

In Silico Design and Analysis of TGF α L3-SEB Fusion Protein as “a New Antitumor Agent” Candidate by Ligand-Targeted Superantigens Technique

Abbas Ali Imani-Fooladi¹, Forough Yousefi², Seyed Fazloallah Mousavi², Jafar Amani¹

Abstract

Background: Bacterial superantigen Staphylococcal Enterotoxins (SEs), has stimulated polyclonal T cells irrespective of their antigen specificity, resulted a massive release of cytokines, and suggested that they could be assigned as a candidate of new antitumor agents. Recent attempts have done to specifically target superantigens towards tumors, subsequently Monoclonal antibodies and tumor-related ligands have employed as targeting molecules of superantigen for the preclinical treatment of different tumors. Here, we have evaluated TGF α L3-SEB fusion protein as a new antitumor candidate by genetically fusing the third loop of transforming growth factor alpha (TGF α L3) to Staphylococcal Enterotoxin type B.

Methods: An in silico techniques have launched to characterize the properties and structure of the protein, before initiating the experimental study, we have predicted physicochemical properties, structures, stability, MHC binding properties and ligand-receptor interaction of this chimeric protein by means of computational bioinformatics tools and servers.

Results: Our results have indicated codon adaptation index of *tgfa3-seb* fusion gene has increased from 0.5 in the wild type sequences to 0.85 in the chimeric optimized gene. The mfold data has shown the *tgfa3-seb* mRNA was stable enough for efficient translation in the new host. Based on Ramachandran plot TGF α L3-SEB has classified as a stable fusion protein. Our result has shown fusing of TGF α L3 in N-terminal of the TGF α L3-SEB construct, had no effects on MHC binding and subsequently superantigenic activity of SEB. Finally based on ligand-receptor docking the binding ability of TGF α L3 was strong enough to its receptor, so TGF α L3-SEB could be assigned as a new antitumor candidate in cancer immunotherapy.

Conclusion: Our results have proposed that TGF α L3-SEB was a stable fusion protein with proper affinity to its receptor that overexpressed in various human carcinomas, so it could generate potent immune response towards tumors.

Keywords: Enterotoxin type B; Growth Factor α ; Cancer Immunotherapy, in silico modeling

Please cite this article as: Imani Fooladi AA, Yousefi F, Mousavi SF, Amani J. In Silico Design and Analysis of TGF α L3-SEB Fusion Protein as “a New Antitumor Agent” Candidate by Ligand-Targeted Superantigens Technique. *Iran J Cancer Prev.* 2014; 7(3):152-64.

Introduction

Activation of the patients' immune system is one of several promising therapeutic methods for controlling cancer progression, because tumor cells have often avoided presenting their own antigens to T cells. So although tumors in cancer patients, which have treated by chemotherapy, surgery, and radiotherapy, regress, in many cases, they have also

often metastasized. One of the major goals of tumor immunotherapy was generating tumor-specific T cells that finally contributed to the tumors eradication. Superantigens (SAGs) were bacterial and viral proteins that could activate a large number of T cells irrespective of their antigen specificity, resulted in a massive release of cytokines from T cells and monocytes, they have increased the

1. Applied Microbiology Research center, Baqiyatallah University of Medical Sciences, Tehran, Iran
2. Dept. of Bacteriology, Pasteur Institute of Iran, Tehran, Iran

Corresponding Author:

Jafar Amani, PhD;
Assistant Professor of Molecular Genetics
Tel: (+98) 21 82482568
Email: Jafar.amani@gmail.com
Received: 22 Apr. 2014
Accepted: 25 June 2014
Iran J Cancer Prev. 2014; 3:152-64

antitumor activity of the immune system and prevent tumor growth and metastasis [1].

To date, there have been 21 identified SE and Staphylococcal-like Enterotoxin (SE 1) genes, including sea to see, seg to sev [2]. In this study, Among Sags, we have chosen the bacterial superantigen Staphylococcal Enterotoxin B (SEB) that was a potent inducer of cytotoxic T-cell activity and cytokine production in vivo [1]. Most anticancer agents, such as radiation and cytostatic drugs, have worked by affecting or preventing cell division. Because they were nonspecific, all dividing cells have affected the results in adverse side effects, that occur as a result of toxicities to normal tissues [3] So anticancer chemotherapeutics have often given at suboptimal doses, resulted the eventual failure of therapy, that has often accompanied by (go along with or cause) the drug resistance development, and tumor metastasis [4]. Several approaches for improving the selective toxicity of anticancer therapeutics, and reducing their side effect have been following at present, such as delivery of anti-neoplastic drugs to cancer cells, by associating the drugs with molecules (mAbs) that bound to antigens or ligands, and bound to receptors that are either uniquely expressed or overexpressed on the target cells relative to normal tissues. This has allowed a specific delivery of drugs to the cancer cells. Ligand-Targeted Therapeutics (LTTs) had advantages to mAbs, and then tumor-related ligand was less antigenic than mAbs, as well as non-antibody ligands were often readily available, inexpensive to manufacture and easy to handle [4], then facilitate drug penetration into solid tumors [5]. Generally, the targeted antigen or receptor should have a high density on the surface of the target cells [4] so we have chosen the Epidermal Growth Factor Receptor (EGFR), as a suitable receptor for the design of ligand-targeted therapeutics in cancer immunotherapy. EGFR was a 170kDa transmembrane protein consisting of an extracellular EGF binding domain, a short transmembrane region, and an intracellular domain with ligand-activated tyrosine kinase activity. EGFR could be activated by two ligands: Epidermal Growth Factor (EGF) and Transforming Growth Factor-alpha (TGF α). Ligand binding to EGFR has resulted in receptor homo- or hetero-dimerization (with one of the HER family of receptor tyrosine kinases) that has followed by auto phosphorylation of the tyrosine kinase domain, then phosphorylated tyrosine residues have served as binding sites for the recruitment of signal

transducers and activators of intracellular substrates. The phosphatidyl inositol 3' kinase pathway, and the Ras-Raf mitogen-activated protein kinase, and Akt pathway were the major signaling routes for the HER (human epidermal growth factor receptor) family, including EGFR. These pathways have controlled several important biological events, including cellular proliferation, angiogenesis and inhibition of apoptosis [6, 7]. Overexpression of EGFR protein has described in various human carcinomas including breast, head, neck, esophageal, gastric, pancreatic, colorectal, prostate, bladder, renal, ovarian and Non-Small Cell Lung Cancer (NSCLC) [8] and has generally reported as an adverse prognostic marker [9-11] Moreover, the degree of EGFR over-expression has associated with an advanced tumor stage, and then resistance to standard therapies.

