**Introduction**

Dengue viruses are belonging to genus *Flavivirus* of the family *Flaviridae*. They are transmitted via mosquito *Aedes aegypti*. The dengue virus group comprises an independent serogroup of four closely related but antigenically distinct serotypes 1, 2, 3, and 4 [1]. It is estimated that more than 2.5 billion people are at the risk of infection by dengue virus. Dengue fever, dengue hemorrhagic fever (DHF), and dengue shock syndrome are important viral diseases caused by any of the four serotypes [2]. There is still no approved vaccine or drug for this virus. One of the challenges of dengue infection is that, although antibodies produced against the infecting serotype confer lifelong immunity against the infecting serotype, re-infection by another serotype has been associated with severe disease. Individuals with secondary dengue infection produce cross-reactive and non-neutralizing antibodies and also memory T cells, which can enhance
Hossein Fahimi et al • An in silico dengue vaccine

virus replication and over production of cytokines, that in turn are risk factors for DHF [3]. The leading hypothesis to explain pathogenesis of dengue virus is the antibody-dependent enhancement of infection; which suggests cross reactive sub-neutralizing antibodies may enhance the severity of infection [2]. Therefore, an effective vaccine against dengue virus must induce a simultaneous and strong immune response against all four serotypes.

Dengue virus genome comprised of an 11-kb plus-sensed RNA, surrounded by an icosahedral scaffold and covered by a lipid envelope. The genome contains one open reading frame encoding a poly-protein, which after cleavage by host and viral enzymes, produces three structural and seven non-structural proteins (Fig. 1) [4]. The envelope protein (E protein) is the major component of the virus, which contains specific sites for host cell attachment and fusion. This protein is a dominant antigen, which is potent for induction of immune response in infected hosts and eliciting virus-neutralizing antibodies [5,6].

The E protein contains three domains (EDI, EDII, and EDIII) (Fig. 2). The EDIII contains C-terminal -100 amino acids and forms an immunoglobulin-like β-barrel structure composed of 10 anti-parallel beta-strands [7], which contains the virus receptor-binding sites [8]. Peptides of the EDIII have been used with promising results as antigens for serologic diagnosis and immunization against these viruses. The EDIII participates in major activities of the E protein and it is also the major antigenic domain of the whole protein [9]. These findings put together justify the choice of domain III for tetravalent vaccine design.

Several types of vaccine candidates are developed against dengue virus, that most of them are monovalent vaccines and mixed physically after production in order to formulate as a tetravalent vaccine [10]. Therefore, we attempted to design a multi-component antigen, which is expected to induce immune system against all four serotypes, simultaneously. This is very important for equal induction of immune system against all serotypes, and prevention of unbalanced immunity. Unequal immunity cause undesirable reactions such as antibody depended enhancement of disease. This construct should contain EDIII domains from all four serotypes of dengue virus. Therefore, in order to designing and construction of a tetravalent dengue vaccine we developed four consensus EDIII sequences, corresponding to four serotypes of the virus. Then, we have designed a new tetravalent protein by fusing four consensus sequences together with appropriate hydrophobic linkers. With this aim, a rational and novel in silico approach have used for prediction of structural and immunological properties of the fusion protein as a dengue vaccine candidate.

Materials and Methods

Sequence selection and multiple alignments

In the first step, the corresponding sequences of EDIII (103 amino acids) from serotype 1 (48 sequences), serotype 2 (53 sequences), serotype 3 (45 sequences), and serotype 4 (29 sequences) were collected from Genbank at NCBI website. Multiple sequence alignments using ClustalW method of MegAlign software were performed in order to identify a common fragment to all EDIII sequences from each serotype.
Construct design
Net servers and several types of bioinformatics software are applied in our study [11-22]. The *in silico* gene analysis and optimization of the chimeric gene was performed using GeneRunner, EditSeq and online software such as Optimizer [11] and NEB cutter [12]. The mRNA secondary structure of the chimeric gene was predicted using Mfold program [13].

