

Bioinformatic development of a recombinant trivalent synthetic protein construct using PTXa, Tox, and TetX toxins as a DTP vaccine candidate

Salahi, Z¹, Noofeli, M², Ranjbar, MM², Bagheri, M², Esmaelizad, M², Niakan, M^{1*}

1. Department of Microbiology, Faculty of Medicine, Shahed University, Tehran, Iran
2. Razi Vaccine and Serum Research Institute (RVSRI), Agricultural Research, Education and Extension Organization (AREEO), Karaj, Iran

How to cite this article: Salahi Z, Noofeli M, Ranjbar MM, Bagheri M, Esmaelizad M, Niakan M. Bioinformatic development of a recombinant trivalent synthetic protein construct using PTXa, Tox, and TetX toxins as a DTP vaccine candidate. *Archives of Razi Institute Journal*. 2024;79(4):777-788. DOI: 10.32592/ARI.2024.79.4.777



Copyright © 2023 by



Razi Vaccine & Serum Research Institute

ABSTRACT

Traditionally, diphtheria-tetanus-pertussis (DTwP or DTaP) as pediatric vaccines are produced from the corresponding inactivated toxins or whole -cell pathogenic bacteria of *Corynebacterium diphtheria* toxin (Tox), *Clostridium tetani* toxin (TetX) and *Bordetella pertussis*. There are major concerns in the classic or acellular DTP (DTaP) vaccine production processes from native live bacterial sources as it may raise concerns on adverse effects and safety issues, complexity of the purifications for each agent as well as cost. Here, we designated a recombinant multi-epitope vaccine candidates by vaccinoinformatics study to address the mentioned issues and to develop a single trivalent fusion protein as a potent recombinant DTP vaccine. To achieve these goals, stages of immunobioinformatics were retrieved using proteinaceous toxins sequences, predicting secondary/tertiary structure and transmembrane topology, energy minimization, and model validation. Then, conformational and linear Bcell epitope prediction by several servers, mapping of consensus linear/discontinuous immunogenic regions and construction synthetic fusion vaccine candidates in respect to optimal immunogenic, physicochemical properties and high expression in prokaryote host were achieved. Finally, reverse translation, codon optimization, addition of cloning tags for pet 28a vector and optimization of physicochemical properties of synthetic trivalent fusion protein were performed. Through various hybrid immuno-informatics and structural bioinformatics analysis of predicted and experimental epitopes finally, 12 new consensus highly immunogenic linear and discontinuous epitopes in Tox, TetX and PTXa proteins were selected. The peptide sequences of these immunogenic regions were as follows: PTXa (AA34-64, AA184-256 and AA98-116), Tox (AA47-76, AA117-159, AA515-557 and AA245-265) and TetX (AA226-249, AA819-844, AA923-967, AA1009-1067 and AA1225-1315). In addition, the characteristics of the recombinant trivalent fusion construct were; 546 residue length, soluble (Grand average of hydropathicity (GRAVY) was -0.475), estimated half-life was >10 hours in *Escherichia coli*, pI 5.94 (a little acidic), stable protein (The instability index (II) 35.58) as well as thermally stable (Aliphatic index (AI) 71.67). The putative antigenic epitopes from different organisms in a single protein, as in the current study, will possibly improve the protective efficacy as novel potent, safe, cheap and broad-spectrum vaccines for better prevention of diphtheria, tetanus and pertussis infections in the future.

Keywords: *Bordetella Pertussis*, *Corynebacterium Diphtheria*, *Clostridium Tetani*, Epitopes, Recombinant Vaccine.

Article Info:

Received: 11 November 2023

Accepted: 7 February 2024

Published: 31 August 2024

Corresponding Author's E-Mail:
niakan@shahed.ac.ir

1. Introduction

The DTaP vaccine is a traditional polyvalent vaccine which helps to prevent and protect against the life-threatening diseases of diphtheria, tetanus, and pertussis in infants and young children (1,2). Diphtheria disease caused by infection with toxin-producing strains of *Corynebacterium diphtheriae*, a gram-positive bacillus. Tetanus disease caused by neurotoxins secreted by the Gram-positive bacillus, *Clostridium tetani* (1,2). Moreover, Pertussis is an acute and prolonged infectious cough illness and highly contagious human respiratory disease caused by *Bordetella pertussis* (*B. pertussis*), a fastidious gram-negative coccobacillus (3, 4). Because of the possible side effects of DTwP vaccine including severe allergic reactions following vaccination which can be life-threatening presenting with swelling of the face and throat, difficulty breathing, fast heartbeat, dizziness and weakness with cyanosis; since 1991, a promising DTaP vaccine containing acellular pertussis in place of killed pertussis has been licensed. However, countries using DTaP are now experiencing a resurgence of whooping cough, with the characteristic peak every 2–5 years, similar to the pre-vaccine era. The less effective long-term protection (waning after 3–5 years) of DTaP has allowed the pertussis epidemic cycles to resume. Accordingly, pertussis continues to be a global concern, with widely increased incidence in spite of high vaccination coverage rates with DTaP in developed countries, mainly due to waning immunity and pathogen adaptation. Therefore, there is a need to develop a new generation of vaccine(s) capable of overcoming the weaknesses related to the current vaccines and improving disease control and reducing the incidence of disease (5). Pertussis toxin (PT) is an AB toxin produced by *Bordetella pertussis* with the A protomer (S1 subunit) as the toxic subunit and B oligomer as the pentamer that binds to the cell receptors. The S1 subunit consists of 234 amino acids and is immunodominant. Antibodies to the S1 subunit have been shown to neutralize pertussis toxin *in vitro* and to protect mice against *B. pertussis* infection through aerosol. The oligomer is composed of one each of the S2, S3 and S5 subunits and two S4 subunits. S2 and S3 mediate adherence of the toxin to host cells. Similar to the S1 subunit, antibodies against oligomer or S2 and S3 subunits confer protection against *B. pertussis* infection in animal models, but are less effective than antibodies to S1. Diphtheria toxin (DT) produced by *Corynebacterium diphtheriae* is a secreted 535 amino acid protein and responsible for the symptoms of the diphtheria disease. After synthesis, DT is proteolytically cleaved into two fragments, A and B. The catalytic domain is located on fragment A and the receptor binding and translocation domains are located on fragment B. Fragment B is responsible for the binding of DT to specific cell surface receptors and the translocation of fragment A into the cytosol. Fragment A catalyzes the ADP-ribosylation of elongation factor-2 resulting in inhibition of protein synthesis and cell death. Tetanus toxin (TT) is a 150 kDa protein produced by *C. tetani* as a single

polypeptide which is subsequently cleaved into a 100 kDa heavy B chain and a 50 kDa light A chain linked by two disulfide bonds. The carboxyl-terminal portion (TTC, 50kDa) of the heavy chain binds to GT1b gangliosides on neuronal cells. The light chain, a zinc endopeptidase, is then internalized and transported to the central nervous system and released at the presynaptic nerve terminals. TT acts by blocking the release of neurotransmitters to inhibitory synapses, thus causing excitatory circuits to become resulting in tetanic spasms. Toxin-neutralizing antibodies mediate immune protection against tetanus disease (6). DT and TT are well-characterized antigens currently used in licensed diphtheria-tetanus-inactivated and acellular pertussis (DTwP & DTaP) vaccines. In DTaP vaccines, the pertussis antigens included are a combination of pertussis toxin (PT), filamentous hemagglutinin (FHA), pertactin and fimbriae. FHA consists of two immunodominant domains, named as type 1 and type 2, corresponding to the carboxyl and aminoterminals of the protein, respectively. Vaccines containing one or more of the PT and FHA antigens have been shown to be very similar in efficacy to those containing all four pertussis antigens (6). However, while the efficacy of DTaP vaccines is well documented, the vaccines are expensive to produce because the antigens are derived and purified from three different bacteria. Therefore, developing antigens that can be produced by a single organism will simplify vaccine production and reduce cost. The production of the recombinant compounds can also be considered in terms of reducing the side effects of chemicals used, such as formalin, in toxin inactivation (7). The aim of this project was to design a single protein containing three immunodominant parts from *Corynebacterium diphtheriae*, *Clostridium tetani* and *Bordetella pertussis* toxins with respect to a higher immune response and fewer side effects. The current study generated a fusion protein containing the N-terminal 179 amino acid fragment of the PT (S1 subunit), which has been shown to be protective in an animal infection model the full-length genetically detoxified DT, and the 50kDa tetanus toxin fragment C (TTC), which has also been shown to confer protection against toxin challenge assays in animal models. The work represents a first step in demonstrating a proof of principle (concept) that such an approach could be explored to introduce a new generation of DTaP vaccines.

