Targeting the EGFR in cancer cells by fusion protein consisting of arazyme and third loop of TGF-alpha: an in silico study

Abdolamir Ghadaksaz, Abbas Ali Imani Fooladi, Hamideh Mahmoodzadeh Hosseini, Taher Nejad Sataria, and Mohsen Amin

Department of Biology, Science and Research Branch, Islamic Azad University, Tehran, Iran; Applied Microbiology Research Center, Systems Biology and Poisonings Institute, Baqiyatallah University of Medical Sciences, Tehran, Iran; Department of Drug and Food Control, Faculty of Pharmacy, and The Institute of Pharmaceutical Sciences, Tehran University of Medical Sciences, Tehran, Iran

Communicated by Ramaswamy H. Sarma

ABSTRACT

The anticancer effects of arazyme, a bacterial metalloprotease, have been revealed in previous studies. Because of the overexpression of epidermal growth factor receptor (EGFR) in tumor cells, targeting this receptor is one of the approaches to cancer therapy. In the present study, we designed fusion protein by using a non-mitogenic binding sequence of TGFα, arazyme, and a suitable linker. The I-TASSER and Robetta web servers were employed to predict the territory structures of the Arazyme-linker-TGFα-L3, and TGFα-L3-linker-Arazyme. Then, models were validated by using PROCHECK, ERRAT, ProSA, and MolProbity web servers. After docking to EGFR, Arazyme-linker-TGFα-L3 showed a higher binding affinity and was selected to be optimized through 100 ns Molecular dynamic (MD) simulation. Next, the stability of ligand-bound receptor was examined utilizing MD simulation for 100 ns. Furthermore, the binding free energy calculation and free energy decomposition were carried out employing MM-PBSA and MM-GBSA methods, respectively. The root mean square deviation and fluctuation (RMSD, RMSF), the radius of gyration, H-bond, and binding free energy analysis revealed the stability of the complex during simulation time. Finally, linear and conformational epitopes of B cells, as well as MHC class I and MHC class II were predicted by using different web servers. Meanwhile, the potential B cell and T cell epitopes were identified throughout arazyme protein sequence. Collectively, this study suggests a novel chimera protein candidate prevent cancer cells potentially by inducing an immune response and inhibiting cell proliferation.

KEYWORDS

Arazyme; EGFR; TGFα; targeted drug delivery; molecular dynamics simulation

1. Introduction

Targeted drug delivery which is a new aspect of anti-cancer drug design shows an optimized therapeutic index owing to the focused drug activity on the site of tumor cells. By selecting a suitable ligand as a specific cancer cell binder, the targeted receptor on the only cancer cell surface is recognized and the carried drug is delivered to the cells. As a result, the normal cells are not recognized (Das et al., 2009).

Several ligands such as TGFα and EGF can activate the epidermal growth factor receptor (EGFR). EGFR induces the important cell process, cell proliferation, differentiation, survival, and motility which are related to the activation of different signaling cascades, including Ras/Raf/MAPK, PI3K/Akt, and phospholipase Cγ pathways (Appert-Collin et al., 2015). High density and overexpression of epidermal growth factor receptor (EGFR) and its related ligands were demonstrated in tumor cells, as well (Wykosky et al., 2011; Yotsumoto et al., 2009). Therefore, EGFR-targeted therapeutics can overcome tumor cell growth and its resistance to apoptosis (Coleman & Tsongalis, 2016; Wykosky et al., 2011). EGFR (HER1 or ErbB1) which belongs to the ErbB family of tyrosine kinase receptors comprises L1, L2, and CR1 as three structural domains. A back-to-back dimer interaction by the CR1 domain of each receptor was suggested. The mature structure of TGFα consists of 50 amino acids, from which the residues 15–22 and
In the current study, we designed fusion proteins, Arazyme-linker-TGFαL3, and TGFαL3-linker-Arazyme to target cancer cells by considering the potential anti-cancer activity of arazyme and binding affinity of the third disulfide loop of TGFα to EGFR. The three-dimensional structure of fusion proteins was predicted and validated. Then, the chimeric proteins were docked to the receptor and their binding affinities were compared. Afterward, MD simulation was employed to explore interactions in atomic details. Finally, the immunological properties were evaluated.

