivated mature peptide or prototoxin. Analyzing nucleotide sequences of beta toxin gene showed that this gene codes for a protein containing of 336 amino acids. A fragment of 27 amino acids at the N-terminal is a signal peptide and the remained 309 amino acid fragment is a mature peptide.

3.2 Fusion protein structural prediction

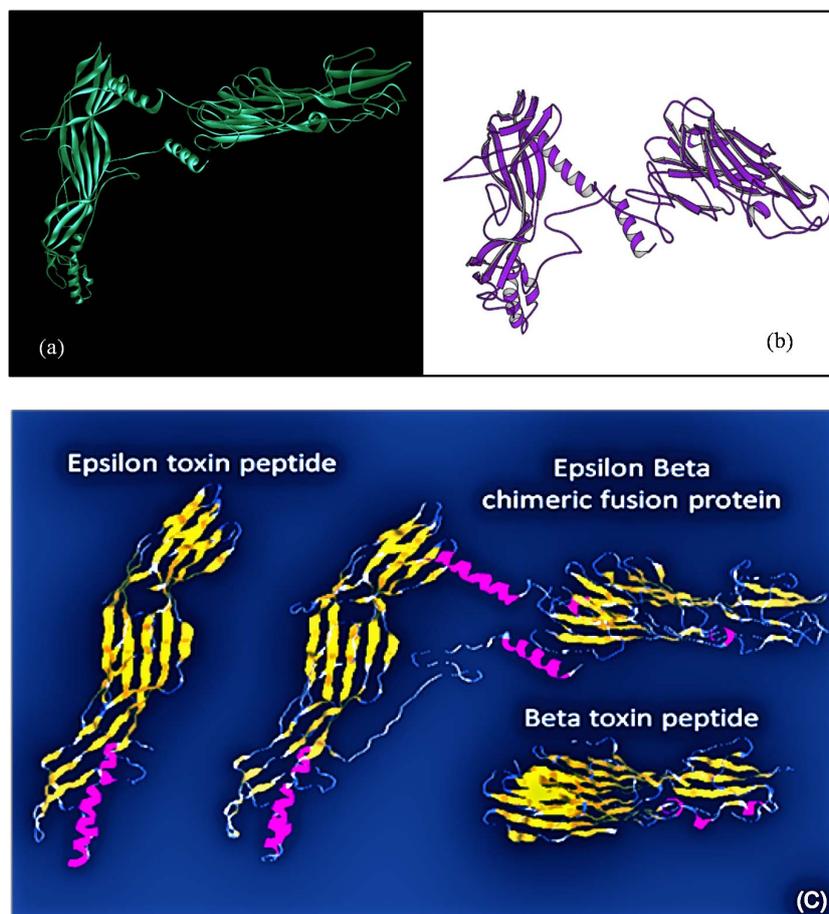
The designed fusion gene length is 1947 bp, which within 1 to 984 is belongs to epsilon toxin gene, 985 to 1020 (36 bp) is linker sequence which is optimized for *E.coli*. Base pairs 1021 to 1947 (927 bp) is belongs to beta toxin gene.

3.3 Primary structure and Physico-chemical parameters of chimeric fusion protein sequence

Primary structural analysis showed that the whole structure of the fusion protein construction is made up of, *Clostridium* epsilon toxin ETX/*Bacillus mosquitocidal* toxin MTX2, family member which is combined with Leucocidin/ASH4 hemolysin domain and also *Staphylococcal* bi-component toxin, family members. Physico-chemical parameters of the primary protein structure showed that it's formula is $C_{3187}H_{5033}N_{845}O_{1034}S_{16}$. Molecular weight computed to be 72244.0 and theoretical pI is 6.66. Total number of negatively charged residues (Asp + Glu) is 71, and total number of positively charged residues (Arg + Lys): are 70, extinction coefficients are in units of $M^{-1} cm^{-1}$, at 280 nm measured in water. Extinction coefficient 78285, Abs 0.1% (=1 g/l) 1.084, assuming all pairs of Cys residues form cystines. Extinction coefficient 78160 Abs 0.1% (=1 g/l) 1.082, assuming all Cys residues are reduced. The instability index is computed to be 28.38; this classifies the protein as stable. The Grand average of hydropathicity (GRAVY) is -0.499.