The conjugate composed of ligand and superantigen has presumed to kill several types of tumor [8]. Human transforming growth factor alpha (hTGF α) was a native ligand co-overexpressed with its receptor EGFR in many human tumors; which had three isoform 1, 3 and 4, that have expressed in keratinocytes and tumor-derived cell lines. hTGF α has consisted of three loops, the third of which (TGF α L3) retained the binding ability to EGFR. In this study, after selecting TGF- α as a legend, for preventing ligand/receptor induced internalization, we have bound SEB superantigen to third loop of TGF α (TGF α L3) that retained the binding ability to EGFR. Moreover, in comparison to mAbs, TGF α L3 was presumably less antigenic, thereby maintaining a longer circulating half-life. These properties have enabled TGF α L3 to be an attractive targeting molecule for the superantigens, also which has only occurred in time of presence on the surface of the cells [8].

We have used in silico techniques to design construct, optimized for expression in suitable host, predicted physicochemical and structural properties and stability, then identified MHC binding and T cell epitopes to allow accelerating the strong antigenic and immune responses, and finally ligand-receptor interaction. An unguided experimental has searched for antigenic and immunogenic regions was basically laborious and resource intensive. The computational approaches could speed up the process, then with the potential for simplifying the evaluation presses to a great extent [12]. Hence, the novel TGF α L3-SEB fusion protein has determined as a candidate for cancer immunotherapy, could

Table 1. Prediction of binding affinity of TAP binder by TAPPred in TGF α L3-SEB fusion protein

Peptide Rank	Start Position	Sequence	Score	Predicted Affinity
1	45	LMENMKVLY	8.103	High
2	112	ANYYYQCYF	7.893	High
3	9	*VRCEHADLL	7.273	High
4	98	LADKYKDKY	7.183	High
5	70	LYFDLIYSI	7.078	High
6	152	DKYRSITVR	6.835	High
7	111	GANYYYQCY	6.744	High
8	220	SFWYDMMMPA	6.725	High
9	64	KSIDQFLYF	6.701	High
10	113	NYYYQCYFS	6.566	High

*There was only one additional TAP binding sequence in TGF α L3-SEB fusion protein, belonged to ligand part in comparison with SEB protein.

rapidly identified in silico, then have subjected to in vitro and in vivo confirmatory studies.

Materials and Methods

Protein retrieval and sequence analysis

The protein sequence of SEB and TGF α L3 protein has retrieved from Uniprot Knowledgebase data, have also used accession no. P01552 and P01135 respectively.

Design of the construct and gene optimization

Recombinant TGF α L3-SEB and SEB-TGF α L3 sequences have constructed by fusing the C-terminal of seb and the N-terminal of tgfa3 (TGF α L3-SEB) and N-terminal of seb and the C-terminal of tgfa3 (SEB-TGF α L3), have been using hydrophobic GGSGSGGG amino acid linker. To optimize the multiparameter chimeric gene, the in silico analysis has used online data bases such as Gene bank codon data base, the codon database, Swissprot reverse translation online tool [13, 14], and stand-alone softwares such as DNAsis MAX (Hitachi Software). After verification of the construct's properties by Gen-Script (NJ, USA), the chimeric gene has synthesized by ShineGene Molecular Biotech, Inc. (Shanghai, China)

MRNA structure prediction

The messenger RNA secondary structure of the chimeric gene has analyzed by the program mfold [15].

Primary structure prediction

For physicochemical characterization, theoretical isoelectric point (pI), molecular weight, total number of positive and negative residues, extinction coefficient, instability index, aliphatic index and Grand Average hydropathy (GRAVY) have computed using the Expasy ProtParam server [16].

Secondary structure prediction

GOR secondary structure prediction method version IV has employed for computing and analyzing the secondary structural features of TGF α L3-SEB fusion protein sequence [17].

3D structure prediction using homology approach

The 3D model of the recombinant TGF α L3-SEB protein has generated using the I-TASSER online server [18] which generates 3D models along with their confidence score (C-Score).

Energy minimization has determined by analysis of 3D structural stability of the chimeric protein using Swiss-PdbViewer [19]. Solvent accessibilities of the protein residues have evaluated with the online program ASA, [20].

Evaluation of model stability

After generating 3D model, energy minimization has performed by GROMOS96 force field in a Swiss-PdbViewer. Structural evaluation and stereo chemical analyses have performed by using ProSA-web, Z-scores and Procheck Ramachandran plot [21]. Furthermore, superimposition of query and template structure, and

Table 2. NetCTL-1.2 predictions using MHC supertype A1. Threshold 0.750000. TGF α L3-SEB fusion protein, Number of MHC ligands 15 identified. Number of peptides 256. There were no differences between epitopes identified in SEB protein (data not shown) and TGF α L3-SEB fusion protein.

Position	Sequence	aff	aff_rescale	cle	Tap	COMB
106	YVDVFGANY	0.7153	3.0371	0.7643	2.9680	3.3002
98	LADKYKDKY	0.5768	2.4489	0.4118	2.6660	2.6440
45	LMENMKVLY	0.5515	2.3418	0.9257	2.8760	2.6244
184	ELDYLTRHY	0.4247	1.8030	0.7895	2.5150	2.0472
111	GANYYYQCY	0.2489	1.0568	0.8893	2.5990	1.3202
234	DQSKYLMMY	0.2391	1.0153	0.8626	2.5750	1.2735
250	SKDVKIEVY	0.2221	0.9428	0.9397	2.7120	1.2194
179	KVTAQELDY	0.1841	0.7818	0.8676	3.2350	1.0737
78	IKDTKLGNY	0.1848	0.7845	0.6426	2.8750	1.0247
131	QTDKRKTCM	0.1995	0.8469	0.6038	0.0970	0.9423
199	LYEFNNSPY	0.1499	0.6364	0.9655	3.1500	0.9388
108	DVFGANYYY	0.1262	0.5360	0.9723	2.9060	0.8272
143	VTEHNGNQL	0.1543	0.6550	0.7051	0.8210	0.8019
63	VKSIDQFLY	0.1346	0.5715	0.3596	3.1730	0.7841

Table 3. MHC Restriction of CTL Epitope prediction by CTLpred based on Artificial Neural Network in TGF α L3.

Peptide Rank	Start Position	Sequence	Score	MHC Restriction
1	16	LLGGSGSGG	1.00	HLA-A3, HLA-Cw*0401 HLA-A*3301, HLA-A*6801
2	73	DLIYSIKDT	0.990	HLA-A*0201, HLA-B8 HLA-Cw*0401
3	83	LGNYDNVRV	0.990	HLA-B*51, HLA-Cw*0401

visualization of generated models have performed by using the Swiss-PdbViewer.

