Research in Pharmaceutical Sciences, April 2023; 18(2): 159-176
Received: 17-05-2022
Peer Reviewed: 22-08-2022
Revised: 10-11-2022
Accepted: 06-12-2022
Published: 19-01-2023

School of Pharmacy & Pharmaceutical Sciences
Isfahan University of Medical Sciences

Original Article

A comparative study of the arazyme-based fusion proteins with various ligands for more effective targeting cancer therapy: an *in-silico* analysis

Rezvan Mehrab[1], Hamid Sedighian[2], Fattah Sotoodehnejadnematalahi[1], Raheleh Halabian[2], and Abbas Ali Imani Fooladi[2,*]

[1]Department of Biology, Science and Research Branch, Islamic Azad University, Tehran, I.R. Iran.
[2]Applied Microbiology Research Center, Systems Biology and Poisonings Institute, Baqiyatallah University of Medical Sciences, Tehran, I.R. Iran.

Abstract

Background and purpose: Recently, the use of immunotoxins for targeted cancer therapy has been proposed, to find new anticancer drugs with high efficacy on tumor cells with minimal side effects on normal cells. we designed and compared several arazyme (AraA)-based fusion proteins with different ligands to choose the best-targeted therapy for interleukin 13 receptor alpha 2 (IL13Rα2)-overexpressed cancer cells. For this purpose, IL13Rα2 was selected as a receptor and IL13 and IL13.E13K were evaluated as native and mutant ligands, respectively. In addition, Pep-1 and A2b11 were chosen as the peptide ligands for targeted cancer therapy.

Experimental approach: Several bioinformatics servers were used for designing constructs and optimization. The structures of the chimeric proteins were predicted and verified by I-TASSER, Q-Mean, ProSA, Ramachandran plot, and Verify3D program. Physicochemical properties, toxicity, and antigenicity were predicted by ProtParam, ToxinPred, and VaxiJen. HawkDock, LigPlot^{+}, and GROMACS software were used for docking and molecular dynamics simulation of the ligand-receptor interaction.

Findings/Results: The *in silico* results showed AraA-A2b11 has higher values of confidence score and Q-mean score was obtained for high-resolution crystal structures. All chimeric proteins were stable, non-toxic, and non-antigenic. AraA-(A(EAAAK)$_4$ALEA(EAAAK)$_4$A)$_2$-IL13 retained its natural structure and based on ligand-receptor docking and molecular dynamic analysis, the binding ability of AraA-(A(EAAAK)$_4$ALEA(EAAAK)$_4$A)$_2$-IL13 to IL13Rα2 was sufficiently strong.

Conclusion and implications: Based on the bioinformatics result AraA-(A(EAAAK)$_4$ALEA(EAAAK)$_4$A)$_2$-IL13 was a stable fusion protein with two separate domains and high affinity with the IL13Rα2 receptor. Therefore, AraA-(A(EAAAK)$_4$ALEA(EAAAK)$_4$A)$_2$-IL13 fusion protein could be a new potent candidate for target cancer therapy.

Keywords: Arazyme, Cancer; *In silico*; Interleukin 13; Targeting therapy; Peptide ligand.

INTRODUCTION

An applied approach in the clinical microbiology field is the study of the effect of microbial metabolites, especially microbial toxins and enzymes for the treatment of various types of cancer (1,2). Moxetumomab pasudotox is the generic name of an anticancer drug, which is a fusion protein consisting of the catalytic fragment of *Pseudomonas* exotoxin-A (PE38) and the fragment variable (Fv) portion of the anti-CD22 antibody, approved by the United States Food and Drug Administration (FDA) for the hairy cell leukemia treatment in 2018 (3).

Access this article online	
	Website: http://rps.mui.ac.ir
	DOI: 10.4103/1735-5362.367795

*Corresponding author: A.A. Imani Fooladi
Tel & Fax: +98-2188068924
Email: imanifouladi.a@bmsu.ac.ir

Another drug for which the FDA has made recommendations specifically for additional clinical/statistical data and analysis, in addition to chemistry, manufacturing, and control issues related to a recent pre-approval inspection and product quality, is oportuzumab montox (Vicineum™), contains the PE and anti-EpCAM single chain antibodies to treat bladder cancer (4,5). MT-3724 is an engineered form of Shiga-like toxin A subunit in phase II clinical trials for B-cell lymphoma (6). E7777 is a recombinant cytotoxic protein composed of the human interleukin (IL) 2 and diphtheria toxin, and E7777 is a modified-processes Ontak (*Denileukin diftitox*). This modified-processes toxin to a reduced level of aggregated and/or misfolded protein impurities and the percentage of active proteins increased. In phase II clinical trials in Japanese patients, sufficient efficacy with a manageable safety profile of E7777 was confirmed for the treatment of cutaneous and peripheral T-cell lymphoma (7).

Proteases are microbial metabolites with many biological functions. Arazyme (AraA), an extracellular metalloproteinase secreted by *Serratia proteamaculans*, has strong proteolytic activity on various protein substrates such as albumin, elastin, creatinine, and collagen in a wide pH and temperature range (8,9). Several studies have reported the strong anticancer activity of AraA, especially its antimetastatic effect on various cancers such as colon, ovarian, and melanoma cancer cell lines (8,10,11). Ghadaksaz *et al.* have investigated a fusion protein consisting of transforming growth factor alpha third loop (TGFαL3)-targeted AraA for breast cancer therapy (12). The main mechanism of AraA for cancer treatment is its protease-dependent action. In addition, this enzyme has non-proteolytic antitumor effects which are mostly produced by the induction of cytotoxic T lymphocytes (10). The hepatoprotective anti-inflammatory activity (13) and the inhibitory effect of this enzyme on allergic inflammation have also been demonstrated (14). Moreover, AraA induces antioxidant signaling and inhibits the release of T helper cell type 2 cytokine (15).

For many years, ligand-targeted therapeutics have been studied to improve the efficacy of cancer treatments, but in recent years these strategies have changed significantly (16). In general, the density of the target receptor on the surface of tumor cells must be higher than normal cells or overexpressed (17). IL13 is secreted by T-helper 2 cells, mast cells, and CD4 cells such as macrophages (18). This cytokine plays an important role in immune responses to cancer, inflammation, and infection and binds to both IL13 receptor alpha 1 (IL13Rα1) and IL13Rα2 receptors (19). The IL13Rα2 has a high affinity for IL13 and is overexpressed in various cancer, compared to low or no expression in natural tissues. High-affinity binding of IL13Rα2 for IL13 induces signals *via* an activator protein 1-dependent pathway independent of STAT-6, which results in increased activity of TGF-β, but it does not induce signal transmission through the JAK/STAT6 pathway like IL13Rα1 (20). IL13.E13K is a modified IL13 with a specific high affinity for IL13Rα2 containing a modified glutamic acid in the 13th amino acid instead of lysine (21).

Recently, linear peptide ligands have been considered due to their small size, low molecular weight, high flexibility, high specificity, low antigenicity, and non-immunogenicity (22,23). One of the peptide ligands which binds specifically to the IL-13Ra2 is Pep-1, a linear peptide with CGEMGWVRC synthetic sequence (22). The A2b11 peptide (WALRVKAG), which was a T7 random peptide phage display, was shown to bind to IL13Ra2 with the highest specificity (24).

Based on the findings of the *in silico* study, to obtain an effective fusion protein, comparison of the designed proteins is easier and rapid analysis before starting the laboratory experiments (25,26). Moreover, the pre-experimental study utilizing bioinformatics tools is cost-effective and fast in contrast to the experimental methods. In a recent study, several novels engineered analogs of serratiopeptidase were computationally designed and molecular dynamically simulated. Based on the results of the *in silico* study, two engineered analogous of serratiopeptidase (T344 and T380) were chosen as appropriate candidates for the experimental study (26).

Bioinformatics tools (such as servers, software, etc.) are designed and developed to extract information such as sequence or retrieve data from genomic sequence databases as well as predict biomolecule structures including DNA, RNA, proteins, etc. Also, this field of science is widely used in different studies of biological research such as the design and discovery of new drugs and vaccines plus structure and function prediction (27). In addition, these tools provide access to scientific databases in various fields of the life sciences, including genomics, proteomics, systems biology, phylogeny, population genetics, etc.

Since it is necessary to find anticancer agents with high efficacy on tumor cells and minimal side effects on normal cells, here, we have tried to design and compare several AraA-based fusion proteins with different ligands to choose more effective targeted therapy for some cancers with overexpression of IL13Rα2 using bioinformatics tools. Consequently, IL13Rα2 was selected as the receptor and IL13 as well as IL13.E13K were considered native and mutant ligands, respectively, for targeted cancer therapy. In addition to IL13 and IL13.E13K, we also studied Pep-1 and A2b11 as peptide ligands for targeted drug delivery to IL13Ra2-overexpressed cancer cells. Our next study will confirm these *in silico* results using *in vitro* and *in vivo* studies.

MATERIALS AND METHODS

Designing the chimeric constructs

The protein sequence of AraA (*Serratia proteamaculans*) and IL13 protein were obtained from Uniprot (with accession numbers Q2VU40 and P35225, respectively) and GenBank database (with accession number AAX21094.1 and X69079.1, respectively). In this study, the main component of chimeric constructs was the AraA protein as a functional site that fused with various ligands including IL13, IL13.E13K, Pep-1 and A2b11 for targeted drug delivery to IL13Rα2-overexpressed cancer cells. Several linkers were used for fusion protein design and tested by the I-TASSER server to obtain the best linker for separating two functional domains of the fusion proteins that do not interfere with each other or are only minimally disturbed, compared to their native 3D structure. Finally, the sequences of the chimeric proteins were codon optimized according to *E. coli* codon usage using the Java Codon Adaptation tool (JCat, http://www.jcat.de/) to increase the protein expression. In addition, GC content and codon adaptation index (CAI) were calculated by this server.

The physicochemical property and secondary-structure prediction

In order to compute the physiochemical characterization of the designed constructs, including molecular weight, theoretical isoelectric point, the total number of positive and negative residues, half-life, instability index, extinction coefficient, aliphatic index, and grand average of hydropathicity, the Expasy ProtParam server (https://web.expasy.org/protparam/) was used. To predict the percentage of secondary structures in the fusion proteins, the GOR IV online database was employed (28).