3.4 Secondary and tertiary structure prediction of the fusion protein

Secondary structural summary of the fusion protein construction prediction by PROMOTIF program is: Strand, 203 (31.3%); alpha helix; 52 (8.0%); 3-10 helix; 13 (2.0%); others; 381 (58.7%). Total residues are 649 which is made up of, 9 sheets, 9 beta hairpins, 1 psi loop, 7 beta bulges, 34 strands, 8 helices, 2 helix-helix interacts, 57 beta turns and 43 gamma turns. GOR analysis results showed that there is a helix pick located between positions 328-340 that is correspond to the linker fragment. According to SignalP most likely cleavage site is located between positions 32-33. Finally HMMTOP determined that fusion protein N-terminus is IN and there is transmembrane helix that is located between amino acids number 8 and 25. Results of tertiary structure of the fusion protein construction using I-TASSER showed a protein with two main domains linked together with a small linker. Figure 1 (a to c) showed different views of fusion protein construction tertiary structure predicted by I-TASSER.



4. Discussion

A recombinant fusion protein is a protein created through genetic engineering of a fusion gene. This typically involves removing the stop codon from a DNA sequence coding for the first protein, then appending the DNA sequence of the second protein in frame through ligation or overlap extension PCR. That DNA sequence will then be expressed by a cell as a single protein. The protein can be engineered to include the full sequence of both original proteins, or only a portion of either. If the two entities are proteins, often linker (or "spacer") peptides are also added which make it more likely that the proteins fold independently and behave as expected. Especially in the case where the linkers enable protein purification, linkers in protein or peptide fusions are sometimes engineered with cleavage sites for proteases or chemical agents which enable the liberation of the two separate proteins. This technique is often used for identification and purification of proteins, by fusing a GST protein, FLAG peptide, or a hexa-his peptide (aka: a 6xhis-tag) which can be isolated using nickel or cobalt resins (affinity chromatography). Chimeric proteins can also be manufactured with toxins or anti-bodies attached to them in order to study disease development.

This is the first time that *Clostridium perfringens* epsilon-beta toxin genes fusion construction is designed and produced. *Clostridium perfringens* is a spore forming bacterium and widely occurring pathogen. Vaccines against epsilon and beta toxins is manufactured using secreted toxin by virulent *Clostridium perfringens* types B and D. Large scale production of vaccines from virulent strains requires stringent safety conditions and costly detoxification and control steps. Therefore it would be beneficial to produce these toxins in a safe production host and in an immunogenic form.

In the present study, the epsilon toxin gene including its signal peptide sequence, for the proper secretion of the fusion protein, and the beta toxin without its signal peptide was used. Two repeats of "EAAAK" as a linker was designed. The synthetic construct epsilon-beta fusion protein gene, partial cds was deposited in GenBank under accession number JF833085. As the classical end to end gene fusion technique has widely been used for monitoring gene expression, biological screening and purification of recombinant proteins this gene fusion technique was considered for design of epsilon-beta fusion gene construction. The InterProScan,