2. Materials and Methods

2.1. Bioinformatics study

This research was conducted in the Research and Development department of Razi Vaccine and Serum Research Institute of Iran. Data preparation and classification, sequencing and repair, and selection of the best analysis methods were performed using BioEdit 6.6.2 and MEGA X software, and then the identification of the linear epitopes, spatial and non-continuous spheres, domains, and immunodominant regions of crystallized toxin proteins were studied in all three bacteria. Experimental spatial and linear epitopes were then

examined from the IEDB site database for B lymphocytes, and collectively agreed epitopes were predicted in a variety of physicochemical and biological systems. DISCOTOPE, ELLIPRO, and BCTOPE servers were used to predict spatial and continuous epitopes. The approved gene structure was constructed by bioinformatics analysis, and then the desired gene was cloned in *Escherichia coli* strain DH5 α (Gibco-BRL). The complete amino acid sequences of *Bordetella pertussis* toxin subunit 1 (P04977), *Corynebacterium diphtheria* toxin (Tox) (Q6KE85) and *Clostridium tetani* toxin (P04958) were obtained from the UniProtKB database (<https://www.uniprot.org/uniprot/>). The sequences were then inspected for signal peptides (SP), chains (domains), disulfide bonds, and as well as binding and active sites by reading annotations (in case of required prediction) and analyzing in BioEdit 6.6.9 software. To find the conformational/discontinuous immunogenic regions, in addition to the sequence, a three-dimensional structure of the proteins was performed using a modeling strategy. PTXA was modeled by SWISSMODEL server (<https://swissmodel.expasy.org/>) based on homology modeling due to the presence of the three-dimensional structure with more than 40% similarity to the patterns in PDB (Protein 3D structure database). GMQE (Global Model Quality Estimation), in this server, is a quality estimation that combines properties from the target-template alignment and the template structure (9). In addition, for each model residue, the expected similarity to the native structure is shown in the "Local Quality" plot. Typically, residues with a score below 0.6 are expected to be of low quality. Moreover, different model chains were displayed in different colors. The local score was reported in the B-Factor column of the PDB file (10, 11), if the model downloaded. The local quality estimation plot and the comparison with the non-redundant set of PDB structures were also provided. In the "Comparison" plot (8), the model quality scores of individual models are related to the scores obtained for experimental structures of similar size. The x-axis is the protein length (number of residues) and the y-axis is the normalized QMEAN score. Each point represents an experimental protein structure. Black dots are experimental structures with a normalized QMEAN score within 1 standard deviation of the mean ($|Z\text{-score}|$ between 0 and 1), and experimental structures with a $|Z\text{-score}|$ between 1 and 2 are gray. Experimental structures that are even further from the mean are light gray. The actual model is shown as a red star. The mean and standard deviation of the experimental structures around the x-location of the star are used to calculate the QMEAN Z-score of the model, i.e. how many standard deviations are away from the model mean scores (10, 11). TetX and Tox 3D proteins were modeled using the threading method with tools available on the server I-TASSER (Iterative Threading Assembly Refinement) server (<https://zhanglab.dcmf.med.umich.edu/I-TASSER/>). It is a hierarchical approach to protein structure prediction and structure-based function annotation. Structural templates

from the PDB were first identified by multiple threading approaches using LOMETS, with full-length atomic models constructed by iterative template-based fragment assembly simulations. Energy minimization processes and model validation were performed using the GROMOS96 method implementation in the SWISS PDB Viewer software version 4 (<http://www.expasy.org/spdbv/>). Structural validation was also analyzed by assessment of the Ramachandran plot (PDB file as input) tool in the PROCHEK server using implementations of the SAVES v6.0 (UCLA-DOE LAB) server (<https://saves.mbi.ucla.edu/>). DiscoTope 2.0 (12) and ElliPro (13) servers were used to predict discontinuous/conformational B cell epitopes from 3D protein structures in PDB format. CBTOPE (14) was also used to predict conformational epitopes using amino acid sequence as an input feature with a prediction accuracy of more than 85%. The CBTOPE server applied the Support Vector Machine (SVM) based method as mentioned above by using the amino acid composition generated from the query sequence(s) with overall accuracy (14). Specifically, Disco Tope was used to calculate and combine contact numbers (synonym for surface accessibility) and the propensity scores of residues in spatial proximity (12). Based on CASP (Critical Assessment of Protein Structure Prediction) experiments, DiscoTope (12) is more valid than other conformational epitope prediction servers. ElliPro has also been used to predict antibody epitopes, using an input protein structure in PDB format, based on solvent accessibility and flexibility with a score that is defined as a Protrusion Index (PI) value averaged over epitope residues (13). In addition, the Kolaskar and Tongaonkar method (15) has been used to predict linear antigenic epitopes based on physicochemical properties of amino acid residues with 75% prediction accuracy. Similarly, Emini Surface Accessibility Prediction was also used for linear epitope prediction. ABCpred server was also used to predict linear epitopes based on machine learning approach (an artificial neural network (ANN) algorithm (https://webs.iitd.edu.in/raghava/abcpred/ABC_method.html)). Finally, for precise determination of immune-dominant epitopes targeted by neutralizing antibodies, predicted epitopes and regions/ residues with immunogenicity potential based on 3D structures of toxins, were analyzed to find consensus immunogenic regions in native proteins. In addition, the IEDB server (<https://www.iedb.org/>) was searched for previous experimental epitopes. The amino acid sequences of selected consensus epitopes (immunogenic regions) for three toxins were manually assembled. Different patterns of glycine and serine (gsgg, ggggs, ggg, ggss, ggssg, ggssgg, ssg) were used as flexible linkers. Eventually, the final preferred epitopes were designed and selected by considering the appropriate immunogenicity through peptide engineering, epitopes binding with appropriate linkers based on various criteria such as greater immunogenicity and physicochemical properties of the vaccine structure such as molecular weight, pH, half-life, solubility, stability etc. It then

evaluated the essential tags, expression and structures in terms of immunogenicity and estimated occurrence of the desired characteristics in the body and proper processing. The order of fusion peptides (poly-epitope construct) was as TetX (tetanus toxin)-Tox (diphtheria toxin)-PTXA (from pertussis toxin A). The construct was also started with a secretory signal peptide attached. Post-translational modifications (PTMs) such as glycosylation, enzymatic cleavage sites, etc. were neglected due to the source of epitopes from bacteria. If it was necessary to maintain the structural conformation of immunogenic regions in the new synthetic protein compared to the native protein, this was considered during the design. Various physical and chemical parameters characterization such as theoretical isoelectric point (IP), molecular weight, total number of positive and negative residues, extinction coefficient, instability index, aliphatic index and Grand Average Hydropathicity (GRAVY) were estimated using the ExPasy ProtParam server (<https://web.expasy.org/protparam>). The protein sequence of the cassette was reverse translated (back-translated) into DNA using online JCAT (Java codon adaptation tool) server (16) without further manipulation. JCAT is a multi-tasking tool for adapting target gene codon usage to the most sequenced prokaryotes and some eukaryotic gene expression. It is used to improve heterologous protein production and does not require the manual definition of the highly expressed genes. The reverse translation tool takes a protein sequence as input and uses a codon usage table to generate a DNA sequence by representing the most likely non-degenerate coding sequence. The settings also take into

account the avoidance of unwanted cleavage sites for restriction enzymes. Finally, the *NcoI* RE (restriction enzyme) site (5'-C/CATGG-3') and *BamHI* RE site (5'-G/GATCC-3') and stop codon were added at the beginning and the end of the annotated cassettes, respectively, for insertion into the expression vector (16).