3D-Structure prediction of fusion proteins

The 3D model of the chimeric proteins was predicted using the I-TASSER (https://zhanglab.ccmb.med.umich.edu/I-TASSER/), which generates 3D models along with a confidence score (C-Score) for the quality of the predicted structure. I-TASSER predicts the 3D structure of the fusion protein and function based on a hierarchical approach from sequence to structural and functional model (28). To investigate the conformational state of the domains in fusion proteins, 3D structures aligned with protein data bank (PDB) file 1SAT (based on the template, Q2VU40) for AraA and 1IK0 solution structure of human IL13 using the template modeling (TM)-Align (29) and CHIMERA software V1.10.2. Finally, the 3D structure of the superior fusion protein was also predicted by Robetta (http://new.robetta.org/) (30) and RaptorX (http://raptorx.uchicago.edu/) (31) online servers and the predicted models were compared by different servers to check the similarities. RaptorX is a server for predicting the 3D structure of the fusion protein without using templates, including solvent accessibility,

secondary structure and disordered regions (30). The Robetta server generates 3D structural models for fusion protein *via* either comparative modeling or *de novo* structure prediction techniques (30).

Refinement and validation of fusion protein structures

In order to improve the 3D protein structure predicted to be closer to the native structure, all fusion proteins were refined using the 3D refine server (http://sysbio.rnet.missouri.edu/3D refine/index.html) (32).

The refined 3D structure of all fusion proteins validated by both QMEAN (https://swissmodel.expasy.org/qmean/) (33) and ProSA-web (https://prosa.services.came. sbg.ac.at/prosa.php) (28). The quality measured as Z-score was computed by ProSA and indicated that the predicted structures were in the score range of the native proteins with similar size.

For more validation, the predicted structures were checked using UCLA-DOE LAB - SAVES v6.0 server for several parameters including ERRAT, VERIFY, PROVE and PROCHECK (https://saves.mbi.ucla.edu/). By ERRAT, nonbonded interactions between different types of atoms were analyzed and showed the value of the error function in comparison with highly refined structures. The overall quality factor for good structure with high resolution and lower resolutions is valued at around ≥ 95% and 91%, respectively. The compatibility of the 3D model with its amino acid sequence was determined by assigning a structural class based on its environment and location (polarity, non-polarity, alpha, beta, loop, etc.) and comparing the results to good structures by the VERIFY3D program. In verified structures, at least 80% of the residues have a 3D/1D score ≥ 0.2.

The PROVE program calculated volumes of atoms in the macromolecules and the Z-score deviation for the predicted 3D structures from highly resolved and refined PDB structures were calculated. If the total buried outlier protein atoms of the predicted structure, was > 5%, the 3D model failed. The PROCHECK program was applied to evaluate the stereochemical quality of the 3D structure. In this program, residue-by-residue geometry and overall structural geometry were analyzed and then compared to the stereochemical parameters of the high-resolution and well-refined PDB structures.

Prediction of cleavage sites

Proteasome cleavage sites in the fusion protein have been predicted by Netchop 3.1. This server is the best neural network for cleavage site predictions of the human proteasome and approximately 75% of the cleavage sites have been correctly predicted with false positives of about 15% (17). The C-term 3.0 network was used for predicting the cytotoxic T lymphocyte epitopes. This network is a database containing 1260 class I major histocompatibility complexes ligands.

B-Cell epitopes prediction

Linear B cell epitopes of the chimeric proteins were predicted by the ABCpred server (https://webs.iiitd.edu.in/raghava/abcpred/AB C_submission.html) with a threshold of 0.8 and 16 mers epitope length (34). Continuous B cell epitopes of fusion proteins were predicted using the BcePred server (https://webs.iiitd.edu.in/raghava/bcepred/bcep red_submission.html) with the 1.9 threshold based on the antigenic propensity (35).

Toxicity and antigenicity prediction

The toxicity of the chimeric proteins has been predicted by ToxinPred. ToxinPred is an *in silico* database server for checking the presence of toxic peptides in proteins. The foremost dataset used in this method consists of 1805 toxic peptides (≤ 35 residues) (36). To identify the antigenicity of the chimeric proteins the VaxiJen V2.0 server was employed. VaxiJen is an *in silico* approach to the prediction of subunit vaccines and protective antigens (28).

Ligand-receptor docking

In this study, AraA-$(A(EAAAK)_4ALEA(EAAAK)_4A)_2$-IL13 fusion protein was selected and further investigated for its natural ligand. Molecular docking simulation is performed using HawkDock web server. HawkDock is a web

server to predict and analyze the protein-protein complex based on computational docking and molecular mechanics-generalized born surface area (MM/GBSA). Our proteins were docked with the HawkDock web server (37) and ranked according to binding free energy scores. The hydrogen bond and hydrophobic interactions between the IL13 and IL13Rα2 of the docked complexes were drawn using the LigPlot$^+$ software.

Among the molecular species docked by the HawkDock server, the best molecular species were extracted and subsequently, were selected as the primary structure to enter the molecular dynamics (MD) simulation step.

MD simulation of complexes

The MD simulation of the best docking pose of top-ranked complexes was accomplished using the GROMACS simulation package (2020) (38). The all-atom optimized potentials for liquid simulations force field were used for simulating our system (39). The protein-protein complexes were solvated in the MD simulation box with transferable intermolecular potential 3P (TIP3P) solvent molecules and then neutralized by adding 0.15 mol/L Na^+/Cl^- ions. Energy minimization of all systems was performed by the steepest descent algorithm. The systems were then equilibrated using NVT and NPT with 400 ps steps, respectively, utilizing a V-rescale Berendsen thermostat and a Parrinello-Rahman (38). The heating of the systems was gradually increased from 0 to 310 K, and the pressure of the systems was set to 1 atm for the NVT and NPT ensembles, respectively. The particle-mesh Ewald (PME) and the LINCS algorithms were applied to assess all electrostatic connections and to restrain all bond lengths in the protein, respectively. Moreover, periodic boundary condition was utilized during the simulation (37).

The MM/GBSA method was also, applied to calculate the binding free energy using the following equations:

$$\Delta G_{bind} = G_{complex} - (G_{receptor} + G_{ligand}) \quad (1)$$

$$\Delta G_{bind\,=} \Delta E_{MM} + \Delta G_{solv} - T\Delta S \quad (2)$$

$$\Delta G_{bind} = \Delta E_{vdW} + \Delta E_{ele} + \Delta G_{pol} + \Delta G_{nonpol} - T\Delta S \quad (3)$$

In these equations, $G_{receptor}$, G_{ligand}, and $G_{complex}$ refer to free energies of the receptor, ligand, and complex, respectively. A sum of changes in molecular mechanical (MM) gas-phase binding energy (ΔE_{MM}), the solvation-free energy (ΔG_{solv}), and entropic ($-T\Delta S$) contributions were applied to compute the binding free energy (ΔG_{bind}). Van der Waals (ΔE_{vdW}) and gas-phase electrostatic (ΔE_{ele}) energies create the molecular mechanical term (ΔE_{MM}). Nonpolar (ΔG_{nonpol}) and polar (ΔG_{pol}) energies create ΔG_{solv}. The ΔG_{nonpol} was calculated using $\Delta G_{nonpol} = \gamma SASA + \beta$ (40).

RESULTS

Designed and optimized codon chimeric constructs

The major component of the chimeric constructs was the Arazyme protein (AraA, *Serratia proteamaculans*) located at the functional domain of the fusion proteins and considered to be an anticancer agent. Several ligands were investigated for targeting cancer therapy to obtain the best ligand with high affinity to IL13ralphα2, including IL13 as a native ligand, IL13.E13K as a modified ligand, Pep-1 and A2b11 as peptide ligands. The Schematic model of designed chimeric proteins is shown in Fig. 1.

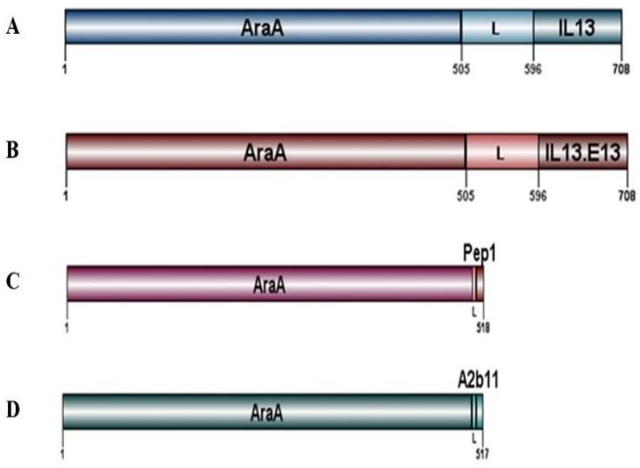

Fig. 1. Schematic model of designed chimeric constructs. (A) AraA + linker + native IL13; (B) AraA + linker + IL13.E13; (C) AraA + linker + Pep1; (D) AraA + linker + A2b11.

Several hydrophobic linkers (34 linkers listed in Table 1) used in the last studies, were tested by the I-TASSER server to obtain two separate functional domains of the fusion proteins do not interfere with each other or are only minimally disturbed, compared to their native 3D structure of the 33 linkers, only linker-31 $(A(EAAAK)_4ALEA(EAAAK)_4A)_2$ was selected as the appropriate hydrophobic linker for AraA-IL13 and AraA-IL13.E13K chimeric constructs with two distinct functional domains. Ultimately, the restriction enzyme *Bam*H1 site at the 5' end and *Hin*dIII site at the 3' end of the construct were regarded to facilitate the cloning in the pET28a+ vector for further cloning purposes (41). Improved GC content and CAI to near 50% and 1 (GC% of *E. coli* is about 50%), respectively. Therefore, the stability of mRNA in the chimeric gene is expected to increase.

The physicochemical properties and secondary structure prediction

The results of the physicochemical properties of the chimeric proteins from the Expasy ProtParam server are presented in Table 2. The instability index of all designed fusion proteins is less than 40, indicating stability. AraA-$(G_4S)_1$-Pep1 and AraA-$(G_4S)_1$-A2b11 fusion proteins have better interaction with water due to their lower grand average of hydropathicity. AraA-$(A(EAAAK)_4ALEA(EAAAK)_4A)_2$-IL13 and AraA-$(A(EAAAK)_4ALEA(EAAAK)_4A)_2$-IL13.E13K with high aliphatic index (73.12) have thermal stability for a wide range of temperatures.

Table 1. List of linkers investigated to obtain fusion proteins with separate domains.