Bioinformatics analysis of recombinant fusion protein
For prediction of primary properties of fusion protein, EditSeq and online ProtParam software were used (http://web.expasy.org/protparam/). Secondary-structure prediction for fusion protein was performed by using the neural-network-based algorithm program PSIPRED (http://bioinf.cs.ucl.ac.uk/psipred/) and GOR4 method [14]. The Chou and Fasman method used for further analysis of secondary structure of protein (http://web.expasy.org/protsec). Tertiary structure prediction was performed by three methods including Modeller (http://salilab.org/modeller/), TASSER (http://zhanglab.ccmb.med.umich.edu/I-TASSER/), and FOLDpro (http://mine10.ics.uci.edu/). The stability of produced structural models was analyzed by Pymol and swiss-pdbviewer. Evaluation of model stability was performed based on energy minimization and Ramachandran plot. The Ramachandran plots were created by Procheck based at the biotech validation site (http://nihserver.mbi.ucla.edu/SAVS1/). Relative solvent accessibility of different residues was evaluated by online SCRATCH server (http://scratch.proteomics.ics.uci.edu/).

The DiNNNA online software was used for disulfide bound prediction (http://clavius.bc.edu/~clotelab/DiANNA/).

Prediction of B-cell epitopes
Web-based algorithm servers, Bcepred (http://www.imtech.res.in/raghava/bcepred/) and Discotope (http://www.cbs.dtu.dk/services/discotop/), were used for prediction of continuous and discontinuie B-cell epitopes, respectively. MHC super type A1 by NetCTL 1.2 Server (http://www.cbs.dtu.dk/services/NetCTL/). Finally, we used the Vaxijen server to predict the immunogenicity of the whole tetravalent fusion protein (http://www.ddg-pharmfac.net/vaxijen/Vaxijen/Vaxijen.html).

Results

Designing and construction of fusion protein
A multi-step immunoinformatic approach was used for designing a multivalent fusion protein from domain III of E protein. Firstly, the amino acid sequences of EDIII from all four serotypes were obtained from Genebank, presenting a variety of geographic and temporal origins. Multiple sequence alignments were performed in order to identify a consensus amino acid sequence for each serotype. Four consensus EDIIIs peptide sequences produced, and named as EDIII1, EDIII2, EDIII3, and EDIII4. Then, in order to identify sequence difference between four consensus EDIIIs, these sequences were aligned and compared, as depicted in Fig. 3. Several different residues between four EDIIIs are depicted in red.

**Fig. 3.** Comparison of amino acid sequences of the four consensus EDIIIs. (A) Sequence alignment of EDIIIs; different amino acid residues between serotypes are indicated in red blocks. The most conserves cysteine residues are showed by green arrows in the positions of 8 and 39. (B) Percentage identity and divergence among EDIIIs of four serotypes.
blocks. As depicted in Fig. 3B, EDIII4 sequence shows maximum divergence with other EDIIIs, whereas the EDIII1 and EDIII3 are the most identical sequences (with 71.8% identity and 35.5% divergence).

Consensus EDIII sequences were used for designing a tetravalent fusion protein, which named as EDIIIF (GenBank accession No. JN985899). In order to efficient separation of the EDIIIs in fusion protein, linkers containing four repeated EAAAK sequences were inserted between EDIII sequences. It has been shown that the salt bridge Glu-Lys between repeated Ala residues can stabilize helix formation [15], which expected to provide independent folding of four EDIII domains. In order to, efficient purification of EDIIIF protein, a 6×-His tag sequence was added to carboxyl terminus of designed protein. The arrangement of fragments junctions and linker sites are shown in Fig. 4.