Prediction of cleavage sites

Proteasome cleavage sites of the chimeric protein have predicted by Netchop 3.1, MAPPP and PCPS. Peptides Binding affinity to TAP protein has computed by TAPPred [22].

Prediction of T-cell epitopes and MHC binding peptides affinity

The amino acid sequence has analyzed by using four web-based T-cell epitope prediction algorithms; NetCTL, SYFPEITHI (<http://www.syfpeithi.de/>), CTLPred and NetMHC [23].

Briefly, at first the chimeric protein has analyzed for MHC-presented epitopes and MHC-specific anchor and auxiliary motifs using NetCTL

[24], SYFPEITHI [25] and CTLPred [26]; then the NetMHC server, which has produced a neural network prediction of binding affinities for MHC [27].

Ligand-receptor Docking using Hex

The docking of TGF α L3 with TGFR has performed by using Hex [28], in order to investigate the protein-ligand interactions and investigate the application of the models for ligand binding potency prediction.

Results

Design and construction of chimeric gene

To evaluate the effect of TGF α L3 fusion on SEB in silico superantigenic activity, two sets of

VCHSGYVGVRCHEADLL **GGSGSGGG**ESQDPKPKDELHKASKFTGLMENMKVL YDDNHVSAINVKSIDQFLYFDLIYSIKD
 TKLGNYDNVRVEFKNKDLADKYKDKYVDVFGANYYYQCYFSKKTNDINSHQTDKRRKTCMYGGVTEHNGNQLDKYRSITVR
 VFEDGKNLLSFDVQTNKKKVTAQELDYLTRHYLVKNKKLYEFNNSPYETGYKFIENENSFWYDMMPAPGDKFDQSKYLM
 MYNDNKMVDSKDVKIEVYLTTKKK



ESQDPKPKDELHKASKFTGLMENMKVL YDDNHVSAINVKSIDQFLYFDLIYSIKD TKLGNYDNVRVEFKNKDLADKYKDK
 YVDVFGANYYYQCYFSKKTNDINSHQTDKRRKTCMYGGVTEHNGNQLDKYRSITVRVFEDGKNLLSFDVQTNKKKVTAQEL
 DYLTRHYLVKNKKLYEFNNSPYETGYKFIENENSFWYDMMPAPGDKFDQSKYLM MYNDNKMVDSKDVKIEVYLTTKKK
GGSGSGGGVCHSGYVGVRCHEADLL

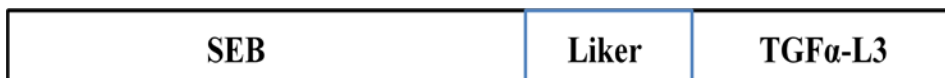


Figure 1. Sequence and Schematic model which has shown the construct of TGFαL3 and SEB bound together by the GGSGSGGG linker.

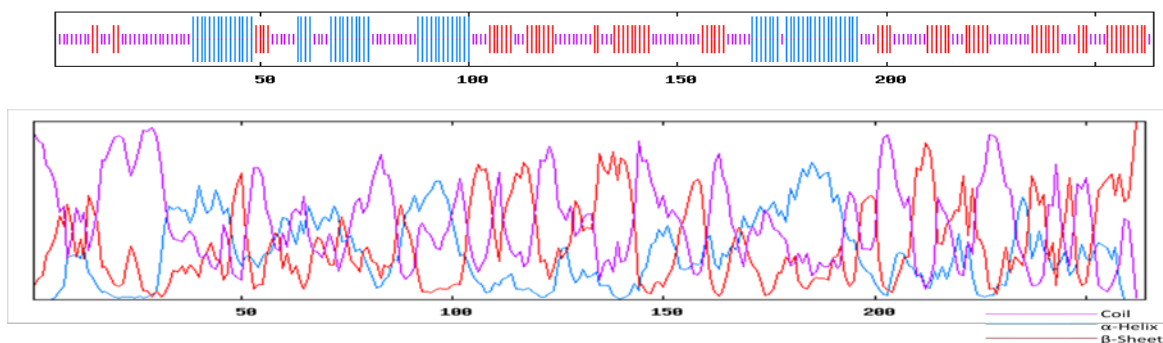


Figure 2. Graphical Representation of Secondary Elements in chimeric TGFαL3-SEB protein.

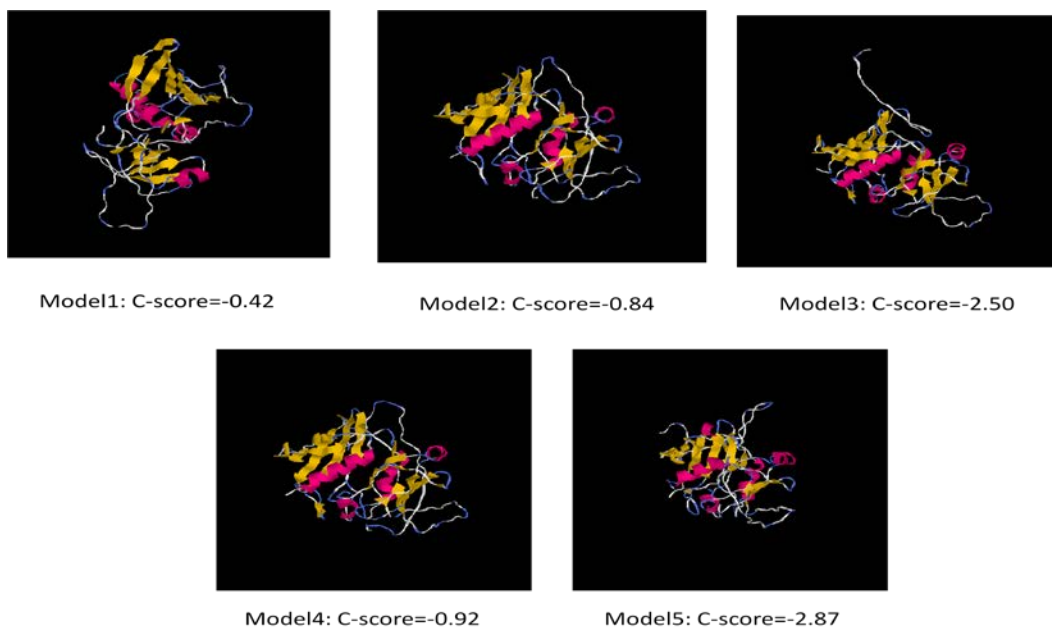


Figure 3. I-TASSER server has used to predict the tertiary structure of the chimeric protein, TGFαL3-SEB. The result has viewed by Swiss-PdbViewer.

amine-terminus of third loop of Transforming Staphylococcal Enterotoxin type B, 239 amino acids

Table 4. Predictions of MHC-binding peptide affinity for the SEB and TGF α L3-SEB construct by NetMHC version 3.0. Server using ANNs approximation. Strong binder threshold score 50 nM. Weak-binder threshold score 500 nM. (HLA-A0211, HLA-B1517, HLA-A8001, HLA-A0212, HLA-A0211, HLA-A2902, HLA-A2403). There were no differences between SEB and TGF α L3-SEB fusion protein in binding to MHC..