No.	Linker sequences	No.	Linker sequences
1	$(EAAAK)_1$	**18**	$A(EAAAK)_4ALEA(EAAAK)_4A$
2	$(EAAAK)_2$	**19**	AEAAAKEAAAKA
3	$(EAAAK)_3$	**20**	GAAPAAAPAKQEAAAPAPAAKAEAPAAAPAAKA
4	$(G_4S)_1$	**21**	GGSGSGGG
5	$(G_4S)_2$	**22**	GSGGSGGSGGSG
6	$(G_4S)_3$	**23**	GSGGSGGSGGSGAEAAAKEAAAKAGSGGSGGSGGSG
7	$(G_4S)_4$	**24**	AEAAAKEAAAKAGSGGSGGSGGSGAEAAAKEAAAKA
8	$(G_4S)_6$	**25**	PTPPTTPTPPTTPTPTP
9	$G_2SG_3SG_2$	**26**	GGGGSGGGGSGGGGSGGGGSGGGGSAS
10	G_8	**27**	GGGGSLVPRGSGGGGS
11	GSA	**28**	GGSGGHMGSGG
12	P_6	**29**	GSADGGSSAGGSDAG
13	PAPAP	**30**	AEAAAKEAAAKEAAAKEAAAKA
14	$(AP)_8$	**31**	$(A(EAAAK)_4ALEA(EAAAK)_4A)_2$
15	$A(EAAAK)_{16}A$	**32**	$(A(EAAAK)_2ALEA(EAAAK)_2A)_2$
16	$(AP)_{20}$	**33**	$(A(EAAAK)_3ALEA(EAAAK)_3A)_2$
17	$(GGS)_{12}$		

Table 2. Physical and chemical parameters of designed fusion proteins.

Constructs	N. a.a	MW	PI	a.a-	a.a+	II	AI	GRAVY
AraA-Pep1 (without linker)	513	54.87	4.60	57	29	26.57	67.76	-0.393
AraA-$(EAAAK)_1$-Pep1	518	55.34	4.62	58	30	26.41	67.68	-0.393
AraA-$(G_4S)_1$-Pep1	518	55.19	4.60	57	29	27.59	67.10	-0.394
AraA-GSA-Pep1	516	55.08	4.60	57	29	27.16	67.56	-0.390
AraA-$(A(EAAAK)_4ALEA(EAAAK)_4A)_2$-Pep1	605	63.45	4.76	75	45	24.78	68.00	-0.361
AraA-A2b11 (without linker)	512	54.73	4.66	56	30	26.70	69.04	-0.393
AraA-$(EAAAK)_1$-A2b11	517	55.20	4.68	57	31	26.54	68.96	-0.393
AraA-$(G_4S)_1$-A2b11	517	55.04	4.66	56	30	27.09	68.38	-0.394
AraA-GSA-A2b11	515	54.94	4.66	56	30	26.44	68.83	-0.390
AraA-$(A(EAAAK)_4ALEA(EAAAK)_4A)_2$-A2b11	604	63.31	4.81	74	46	24.16	69.09	-0.361
AraA-$(A(EAAAK)_4ALEA(EAAAK)_4A)_2$-IL13.E13K	708	74.75	5.07	82	57	25.44	73.12	-0.306
AraA-$(A(EAAAK)_4ALEA(EAAAK)_4A)_2$-IL13	708	74.76	5.01	83	56	26.64	73.12	-0.305
Constituents								
AraA	504	53.85	4.59	56	28	26.94	68.39	-0.403
IL13	112	12.34	8.81	9	12	38.88	97.50	0.036
IL13.E13K	112	12.34	9.27	8	13	31.29	97.50	0.032

AraA, Arazyme; N. a.a, number of amino acids; MW, molecular weight; pI, theoretical isoelectric point; a.a-: total number of negative residues; a.a+, total number of positive residues; II, instability index; AI, aliphatic index; GRAVY, grand average of hydropathicity; IL, interleukin.

Table 3. The secondary structure of fusion proteins predicted by GOR IV.

Constructs	α helix	β Sheet (extended strand)	Random coil
AraA-Pep1 (without linker)	24.17%	21.64%	54.19%
AraA-$(EAAAK)_1$-Pep1	25.87%	19.69%	54.44%
AraA-$(G_4S)_1$-Pep1	23.94%	21.43%	54.63%
AraA-GSA-Pep1	24.03%	21.12%	54.84%
AraA-$(A(EAAAK)_4ALEA(EAAAK)_4A)_2$-Pep1	36.53%	16.86%	46.61%
AraA-A2b11 (Without Linker)	25.00%	21.09%	53.91%
AraA-$(EAAAK)_1$-A2b11	26.31%	20.12%	53.58%
AraA-$(G_4S)_1$-A2b11	23.98%	21.86%	54.16%
AraA-GSA-A2b11	24.08%	21.94%	53.98%
AraA-$(A(EAAAK)_4ALEA(EAAAK)_4A)_2$-A2b11	36.92%	17.22%	45.86%
AraA-$(A(EAAAK)_4ALEA(EAAAK)_4A)_2$-IL13.E13K	36.72%	15.96%	47.32%
AraA-$(A(EAAAK)_4ALEA(EAAAK)_4A)_2$-IL13	36.72%	15.96%	47.32%
Constituents			
AraA	24.60%	21.03%	54.37%
IL13	34.82%	11.61%	53.57%
IL13.E13K	35.71%	11.61%	52.68%

AraA, Arazyme; IL, interleukin.

The results of the secondary structure of fusion proteins using the GOR IV server are shown in Table 3. The most common secondary structural elements are the random coils and then alpha helix. The beta sheets are considered to be the third common secondary structure.

The tertiary structure of designed fusion proteins

The 3D models of the designed fusion proteins were predicted using the I-TASSER (https://zhanglab.ccmb.med.umich.edu/I-TASSER/) online server, which generates 3D models along with a C-score showing the quality of the predicted structure. I-TASSER-predicted Model.1 of five probable tertiary structures is shown in Fig. 2. Afterward, the 3D model of the superior fusion protein was predicted using Robetta and RaptorX to confirm the predicted models of I-TASSER (Fig. 3).

3D Models of chimeric proteins predicted by I-TASSER showed the chimeric proteins with two separate main domains linked together with the linker. The C-score of model 1 was the maximum when compared to other models. The C-score is generally reported in the range of -5 to 2 and a higher C-score indicates higher confidence in the model. In addition, the TM-score and root-mean-square deviation (RMSD) for model 1 were reported in Fig. 2.

In the next step, model 1 of the predicted 3D structure of the selected fusion protein by I-TASSER was aligned with the PDB file of AraA and IL13 using the TM-Align server. TM-Score for AraA and IL13 aligned with AraA-IL13 were 0.887 and 0.866, respectively ($0.0 < \text{TM-score} < 0.30$, random structural similarity and $0.5 < \text{TM-score} < 1.00$, in about the same fold) (28) which indicates that the protein fusion components are similar to their native structure (Fig. 3E).

Refinement and validation of fusion protein structures

To improve the predicted 3D protein structure to be similar to the native structure, all fusion proteins were refined using the 3D refine server. Then, the refined 3D structure of all fusion proteins was confirmed with the output of both QMEAN (https://swissmodel.expasy.org/qmean/) and ProSA (https://prosa.services.came.sbg.ac.at/prosa.php). The higher value of the C-score signifies a higher confidence score and *vice versa*. AraA-$(G_4S)_1$- Pep1 and AraA-A2b11 (without linker) have higher C-score and Q-mean scores. The 3D structures of all designed fusion proteins were analyzed by ProSA-web to detect errors in the 3D model of fusion proteins. Fusion proteins with $(A(EAAAK)_4ALEA(EAAAK)_4A)_2$ linker are out of the native protein range (Table 4).

For more validation, the predicted structures were checked using UCLA-DOE LAB - SAVES v6.0 server with several parameters including ERRAT, VERIFY, PROVE, and PROCHECK (https://saves.mbi.ucla.edu/). Unfortunately, all overall quality factors for all designed fusion proteins were less than 90%, while overall quality factors for good structure with high resolution and lower resolutions were calculated to be about $\geq$ 95% and 91%, respectively.

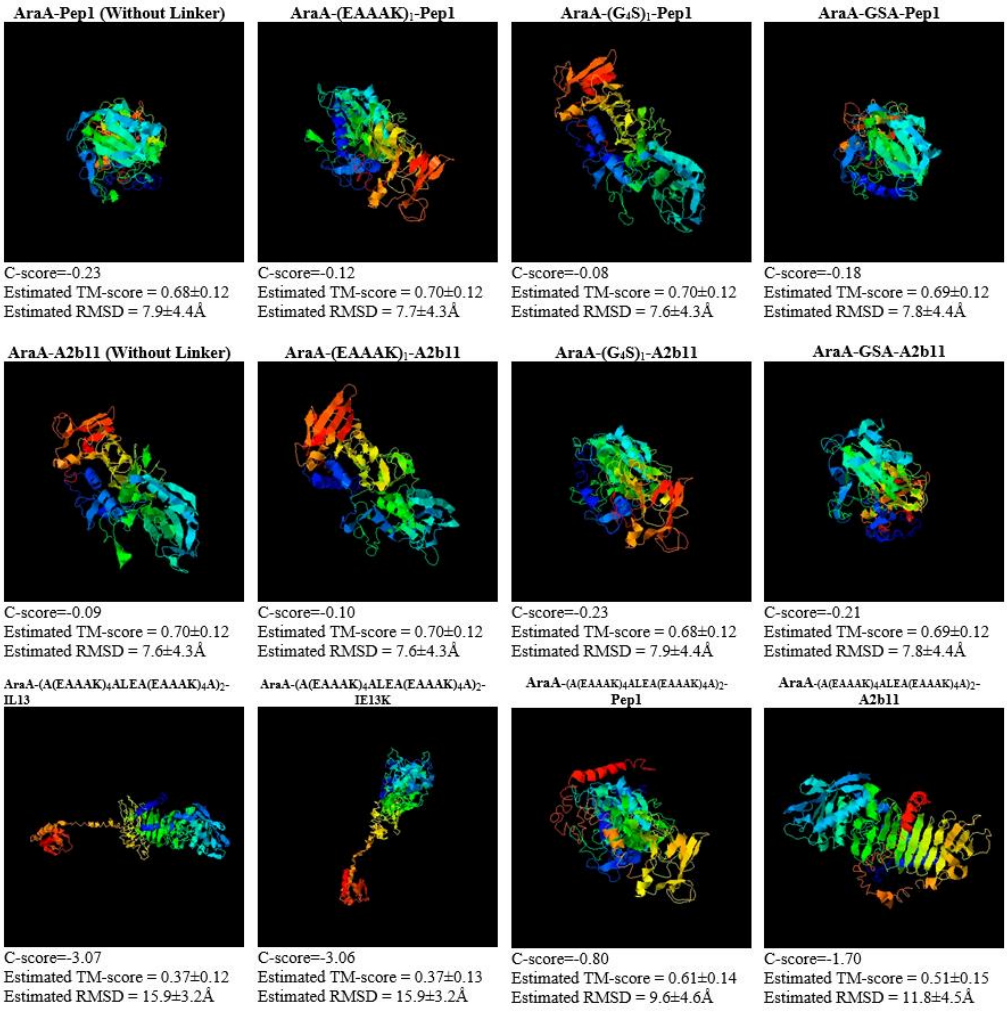

Fig. 2. Model.1 of 3D designed structures was predicted by I-TASSER.