### Prediction of primary and secondary structure of EDIIIF protein

On the basis of data obtained from prediction of EDIIIF properties by EditSeq software and online software ProtParam (in ExPASy tools), it is an approximately 52 kD protein (478 amino acid) with pH isoelectric 7.30. The instability index for this protein was computed to be 31.87, which classifies the protein as a stable protein in *Escherichia coli*. A similar analysis carried out using DiNNNA online software for disulfide bound prediction (Table 1). As depicted in the Table 1, prediction for numbers and positions of disulfide bonds in EDIIIF protein, showed that all of eight cysteine residues participate in disulfide bond formation. Clearly, there is one intra-domain disulfide bridge in each domain and no any inter-domain disulfide bond was predicted. These properties are essential for accurate conformation of EDIIIF protein, which predicted by temp-prot server analysis as well. Furthermore, predicted probability of antigenicity according the analysis by this server was 94%, which showed the EDIIIF has enough antigenicity as a vaccine candidate. The predicted antigenicity for the chimeric protein was higher than the predicted antigenicity for each EDIII proteins in separate form.

The secondary structure of the EDIIIF protein was predicted by online software. Several prediction methods were used and the results were compared for evaluation of the structure of this protein. These results indicated that helix structures lie in the regions of amino acid 104 to 123, 227 to 246, and 350 to 369, which are related to the hydrophobic amino acids entered between EDIII domains (Fig. 5).

Furthermore, the accuracy of this secondary structure prediction confirmed using ProtScale server. These properties include the scores for α-helix structures, the average flexibility, the beta-sheets, and the beta turns (Fig. 6).

### Prediction of tertiary structure of the EDIIIF protein

Comparative and homology modeling of the EDIIIF sequence was exploited to produce three dimensional models of the fusion protein. We used three methods for construction of three dimensional models of protein. The Modeller,
Tasser1, and FOLDpro methods are very useful for this purpose. In order to precise evaluation of new EDIIIF protein several three dimensional models were generated. The models analyzed using Swiss-pdbViewer and Pymol software in order to determine the final structure of the protein. The evaluations showed that, there were three \(\alpha\)-helixes and several \(\beta\)-sheets and \(\beta\)-turns, which were consistent with the results of secondary structure analyses, and also with the native structure of dengue virus envelope protein domain III. Formation of four separate domains in the recombinant fusion protein was approved by the results of tertiary structure prediction (Fig. 7).

In order to evaluation of EDIIIF protein stability, the profile of energy minimization was calculated by spdv software (-8,593.323 kJ/mol), which indicated the acceptable stability of recombinant protein. Additionally, the data generated by a Ramachandran plot confirmed the structural stability of the protein (Fig. 8).

Furthermore, the distribution of accessible regions to solvent (relative solvent accessibility), were predicted using the major hydrophobic and polarity properties of residual patterns. These patterns showed the acceptable solvent accessibility properties of the EDIIIF protein (Fig. 9). Accessible residues are distributed along all four domains of EDIIIF.

**Prediction of B-cell epitopes**

The epitopes located on the protein should be flexible and exposed to interact easily with antibodies. The Bcepred software was used to determine the continuous B-cell epitopes based on some characters including accessibility, hydrophilicity, flexibility, antigenicity, polarity, and exposed surface. There are several linear epitopes in each domain of EDIIIF, which are very important for antigenic properties of the protein (Table 2).
**Bioinformatics analysis of the optimized synthetic gene**

A synthetic sequence encoding the EDIIIIF protein was designed using *E. coli* codon usage. The coding sequence was analyzed for codon usage and GC content by online Optimizer software. This optimization is reflected by the codon adaptation index (CAI). The CAI is a useful measure for evaluation of the optimized synthetic gene.

**Furthermore, the constructed 3D models were used to predict the conformational B-cell epitopes.** The 49 spatial epitopes were predicted by Discotope server (Table 3). By using the constructed 3D models we also predicted the MHC super type A1 epitopes. The most likely predicted MHC epitopes in EDIIIIF protein were GTVLVQVKY, AAAKGMRY, GTIVRQVY, AAAKGMRY, CTNTFVLKK, VSETQHGTI, GTILKVEY, AAAKGMRY, and GTTVKVKY.

**Bioinformatics analysis of the optimized synthetic gene**

A synthetic sequence encoding the EDIIIIF protein was designed using *E. coli* codon usage. The coding sequence was analyzed for codon usage and GC content by online Optimizer software. This optimization is reflected by the codon adaptation index (CAI). The CAI is a useful measure for evaluation of the optimized synthetic gene.