3. Results

3.1. Modeling, Energy Minimization and Toxin Validation

The 3D structural PTXA model was generated using the Swiss Model Server based on 99.15% identity to the 1prt PDB ID template. The GMQE and QMEAN for PTXA toxin were 0.91 and -1.68, respectively. As the GMQE score is near 1, it shows a high degree accuracy of the modeled structure (Figure 1). Also, the protein size (residues) is in good position compared to other characterized 3D protein structures. The TetX and Tox sequences were modeled using the I-TASSER server, as the similarity between these sequences with templates in the PDB database was less than 36.97% (PDB ID=3v0a) and 22% (PDB ID=6dkk), respectively. The predicted modeled structures (Figure 2) were further subjected to the constrained energy minimization procedure with a harmonic constraint of 100 kJ.mol⁻¹.Å⁻². It was applied to all the protein atoms using the steepest descent and conjugate gradient technique with Gromos96 43B1 parameter set, implemented in Swiss-PdbViewer, to eliminate the bad contacts between structural water molecules and protein atoms and to improve the stereochemistry of the model.

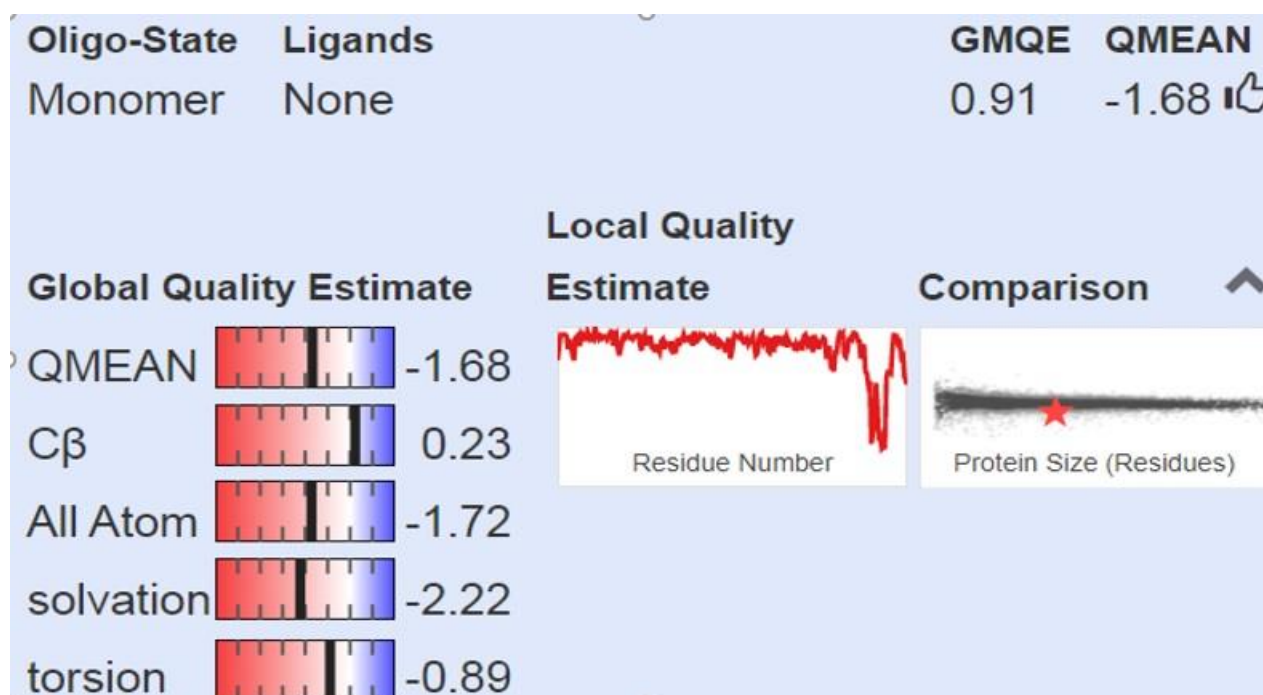


Figure 1. Quality assessment of PTXA modeled structure. The top-ranked templates and alignments were further analyzed and sorted according to the expected quality of the resulting models, as estimated by GMQE.

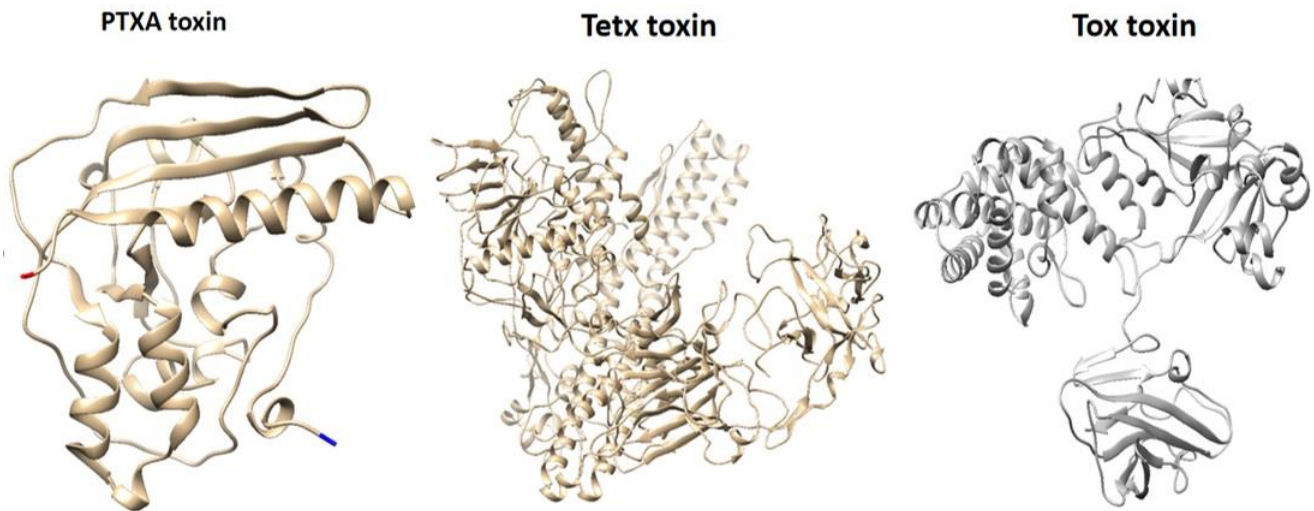


Figure 2. Modeled, validated and energy minimized tertiary structures of PTXA, Text and Tox toxins are illustrated. For all three proteins, substructures (coil, helix and β -sheet) are present in each structure. The structures are visualized and analyzed using the Chimera UCSF software (<https://techtransfer.universityofcalifornia.edu>). The beginning and the end of the structure are marked in red and blue, respectively.