	Peptide	logscore	affinity(nM)	Binding Level
1	MMYNDNKMV	0.928	2	SB
2	KSIDQFLYF	0.879	3	SB
3	KVTAQELDY	0.856	4	SB
4	GLMENMKVL	0.852	4	SB
5	FLYFDLIYS	0.840	5	SB
6	QFLYFDLIY	0.829	6	SB
7	LYFDLIYSI	0.817	7	SB

Table 5. MHC Class-II Binding Peptide Prediction in TGF α L3-SEB fusion protein Results with 51 alleles query by Propred I online server. There were no differences between SEB (data not shown) and TGF α L3-SEB fusion protein in binding to MHC Class-II. (ALLELE: DRB1_0701, ALLELE: DRB1_1502, ALLELE: DRB1_0301, ALLELE: DRB1_1501, ALLELE: DRB1_0301, ALLELE: DRB1_0817, ALLELE: DRB1_0817, ALLELE: DRB1_1501).

Rank	Sequence	At position	Score
1	YRSITVRVF	153	6.4000
2	FLYFDLIYS	68	5.6000
3	MYNDNKMVD	240	5.5000
4	LMMYNDNKM	238	5.3800
5	LYFDLIYSI	69	5.2500
6	LVKNKKLYE	192	5.2000
7	YLVKNKKLY	191	5.0000
8	LGNYDNVRV	82	4.8000

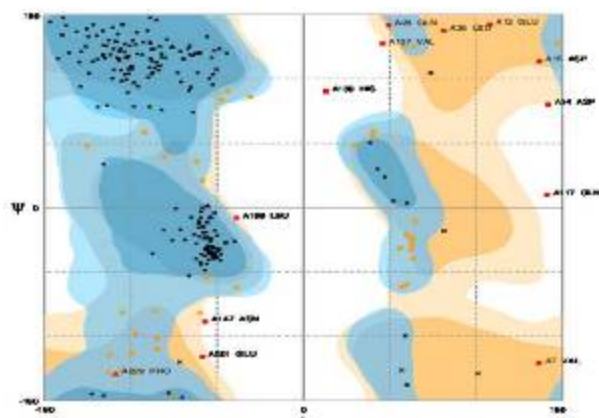


Figure 4. Evaluation of model stability has based on a Ramachandran plot.

(lacking 27 amino acids of signal sequence from the N-terminal of the protein) has selected and inserted one time in the carboxy-terminus of chimeric fusion protein to design TGF α L3-SEB, and then another time in the amino-terminus of chimeric fusion protein to design SEB-TGF α L3 construct. These two parts have joined by a linker consisting of 8 amino acids (GGSGSGGG). The amino acid composition of TGF α L3-SEB and SEB-TGF α L3 sequences have computed by using the tool CLC free Workbench.

Codon adaptation analysis of the wild type and optimized synthetic gene

Both the wild type and the synthetic chimera have analyzed for their codon bias and GC content. The optimized gene has shown a codon bias for E.coli, bacterial expression host, and has contained

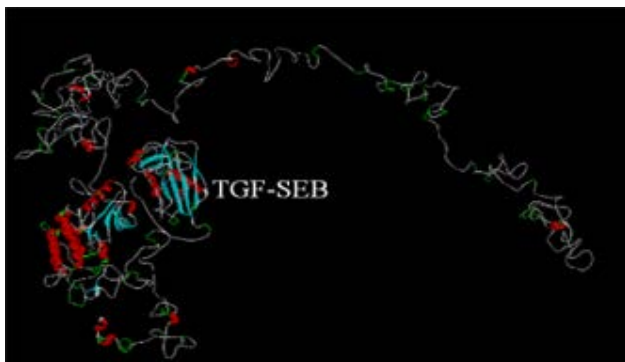


Figure 5. Docking of TGF α L3 with TGFR using Hex. To examine the protein-ligand interactions the models for ligand binding potency has predicted.

no rarely used codon. This has also reflected by the codon adaptation index (CAI), which was a measurement of the relative adaptiveness of the codon usage of a gene, that in comparison with the codon usage of highly expressed genes [29, 30]. The chimeric gene has shown a CAI of 0.85, that has compared to the wild type gene, which was only 0.5. The overall GC content has reduced from 45.83 to 44.06%, which could increase the overall stability of mRNA from the synthetic gene.

Furthermore, the necessary restriction enzyme sites (*Bam*HI and *Hind*III) have introduced at the ends of the sequence for cloning purpose.

MRNA structure prediction

A genetic algorithm-based RNA secondary structure prediction has combined with comparative sequence analysis to determine the potential folding of the chimeric gene. The 5' terminus of the gene has folded in the way typical of all bacterial gene structures. The minimum free energy for secondary structures that has formed by RNA molecules has also predicted. All 29 structural elements that have obtained in this analysis have revealed folding of the RNA construct. The data has shown the mRNA was stable enough for efficient translation in the new host.

Primary structure prediction

ProtParam has used to find out the physicochemical properties of a protein sequence. The physicochemical properties of the protein have revealed the number of amino acids to be 264, molecular weight: 30708.5 and theoretical isoelectric pointed as 7.72. The maximum number of amino acids, which have been present in the sequence, has been the Lysine (12.5%), and the least was the Pyl (O) and Sec (U) (0.0%).

The total number of positively charged residues (Arg+ Lys) was 39, and total number of negatively charged residues (Asp+ Glu) was 38. The instability index of the protein has computed to be 30.96. This fact has classified the protein as stable protein. The N-terminal of the sequence has considered as the F (Phe). Therefore estimated half-life was; 100 hours (mammalian reticulocytes, in vitro), >20 hours (yeast, in vivo), and finally >10 hours (Escherichia coli, in vivo). The grand hydropathicity has calculated to be -0.821.