**Bioinformatics analysis of the optimized synthetic gene**

A synthetic sequence encoding the EDIIIIF protein was designed using *E. coli* codon usage. The coding sequence was analyzed for codon usage and GC content by online Optimizer software. This optimization is reflected by the codon adaptation index (CAI). The CAI is a useful measure for evaluation of the optimized synthetic gene.
of the relative adaptiveness of desired sequences with the sequence of highly expressed genes in the same host. The optimized gene (edIIIf) showed a CAI of 1.0, compared to the primary deduced gene, which was only 0.7. The GC content of this sequence optimized on about 52%, which is appropriate for efficient expression in E. coli host. Furthermore, the necessary restriction enzyme sequence sites (NdeI and XhoI) were introduced at the ends of the coding sequence for cloning purposes.

Finally, secondary structure of corresponding mRNA for edIIIf gene was predicted. As the folding pattern of mRNA is very important factor affecting translation efficiency, we used the mfold web based software, in order to determine the potential folding pattern of mRNA. The minimum free energy for secondary structure formed by mRNA molecule was also predicted (dG = -17.12). The results of this prediction showed

![Ramachandran plot file](image)

**Fig. 8.** Evaluation of model stability by using the Ramachandran plot. According to the plot statistics, more than 82% of amino acid residues are in the most favored regions (A, B, L), and 13% are in additional allowed regions (a, b, l, p); whereas only 2.7% are in generously allowed (-a, -b, -l, -p) and 1.7% are in disallowed regions. Accordingly, the constructed model has a good quality.

**Table 2.** The linear epitopes predicted in EDIIIF protein based on different parameters by using Bcepred software

| Prediction parameter | Epitope peptide sequence |
|----------------------|--------------------------|
| Flexibility          | IPFSTQDE, PVVTKE, LSWFKKGSS, VOYEGDG, PVVTKE, LNWFKKGSS, PVVTKE, INWYKKGSS, EIRDVVK, LHWFKGSS |
| Accessibility        | SPKLEKEVAETO, OVKYEGTADP, PFDSTDEKVGTQVN, IVTDKEPVNIEAEP, KLFWKGGQSS, KFKVKEIAETO, RVQYEGDGQ, MOLERKMH, VTEKIDSNVIAEAPP, KLNWFKKGSSEA, KEEAAAKEAAAKKGMSYA, LKVEKSETQHGT, KVEKYGEDAP, TDEGQKAHANGR, NYPFVTKEEPVIAEAPP |
| Antigenic propensity | GHGTAVCLVKVE, PCKIPFS, GESYIVVG, FVKKKEI, GTIVIRVQYE, SPCKIPFIEI, LEKRHVLQRTWNPV, SYVIGVEP, CTNTIFVAMKVES, GTLIVKVEYK, PCKIPFES, GTTVKVKYVE, PCKVIEI, NVNEKVRGIRIS, SYVGA |
| Exposed surface      | SPKLEKE, VTDKEKVNV, KFKVKE, KLVWFKE, LKKEVSET, KVEYKGE, VVTKEEPVN, KINWVYKGSSE, KEAAAKK, RDVNEKEV |
| Hydrophilicity       | EKEVATEIQHT, KVEGTAPCK, VTDKEKPVN, PSTDEKVGTQVN, KGLSSSEA, AETQHGT, OYEGDGSPC, VTEKIDSNV, KKEVSETQHGT, VEYKGEDAPC, FSTEDGGKAKHNGR, KKEEPVN, YKGGSSSEAAKEAAAKEAAAKE, DKEMAETQHGT, RDVKEK, AENTNSVTN, RKGSSSH, KINWYKGSSE, KFSIDKEMAETQHGT, KKVKEGA, EIRDWNEKVKVGR, PFANEINTSVN, FRKGSSSH |
| Polarity             | SPKLEKEVAETO, IVTDKEKPVNIE, KLFWKGGQSS, KFKVKEIAETO, EMDLEKRRVHLGRL, KLNWFKKGSSEA, KEAAAKEAAAKEAAAKE, LKKEVSETOH, KVEYKGEA, GKAHNGRL, PVVTKEEPVNIE, KINWYKGSSE, KFSIDKEMAETQHGT, KVPKIRDNKEKVGR, TGLHWFKGSSHHHHHH |

4A linear epitope, AENTNSVTN, was also predicted in a turn.