In addition, the stereochemical quality of the TetX and Tox structures was evaluated by analyzing the geometry of each residue and the overall structure geometry using the PROCHECK program. Ramachandran plot results showed that 90.9% of residues were in favorable region, 8.6% in allowed region and 0.4% (27 labeled residues (out of 1313)) in outlier region (Figure 3). In parallel to Text, the stereochemical quality of Tox structure in Ramachandran plot showed 91.8% of residues in favorable region, 6.9% in allowed region and 1.1% (14 labeled residues (out of 511)) in outlier Figure 4. Taken together, these results showed that all modeled structures were suitable and valid. Energy minimization computation was performed in vacuum with GROMOS 96 43BQ parameters set without reaction field for three above modeled bacterial toxins. After energy minimization, the calculated total energy of the

model for PTXA, TetX and Tox were as $E = -23447.215$ kJ/mol, $E = -78556.586$ kJ/mol and $E = -23549.840$ kJ/mol, respectively. Since the optimization was energetic from the initial state (negative ΔG), the free energy of the predicted structures is correctly released.

3.2. Conformational and Linear B-cell Epitope Prediction by Hybrid Approach

Tables 1-3 show the results of B-cell epitope residues identified by different methods. Figure.1 also shows the results of the SEPPA and ElliPro servers visualized and represented in 3D modeled structures. Informative 3D visualization of predicted consensus immunogenic regions in alpha and NetB toxin by ribbon and surface representations is shown in Figure 2.

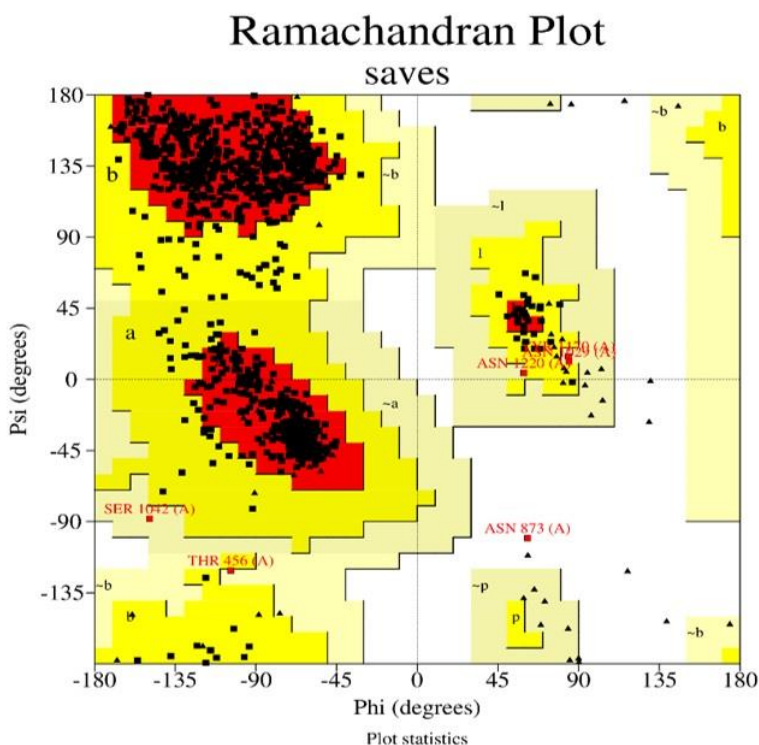


Figure 3. Text predicted model validation by Ramachandran plot in PROCHECK program (<https://saves.mbi.ucla.edu/>).

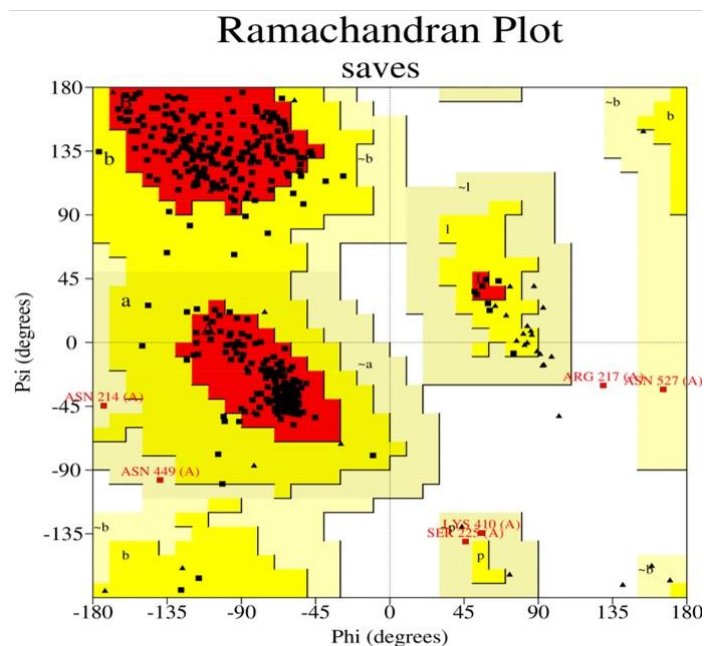


Figure 4. Validation of the predicted Tox protein model by Ramachandran plot in the PROCHECK tool (<https://saves.mbi.ucla.edu/>).

Table 1. Predicted linear and conformational (discontinuous) B-cell epitopes of the PTXA toxin by physicochemical properties using IEDB server and Machine Learning approaches. Corresponding boxes are based on their scores obtained from different servers.

Type of prediction	Server Name	Signal peptide & Transmembrane Topology prediction	PTXA epitopes positions (Immunogenic regions)
Linear epitope Prediction (Machine learning approach)	ABCpred (Threshold setting (Default value is 0.80))	S1 subunit Signal peptide: 1-35 Binding site: AA60 Active site: AA69, 163 Disulfide bond: 75 ↔ 235	88-112, 127-141, 170-184, 31-45, 3-17, 177-191, 163-177, 143-157
Linear epitope (Physico-chemical properties approaches, by IEDB server)	Kolaskar and Tongaonkar Antigenicity prediction		39-44, 72-78, 82-88, 94-99, 118-125, 134-144, 152-169, 198-204, 2015-237
	Emini surface accessibility prediction		42-49, 88-94, 125-130, 189-197, 201-215, 238-243
Predicting of conformational/ discontinuous epitopes	CBTOPE (Sequence based method, SVM threshold -0.1)		65-76, 149-164, 206-2018
	DiscoTope 2.0		38-53, 96, 147-153
	ElliPro		Epitope 1: 36-39, 47-58, 103-117, 126-130, 140-153, 169-224, 249-251, 260-269 Epitope 2: (with lower rank): 253-257

Table 2. Predicted linear and conformational (discontinuous) B-cell epitopes of the TetX toxin by physicochemical properties using IEDB server and Machine Learning approaches

Type of prediction	Server name	Signal peptide & Transmembrane Topology prediction	TetX epitopes positions (Immunogenic regions)
Linear epitope Prediction (Machine learning approach)	ABCpred (Threshold setting (Default value is 0.87))	TetX Signal peptide: 1-30 Metal binding: 233 , 237 Active site: 234 Tetanus toxin light chain: 2-457 Tetanus toxin heavy chain: 458-1315 Disulfide bond: 439 ↔ 467 (Interchain (between light and heavy chains)), 1077 ↔ 1093	268-282, 859-873, 372-386, 922-936, 551-565, 405-419, 174-188, 159-173
Linear epitope (Prediction by physico-chemical properties approaches, by IEDB server)	Kolaskar & Tongaonkar Antigenicity Prediction		24-42, 107-122, 163-173, 176-182, 195-210, 220-262, 301-309, 311-316, 376-382, 427-445, 506-520, 562-570, 587-618, 635-657, 664-694, 713-728, 741-759, 850-866, 881-886, 901-922, 925-933, 953-971, 985-990, 1025-1030, 1043-1051, 1086-1112, 1124-1140, 1167-1174, 1191-1197, 1200-1206, 1227-1232, 1234-1264, 1296-1305
Predicting of conformational/ discontinuous epitopes	CBTOPE (Sequence based method, SVM threshold -0.3)		18-22, 69-76, 136-143, 177-181, 194-196, 208, 306-307, 343-344, 425-426, 441, 505-508, 563, 601-610, 735-737, 865-867, 963-965, 1114-1117, 1206, 1246-1259, 1275
	DiscoTope 2.0 (-2.0 Threshold, Structural based method)		54-82, 141-149, 179-181, 205-208, 217-224, 293-348, 380-431, 447-456, 495-502, 541-546, 606-625, 693-709, 757-787, 837-850, 962-971, 1155-1165, 1181-1188, 1203-1235, 1274-1280, 1313-1315
	ElliPro (Structural based method)		486-500, 678-711, 812-859, 927-1315