Protein secondary structure prediction

The secondary structure of the chimeric protein has predicted by online software and random coils has found to be frequent (46.59%), followed by Extended strand (28.03%) and alpha helix has found to be less frequent (25.38%). This was graphically represented in Figure 2.

Tertiary structural prediction for the chimeric protein

Chimeric protein 3D models, have been produced by i-Tasser (Figure 3), uploaded to the Swiss-PdbViewer server to depict the tertiary structural illustrations [31].

Evaluation of model stability

The profile of energy minimization has calculated by spdbv (Swiss-PdbViewer) – 6107.159 Kcal/mol indicating that the recombinant protein had acceptable stability. Furthermore, the structural stability of the chimeric protein has confirmed based on data generated by a Ramachandran plot (Figure 4).

Solvent accessibility prediction

The solvent accessibility distributions have characterized using the major hydrophobic and polarity properties of residual patterns. These patterns have shown that the mean residue accessible surface area (ASA) have given a high solvent accessibility value, approximately fifty percent (Data have not shown).

Prediction of the cleavage site

Cleavage site analysis on the construct protein has performed using Net Chop server, an improved neural network training strategy. This server has produced neural network predictions for cleavage sites of the human proteasome using two different methods; C-term 3.0 and 20S 3.0 [32]. C-term 3.0 has used here. The Net Chop neural network-based method was the best presently-available system for cleavage site predictions. The new version of Net Chop has predicted approximately 75% of cleavage

sites correctly with false positives near 15%. Cleavage sites on the construct protein have analyzed with Net Chop (data not shown). Prediction of binding affinity of TAP binder in TGF α L3-SEB fusion protein has done by TAPPred that is a SVM based quantitative method for predicting peptide TAP affinity binding (Table 1)

Prediction of T-cell epitopes

NetCTL 1.2 server has predicted CTL epitopes in the chimeric protein sequence. The server has predicted CTL epitopes, restricted to 12 MHC class I supertypes using ANNs [24]. The scores from the individual prediction methods have integrated, and thresholds for the integrated scores of each peptide have translated into sensitivity and specificity values (Table 2). Also the SYFPEITHI epitope prediction algorithm has used. This server has allowed quantification of the ligation strength to a defined HLA type for a sequence of amino acids, and the probability of the peptide being processed and presented has given in order to predict T-cell epitopes [25]. The scoring system of SYFPEITHI has evaluated each amino acid in the peptides. The maximum score for HLA-A*0201peptides was 27 (data have not shown). CTLPred, a direct method for prediction of CTL epitopes, has also used. This method has based on elegant machine learning techniques like an ANN and support vector machine [26]. The scores of CTLPred-predicted epitopes for the chimeric protein have shown in Table 3. The default cutoff score was 0.51, in which the sensitivity and specificity of prediction methods were highly similar.

MHC binding peptides affinity

T lymphocytes have played a central role in the generation of a protective immune response in many microbial infections. The binding strength of T cell epitopes to major histocompatibility complex (MHC or HLA) molecules was a key determinant in T cell epitope immunogenicity. This has allowed the epitopes with higher binding affinities, to be more likely to be displayed on the surface of the cells, where they have recognized by their corresponding T cell receptor (TCR) [12].

NetMHC 3.2 server has predicted peptide binding to a number of different HLA alleles using ANNs.

For ANN prediction, values have given in nM IC50 values so that high-binding peptides had IC50 values below 50 nM, and weakly-binding peptides had IC50 values below 500 nM (22). The results

have summarized in Table 4. MHC Class-II Binding Peptide Prediction in TGF α L3-SEB fusion protein (with 51 alleles query) has done by ProPred I online server (Table 5).

Ligand docking

Docking has performed using Hex server. This server could calculate protein ligand docking. We have uploaded a pair of epidermal growth factor receptor and TGF α L3-SEB fusion protein as a ligand structures in PDB format in Hex server. Default parameters have used for carrying out the jobs. To be able to analyze the docking, the e-values have obtained using the Hex software [28]. The docking process has been more efficient, related to the negative e-value. When we have viewed the visualization tool like SPV BV, the docking between receptors of proteins and the ligand could be clearly observed as shown in Figure 5.

Discussion

The Staphylococcal SAGs were potent T cells mitogens (27), Antitumor activity of superantigens have proven in many studies [1, 8, 33-35] and Staphylococcal Enterotoxins, especially type B (SEB), were classic models of superantigens (SAGs) [1]. So in this study we have chosen SEB as an antitumor agent. Furthermore, avoiding the side effects that have occurred as a result of toxicities to normal tissues, we have brought the SEB on the surface of tumor cells by ligand-targeted technique. Ligand-targeted therapy has made possible tumor specificity, and limited toxicity, and has shown promise in the development of cancer novel therapies. It could carry higher doses of a drug to the tumor tissue and might overcome obstacles presented by cytotoxic chemotherapy [5].

Since EGFRs have over-expressed in a variety of human tumor cells, including breast, head, neck, gastric, colorectal, esophageal, prostate, bladder, renal, pancreatic, ovarian and Non-Small Cell Lung Cancer (NSCLC) Moreover, the degree of EGFR over-expression has associated with an advanced tumor stage and resistance to standard therapies [8], we have selected its ligand (TGF α) to fuse with SEB by genetically fusing the third loop of transforming growth factor alpha (TGF α L3) to Staphylococcal Enterotoxin B as a new antitumor candidate. Due to the limitations in experimental methods for determining binary interactions and structure determination of protein complexes, the request has existed for computational models to fill the

increasing gap between genome sequence information and protein annotation, so before starting experimental study by the aim of in-silico techniques, we have predicted physicochemical properties, structures, stability, MHC binding properties and ligand-receptor interaction of this chimeric protein by means of computational bioinformatics tools and servers. Here the recombinant TGF α L3-SEB and SEB-TGF α L3 sequences have constructed by fusing the C-terminal of seb and the N-terminal of tgfa3 (TGF α L3-SEB) and N-terminal of seb and the C-terminal of tgfa3 (SEB-TGF α L3) using hydrophobic GGSGSGGG amino acid linker.