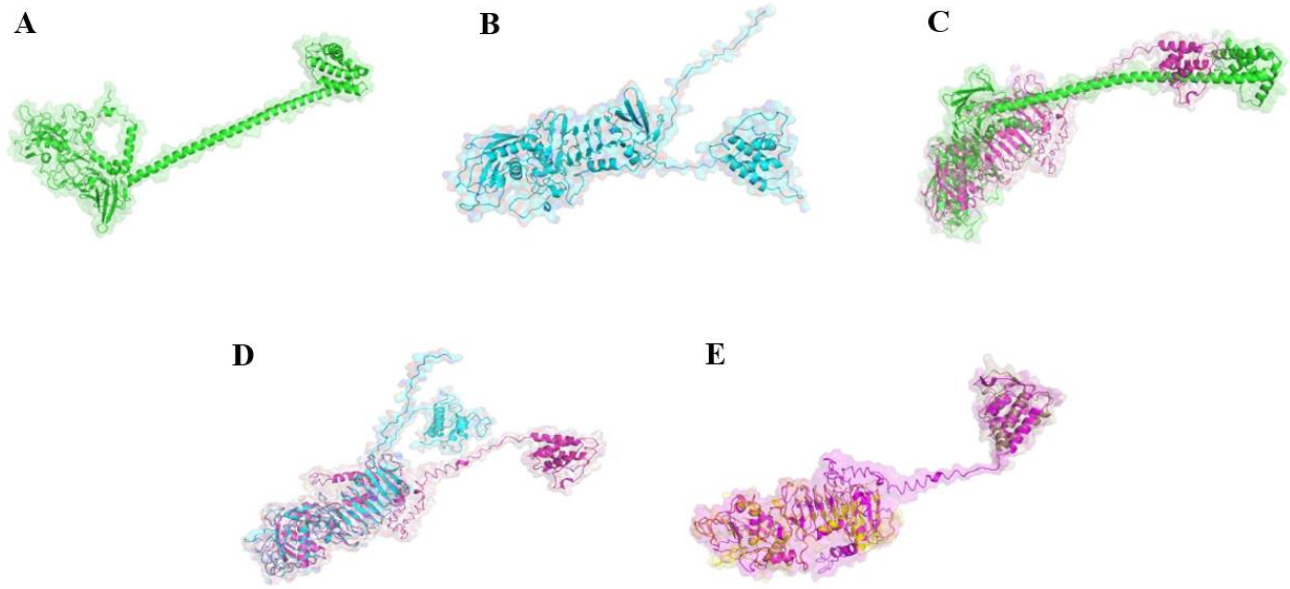

Fig. 3. 3D structures of the selected fusion protein were predicted by (A) Robetta and (B) RaptorX; (C) Robetta model aligned with I-TASSER model; (D) model RaptorX aligned with I-TASSER model; and (E) the results obtained from TM-align server for aligned AraA-IL13 fusion protein with protein data bank files of AraA and IL13 using Chimera software 1.10.2. AraA, Arazyme; IL, interleukin.

Table 4. Structure refinement and scores of Q-mean and ProSA for the designed fusion proteins.

Constructs	3D structure I-TASSER			Refinement
	C-Score Model1	ProSA Z-Score	Q-Mean Score	ProSA after refinement by 3D refine
AraA-Pep1 (Without Linker)	-0.23	-6.63	0.79	-7.49
AraA-$(EAAAK)_1$-Pep1	-0.12	-6.67	0.79	-7.04
AraA-$(G_4S)_1$-Pep1	-0.08	-6.49	0.80	-6.68
AraA-GSA-Pep1	-0.18	-6.64	0.79	-7.62
AraA-$(A(EAAAK)_4ALEA(EAAAK)_4A)_2$-Pep1	-0.80	-3.38*	0.70	-4.67*
AraA-A2B1 (Without Linker)	-0.09	-6.63	0.82	-7.2
AraA-$(EAAAK)_1$-A2b11	-0.10	-6.1	0.81	-6.9
AraA-$(G_4S)_1$-A2b11	-0.23	-6.69	0.79	-7.27
AraA-GSA-A2b11	-0.21	-6.21	0.81	-7.13
AraA-$(A(EAAAK)_4ALEA(EAAAK)_4A)_2$-A2b11	-1.70	-4.16*	0.66	-5.41*
AraA-$(A(EAAAK)_4ALEA(EAAAK)_4A)_2$-IL13.E13K	-3.06	-4.24*	0.61	-5.76*
AraA-$(A(EAAAK)_4ALEA(EAAAK)_4A)_2$-IL13	-3.07	-5.55*	0.59	-4.99*

* Out of native protein range

Table 5. The assessment of a 3D refined model of designed fusion proteins

Constructs	PROSESS	ERRAT	VERIFY	PROVE	PROCHECK	
					Errors, warnings, pass	Favored regions
AraA-Pep1 (Without Linker)	3.5	77.8218	88.30% Pass	4.5% Warning	7, 2, 0	78.3%
AraA-$(EAAAK)_1$-Pep1	3.5	83.9216	86.68% Pass	4.6% Warning	6, 1, 1	81.7%
AraA-$(G_4S)_1$-Pep1	3.5	78.0392	90.35%	4.2% Warning	6, 1, 1	79.9%
AraA-GSA-Pep1	3.5	75.7874	89.92% Pass	3.7% Warning	6, 2, 0	81.3%
AraA-$(A(EAAAK)_4ALEA(EAAAK)_4A)_2$-Pep1	2.5	80.9045	77.02% Fail	5.0% Warning	7, 1, 0	77.3%
AraA-A2b11 (Without Linker)	3.5	85.9127	92.38% Pass	4.5% Warning	6, 1,1	82.9%
AraA-$(EAAAK)_1$-A2b11	3.5	81.336	86.85% Pass	5.2% Fail	6, 2, 0	82.2%
AraA-$(G_4S)_1$-A2b11	3.5	86.6405	92.26% Pass	3.6% Warning	6, 1, 1	80.9%
AraA-GSA-A2b11	3.5	87.7712	85.44% Pass	4.0% Warning	6, 3, 0	81.1%
AraA-$(A(EAAAK)_4ALEA(EAAAK)_4A)_2$-A2b11	2.5	78.8591	75.00% Fail	5.0% Warning	7, 1, 0	75.1%
AraA-$(A(EAAAK)_4ALEA(EAAAK)_4A)_2$-IL13.E13K	2.5	81.1429	79.94% Fail	4.7% Warning	7, 2, 0	73.3%
AraA-$(A(EAAAK)_4ALEA(EAAAK)_4A)_2$-IL13	2.5	75.7143	82.77% Pass	8.1% Fail	7, 2, 0	75.4%

AraA, Arazyme; IL, interleukin.

In structures verified by the Verify3D program, at least 80% of the residues have a 3D/1D score ≥ 0.2. Designed fusion proteins with $(A(EAAAK)_4ALEA(EAAAK)_4A)_2$ linker have less than 80% of the residues with the averaged 3D-1D score ≥ 0.2, except AraA-$(A(EAAAK)_4ALEA(EAAAK)_4A)_2$-IL13 that has 82.77% of the residues with scored ≥ 0.2 in the 3D/1D profile. In the PROVE program, buried outlier protein atoms total for the 3D model of AraA-$(EAAAK)_1$-A2b11, AraA-$(A(EAAAK)_4ALEA(EAAAK)_4A)_2$-IL13 fusion proteins were calculated to be > 5%, therefore these structures have failed.

In the Ramachandran plot presented by the PROCHECK program, a good-quality model would be expected to have over 90% amino acid residues in the most favored regions. Among the designed structures, AraA-A2B1 (without linker) fusion protein with the highest score has 82.9% residues in the most favored regions with the lowest errors/warnings (Table 5).

Cleavage sites predicted in designed fusion proteins

Proteasome cleavage sites inside the fusion protein were predicted by Netchop 3.1 server. The Net Chop neural network-based method was the most effective presently-available system for the predictions of the cleavage site. The new version of Net Chop has predicted approximately 75% of cleavage sites correctly with false positives near 15%. The number of the cleavage sites on the fusion proteins computed with NetChop 3.1 server using version C-term and threshold 0.800000 has been shown in Table 6.

Table 6. The predicted cleavage sites of the human proteasome for designed fusion proteins by Netchop 3.1.