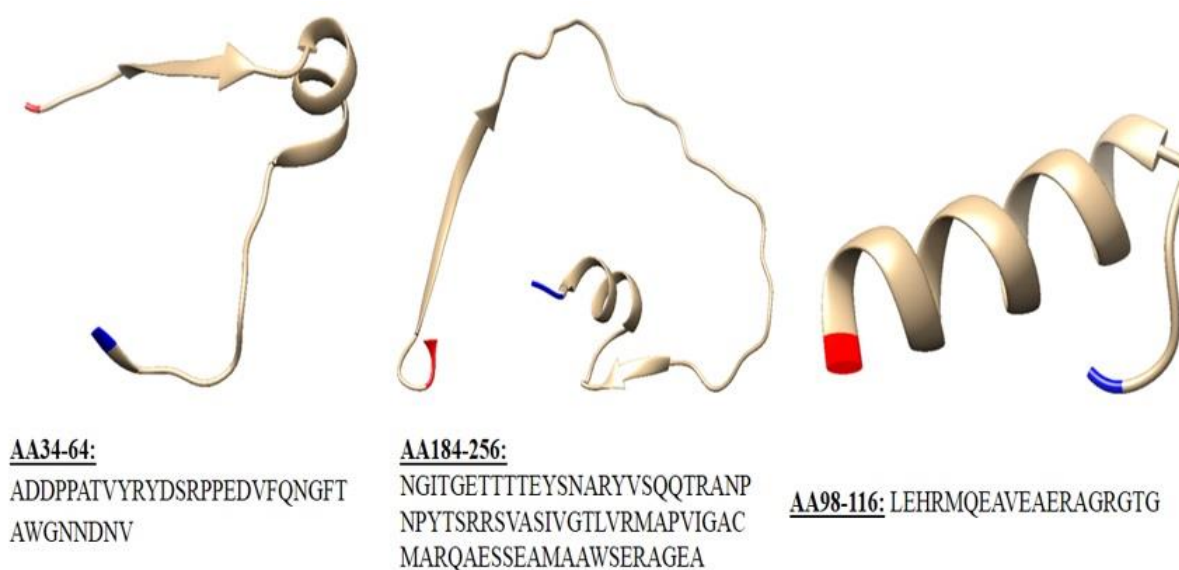
Table 3. Predicted linear and conformational (discontinuous) B-cell epitopes of Tox toxin by physicochemical properties using IEDB server and Machine Learning approaches. Corresponding boxes are based on their scores obtained from different servers.

Type of prediction	Server name	Signal peptide/Transmembrane Topology prediction	Tox epitopes Positions (Immunogenic regions)
Linear epitope Prediction (Machine learning approach)	ABCpred (Threshold setting (Default value is 0.86))	Tox Signal peptide: 1-25 Chain: 26-538 Domain: 26-212 , 225-404 and 406-538	14-28, 414-428, 380-394, 350-364, 50-64, 56-70
Linear epitope (Prediction by physico-chemical properties approaches)	Kolaskar and Tongaonkar Antigenicity Prediction		5-18, 27-36, 50-56, 103-111, 113-121, 157-166, 182-189, 221-232, 305-316, 328-336, 353-381, 383-398, 412-420, 452-463, 493-507, 509-517
Predicting of conformational /discontinuous epitopes	CBTOPE (Sequence based method, SVM threshold -0.3)		60-103, 154-169, 2011-222, 270-271, 315-325, 407-441, 469-502, 530-537
	DiscoTope 2.0 (-2.0 Threshold, Structural based method)	48-73, 92-103, 191-196, 426-438, 517-527	
	EllIPro	433-440, 450-538	

3.3. Final Selection of Consensus Immunogenic Regions

Finally, the consensus immunogenic regions were selected on the three-dimensional structure of the modeled toxins based on the scores reported in the epitope prediction servers, especially servers predicting conformational/discontinuous epitopes, servers agreement degree on the epitopes, the hydrophilicity degree, surface exposure and flexibility, the epitopes configuration or immunogenic region in 3D space, adaptability structure of new synthetic protein and respecting to the known experimental epitopes data. Informative sequence and 3D visualization of predicted consensus immunogenic regions for PTXA, TetX and Tox by ribbon representation are shown in Figures 5, 6

and 7. For the prediction of linear epitopes, different methods were used to increase the prediction accuracy, although the Machine Learning method provides more reliable results. Again, prediction of spatial epitopes is much more important than linear epitopes since, many epitopes (more than 80%) that are immunogenic and recognized by Bcells are conformational/discontinuous compared to linear epitopes (19). The protein sequence of the fusion multi-epitope construct was also shared in the Gene Bank for corresponding similarities (Figure 8). The construct was arranged in the order Tetanus (TetX) - Diphtheria toxin (Tox)-Pertussis toxin (PTXA) from 5' to 3' direction.

**Figure 5.** Sequences and tertiary structures of the predicted consensus epitopes for PTXA. The beginning and end of the structures are marked in red and blue, respectively.

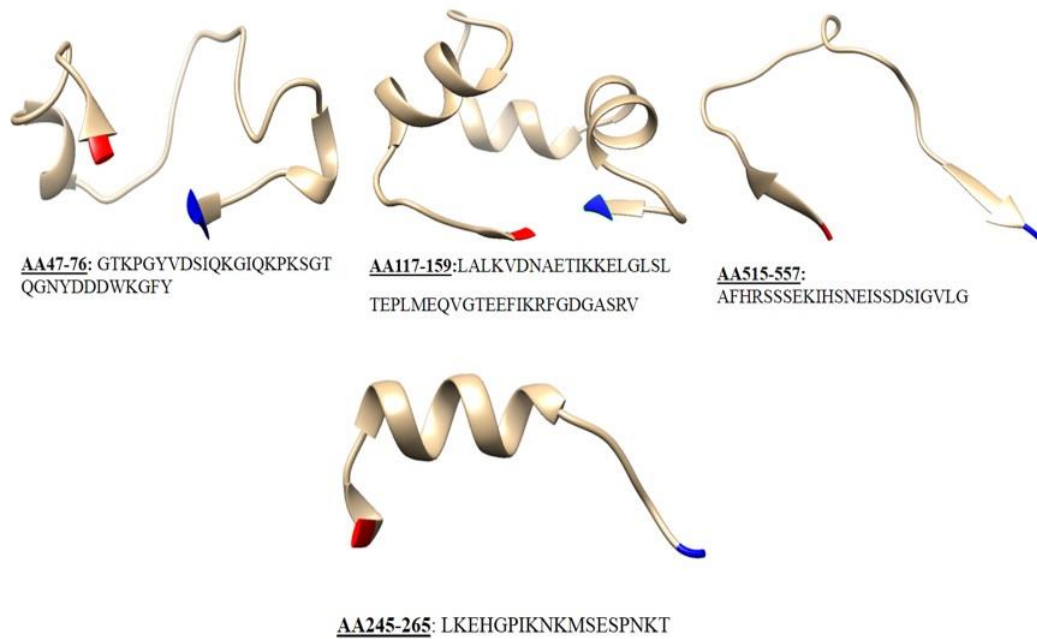


Figure 6. Sequences and tertiary structures of the predicted consensus epitopes for **Tox**.