**Table 3.** The identified 49 B-cell epitope residues out of 478 total residues as conformational B-cell epitopes of EDIIIF, which were predicted by Discotope server

| Amino acid position and sequence | |
|----------------------------------|--|
| 6 (Lys)                          | 143-145 (Glu- Thr-Gln) |
| 20-22 (Glu-Thr-Gln)             | 263 (Glu) |
| 20-22 (Glu-Thr-Gln)             | 292-299 (Glu-Asp-Gly-Gln-Gly-Lys-Ala-His) |
| 83 (Tyr)                         | 258 (Gln) |
| 101 (Gly)                        | 279 (Lys) |
| 126-127 (Met-Ser)               | 281-283 (Glu-Asp-Ala) |

http://dx.doi.org/10.7774/cevr.2016.5.1.41
http://www.ecevr.org/
enough stability of mRNA for efficient translation in bacterial host.

**Discussion**

*In silico* analysis of vaccine candidates is a useful approach for rational design of chimeric vaccines [16]. Bioinformatics tools for predicting antigenic properties and candidate vaccines analysis are now a standard approach. Rational design of vaccines and *in silico* analysis, combined with *in vitro* and *in vivo* verification, accelerate the vaccine discovery process by approximately 10-20 fold. Several bioinformatics software and servers are available that can help in the process for designing of chimeric vaccine design [17].

The minimalistic approach for dengue vaccine comes from the observations that EDIII based vaccines could induce virus neutralizing antibodies which were preventive for virus infection [9]. In the diagnostic context, it has been proposed that recombinant EDIII peptides containing virus-specific epitopes could be used for specific serological diagnosis of infections [18]. Recently, we reported immunogenicity of a recombinant EDIII protein from dengue virus type-3 in mice [19]. On the base of these studies, we conclude that the designed recombinant fusion protein also can be used for serologic diagnosis of dengue viruses from other flaviviruses.

Bioinformatics approach used for rational design of recombinant EDIIIF protein as a multivalent vaccine candidate, and the corresponding synthetic gene was optimized for appropriate protein expression, considering the codon usage and GC content of the expression host genome. Since the efficiency of heterologous protein production can be diminished by biased codon usage [20], the synthetic DNA fragment encoding the EDIIIF protein was constructed based on the codon usage table of *E. coli*. This index was increased from 0.75 in the primary sequence to 0.98 in designed synthetic gene. Furthermore, the percentage and distribution of GC in the coding gene were adjusted on 52%.

In the protein structure prediction, the EDIIIF protein models formed four domains that are separated by three main α-helix structures. In the 3D models of protein the three α-helix structures are related to the repeated EAAAK residues (104-123, 227-246, and 350-369), which are inserted between EDIII domains. The α-helix structures provide independent conformation possibility for all four domains without any interaction and by minimal steric hindrance between domains [21]. With these results we could expect that these α-helical parts could support the stable structure of protein. B-cell epitopes for the fusion protein could be predicted on the basis of the structural predictions and solvent accessibility. Several methods based on accessibility, hydrophilicity, flexibility, antigenicity, and secondary structure of proteins have been developed [22]. The integrated results based on different parameters showed that the most likely B-cell epitopes of this fusion protein were located in four distinct domains selected as the EDIII1, EDIII2, EDIII3, and EDIII4.

In conclusion, our results from bioinformatics analysis of the designed tetravalent EDIIIF, as a model, showed the feasibility of this approach in rational design and *in silico* evaluation of chimeric dengue vaccines.