Constructs	Proteasome cleavage sites using version C-term, threshold 0.8		
	Number of amino acids	Number of cleavage sites	cleavage sites
AraA-Pep1 (without linker)	513	78	CGEMGWVRC*
AraA-$(EAAAK)_1$-Pep1	518	78	EAAAKCGEMGWVRC
AraA-$(G_4S)_1$-Pep1	518	78	GGGGSCGEMGWVRC*
AraA-GSA-Pep1	516	77	GSACGEMGWVRC
AraA-$(A(EAAAK)_4ALEA(EAAAK)_4A)_2$-Pep1	605	79	AEAAAKEAAAKEAAAKEAAAKALEAEAAAKE AAAKEAAAKEAAAKAAEAAAKEAAAKEAAAK EAAAKALEAEAAAKEAAAKEAAAKEAAAKAC GEMGWVRC
AraA-A2b11 (without linker)	512	79	WALRVKAG
AraA-$(EAAAK)_1$-A2b11	517	80	EAAAKWALRVKAG
AraA-$(G_4S)_1$-A2b11	517	78	GGGGSWALRVKAG
AraA-GSA-A2b11	515	78	GSAWALRVKAG
AraA-$(A(EAAAK)_4ALEA(EAAAK)_4A)_2$-A2b11	604	80	AEAAAKEAAAKEAAAKEAAAKALEAEAAAKE AAAKEAAAKEAAAKAAEAAAKEAAAKEAAAK EAAAKALEAEAAAKEAAAKEAAAKEAAAKAW ALRVKAG
AraA-$(A(EAAAK)_4ALEA(EAAAK)_4A)_2$-IL13.E13K	708	100	AEAAAKEAAAKEAAAKEAAAKALEAEAAAKE AAAKEAAAKEAAAKAAEAAAKEAAAKEAAAK EAAAKALEAEAAAKEAAAKEAAAKEAAAKAG PVPPSTALRELIEELVNITQNQKAPLCNGSMVWS INLTAGMYCAALESLINVSGCSAIEKTQRMLSGF CPHKVSAGQFSSLHVRDTKIEVAQFVKDLLLHL KKLFREGRFN
AraA-$(A(EAAAK)_4ALEA(EAAAK)_4A)_2$-IL13	708	101	AEAAAKEAAAKEAAAKEAAAKALEAEAAAKE AAAKEAAAKEAAAKAAEAAAKEAAAKEAAAK EAAAKALEAEAAAKEAAAKEAAAKEAAAKAG PVPPSTALRELIEELVNITQNQKAPLCNGSMVWS INLTAGMYCAALESLINVSGCSAIEKTQRMLSGF CPHKVSAGQFSSLHVRDTKIEVAQFVKDLLLHL KKLFREGRFN
Constituents			
AraA	504	77	MSICLIENNQLMSGIEPMQSTKKAIEITESSLAAA SSAYNAVDDLLHYHERGNGIQVNGKDSFSTEQA GLFITRENQTWNGYKVFGQPVKLTFSFPDYKFSS TNVAGDTGLSKFSAEQQQQAKLSLQSWSDVANI TFTEVGAGQKANITFGNYSQDRPGHYDYDTQA YAFLPNTIYQGQNLGGQTWYNVNQSNVKHPAS EDYGRQTFTHEIGHALGLSHPGDYNAGEGNPTY RDASYAEDTREFSLMSYWSETNTGGDNGGHYA AAPLLDDISAIQHLYGANQTTRTGDTVYGFNSN TGRDFLSTTSNSQKVIFAAWDAGGNDTFDFSGY TANQRINLNEKSFSDVGGLKGNVSIAAGVTIENA IGGSGNDVIVGNAANNVLKGGAGNDVLFGGGG ADELWGGAGKDTFVFSAVSDSAPGASDWIKDF QKGIDKIDLSFFNQGAQGGDQIHFVDHFSGAAG EALLSYNASNNVSDLALNIGGHQAPDFLVKIVG QVDVATDFIV
IL13	112	20	GPVPPSTALRELIEELVNITQNQKAPLCNGSMVW SINLTAGMYCAALESLINVSGCSAIEKTQRMLSG FCPHKVSAGQFSSLHVRDTKIEVAQFVKDLLLH LKKLFREGRFN
IL13.E13K	112	20	GPVPPSTALRELIKELVNITQNQKAPLCNGSMV WSINLTAGMYCAALESLINVSGCSAIEKTQRMLS GFCPHKVSAGQFSSLHVRDTKIEVAQFVKDLLL HLKKLFREGRFN

AraA, Arazyme; IL, interleukin.

Cleavage sites of the human proteasome by AraA in full sequence are shown in Table 6 and AraA sequence for other designed fusion proteins is not shown.

B-Cell epitopes

The linear B cell epitopes predicted for AraA and AraA-(A(EAAAK)$_4$ALEA(EAAAK)$_4$A)$_2$-IL13 have been shown in Table 7. The predicted B-cell epitopes for IL13.E13K (modified IL13) were similar to the native IL13. The continuous B-cell epitopes for other residues in the designed protein were not identified by Bcepreds, except, the TDFIVGGGGSWALRVK sequence in start position 500 of AraA-(G_4S)$_1$-A2b11 fusion protein with the score 0.87 predicted as B-cell epitopes. The continuous B cell epitopes predicted for AraA, the main component of fusion proteins, and IL13 are shown in Table 8.

Table 7. The linear B-cell epitopes predicted in designed fusion proteins using the ABCpred prediction server.

Construct	Rank	Sequence	Start position	Score
AraA	1	GGGGADELWGGAGKDT	393	0.91
	1	QKVIFAAWDAGGNDTF	310	0.91
	1	TRTGDTVYGFNSNTGR	285	0.91
	1	TFGNYSQDRPGHYDYD	149	0.91
	2	GGAGKDTFVFSAVSDS	402	0.89
	3	GGSGNDVIVGNAANNV	366	0.88
	4	TRENQTWNGYKVFGQP	73	0.87
	4	AVSDSAPGASDWIKDF	413	0.87
	4	SETNTGGDNGGHYAAA	251	0.87
	4	ASYAEDTREFSLMSYW	235	0.87
	4	PGHYDYDTQAYAFLPN	158	0.87
	4	LMSGIEPMQSTKKAIE	11	0.87
	5	PGASDWIKDFQKGIDK	419	0.86
	5	EGNPTYRDASYAEDTR	227	0.86
	6	LGLSHPGDYNAGEGNP	215	0.85
	6	HEIGHALGLSHPGDYN	209	0.85
	6	GAGQKANITFGNYSQD	141	0.85
	7	HYHERGNGIQVNGKDS	47	0.83
	7	SGAAGEALLSYNASNN	457	0.83
	7	SSLAAASSAYNAVDDL	30	0.83
	8	GGAGNDVLFGGGGADE	384	0.82
	8	LGGQTWYNVNQSNVKH	181	0.82
	9	ALNIGGHQAPDFLVKI	477	0.81
	9	QSNVKHPASEDYGRQT	191	0.81
	10	AGGNDTFDFSGYTANQ	319	0.80
	10	SLMSYWSETNTGGDNG	245	0.80
AraA-(A(EAAAK)$_4$ALEA(EAAAK)$_4$A)$_2$-IL13	1	CSAIEKTQRMLSGFCP	652	0.92
	5	AGMYCAALESLINVSG	636	0.87
	8	AAKEAAAKEAAAKAGP	583	0.84
	10	TQNQKAPLCNGSMVWS	616	0.82

AraA, Arazyme; IL, interleukin.

Table 8. The continuous B-cell epitopes predicted in chimeric proteins using Bcepred server.

Construct	B-cell epitopes
AraA	MSICLIENNQLMSGIEPMQSTKKAIEITESSLAAASSAYNAVDDLLHYHERGNGIQVNGKDSFSTE QAGLFITRENQTWNGYKVFGQPVKLTFSFPDYKFSSTNVAGDTGLSKFSAEQQQQAKLSLQSWS DVANITFTEVGAGQKANITFGNYSQDRPGHYDYDTQAYAFLPNTIYQGQNLGGQTWYNVNQSN VKHPASEDYGRQTFTHEIGHALGLSHPGDYNAGEGNPTYRDASYAEDTREFSLMSYWSETNTGG DNGGHYAAAPLLDDISAIQHLYGANQTTRTGDTVYGFNSNTGRDFLSTTSNSQKVIFAAWDAGG NDTFDFSGYTANQRINLNEKSFSDVGGLKGNVSIAAGVTIENAIGGSGNDVIVGNAANNVLKGGA GNDVLFGGGGADELWGGAGKDTFVFSAVSDSAPGASDWIKDFQKGIDKIDLSFFNQGAQGGDQI HFVDHFSGAAGEALLSYNASNNVSDLALNIGGHQAPDFLVKIVGQVDVATDFIV
Ara-L31-IL13	GPVPPSTALRELIEELVNITQNQKAPLCNGSMVWSINLTAGMYCAALESLINVSGCSAIEKTQRML SGFCPHKVSAGQFSSLHVRDTKIEVAQFVKDLLLHLKKLFREGRFN

AraA, Arazyme; IL, interleukin.

The predicted continuous B-cell epitope for IL13.E13K (modified IL13) was similar to the native IL13. The continuous B-cell epitopes for other residues in the designed protein were not identified by Bcepreds, except for AraA-Pep1 (without linker) fusion protein that predicted TDFIVCGE sequence as B-cell epitopes.

Toxicity and antigenicity of chimeric proteins

For toxicity prediction, the designed chimeric proteins were scanned with ToxinPred server. This server predicts whether overlapping peptide/analog is toxic or not. Based on the protein scanning using the ToxinPred server, all designed chimeric proteins are considered non-toxic for peptide fragment length 10. To identify the antigenicity, VaxiJen V2.0 was employed. All designed fusion proteins predicted probably are not antigenic for the tumor model with threshold 1 using the VaxiJen server.

Ligand-receptor docking

In this study, AraA-(A(EAAAK)$_4$ALEA(EAAAK)$_4$A)$_2$-IL13 fusion protein was selected and further investigated due to its natural structure of components. Protein-protein docking was done by the HawkDock web server (37) and ranked according to binding free energy scores (Table 9). The schematic 2D view of the IL13-IL13Rα2 and chimeric IL13-IL13Rα2 docked complexes were generated by LigPlot$^+$ software. The hydrogen bond and hydrophobic interactions between complexes are presented in Fig. 4. To validate the docking results, 100 ns MD simulations on the top-ranked binding pose are performed.

Table 9. Protein-protein docking results of chimeric protein with IL13Rα2.

Complexes	Binding energy (kcal/mol)	Complexes	Binding energy (kcal/mol)
Model 1	-18.70	model.6	-33.18
Model 2	-35.87	model.7	-25.17
Model 3	-25.53	model.8.	-5.92
Model 4	-20.57	model.9	-24.96
Model 5	-33.22	model.10	-31.42