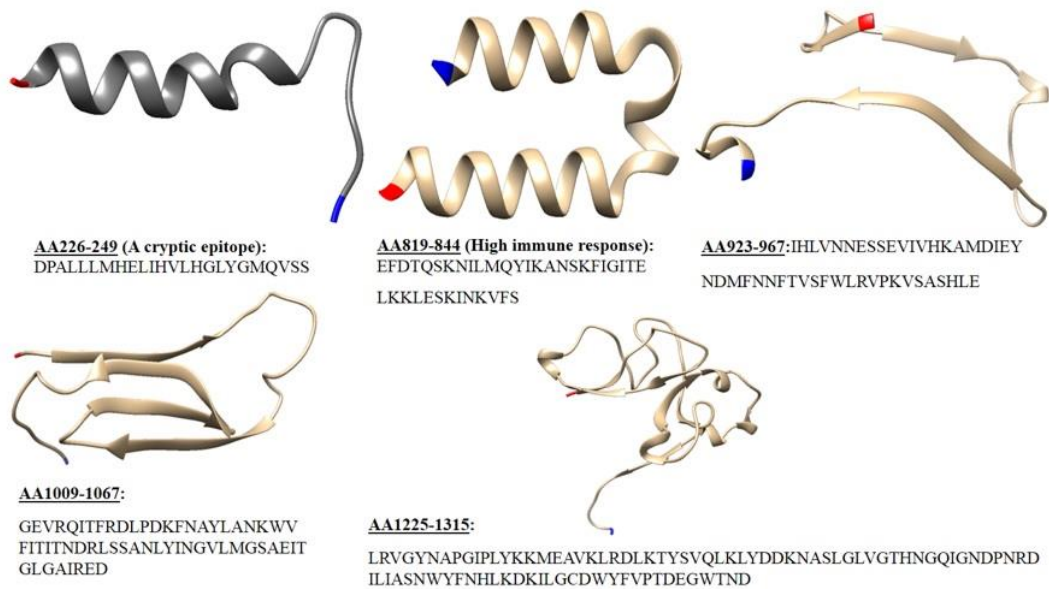


Figure 7. Sequences and tertiary structures of the predicted consensus epitopes for TetX toxin (TetX). The beginning and the end of the structure are marked in red and blue, respectively.

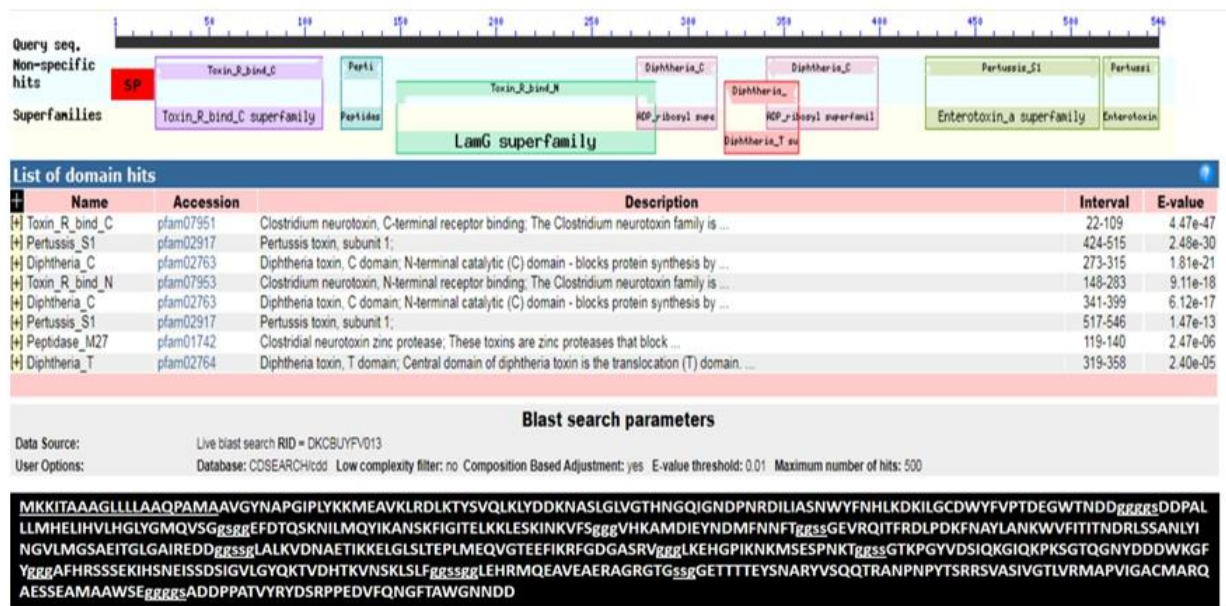


Figure 8. Schematic representation of the synthetic fusion construct and its blasting hits results. SP= Signal peptide. Underlines indicate signal peptide and linker (gs) sequences.

3.4. Prediction of the Physicochemical Properties of Construct

To predict the in vitro function and properties of the candidate vaccine construct, the physicochemical properties were evaluated to make the critical correction(s) (if any). The results showed that the total number of amino acids was 546, the theoretical pI was 5.94, the total number of negatively charged residues (Asp + Glu) was 64, and the total number of positively charged residues (Arg + Lys) was 57. The estimated half-life was 30 hours (mammalian reticulocytes, in vitro), > 20 hours (yeast, in vivo) and >10 hours (*Escherichia coli*, in vivo). The instability index (II) was calculated 35.58 that implies and classifies as a stable

protein, since protein instability index of less than 40 is predicted as stable. Aliphatic index (AI) was computed 71.67 for the fused peptide, which above 40, reflects thermal stability of protein. Grand average of hydropathicity (GRAVY) was -0.475, which shows hydropathic synthetic of protein and reflects suitable solubility for fusion toxins construct.

3.5. Reverse Translation, Codon Optimization and Final Optimizations of Synthetic Fusion Protein

The final fusion construct protein sequence was finally designed as a cassette containing components of tetanus toxin (TetX)–diphtheria toxin (Tox)-pertussis toxin (PTXA) fusion peptides:

Tetanus Toxin C Fragment

“MKKITAAGLLLLAAQPAMAAVGYNAPGIPLYKKMEAVKLRDLKTYSVQLKLYDDKNASLGLVGTHTNGQIGNDPNRDILIASNWYFNHLKDKILGCDWYFVPTDEGWTTNDDggsDDPALLLMHELHVLHGLYGMQVSGgsgEFDTQSKNILMQYIKANSKFIGITELKKLESKINKVFSgggVHKAMDIEYNDMFNNFTggsGEVRQITFRDLPDKFNAYLANKWWFITITNDRLLSSANLYINGVLMGSAEITGLGAIREDDggsLALKVDNAETIKKELGSLTEPLMEQVGTTEFIKRFGDGASRVgggLKEHGPIKNKMSSEPNKTggsGTKPGYVDSIQKGIQPKSGTQGNYYDDWKGfygggAFHRSSSEKIHSNEISSDSIGVLGYQKTVDHTKVNKLSLFGgssgLEHRMQEAVEAERAGRGTgssGETTTTEYSNARYVSQQTRANPNPYTSRRSVASIVGTLVRMAPVIGACMARQAESSEAMAAWSEggsADDPPATVYRYDSRPPEDVFQNGFTA WGNND”

The estimated half-life was 30 hours in mammalian reticulocytes, in vitro, >20 hours in yeast, in vivo and >10 hours in *Escherichia coli*, in vivo. Reverse translation, changing RE codon optimized construct and codon usage adapted to *Escherichia coli* (strain K12) during conversion were performed. A polyhistidinetag Deca (8x)-histidine tag optimized motif was tagged to the 3' end of the construct just before the TAA stop codon. The final Improved nucleotide sequence for construction into expression vector was as follows: GCcontent before and after optimization was 50.734 and 50.183, respectively. Trivalent fusion protein which designed as recombinant poly-epitope vaccine candidate showed optimal physicochemical properties. The construct sequence was further refined for absence of *NcoI*, *BamHI* and some other enzyme restriction sites. Moreover, *NcoI* (5'-C/CATGG-3') and *BamHI* (5'-G/GATCC-3') were added to 5' and 3' ends, respectively. To avoid frameshift problems, the ATG start codon of the signal peptide was replaced by GC nucleotides. Consequently, the final nucleotide construct sequence after reverse translation and optimization was as follows:

4. Discussion

Vaccination of Iranian children began in 1940s with the use of domestically manufactured vaccines against three deadly diseases: diphtheria, tetanus and pertussis. In rare cases of adverse reactions in children, complications may be related to the vaccine formulation methods as well as the cellular components of the bacteria used in the production processes (17, 18). Due to the side effects of using whole-cell or several proteins in vaccine composition including high fever, weakness, lethargy and anaphylactic shock in rare cases; recombinant DNA procedures have allowed fusions of genes in a simple way to concentrate the immune response, enhance easier production, and reduce side effects and costs (19, 20). Henceforth, it was decided to use modern methods of immunobioinformatics and cost-effective knowledge of recombinant protein to design better immunogenic antigens in a safe system, without the disadvantage of the chemicals side effects such as formalin used to inactivate toxins which is often allergenic (3).

Codon Usage adapted to *Escherichia coli* (strain K12)

“ATGAAAAAATCACCGCTGCTGCTGGTCTGCTGCTGGCTGCTCAGCCGGCTATGGCTGCTGTTGGTTACAACGCTCCGGGTATCCCGCTGTACAAAAAATGGAAGCTGTTAACTGCGTGACCTGAAAACCTACTCTGTTTCAGCTGAAACTGTACGACGACAAAAACGTTCTCTGGGTC TGGTTGGTACCCACAACGGTCAGATCGGTAACGACCCGAACCGTGACATCCTGATCGCTTCT AACTGGTACTTCAACCACCTGAAAGACAAAATCCTGGGTTGCGACTGGTACTTCGTTCCGAC CGACGAAGGTTGGACCAACGACGACGGTGGTGGTGGTCTGACGACCCGGCTCTGCTGCTG ATGCACGAACTGATCCACGTTCTGCACGGTCTGTACGGTATGCAGGTTTCTGGTGGTCTGG TGGTGAATTCGACACCCAGTCTAAAAACATCCTGATGCAGTACATCAAAGCTAACTCTAAAT TCATCGGTATCACCGAACTGAAAAAACTGGAATCTAAAATCAACAAAGTTTTCTCTGGTGGT GGTGTTCAAAAGCTATGGACATCGAATACAACGACATGTTCAACAACTTCACCGGTGGTTC TTCTGGTGAAGTTCGTCAGATCACCTTCCGTGACCTGCCGGACAAATTCAACGCTTACCTGG CTAACAAATGGGTTTTTCATCACCATCACCAACGACCGTCTGTCTTCTGCTAACCTGTACATCA ACGGTGTTCTGATGGGTTCTGCTGAAATCACCGGTCTGGGTGCTATCCGTGAAGACGACGGT GGTCTTCTGGTCTGGCTCTGAAAGTTGACAACGCTGAAACCATCAAAAAAGAACTGGGTCT GTCTCTGACCGAACCGCTGATGGAACAGGTTGGTACCGAAGAATTCATCAAACGTTTTCGGTG ACGGTGCTTCTCGTGTTGGTGGTGGTCTGAAAGAACACGGTCCGATCAAAAAACAAAATGTCT GAATCTCCGAACAAAACCGGTGGTCTTCTGGTACCAAACCGGGTTACGTTGACTCTATCCA GAAAGGTATCCAGAAACCGAAATCTGGTACCCAGGGTAACTACGACGACGACTGGAAAGGT TTCTACGGTGGTGGTGCTTTCCACCGTTCTTCTTCTGAAAAAATCCACTCTAACGAAATCTCT TCTGACTCTATCGGTGTTCTGGGTTACCAGAAAACCGTTGACCACACCAAAGTTAACTCTAA ACTGTCTCTGTTCCGGTGGTCTTCTGGTGGTCTGGAACACCGTATGCAGGAAGCTGTTGAAG CTGAACGTGCTGGTCGTGGTACCGGTTCTTCTGGTGGTGAACCACCACCACCGAATACTCT AACGCTCGTTACGTTTCTCAGCAGACCCGTGCTAACCCGAACCCGTACACCTCTCGTCGTTCT GTTGCTTCTATCGTTGGTACCCTGGTTCGATGGCTCCGGTTATCGGTGCTTGCATGGCTCG TCAGGCTGAATCTTCTGAAGCTATGGCTGCTTGGTCTGAAGGTGGTGGTGGTCTGCTGACG ACCCGCCGGCTACCGTTTACCGTTACGACTCTCGTCCGCCAGAGGACGTCTTCCAGAACGGC TTCACCGCTTGGGGTAACAACGACGAC”

Numerous studies have shown that recombinant bacterial vaccines successfully stimulate the immune responses, and their production, in contrast to conventional inactivated antigen processing and purification, allows high expression levels and improves vaccine safety (19, 20). Aminian *et al.* used two copies of the *Bordetella pertussis* toxin S1 fragment, *Corynebacterium diphtheria* (DT) and *Clostridium tetani* (TT) modified toxin fragments as a recombinant fusion protein-based vaccine candidate (6). They cloned the three desired sequences in orders to make a multi-step cassette construct, and then expressed it in an expression vector where the complexity was increased due to the nature of the multi-step procedure. It was concluded that the antisera produced against the recombinant fusion protein recognized the native PT, DT and TT, but the protective efficacy of the fusion protein in animal models was not examined. In another study, Torkashvand *et al.*, used a double combination of the *Bordetella pertussis* toxin and FHA fragments in *E. coli*, which showed successful results in stimulating the immune responses (8). They showed an adequate systemic and mucosal immune response after administration of the recombinant fusion protein (F1S1) via the SC or IN route. Immunization with F1S1 as a fusion protein induced high levels of serum-specific IgG and lung IgA antibodies as well as Th1-specific T cell subpopulation responses which were considered as appropriate immune responses. In addition, the high production of recombinant F1S1 protein in *E. coli* was also reflected as another advantage of this fusion protein (3). Ruth *et al.*, also developed an oral triple vaccine from diphtheria, tetanus and pertussis expressed in tomato plant and showed protective specific antibodies in mice (21). In the current study, the distribution of protein sequences and the finding of similar proteins, the detection of mutation sites along the protein sequence and the removed or deleted sequences were investigated using several databases and tools. Protein sequence annotation was performed and the sequence correctness was checked by UNIPROT. Sequence similarity, such as checking the location and amount of protein change and detection of mutation sites along the protein sequence and the amount of coverage of the removed sequence were investigated. Then, the viability of whole sequence was examined by using the BIOEDIT program and drawing the entropy curve. DISCOTOP, BCTOPT and PDB (protein data base) software were used to study the three-dimensional structure and the final model generated in the PDB format in the working method. Moreover, Expasy software and Kolaskar's algorithm were used to identify the physicochemical properties and investigate in terms of hydrophobicity, hydrophilicity and potential curvature of the fusion protein sequence or chimera in the form of TETx-TOX-PTxA. Modeling was performed using the SWISS MODEL program, and epitope prediction and B cell linearity or spatiality were performed using the IEDB Analysis Resource Program. The epitopes were ranked based on the score and values above the target proteins were selected, and then predicted by the ELIPRA server and conformational DISCOTOP of continuous and discontinuous epitopes were