**ORCID**

Hossein Fahimi  https://orcid.org/0000-0002-1744-8675
Majid Sadeghizadeh  https://orcid.org/0000-0002-2497-3152
Mahshid Mohammadipour  https://orcid.org/0000-0002-9813-9953

**References**

1. Kuno G, Chang GJ, Tsuchiya KR, Karabatsos N, Cropp CB. Phylogeny of the genus Flavivirus. J Virol 1998;72:73-83.
2. Halstead SB. Dengue. Lancet 2007;370:1644-52.
3. Rothman AL. Dengue: defining protective versus pathologic immunity. J Clin Invest 2004;113:946-51.
4. Fields BN, Knipe DM, Howley PM. Fields’ virology. 5th ed. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins; 2007.
5. Crill WD, Chang GJ. Localization and characterization of flavivirus envelope glycoprotein cross-reactive epitopes. J Virol 2001;75:13975-86.
6. Roehrig JT, Bolin RA, Kelly RG. Monoclonal antibody mapping of the envelope glycoprotein of the dengue 2 virus, Jamaica. Virology 1998;246:317-28.
7. Modis Y, Ogata S, Clements D, Harrison SC. Structure of the dengue virus envelope protein after membrane fusion. Nature 2004;427:313-9.
8. Crill WD, Roehrig JT. Monoclonal antibodies that bind to domain III of dengue virus E glycoprotein are the most efficient blockers of virus adsorption to Vero cells. J Virol 2001;75:7769-73.
9. Chavez JH, Silva JR, Amarilla AA, Moraes Figueiredo LT. Domain III peptides from flavivirus envelope protein are...
useful antigens for serologic diagnosis and targets for immunization. Biologicals 2010;38:613-8.
10. Swaminathan S, Batra G, Khanna N. Dengue vaccines: state of the art. Expert Opin Ther Pat 2010;20:819-35.
11. Puigbo P, Guzman E, Romeu A, Garcia-Valve S. OPTIMIZER: a web server for optimizing the codon usage of DNA sequences. Nucleic Acids Res 2007;35:W126-31.
12. Vincze T, Posfai J, Roberts RJ. NEBcutter: a program to cleave DNA with restriction enzymes. Nucleic Acids Res 2003;31:3688-91.
13. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 2003;31:3406-15.
14. Garnier J, Gibrat JF, Robson B. GOR method for predicting protein secondary structure from amino acid sequence. Methods Enzymol 1996;266:540-53.
15. Arai R, Ueda H, Kitayama A, Kamiya N, Nagamune T. Design of the linkers which effectively separate domains of a bifunctional fusion protein. Protein Eng 2001;14:529-32.
16. Amani J, Mousavi SL, Rafati S, Salmanian AH. In silico analysis of chimeric espA, eae and tir fragments of Escherichia coli O157:H7 for oral immunogenic applications. Theor Biol Med Model 2009;6:28.
17. Koraka P, Martina BE, Osterhaus AD. Bioinformatics in new generation flavivirus vaccines. J Biomed Biotechnol 2010;2010:864029.
18. Holbrook MR, Shope RE, Barrett AD. Use of recombinant E protein domain III-based enzyme-linked immunosorbent assays for differentiation of tick-borne encephalitis serocomplex flaviviruses from mosquito-borne flaviviruses. J Clin Microbiol 2004;42:4101-10.
19. Fahimi H, Allahyari H, Hassan ZM, Sadeghizadeh M. Dengue virus type-3 envelope protein domain III; expression and immunogenicity. Iran J Basic Med Sci 2014;17:836-43.
20. Burgess-Brown NA, Sharma S, Sobott F, Loenarz C, Oppermann U, Gileadi O. Codon optimization can improve expression of human genes in Escherichia coli: a multi-gene study. Protein Expr Purif 2008;59:94-102.
21. Tong JC, Tan TW, Ranganathan S. Methods and protocols for prediction of immunogenic epitopes. Brief Bioinform 2007;8:96-108.
22. Kolaskar AS, Tongaonkar PC. A semi-empirical method for prediction of antigenic determinants on protein antigens. FEBS Lett 1990;276:172-4.