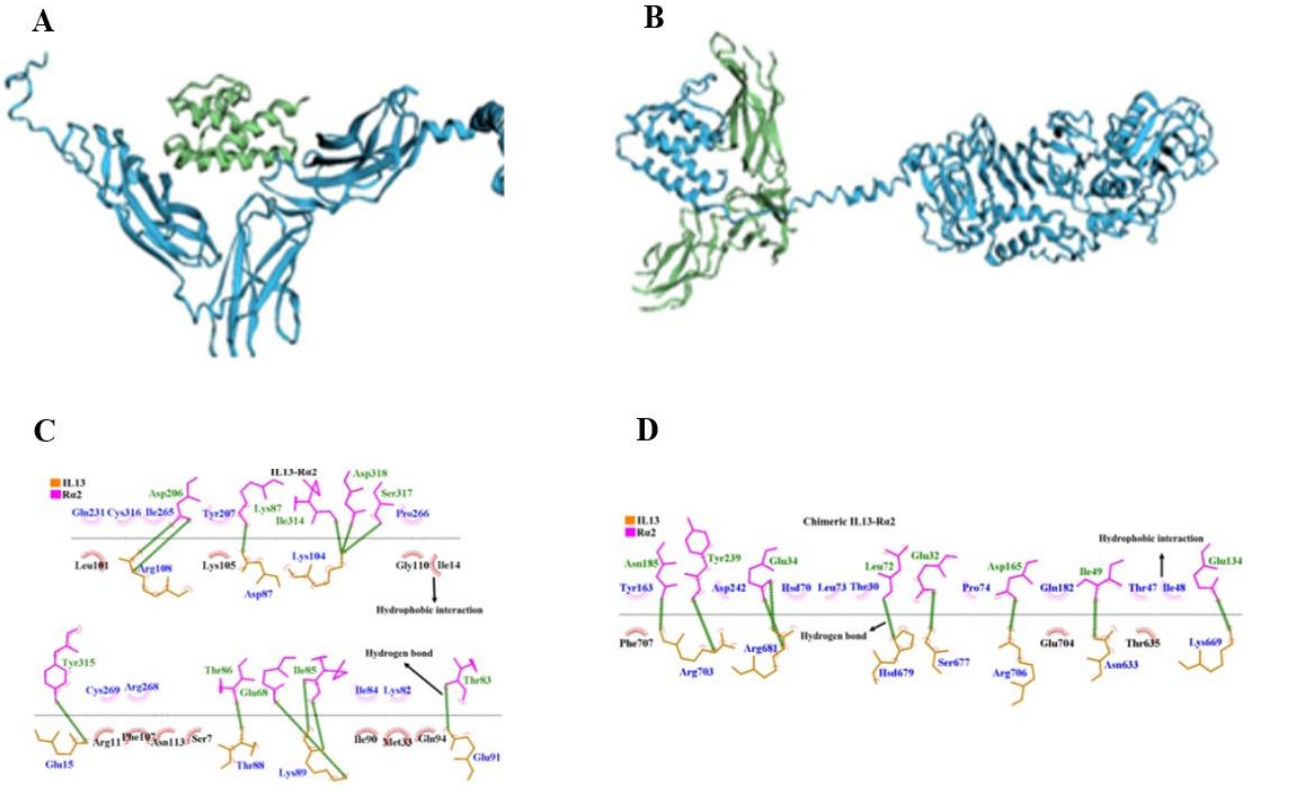

Fig. 4. (A) Interaction between IL13 and IL13Rα2; (B) interaction between chimeric protein and IL13Rα2 receptor; (C) the 2D view of the docked complexes were analyzed using LigPlot$^+$ software. The hydrogen bond (green dashed lines) and hydrophobic interactions (spoked arc) between the IL13-IL13Rα2; (D) chimeric IL13-IL13Rα2. IL, Interleukin.

Molecular dynamics simulation of complexes

The RMSD values for the IL13-Rα2 and chimeric-IL13-Rα2 were calculated as a time-dependent parameter for clarification of the displacement of Cα atoms in the two molecular types during the MD simulation (Fig. 5A). According to the RMSD curve, The IL13-Rα2 structure has a small fluctuation compared with chimeric-IL13-Rα2. The high value of the RMSD's chimeric-IL13-Rα2 can be arising from the big size of the chimeric-IL13. Also, the RMSD values of the IL13 in the two complexes were estimated separately (Fig. 5B). According to Fig. 5B, IL13 has a similar pattern of RMSD values during simulations gradually reaching a stable state with a smaller fluctuation range after 5000 ps. These findings mean that the IL13 has good stability in the two complexes (Fig. 5).

In general, based on the results obtained from the analysis of structure fluctuations, the protein-protein complexes have good structural stability during simulation times. Also, to estimate the strength of bonds for the protein-protein complexes and to clarify the residues which play the important role in protein-protein interactions, free energy calculations and binding energy decompositions per residue were done based on the 100 frames retrieved from the last 10 ns of the MD simulation trajectories. MM/GBSA approach was used to calculate components of the binding energy and to assess the energetic contribution of residue to the binding free energy by energy decomposition scheme (Table 10).

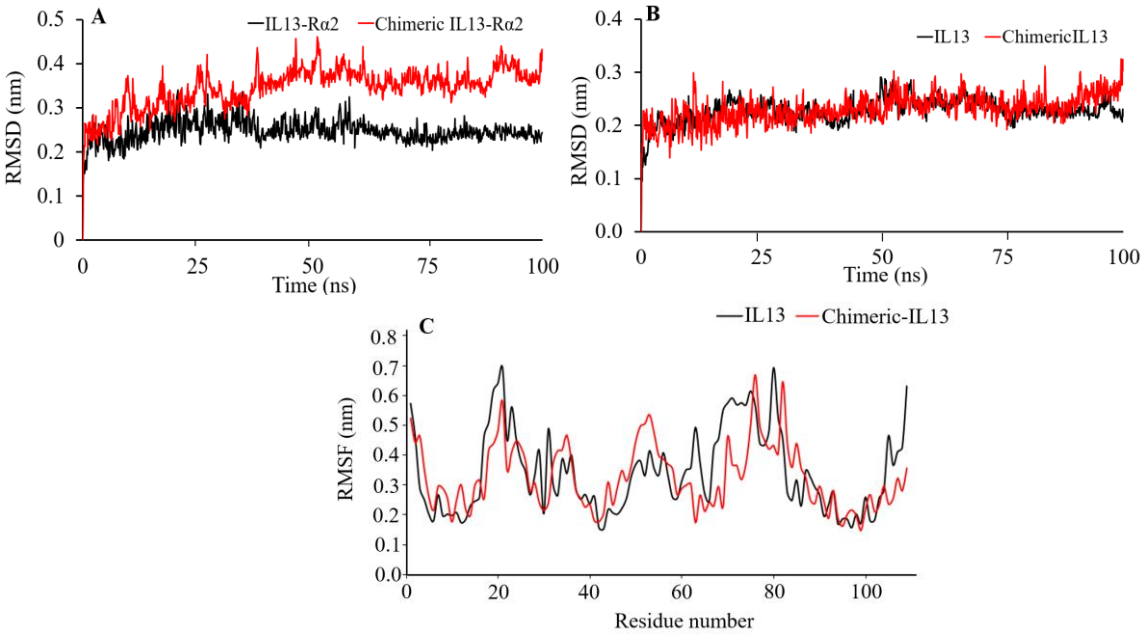

Fig. 5. (A) The Cα-RMSD graph for the IL13Rα2 and chimeric-IL13Rα2 in the complex state; (B) the Cα-RMSD graph for the IL13 and chimeric-IL13 at the complex state; (C) the Cα-RMSF graph for the IL13 in the complexes state. IL13Rα2, Interleukin 13 receptor alpha 2.

Table 10. Contribution of each energy component associated with binding free energy (kcal mol^{-1}).

	IL13Rα2	**IL13Rα2 (chimeric)**
ΔE_{ele}	-667.18	-862.57
ΔE_{vdW}	-102.23	-104.92
ΔG_{GB}	-700.18	-789.01
ΔG_{SA}	-12.50	-12.78
$\Delta E_{non\text{-}polar}$[a]	-114.73	-117.7
ΔE_{polar}[b]	33	73.56
ΔG_{bind}	-81.73	-44.14

ΔE_{ele}, gas-phase electrostatic energy; ΔE_{vdW}, Van der Waals energy; ΔG_{SA}, the nonpolar contribution of desolvation; ΔG_{GB}, generalized born binding free energy; IL13Rα2, Interleukin 13 receptor alpha 2.

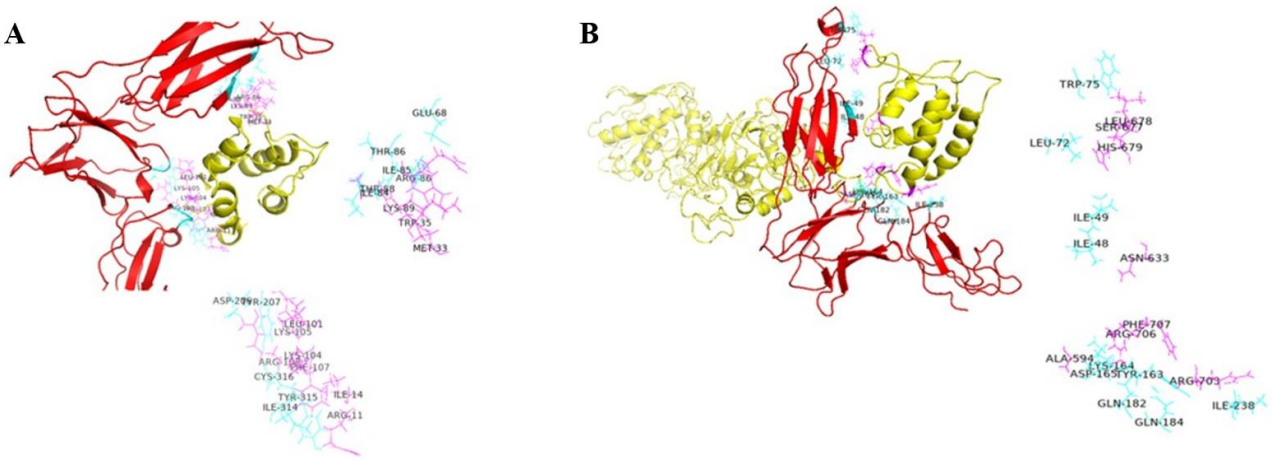

Fig. 6. The 3D view of the final protein-protein complexes. (A) IL13Rα2-IL13; (B) IL13Rα2-IL13 in chimeric state. Residues referring to protein-protein interface sites (cyan and magenta sticks) are labeled. The IL13Rα2 is represented as the red cartoon and IL13 as the yellow cartoon. IL13Rα2, Interleukin 13 receptor alpha 2.

$$\Delta E_{non\text{-}polar} = \Delta E_{vdW} + \Delta G_{SA} \quad (4)$$

$$\Delta E_{polar} = \Delta E_{ele} + \Delta G_{GB} \quad (5)$$

According to Table 10, the non-polar or hydrophobic energies ($\Delta E_{nonpolar}$) have a large contribution to the binding energy. It can be said that these hydrophobic energies are the major force in ligand-receptor binding and electrostatic interactions (ΔE_{polar}) in comparison with hydrophobic interactions, which appear to be unfavorable for ligand binding to the receptor.

Direct intermolecular interactions (ΔE_{ele}) in the IL13Rα2 complex are favorable, but the contribution of these interactions is neutralized by the large component (ΔG_{GB}) which plays a major role in the electrostatic interactions. These interactions (ΔE_{ele}) play an inverse role in the chimeric protein complex with IL13Rα2. However, polar interactions generally play a negative role in the interactions between the two species complexes. To obtain details of the intermolecular interactions, the contribution of each residue in the binding energy between different species of proteins was calculated (data are not shown). Considering the contribution of each component in the free binding energy, the major and desired energies involved in the ligand binding are Van der Waals (vdW) and electrostatic (ele) interactions, while the polar solvation energy (ΔG_{GB}) in binding the ligand to the protein is undesirable. vdW interactions and non-polar solvent energies are responsible for covering hydrophobic groups of proteins.