predicted. Finally, the epitopes attached and designed with serine and glycine linkers were checked and the desired structure was finalized. Once the structure was designed, the hydrophilicity curve and the processing of the chimeric protein in the body and the physicochemical properties of the candidate chimeric fragment were determined and the reverse translation and codon optimization were performed and ultimately the desired structure was ordered. Moreover, a construct design based on three diphtheria, tetanus and pertussis toxin antigens with the accession numbers of TOX (Q6KE85), TEX (P04958), and PTX (P04977), respectively, was performed. It was also included the cleavage sites required for its linkage and the nucleotide sequences encoding glycine + serine and other sequences required for expression and purification (7). The designed sequence was cloned with the *Xba*I and *Xho*I cleavage sites into the expression vector pET-28a (+). In order to position these genes together, researchers have used various methods, including the synthesis of the desired genes in full length, which is not only expensive, but also time-consuming due to the synthesis complexity caused by the full length of the sequence. In another attempt, the initial cloning was performed in a non-expression vector and then subcloned, which was also time-consuming and complex due to the non-specificity of the vectors in the linkage stage. For the first time, using pET vectors and the designed genes, a specific pET sequence vector was aimed and constructed for use in the production of large amounts of the bioactive fusion protein production with tag linkage capability. In addition, the pET vector had a 6×His-Tag sequence, which normally does not affect the structure and activity of the recombinant protein, the g at the end of the C-tag in the expression structure corresponded to the codon of the fusion protein. This structure was consisted of six consecutive amino acids, allowing the recombinant fusion protein to be purified using resins. The use of the pET sequencing in the purification of the recombinant protein using the combined method of histidine tagging and purging significantly reduces the presence of contaminants in the final solution. Also, in this fusion protein construct, the intact protein sequence of *Bordetella pertussis*, *Corynebacterium diphtheria* and *Clostridium tetani* toxins final construct was used to analyze the of probable allergenicity reactions. Due to the side effects of using the whole proteins in vaccine production including local and systemic allergic reactions such as skin rash to anaphylactic shock in rare cases with high mortality rate, the use of bioinformatics solution should be considered in terms of cost effectiveness and speed (19, 22). Also, the use of a modified recombinant fusion protein can be considered in terms of reducing the side effects of chemicals such as formalin used in the inactivation process, which can be a potent local and systemic allergic substance. Finally, a fusion protein construct composed of immunogenic regions from three bacterial organisms in case DTP vaccine is novel potent, safe, cheap and broad-spectrum vaccine and may improve vaccine efficacy / effectiveness for prevention and eradication of the diphtheria, tetanus and pertussis infections in the future.

Acknowledgment

We thank for the sincere cooperation between the faculty of Medicine of Shahed University and Razi Vaccine and Serum Research Institute, Karaj. Iran.

Authors' Contribution

Study concept and design: Z.S, MM.R, M.N, M.N, M.E and M.B.

Data acquisition: Z.S and MM.R

Data analysis and interpretation: Z.S and MM.R

Drafting of the manuscript: Z.S and MM.R

Critical revision of the manuscript for important intellectual content: M.N, M.N, M.E and M.B.

Statistical analysis: Z.S and MM.R

Administrative, technical, and material support: Z.S, M.N and MM.R.

Ethics

Not applicable.

Conflict of Interest

The authors declare that there is no conflict of interest.

Grant Support

There was no grant support for this work.

Data Availability

The data that support the findings of this study are available on request from the corresponding author.

References

- Ogden SA, Ludlow JT, Alsayouri K. Diphtheria Tetanus Pertussis (DTaP) Vaccine. 2022. In: StatPearls. Treasure Island (FL): StatPearls Publishing. 2023.
- Prygiel M, Mosiej E, Górska P, Zasada AA. Diphtheria-tetanus-pertussis vaccine: past, current & future. *Future Microbiol.* 2022; 17:185-197.
- Liang JL, Tiwari T, Moro P, Messonnier NE, Reingold A, Sawyer M, Clark TA. Prevention of Pertussis, Tetanus, and Diphtheria with Vaccines in the United States: Recommendations of the Advisory Committee on Immunization Practices (ACIP). *MMWR Recomm Rep.* 2018;67(2):1-44.
- Howe BJ. Diphtheria-tetanus-acellular pertussis (DTaP) vaccine. *JAMA.* 1996;276(22):1803-4.
- Cherry JD. Comparative efficacy of acellular pertussis vaccines: an analysis of recent trials. *Pediatric Infection Disease J.* 1997; 16:90-96.
- Aminian M, Sivam S, Lee CW, Halperin SA, Lee SF. Expression and purification of a trivalent pertussis toxin-diphtheria toxin-tetanus toxin fusion protein in *Escherichia coli*. *Protein Expr Purif.* 2007; 51: 170-178.
- bocher Ph, stato H, satot,lotchi C. Neutralizing Antibodies and Immunoprotection against Pertussis and Tetanus Obtained by Use of a Recombinant PertussisToxin-Tetanus Toxin Fusion Protein infect and immu, 1994;62(2); 449-456.
- Torkashvand A, Bahrami.F, Adib. M, Ajdary S. Subcutaneous administration of a fusion protein composed of pertussis toxin and filamentous hemagglutinin from *Bordetella pertussis* induces mucosal and systemic immune responses. 2018; 21:753-759.
- Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, Lepore, R., Schwede, T. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* 2018; 46: 296-303
- Studer, G., Rempfer, C., Waterhouse, A.M., Gumienny, R., Haas, J., Schwede, T. QMEANDisCo - distance constraints applied on model quality estimation. *Bioinformatics.* 2020; 36: 1765-1771.
- Benkert, P., Biasini, M., Schwede, T. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics.* 2011;27: 343-350.
- Rinaudo CD, Telford JL, Rappuoli R, Seib KL J. Vaccinology in the genome era. *Clin Invest.* 2009;119(9):2515-25.
- Capecchi B, Serruto D, Adu-Bobie J, Rappuoli R, Pizza M. The genome evolution in vaccine research. *Curr Issues Mol Biol.* 2004;6(1):17-27.
- Ansari HR, Gajendra Ps Raghava. Identification of conformational B-cell Epitopes in an antigen from its primary sequence. 2010; 6:6.
- Kolaskar AS, Tongaonkar PS. A semi-empirical method for prediction of antigenic determinants on protein antigens; 1990;276(1-2):172-174.
- Grote A, Hiller K, Scheer M, Münch R, Nörtemann B, Hempel DC .et.al. a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic acids research*, 33(Web Server issue). 2005;526-531.
- Zarei S, Akhondi M, Zamani A.H, Zeraati H, Bonakdar S, Ghazanfari M et al . Short term reactogenicity of cellular DTP vaccine in 4-6 year old children in Tehran, Iran. *Payesh.* 2007;6(3).
- arami M, Ameri P, Bathaei J. et al. Adverse events following immunization with pentavalent vaccine: experiences of newly introduced vaccine in Iran. *BMC Immunol* 2017; 18: 42.
- Ranjbar M, Ebrahimi M, Shahsavandi S, Farhadi T, Mirjalili A, Tebianian M, et al. Novel Applications of Immuno-bioinformatics in Vaccine and Bio-product Developments at Research Institutes. *Archives of Razi Institute.* 2019;74(3):219-233.
- Bahadori Z, Shafaghi M, Sabzevari J, Madanchi H, Ranjbar MM, Mousavi SF, Shabani AA. Design, development, and assessment of a novel multi-peptide vaccine targeting PspC, PsaA and PhtD proteins of *Streptococcus pneumoniae*. *Int J Biol Macromol.* 2023;258(2):128924.
- Ruth E. Soria et al. Oral immunogenicity of tomato-derived sDPT polypeptide containing *Corynebacterium diphtheriae*, *Bordetella pertussis* and *Clostridium tetani* exotoxin epitope. 2011; 30:417-424.
- Rawat SS, Keshri AK, Kaur R, Prasad A. Immunoinformatics Approaches for Vaccine Design: A Fast and Secure Strategy for Successful Vaccine Development. *Vaccines (Basel).* 2023;11(2):221.