(Zdobnov 2001), search identifies sequence motifs from several databases According to these data bases, the whole structure of the fusion protein construction is made up of, *Clostridium* epsilon toxin ETX/ MTX2, family member which is combined with Leucocidin/ASH4 hemolysin domain. The nucleotide and derived amino acid sequences of the epsilon toxin gene are located between the start codon at the base 188 and the stop codon at the base 1174 (986 bp), (Hunter et al 1992). These data demonstrated that the amino acid sequence from the start Met to Lys (base 284) about 32 amino acid, contained residues characteristic of a signal peptide derived from the *etx* gene suggesting that Lys may be the first residue of the mature, inactivated toxin. In another study it has been shown that the beta toxin which is produced by *Clostridium perfringens* types B and C is synthesized as a 336 amino acid protein, the first 27 residues of which constitutes a signal peptide, (Nagahama et al 2003). For confirmation of this finding, secondary structures of epsilon and beta toxins and the fusion chimeric protein were predicted by PROMOTIF online program software. Our data demonstrated similar characteristics of each of the epsilon and beta fragments comprised to epsilon-beta fusion protein constructions. The data confirms that amino acid Lys number 33 is the first residue of the epsilon mature, inactivated toxin. The trypsin cleavage site in epsilon-prototoxin located between Lys-14 and Ala-15. (Brown et al 1977) Data derived from the epsilon toxin gene indicated that, the residue preceding Lys-14 was also Lys. Since trypsin cleaves proteins on the C-terminal side of basic amino acids, this suggested that there were two possible sites for cleavage and this could explain this finding that the N-terminal residue of epsilon-toxin was Lys-14, not Ala-15. So from Lys (base 284) up to Lys (base 323) about 13 amino acids would be removed after cleavage.

According to the aim of the next step of the research, the designed chimeric gene should be ligated into PET expression vector and then *E.coli* strain *BL21* would be transformed by this DNA insert, so the *NdeI* and *XhoI* restriction sites and their flanking regions at the 3' end of epsilon and 5' end of beta toxin genes were designed, respectively. These sequences are necessary for inserting of fusion gene into PET expression vector. Translation of the fusion gene in the cytoplasm of transformed host cell would produce a 649 amino acids chimeric fusion protein. This protein is consist of epsilon toxin signal peptide (32 amino acids), mature inactivated peptide (296 amino acids) the small hydrophobic linker (12 amino acids) and beta toxin mature peptide (309 amino acids).

Tertiary structure predicted by I-TASSER, is shown in figure 1 demonstrate a protein with two main domains linked together with a small linker. The hydrophobic linker is introduced between two domains can provide suitable flexibility and separation. After expression these constructions can provide suitable accumulation of fusion protein in the periplasmic space of *E.coli*. This will happened because there is a signal peptide at the N terminal of the epsilon toxin of chimeric protein which will allow it to cross the cytoplasmic membrane. Protein models which predicted by I-TASSER are generated by combining the methods of threading, ab initio modeling and structural refinement. The procedure is fully automated and the final model is generated as PDB format. We uploaded this model to the profunc server which had been developed to help identify the likely biochemical function of a protein from it's three-dimensional structure. The Profunc server uses both sequence and structure-based methods to try for providing clues as the protein's likely or possible function, (Laskowski et al 2005, 2005). Often, where one method fails to provide any functional insight, another may be more helpful. Using a prediction program can only get closed to the final properties and specificities of a desired designed protein and it does not exactly reflect the absolute and definite properties and specificities of that protein. Beside of this statement in silico study of a functional protein folding and construction is the most important and the best way to understand of it's structure and functions.

Epsilon-beta fusion gene construction and it's cloning in *E.coli* is reported previously based on these dada (Pilehchian et al 2011), Results of these two studies clearly demonstrated that the designed and cloned epsilon-beta fusion gene construction could be expressed in a suitable host cell, and could serve for development and construction of novel fusion proteins for epsilon, beta and other potential toxins. These findings also indicated that the designed fusion protein could be used for, vaccine production and preventing toxicosis of *Clostridium perfringens* epsilon and beta toxin.

5. Acknowledgment

This study was performed due to a research project with financial and scientific support of National institute of genetic engineering and biotechnology and also Razi vaccine and serum research institute.