The folding of two structures (TGF α L3-SEB and SEB-TGF α L3) has analyzed and the TGF α L3-SEB construct have shown that TGF α L3 was more accessible and has not hidden in SEB structure, so we have continued our study on TGF α L3-SEB fusion protein. In silico studies have confirmed efficient transcriptional and translational, as well as the quality expression of the proposed construct in host expression vectors. Codon Adaptation Index (CAI) was the major factor that has used for gene optimization (24), with a range of 0-1, and an ideal value of 1.0. Since our objective was to design a fusion protein that expressed in *Escherichia coli* as a host expression vector so codon usage table of *Escherichia coli* has selected for the back-translation of the sequence and optimal expression of the construct. In our gene CAI index has increased from 0.5 in the wild type sequences to 0.85 in the chimeric optimized gene. Moreover, the overall GC content has reduced from 45.83 to 44.06%, which should increase the overall stability of mRNA from the synthetic gene. In addition, the required restriction enzyme sites have added to the ends of the designate gene for future assays. Codon optimization has given us assurance that synthetic construct expressed well in desired host vector.

The mRNA structure has optimized based on low ΔG and energy of the start codon. This character could help ribosome binding and translation initiation. For prediction of RNA secondary structure, a genetic algorithm-based RNA secondary structure prediction has combined with comparative sequence analysis to determine the potential folding of the chimeric gene. The 5' terminus of the gene has folded in the way typical of all bacterial gene structures. The minimum free energy for secondary structures has formed by RNA molecules have also predicted. The messenger RNA secondary structure

of the chimeric gene has analyzed by the program mfold with the parameters: Linear RNA folding at 5%, window = 12, max folds = 50. All 29 structural elements have obtained in this analysis have revealed folding of the RNA construct at 37°C with Initial ΔG ranging from -225.00 to -214.00 Kcal/mol. The best structure that had $\Delta G = -225.00$ Kcal/mol. The data has shown the mRNA was stable enough for efficient translation in the new host.

ProtParam [16] has used to find out the physicochemical properties of a protein sequence. The results of primary structure analysis have suggested that TGF α L3-SEB fusion protein was hydrophilic in nature due to the presence of high polar residues content. The presence of 4 Cys residues in TGF α L3-SEB has indicated the presence of disulphide bridges (SS bonds) in this fusion protein. Moreover, the primary structure analysis has suggested that the average molecular weight of TGF α L3-SEB has calculated 30.708 kDa. Isoelectric point (pI) was the pH at which the surface of protein has covered with charge, but net charge of the protein is zero. At pI proteins were stable and compact. The computed pI value of TGF α L3-SEB was 7.72 (pI >7), that has indicated that this fusion protein was basic in character. The computed isoelectric point (pI) would be useful for developing buffer systems for purification by isoelectric focusing method. Although ExPASy's ProtParam has computed the extinction coefficient for a range of (276, 278, 279, 280 and 282 nm) Extinction coefficient of TGF α L3-SEB at 280 nm was 38530 M⁻¹ cm⁻¹ with respect to the high concentration of Cys, Trp and Tyr, indicated that this fusion protein could be analyzed using UV spectral methods. Both of the computed protein concentration, and extinction coefficients could help in the quantitative study of protein-protein and protein-ligand interactions in solution. The bio computed half-life of most of the TGF α L3-SEB was greater than 10 h. On the basis of instability index ExPASy's ProtParam have classified the TGF α L3-SEB fusion protein as stable (Instability index < 40). The Aliphatic Index (AI) which has defined as the relative volume of a protein occupied by aliphatic side chains (A, V, I and L) has regarded as a positive factor for the increase of thermal stability of globular proteins. The lower thermal stability of TGF α L3-SEB was indicative of a more flexible structure with comparison to very high aliphatic index that has inferred stability for a wide range of temperature. Grand Average hydropath (GRAVY) Index of

TGF α L3-SEB was -0.821. The very low GRAVY index of this fusion protein infers that TGF α L3-SEB could result in a better interaction with water. The secondary structural analysis of the protein has done with the help of GOR IV program (Figure 2) and random coil has found to be most frequent (46.59%), followed by extended strand (Ee) (28.03%) alpha helix that has found to be less frequent (25.38%).

The very high coil structural content of TGF α L3-SEB (46.59%) was due to the rich content of more flexible glycine and hydrophobic proline amino acids. Proline had a special property of creating kinks in polypeptide chains and disrupting ordered secondary structure.

The three-dimensional (3D) structure details of proteins were of major importance in providing insights into their molecular functions. The three-dimensional model of the recombinant TGF α L3-SEB protein has generated using the I-TASSER online server [18] which generates 3D models along with their confidence score (C-Score). Five models have generated by this server with C-Scores: -0.42, -0.84, -0.92, -2.50, -2.87 respectively, among the 5 models, model 1 has selected for further analysis as it contained the highest C-Score. After generating 3D model, structural evaluation and stereochemical analyses have performed using Procheck Ramachandran plot [21]. Energy minimization has determined by analysis of 3D structural stability of the chimeric protein using Swiss-PdbViewer.

The percentage of residues was 80.5% favored region, 14.5% allowed, and 5.0% in outlier region so evaluation of model stability by Ramachandran plot have shown that most residues of the chimeric model were in a stable zone. The model has analyzed by different protein analysis programs including PROCHECK for the evaluation of the Ramachandran plot quality.

CTLs distinguish small peptides eight to ten amino acids long. These epitope peptides have generated by the proteasome system. Protease was responsible for intracellular protein degradation. The proteasome has produced the exact C-terminus of CTL epitopes and the N-terminus with a possible extension [36]. CTL responses could be reduced if the epitopes have destroyed by proteasomes; therefore, prediction of proteasome cleavage sites was valuable for identification of potential immunogenic regions in the chimeric protein. Based on these rules we have designed the chimeric protein, and then predicted its proteasome cleavage sites using web-based software. The result has

shown that the highest-scored cleavage positions have located at overall the whole fusion protein (Data not shown).