Therefore, it can be stated that hydrophobic residues play an important role in forming the binding envelope between the studied protein complexes and assure the stability of the complex system during the simulation. These residues also prevent the breakdown of the protein species complexes (Fig. 6).

DISCUSSION

Common treatments for cancer are chemotherapeutic agents, radiation therapy and *etc.* that kill both tumor cells and normal cells. Therefore, it may lead to some very unbearable side effects (42). In different types of cancers, several receptors and enzymes are overexpressed. Therefore, in the last decade, one of the new strategies for target cancer therapy is the identification of overexpressed receptors (42). Techniques and applications of bioinformatics provide computational methods for designing new drugs with specific ligands for overexpressed receptors. Previous studies have shown that IL13Rα2 is overexpressed on several cancer cells (43). In addition, in recent studies, AraA has been introduced as a promising new antitumor chemotherapy agent (8,15). Ghadaksaz *et al.* used AraA as an effective domain in AraA-linker-TGFαL3 chimeric proteins (12). In the current study, we have also selected AraA as the catalytic domain in the designed fusion proteins. These fusion proteins were designed to achieve the highest efficiency of the AraA for targeted cancer therapy and minimal side effect on normal cells

(10). Afterward, various ligands with different linkers were investigated to obtain the best ligand with high affinity to IL13Rα2, including IL13 as native ligand, IL13.E13K as modified ligand, Pep-1 and A2b11 as peptide ligands for targeting IL-13Rα2-overexpressed cancer cell.

At the beginning of the *in silico* analysis, physicochemical parameters were computed by ProtParam server. The instability index for all designed fusion proteins was less than 40, therefore predicted as stable. The high aliphatic index in the range of 67.10 to 73.12 indicates the thermostability of all designed fusion proteins with a high number of hydrophobic amino acids. In order to increase the mRNA stability and high expression of the fusion proteins, nucleic acid codons were optimized based on the *E. coli* host.

After predicting the refined 3D structure of all designed fusion proteins, the structure was validated by both QMEAN and ProSA. For more validation predicted structures were checked by UCLA-DOE LAB - SAVES v6.0 server using several parameters including ERRAT, VERIFY, PROVE, and PROCHECK. Among designed fusion proteins, AraA-$(G_4S)_1$-Pep1 and AraA-A2b11 (without linker) have higher values of C-score model 1 of 3D structure predicted by I-TASSER. Moreover, these fusion proteins have high Q-mean scores with high-resolution crystal structures. By more validation of predicted structures, it was determined that AraA-A2b11 (without linker) had highly favored regions with 82.9% in the Ramachandran plot and had verified structure by Verify3D program with 92.38% of the residues and averaged 3D-1D score $\geq$ 0.2. Based on the protein scanning, all designed chimeric proteins are considered non-toxic and non-antigenic.

The best linker was considered to create the maximum distance between the two fused parts so that the folding of the two structures is close to the original structure. Based on I-TASSER results, among 33 linkers, only $(A(EAAAK)_4ALEA(EAAAK)_4A)_2$ linker was selected as an appropriate rigid linker in AraA-IL13 and AraA-IL13.E13K chimeric constructs. It is due to the huge structure of the AraA that requires maximum distance from the other parts of the molecule for the best folding and for the activity with enhanced stability. Furthermore, due to their small size and high flexibility, peptide ligands may lead to folding in the structure of AraA by fusing to this large enzyme. It may lead to reduced accessibility and prevent their binding to the receptor. Therefore, we selected the AraA-$(A(EAAAK)_4ALEA(EAAAK)_4A)_2$-IL13 fusion protein for further studies, because the native ligand (IL13) and two separate domains including $(A(EAAAK)_4ALEA(EAAAK)_4A)_2$ rigid hydrophobic linker improve the biological activity of the protein, increase the stability of the structure and lead to a high expression due to the correct protein folding (33,44). Tertiary structures of AraA-$(A(EAAAK)_4$ $ALEA(EAAAK)_4A)_2$-IL13 chimeric protein have two separate functional domains including IL13 as a specific ligand and AraA as an anticancer agent. The 3D structure of this chimeric protein was predicted by the I-TASSER, Robetta, and RaptorX servers with $(A(EAAAK)_4ALEA(EAAAK)_4A)_2$ linker that were approximately similar and each domain was analogous to the native component, indicating the preservation of the functional structure. Ghadaksaz *et al.* used the $A(EAAAK)_4ALEA(EAAAK)_4A$ linker for their construct (33), but we used two series of this linker so that two separate domains do not interfere with each other or are only minimally disturbed, compared to their native 3D structure.

The analysis of physicochemical parameters showed that this fusion protein had an acidic nature with an isoelectric point of 5.01. Our fusion protein had 708 amino acids and its molecular weight was estimated to be 74.76 KDa. Furthermore, this chimeric protein is probably non-antigenic and non-toxic.

The HawkDock server was used as a molecular docking method to model the interactions between proteins at the atomic level, which allows us to clarify the behavior of interface residues in the binding pocket of target proteins (37). Protein-protein complex with a high-affinity score was selected as the best pose mode. The MD simulations were used to evaluate the stability of the ligand (IL13) in the binding state. The stability of the ligand binding state in the binding pocket indicates the

accuracy and stability of the ligand binding state. To determine the stability of the ligand binding state in the binding pocket, parameters such as the Cα-RMSD and Cα-root-mean-square fluctuation (RMSF) values of the protein-protein complexes were calculated. Simulation can be performed to determine the ligand bindings to the receptor over time. RMSD analysis can reveal Cα atoms replacement in ligand binding states. By examining this parameter, it is possible to observe the changes in the complex state gradually during MD simulation times, and therefore, the structure corresponding to this change from the simulation film may be extracted.

The current study showed that after an increase in the RMSD graph, the structural fluctuations reach a plateau state which was confirmed by the evaluation of complexes up to 60 ns showing that the pattern of changes at 30 and 60 ns is similar. Thus, based on a comparison between the general and minor fluctuations of the ligand (IL13) structure in the monomeric and chimeric states incorporated into the IL13Rα2 junction envelope, it can be clearly seen that the ligand in the junction envelope exhibits a steady state and completely adapts to the binding site and does not dissociate. Also, the partial structural fluctuations of the IL13 in native and recombinant molecules showed that the pattern of each molecular species is almost similar, indicating the stability of the related molecular species during the simulation. Therefore, based on the results of the present *in silico* study, we predicted that AraA-$(A(EAAAK)_4ALEA(EAAAK)_4A)_2$-IL13 fusion protein tightly binds to IL13Rα2 through IL13 domain and can exhibit potent anticancer function through the AraA domain. In the next study phase, *in vivo,* the anticancer activity of this fusion protein will be investigated on different cell lines.

CONCLUSION

The main purpose of this study is to introduce a new potent fusion protein, which can be a candidate for cancer therapy. In order to find a new anti-cancer drug, we investigated the structure and function of the several fusion proteins that contain AraA for the elimination of cancer cells and a ligand that can be attached to the IL13Rα2 receptor by various bioinformatics tools. The results of this study demonstrated that AraA-IL13 fusion protein with native ligand was a stable chimeric protein with two separate domains and high affinity to IL13Rα2 receptor that is overexpressed in various human tumor cells. Therefore, the bioinformatics results of the present study demonstrated although some considerate fusions protein also show some desirable characteristics as well but AraA-$(A(EAAAK)_4ALEA(EAAAK)_4A)_2$-IL13 fusion protein could be a new potent candidate for *in vitro* and *in vivo* studies in order to target cancer therapy.

Acknowledgments

The authors would like to thank Dr. Abdolamir Ghadaksaz for his technical help.

Conflict of interest statement

The authors declared no conflict of interest in this study.

Authors' contribution

A. A. Imanifooladi and R. Halabian supervised the project; A. A Imanifooladi and H. Sedighian conceived the methodology and validation; R. Mehrab performed the experimental parts of the study; R. Mehrab, H. Sedighian, and R. Halabian analyzed the data; R. Mehrab wrote the original draft of the manuscript and provided grammatical revisions to the manuscript by F. Sotoodehnejadnematalah. The final version of the manuscript was approved by all authors.

REFERENCES

1. Maleki F, Sadeghifard N, Sedighian H, Bakhtiyari S, Hosseini HM, Imani Fooladi AA. TGFαL3-SEB fusion protein as an anticancer against ovarian cancer. Eur J Pharmacol. 2020;870:172919,1-8. DOI: 10.1016/j.ejphar.2020.172919.
2. Mohseni Moghadam Z, Halabian R, Sedighian H, Behzadi E, Amani J, Imani Fooladi AA. Designing and analyzing the structure of DT-STXB fusion protein as an anti-tumor agent: an *in silico* approach. Iran J Pathol. 2019;14(4): 305-312. DOI: 10.30699/ijp.2019.101200.2004.