6. References

- [1] A.S. Brown, A.F.S.A. Habeeb. Structural studies on epsilon toxin of *Clostridium perfringens* type D. Localisation of the site of tryptic scission necessary for activation of epsilon toxin. *Biochemical and Biophysical Research Communications*, 1977, 73: 889-896.
- [2] M.T. Dertzbaugh. Genetically engineered vaccines: an overview. *Plasmid*, 1998, 39: 100-113.
- [3] N. Doi, H. Yanagawa. Insertional gene fusion technology. *FEBS Letters* 1999, 457 1-4: 3967-3979.
- [4] J..F Garnier, B. Gibrat, G.O.R. Robson. Secondary structure prediction method version IV, *Methods in Enzymology*, RF Doolittle Ed, 1996, 266: 540-553.
- [5] A.F.S.A. Habeeb, C.L. Lee, M.Z. Atassi. Conformational studies on modified proteins and peptides VII Conformation of epsilon prototoxin and epsilon toxin from *Clostridium perfringens*. Conformational changes associated with toxicity. *Biochemica et Biophysica Acta*, 1973, 322: 245-250.
- [6] P. Hogh. Experimental studies on serum treatment and vaccination against *Clostridium perfringens* type C infection in piglets. *Dev. Bio. Stand*, 1976, 32: 6-76.
- [7] C. L. Hatheway, *Clin. Microbiol. Rev.* 3:66-98 (1990)
- [8] S.E.C. Hunter, I.N. Clark, D.C. Kelly, R.W. Titball. Cloning and Nucleotide Sequencing of the *Clostridium perfringens* Epsilon-Toxin Gene and Its Expression in *Escherichia coli* ; *Infect Immun*, 1992. 60: No 1. 102-110
- [9] S.E.C. Hunter, E. Brown, P.C.F. Oyston, J. Sakurai, R.W. Titball. Molecular genetic analysis of beta-toxin of *Clostridium perfringens* reveals sequence homology with alpha-toxin, gamma-toxin, and leukocidin of *Staphylococcus aureus*. *Infect Immun*, 1993, 61: 3958–3965.
- [10] E.G. Hutchinson , J.M. Thornton."PROMOTIF - A program to identify structural motifs in proteins" *Protein Science*, 1996, 5:212-220.
- [11] R.A. Laskowski, J.D. Watson, J.M. Thornton. ProFunc: a server for predicting protein function from 3D structure. *Nucleic Acids Res*, 2005, 33: W89-W93.
- [12] R.A. Laskowski, J.D. Watson, J.M. Thornton. Protein function prediction using local 3D templates. *J. Mol. Biol*, 2005, 351: 614-626
- [13] B.A. McClane. *Clostridium perfringens*. Doyle MP, editor. Food microbiology: *fundamentals and frontiers*. Washington, DC: ASM Press, p. 351–72, 2001
- [14] J.L. McDonel. Toxins of *Clostridium perfringens* type A, B, C, D, and E, Dorner F, Drew H, editors. *Pharmacology of bacterial toxins*. Oxford: Pergamon Press; p. 447–517, 1986
- [15] M. Nagahama, S. Hayashi, S. Morimitsu, J. Sakurai: Biological activities and pore formation of *Clostridium perfringens* beta toxin in HL 60 cells. *J Biol Chem*, 2003, 278:36934–36941
- [16] L. Petit, M. Gibert, M.R. Popoff. *Clostridium perfringens*: toxinotype and genotype. *Trends Microbiol*, 1999, 7: 104–110.
- [17] R. Pilehchian Langroudi, K. Aghaei Pour, M. Shamsara, A.R. Jabbari, G.R. Habibi, H. Goudarzi, S.A. Ghorashi. Fusion of *Clostridium perfringens* type D and B epsilon and beta toxin genes and it's cloning in *E.coli*. *Archives of Razi Institute*, 2011, 66: No. 1, 1-10
- [18] S.A. Sheedy, A.B. Ingham, J.I. Rood, R.J. Moore. Highly conserved alpha-toxin sequences of avian isolates of *Clostridium perfringens*. *J Clin Microbiol*. 2004, 42:1345-1347
- [19] P. H. A. Sneath et al., "*Clostridium*," *Bergey's Manual® of Systematic Bacteriology*, Vol. 2, pp. 1141-1200, Williams & Wilkins, 1986
- [20] Y. Zhang. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*, 2008, 9:40.
1. Y. Zhang. I-TASSER: Fully automated protein structure prediction in CASP8. *Proteins*, 2009, S9: 100-113.
- [21] E.M. Zdobnov, R. Apweiler. InterProScan an integration platform for the signature-recognition methods in InterPro, *Bioinformatics*, 2001, 17: 847-848