Prediction of binding affinity of TAP binder in TGF α L3-SEB fusion protein has done by TAPPred. There was only one additional TAP binding sequence in TGF α L3-SEB fusion protein belongs to ligand part in comparison with SEB protein (Table 1). NetCTL 1.2 server has predicted CTL epitopes in the chimeric protein sequence. The server has predicted CTL epitopes restricted to 12 MHC class I super types using ANNs [24]. The scores from the individual prediction methods have integrated, and thresholds for the integrated scores of each peptide have translated into sensitivity and specificity values (Table 2). 15 same MHC ligands have identified in both SEB and TGF α L3-SEB fusion protein by this server. Also the SYFPEITHI epitope prediction algorithm has used. This server has allowed quantification of the ligation strength to a defined HLA type for a sequence of amino acids, and the probability of the peptide being processed and presented has given in order to predict T-cell epitopes [19]. Because of highly polymorphic nature of MHC, different patients typically have bounded different repertoires of peptides; hence it was crucial to identify the optimal set of peptides for a vaccine, given constraints such as MHC allele probabilities in the target population and maximum number of selected peptides. It has investigated that the most common HLA in the general population is HLA-A*0201, which accounts for 30–40% of the major ethnicities [12]. The scoring system of SYFPEITHI has evaluated each amino acid in the peptides. The maximum score for HLA-A*0201 peptides was 36 and the maximum scores for epitopes of both SEB and TGF α L3-SEB chimeric protein was 27 (data not shown). CTLPred, a direct method for prediction of CTL epitopes, has also used. This method has based on elegant machine learning techniques like an ANN and support vector machine [26]. The scores of CTLPred-predicted epitopes for the chimeric protein have shown in Table 3. The default cutoff score was 0.51 (at which the sensitivity and specificity of prediction methods were highly similar). Superantigens (SAGs) were microbial proteins with the capacity to activate a large fraction of T cells. The cellular receptors for SAGs were major histocompatibility complex (MHC) class II molecules and T-cell antigen receptors (TCR). SAGs could bind to the TCR b subunit, and then could activate T cells independently of their CD4 or CD8

phenotype, when presented by MHC class II molecules. Activated T cells have secreted a variety of cytokines, such as TNF α , INF γ , IL-1, IL-2, IL-6, IL-8 and IL-12 [8].

In this assay, to determine whether fusing of TGF α L3 in N-terminal of the TGF α L3-SEB construct has negatively affected on MHC binding, and subsequently super antigenic activity, we have predicted the binding affinity of "TGF α L3-SEB fusion protein" to MHC, in comparison to SEB, as a classic superantigen. NetMHC 3.2 server has predicted peptide binding to a number of different HLA alleles using artificial neural networks (ANNs) trained on C terminals of known epitopes. For ANN analysis, predicted MHC/peptide binding was a log transformed value, has related to the IC50 values in nM units so that high-binding peptides had IC50 values below 50 nM, and weakly-binding peptides had IC50 values below 500 nM [27]. Seven same peptide sequences with high log score have identified as strong MHC binder in both SEB and TGF α L3-SEB fusion protein. These peptides had strong binding affinity to HLA-A0211, HLA-B1517, HLA-A8001, HLA-A0212, HLA-A0211, HLA-A2902, HLA-A2403 alleles. These MHC binding peptides were sufficient for eliciting the desired immune response. The results have summarized in Table 4. Also ProPred, a graphical web tool for predicting class II binding regions in antigenic protein sequences has also accessed by selecting all the 51 alleles present in the tool. The sequence in single letter amino acid code has given as input by using default parameters of the server for the prediction of class-II epitopes. Eight same peptide sequences in SEB and TGF α L3-SEB fusion protein with the highest score in binding to MHC II alleles (ALLELE: DRB1-0701, ALLELE: DRB1-1502, ALLELE: DRB1-0301, ALLELE: DRB1-1501, ALLELE: DRB1-0301, ALLELE: DRB1-0817, ALLELE: DRB1-0817, ALLELE: DRB1-1501) have tabulated in Table 5.

Our result have shown not only fusing of TGF α L3 in N-terminal of the TGF α L3-SEB construct, had no effects on MHC binding and subsequently superantigenic activity, but also based on the prediction results, the selected epitopes of our chimeric construct also have shown high-affinity binding to MHC molecules, and acceptable sensitivity and specificity have been recognized by CTLs. Epitope binding to MHC and recognition of such complexes (epitope/MHC) by CTLs was a

critical step for inducing a significant immune response.

To investigate whether the third (TGF α L3) of human transforming growth factor alpha (hTGF α) that was a native ligand co-overexpressed with its receptor EGFR in many human tumors, has retained its binding ability to bring SEB to tumors over expressing EGFR, we have checked its binding ability by ligand-receptor docking. Molecular Docking has performed using Hex server. Hex was an interactive molecular graphics program for calculating and displaying feasible docking modes of pairs of protein and DNA molecules and also calculates protein ligand docking, assuming the ligand was rigid, and it could superpose pairs of molecules using only knowledge of their 3D shapes. Hex was still one of the few docking programs which have built-in graphics to view the results. Also, it was the first protein docking program to be able to use modern graphics processor units (GPUs) to accelerate the calculation. The Hex software has given corresponding evaluates for each docking. More negative the evaluate more efficient was the docking [28]. TGF α L3 has shown high affinity towards the EGFR. It has given an e-value of -119.96 which was an acceptable e-value for docking results. Our result has shown the binding ability of TGF α L3 was strong enough to its receptor, so TGF α L3-SEB could be a new antitumor candidate in cancer immunotherapy.

Conclusion

Multiple different approaches have been used to activate the immune system against breast cancer. Here we have evaluated the ability of TGF α L3-SEB fusion protein as a new antitumor candidate. Since it was important to establish the structure-function relation of TGF α L3-SEB fusion protein before starting experimental studies, the TGF α L3-SEB fusion protein has analyzed by various tools and softwares.

Our results propose TGF α L3-SEB was a stable fusion protein with proper affinity to its receptor that overexpress in various human carcinomas so it could generate potent immune response towards tumors.

Acknowledgment

This work has supported by Baqiyatallah University of Medical Sciences, Applied Microbiology Research Center.

Conflict of Interest

The authors have declared that they had no competing interest.

Authors' Contribution

Authors have contributed equally.