3. Mazor R, Pastan I. Immunogenicity of immunotoxins containing pseudomonas exotoxin a: causes, consequences, and mitigation. Front Immunol. 2020;11:1261,1-12.
DOI: 10.3389/fimmu.2020.01261.
4. Kaplon H, Muralidharan M, Schneider Z, Reichert JM. Antibodies to watch in 2020. MAbs. 2020;12(1):1703531,1-55.
DOI: 10.1080/19420862.2019.1703531.
5. Havaei SM, Aucoin MG, Jahanian-Najafabadi A. *Pseudomonas* exotoxin-based immunotoxins: over three decades of efforts on targeting cancer cells with the toxin. Front Oncol. 2021;11:781800,1-17.
DOI: 10.3389/fonc.2021.781800.
6. Duque AED, Perekhrestenko T, Musteata V, Zodelava M, Guthrie TH, Strack T, *et al*. A phase II study of MT-3724, a novel CD20-targeting engineered toxin body, to evaluate safety, pharmacodynamics, and efficacy in subjects with relapsed or refractory diffuse large B-cell lymphoma. J Clin Oncol. 2020;38(15_suppl): TPS8074.
DOI: 10.1200/JCO.2020.38.15_suppl.TPS8074.
7. Kawai H, Ando K, Maruyama D, Yamamoto K, Kiyohara E, Terui Y, *et al*. Phase II study of E7777 in Japanese patients with relapsed/refractory peripheral and cutaneous T-cell lymphoma. Cancer Sci. 2021;112(6):2426-2435.
DOI: 10.1111/cas.14906.
8. Pereira FV, Ferreira-Guimaraes CA, Paschoalin T, Scutti JA, Melo FM, Silva LS, *et al*. A natural bacterial-derived product, the metalloprotease arazyme, inhibits metastatic murine melanoma by inducing MMP-8 cross-reactive antibodies. PloS One. 2014;9(4):e96141,1-11.
DOI: 10.1371/journal.pone.0096141.
9. Kwak J, Lee K, Shin DH, Maeng JS, Park DS, Oh HW, *et al*. Biochemical and genetic characterization of arazyme, an extracellular metalloprotease produced from Serratia proteamaculans HY-3. J Microbiol Biotechnol. 2007;17(5):761-768.
PMID: 18051297.
10. Pereira FV, Melo AC, de Melo FM, Mourao-Sa D, Silva P, Berzaghi R, *et al*. TLR4-mediated immunomodulatory properties of the bacterial metalloprotease arazyme in preclinical tumor models. Oncoimmunology. 2016;5(7):e1178420,1-52.
DOI: 10.1080/2162402x.2016.1178420.
11. Amjadi G, Parivar K, Mousavi SF, Imani Fooladi AA. Effect of metalloprotease arazyme on the expression of MMP2 and MMP9 genes in metastasis of colon and ovarian cancer cell lines. Thrita. 2020;8(2):e100004,1-6.
DOI: 10.5812/thrita.100004.
12. Ghadaksaz A, Imani Fooladi AA, Mahmoodzadeh Hosseini H, Nejad Satari T, Amin M. ARA-linker-TGFαL3: a novel chimera protein to target breast cancer cells. Med Oncol. 2021; 38(8): 96. DOI: 10.1007/s12032-021-01546-2.
13. Kim IS, Yang EJ, Shin DH, Son KH, Park HY, Lee JS. Effect of arazyme on the lipopolysaccharide induced inflammatory response in human endothelial cells. Mol Med Rep. 2014;10(2):1025-1029.
DOI: 10.3892/mmr.2014.2231.
14. Kim IS, Kim MJ, Shin DH, Son KH, Park HY, Lee JS. Arazyme inhibits cytokine expression and upregulates skin barrier protein expression. Mol Med Rep. 2013;8(2):551-556.
DOI: 10.3892/mmr.2013.1520.
15. Kim IS, Lee NR, Baek SY, Kim EJ, Kim JS, Jeong TS, *et al*. Inhibitory effect of arazyme on the development of atopic dermatitis-like lesions in BALB/c and Nc/Nga mice. Mol Med Rep. 2015;11(5):3995-4001.
DOI: 10.3892/mmr.2015.3225.
16. Yousefi F, Mousavi SF, Siadat SD, Aslani MM, Amani J, Rad HS, *et al*. Preparation and *in vitro* evaluation of antitumor activity of TGFalphaL3-SEB as a ligand-targeted superantigen. Technol Cancer Res Treat. 2016;15(2):215-226.
DOI: 10.1177/1533034614568753.
17. Imani Fooladi AA, Yousefi F, Mousavi SF, Amani J. *In silico* design and analysis of TGFalphaL3-SEB fusion protein as "a new antitumor agent" candidate by ligand-targeted superantigens technique. Iran J Cancer Prev. 2014;7(3):152-64.
18. Ahn SW, Lee CM, Kang MA, Hussein UK, Park HS, Ahn AR, *et al*. IL4Rα and IL13Rα1 are involved in the development of human gallbladder cancer. J Pers Med. 2022;12(2):249,1-20.
DOI: 10.3390/jpm12020249.
19. Mao YM, Zhao CN, Leng J, Leng RX, Ye DQ, Zheng SG, *et al*. Interleukin-13: a promising therapeutic target for autoimmune disease. Cytokine Growth Factor Rev. 2019;45:9-23.
DOI: 10.1016/j.cytogfr.2018.12.001.
20. Xie M, Wu XJ, Zhang JJ, He CS. IL-13 receptor α2 is a negative prognostic factor in human lung cancer and stimulates lung cancer growth in mice. Oncotarget. 2015;6(32):32902-32913.
DOI: 10.18632/oncotarget.5361.
21. Sharma P, Sonawane P, Herpai D, D'Agostino R, Rossmeisl J, Tatter S, *et al*. Multireceptor targeting of glioblastoma. Neuro Oncol Adv. 2020; 2(1):1-11.
DOI: 10.1093/noajnl/vdaa107.
22. Pandya H, Gibo DM, Garg S, Kridel S, Debinski W. An interleukin 13 receptor α 2-specific peptide homes to human Glioblastoma multiforme xenografts. Neuro Oncol. 2012;14(1):6-18.
DOI: 10.1093/neuonc/nor141.
23. Le Joncour V, Laakkonen P. Seek & Destroy, use of targeting peptides for cancer detection and drug delivery. Bioorg Med Chem. 2018;26(10):2797-2806.
DOI: 10.1016/j.bmc.2017.08.052.
24. Kurihara R, Horibe T, Shimizu E, Torisawa A, Gaowa A, Kohno M, *et al*. A novel interleukin-13 receptor alpha 2-targeted hybrid peptide for effective glioblastoma therapy. Chem Biol Drug Des. 2019;94(1):1402-1413.
DOI: 10.1111/cbdd.13517.
25. Ataee MH, Mirhosseini SA, Mirnejad R, Rezaie E, Mahmoodzadeh Hosseini H, Amani J. Design of two immunotoxins based rovalpituzumab antibody

against DLL3 receptor; a promising potential opportunity. Res Pharm Sci. 2022;17(4):428-444.
DOI: 10.4103/1735-5362.350243.
26. Rouhani M, Valizadeh V, Ahangari Cohan R, Norouzian D. Computational design, structure refinement and molecular dynamics simulation of novel engineered serratiopeptidase analogs. J Biomol Struct Dyn. 2019;37(16):4171-4180.
DOI: 10.1080/07391102.2018.1540361.
27. Jahangiri A, Amani J, Halabian R, Imani Fooladi AA. *In Silico* analyses of *Staphylococcal* enterotoxin B as a DNA vaccine for cancer therapy. Int J Pept Res Ther. 2018;24(1):131-142.
DOI: 10.1007/s10989-017-9595-3.
28. Rezaie E, Bidmeshki Pour A, Amani J, Mahmoodzadeh Hosseini H. Bioinformatics predictions, expression, purification and structural analysis of the PE38KDEL-scfv immunotoxin against EPHA2 receptor. Int J Pept Res Ther. 2020;26(2):979-996.
DOI: 10.1007/s10989-019-09901-8.
29. Zhang Y, Skolnick J. TM-align: a protein structure alignment algorithm based on the TM-score. Nucleic Acids Res. 2005;33(7):2302-2309.
DOI: 10.1093/nar/gki524.
30. Källberg M, Wang H, Wang S, Peng J, Wang Z, Lu H, *et al*. Template-based protein structure modeling using the RaptorX web server. Nat Protoc. 2012;7:1511-1522.
DOI: 10.1038/nprot.2012.085.
31. Kim DE, Chivian D, Baker D. Protein structure prediction and analysis using the Robetta server. Nucleic Acids Res. 2004;32(Web Server issue):W526-W531.
DOI: 10.1093/nar/gkh468.
32. Bhattacharya D, Nowotny J, Cao R, Cheng J. 3Drefine: an interactive web server for efficient protein structure refinement. Nucleic Acids Res. 2016;44(W1): W406-W409.
DOI: 10.1093/nar/gkw336.
33. Ghadaksaz A, Imani Fooladi AA, Mahmoodzadeh Hosseini H, Nejad Satari T, Amin M. Targeting the EGFR in cancer cells by fusion protein consisting of arazyme and third loop of TGF-alpha: an *in silico* study. J Biomol Struct Dyn. 2021;1-14.
DOI: 10.1080/07391102.2021.1963318.
34. Saha S, Raghava GP. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. Proteins. 2006;65(1):40-48.
DOI: 10.1002/prot.21078.
35. Saha S, Raghava GPS. BcePred: prediction of continuous B-cell epitopes in antigenic sequences using physico-chemical properties. In: Nicosia G, Cutello V, Bentley PJ, Timmis J, editors. artificial immune systems. Springer, Berlin, Heidelberg. 2004;3239. pp. 197-204.
DOI: 10.1007/978-3-540-30220-9_16.
36. Gupta S, Kapoor P, Chaudhary K, Gautam A, Kumar R, Open Source Drug Discovery Consortium, *et al*. *In silico* approach for predicting toxicity of peptides and proteins. PloS One. 2013;8(9):e73957,1-10.
DOI: 10.1371/journal.pone.0073957.
37. Weng G, Wang E, Wang Z, Liu H, Zhu F, Li D, *et al*. HawkDock: a web server to predict and analyze the protein-protein complex based on computational docking and MM/GBSA. Nucleic Acids Res. 2019;47(W1):W322-W330.
DOI: 10.1093/nar/gkz397.
38. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ. GROMACS: fast, flexible, and free. J Comput Chem. 2005;26(16):1701-1718.
DOI: 10.1002/jcc.20291.
39. Guvench O, MacKerell Jr AD. Comparison of protein force fields for molecular dynamics simulations. Methods Mol Biol. 2008;443:63-88.
DOI: 10.1007/978-1-59745-177-2_4.
40. Genheden S, Ryde U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. Expert Opin Drug Discov. 2015;10(5):449-461.
DOI: 10.1517/17460441.2015.1032936.
41. Poorhassan F, Nemati F, Saffarian P, Mirhosseini SA, Motamedi MJ. Design of a chitosan-based nano vaccine against epsilon toxin of *Clostridium perfringens* type D and evaluation of its immunogenicity in BALB/c mice. Res Pharm Sci. 2021;16(6):575-585.
DOI: 10.4103/1735-5362.327504.
42. Shokri B, Zarghi A, Shahhoseini S, Mohammadi R, Kobarfard F. Design, synthesis and biological evaluation of peptide-NSAID conjugates for targeted cancer therapy. Arch Pharm (Weinheim). 2019;352(8): e1800379,1-8.
DOI: 10.1002/ardp.201800379.
43. Tu M, Wange W, Cai L, Zhu P, Gao Z, Zheng W. IL-13 receptor α2 stimulates human glioma cell growth and metastasis through the Src/PI3K/Akt/mTOR signaling pathway. Tumour Biol. 2016;37(11):14701-14709.
DOI: 10.1007/s13277-016-5346-x.
44. Patel DK, Menon DV, Patel DH, Dave G. Linkers: a synergistic way for the synthesis of chimeric proteins. Protein Expr Purif. 2021;191:106012.
DOI: 10.1016/j.pep.2021.106012.