References

1. Fooladi AA, Sattari M, Hassan ZM, Mahdavi M, Azizi T, Horii A. In vivo induction of necrosis in mice fibrosarcoma via intravenous injection of type B staphylococcal enterotoxin. *Biotechnology letters*. 2008; 30(12): 2053-9.
2. Ortega E, Abriouel H, Lucas R, Galvez A. Multiple roles of *Staphylococcus aureus* enterotoxins: pathogenicity, superantigenic activity, and correlation to antibiotic resistance. *Toxins*. 2010; 2(8): 2117-31.
3. Sundstedt A, Celander M, Ohman MW, Forsberg G, Hedlund G. Immunotherapy with tumor-targeted superantigens (TTS) in combination with docetaxel results in synergistic anti-tumor effects. *International immunopharmacology*. 2009; 9(9): 1063-70.
4. Allen TM. Ligand-targeted therapeutics in anticancer therapy. *Nature reviews Cancer*. 2002; 2(10): 750-63.
5. Wu H-C, Chang D-K, Huang C-T. Targeted-therapy for cancer. *Journal of cancer molecules*. 2006; 2(2): 57-66.
6. Bhargava R, Gerald WL, Li AR, Pan Q, Lal P, Ladanyi M, et al. EGFR gene amplification in breast cancer: correlation with epidermal growth factor receptor mRNA and protein expression and HER-2 status and absence of EGFR-activating mutations. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc*. 2005; 18(8): 1027-33.
7. Meche A, Cimpean AM, Raica M. Immunohistochemical expression and significance of epidermal growth factor receptor (EGFR) in breast cancer. *Romanian journal of morphology and embryology = Revue roumaine de morphologie et embryologie*. 2009; 50(2): 217-21.
8. Xu Q, Zhang X, Yue J, Liu C, Cao C, Zhong H, et al. Human TGF α -derived peptide TGF α L3 fused with superantigen for immunotherapy of EGFR-expressing tumours. *BMC biotechnology*. 2010; 10: 91.
9. Dua R, Zhang J, Nhonthachit P, Penuel E, Petropoulos C, Parry G. EGFR over-expression and activation in high HER2, ER negative breast cancer cell line induces trastuzumab resistance. *Breast cancer research and treatment*. 2010; 122(3): 685-97.
10. Grupka NL, Lear-Kaul KC, Kleinschmidt-DeMasters BK, Singh M. Epidermal growth factor receptor status in breast cancer metastases to the central nervous system. Comparison with HER-2/neu status. *Archives of pathology & laboratory medicine*. 2004; 128(9): 974-9.
11. Aziz SA, Pervez S, Khan S, Kayani N, Rahbar MH. Epidermal growth factor receptor (EGFR) as a prognostic marker: an immunohistochemical study on 315 consecutive breast carcinoma patients. *JPMMA The Journal of the Pakistan Medical Association*. 2002; 52(3): 104-10.
12. Baloria U, Akhoun BA, Gupta SK, Sharma S, Verma V. In silico proteomic characterization of human epidermal growth factor receptor 2 (HER-2) for the mapping of high affinity antigenic determinants against breast cancer. *Amino acids*. 2012; 42(4): 1349-60.
13. Stothard P. The Sequence Manipulation Suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques*. 2010; 28(6): 1102-4.
14. Ghasemi A, Ranjbar R, Amani J. In silico analysis of chimeric TF, Omp31 and BP26 fragments of *Brucella melitensis* for development of a multi subunit vaccine candidate. *Iranian Journal of Basic Medical Sciences*. 2014; 17(3): 173-80.
15. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic acids research*. 2003; 31(13): 3406-15.
16. Amala S. In silico Analysis and 3D Modeling of ASAH1 Protein in Farber Lipogranulomatosis. *Advanced Biotech*. 2010; 10(6): 6-8.
17. Garnier J, Gibrat JF, Robson B. *Methods in Enzymology*. Ed RFD, editor1996.
18. Zhang Y. I-TASSER server for protein 3D structure prediction. *BMC bioinformatics*. 2008; 9: 40.
19. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis*. 1997; 18(15): 2714-23.
20. Ahmad S, Gromiha M, Fawareh H, Sarai A. ASAVIEW: database and tool for solvent

accessibility representation in proteins. *BMC bioinformatics*. 2004; 5: 51.

21. Laskowski RA, Macarthur MW, Moss DS, Thornton JM. {PROCHECK}: a program to check the stereochemical quality of protein structures. *J Appl Cryst*. 1993; 26: 283-91.

22. Bhasin M, Raghava GP. Prediction of CTL epitopes using QM, SVM and ANN techniques. *Vaccine*. 2004; 22(23-24): 3195-204.

23. Larsen MV, Lundegaard C, Lamberth K, Buus S, Brunak S, Lund O, et al. An integrative approach to CTL epitope prediction: a combined algorithm integrating MHC class I binding, TAP transport efficiency, and proteasomal cleavage predictions. *Eur J Immunol*. 2005; 35(8): 2295-303.

24. Larsen MV, Lundegaard C, Lamberth K, Buus S, Lund O, Nielsen M. Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC bioinformatics*. 2007; 8: 424.

25. Rammensee H, Bachmann J, Emmerich NP, Bachor OA, Stevanovic S. SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics*. 1999; 50(3-4): 213-9.

26. Buus S, Lauemoller SL, Worning P, Kesmir C, Frimurer T, Corbet S, et al. Sensitive quantitative predictions of peptide-MHC binding by a 'Query by Committee' artificial neural network approach. *Tissue antigens*. 2003; 62(5): 378-84.

27. Yaraguppi DA, Udupudi BB, Patil LR, Hombalimath VS, Shet AR. In-silico analysis for predicting protein ligand interaction for snake venom protein. *Journal of Advanced Bioinformatics Applications and Research*. 2012; 3(3): 345-56.

28. 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. *Theoretical biology & medical modelling*. 2009; 6: 28.

29. Nazarian S, Mousavi Gargari SL, Rasooli I, Amani J, Bagheri S, Alerasool M. An in silico chimeric multi subunit vaccine targeting virulence factors of enterotoxigenic *Escherichia coli* (ETEC) with its bacterial inbuilt adjuvant. *J Microbiol Methods*. 2012; 90(1): 36-45.

30. Edwards YJ, Cottage A. Bioinformatics methods to predict protein structure and function. A practical approach. *Molecular biotechnology*. 2003; 23(2): 139-66.

31. Nielsen M, Lundegaard C, Lund O, Kesmir C. The role of the proteasome in generating cytotoxic T-cell epitopes: insights obtained from improved predictions of proteasomal cleavage. *Immunogenetics*. 2005; 57(1-2): 33-41.

32. Perabo FG, Willert PL, Wirger A, Schmidt DH, Wardelmann E, Sitzia M, et al. Preclinical evaluation of superantigen (staphylococcal enterotoxin B) in the intravesical immunotherapy of superficial bladder cancer. *International journal of cancer Journal international du cancer*. 2005; 115(4): 591-8.

33. Perabo FG, Willert PL, Wirger A, Schmidt DH, Von Ruecker A, Mueller SC. Superantigen-activated mononuclear cells induce apoptosis in transitional cell carcinoma. *Anticancer research*. 2005; 25(5): 3565-73.

34. Xu M, Wang X, Cai Y, Zhang H, Yang H, Liu C, et al. An engineered superantigen SEC2 exhibits promising antitumor activity and low toxicity. *Cancer immunology, immunotherapy : CII*. 2011; 60(5): 705-13.

35. Eggers M, Boes-Fabian B, Ruppert T, Kloetzel PM, Koszinowski UH. The cleavage preference of the proteasome governs the yield of antigenic peptides. *The Journal of experimental medicine*. 1995; 182(6): 1865-70.