2. Materials and method

2.1. Fusion protein construction

The protein sequence of arazyme (Accession No. AA21094) and four isoforms of human protransforming growth factor-alpha isoform preproprotein (Accession No. NP_003227, NP_001093161, NP_001295087 XP_005264579, and NP_001295088 XP_011531378) have been obtained from the NCBI protein database (https://www.ncbi.nlm.nih.gov/protein/). All isoforms have an identical amino acid sequence in the third loop of human TGFα by using multiple sequence alignment. The Arazyme-linker-TGFαL3 and TGFαL3-linker-Arazyme chimeric sequences were constructed with A (EAAAK)4ALEA(EAAAK)4A linker (Figure 1).

2.2. Protein physicochemical analysis

Expasy’s ProtParam (https://web.expasy.org/protparam/) was utilized to calculate molecular weight, protein half-time, Instability index, Aliphatic index, Grand average of hydropathicity (GRAVY), and theoretical pi (Gasteiger, 2005).

2.3. Protein structure prediction

Tertiary structure models of Arazyme, Arazyme-linker-TGFαL3, and TGFαL3-linker-Arazyme were built utilizing I-TASSER (https://zhanglab.ccbmed.uchicago.edu/I-TASSER/) and Robetta (https://robetta.bakerlab.org/). I-TASSER and Robetta servers employ ab initio modeling if non-homologous regions are found by using the threading alignment method. Furthermore, structure refinement is performed to obtain high-resolution models. The I-TASSER models are selected based on C-score ranged from −5 to 2, which the best model has a higher c-score. Meanwhile, Rosetta all-atom energies are employed to rank the Robetta models (Raman et al., 2009; Roy et al., 2010; Song et al., 2013).

Then, by using UCSF Chimera and Discovery Studio Visualizer softwares, the best-predicted models were aligned to both sequence and secondary structure of the template used in homology modeling. In addition to RMSD, UCSF Chimera calculates Structural Distance Measure (SDM) and the Q-score to yield more accurate sequence alignment. The SDM value equal to one indicates the identical structures and the higher values show the lower similarity, while the higher Q-scores (ranged between zero and one) values mean the stronger similarity between structures (Johnson et al., 2010).
A. GHADAKSAZ ET AL.

1990; Krissinel & Henrick, 2004). Furthermore, the similarity between predicted structures was evaluated using the IPBA web server (https://www.dsimb.inserm.fr/dsimb_tools/ipba/). This server compares protein structures by the improved version of Protein Blocks alignment and provides RMSD whose lowest value indicates greater conformational similarity between the proteins (Gelly et al., 2011).

2.4. Protein structure validation

The quality of generated models was validated with Qmean (https://swissmodel.expasy.org/qmean/), PROCHECK Ramachandran plot (https://services.mbi.ucla.edu/PROCHECK), ERRAT (https://services.mbi.ucla.edu/ERRAT/), ProSA (https://ProSA.services.came.sbg.ac.at/ProSA.php) and MolProbity (http://molprobity.biochem.duke.edu/) servers. Qmean expresses Z-scores by comparing protein structure models with scores calculated from high-resolution crystal structures (Benkert et al., 2011). ERRAT uses a statistics database of six pairwise bonded interactions from a nine-residue sliding window in high reliable protein structures to analyze errors of query structure models (Colovos & Yeates, 1993). ProSA obtains the potential of mean force of the known protein structures to validate 3D models of protein structures (Wiederstein & Sippl, 2007). MolProbity checks the quality of macromolecules based on both global and local levels and analyses contacts and geometry of all-atoms by evaluating Ramachandran, rotamer, and Covalent-geometry (Prisant et al., 2020). Several stereochemical parameters, such as dihedral angles and chirality are utilized by Procheck in addition to Ramachandran-based structure validation (Laskowski et al., 1993).

2.5. Molecular docking

At first, to compare the binding energy of different regions of TGFα with EGFR, protein structures were prepared to employ the software MGL Tools 1.5.6 and Discovery Studio Visualizer. To prepare the ligand structure and EGFR, polar hydrogen and Kolleman charges were added and then all heteroatoms and water molecules were removed. The grid box was set to cover the entire ligand surface. Then docking was carried out by autodock vina software (Trott & Olson, 2010). Afterwards, the interaction of the 3D structure model of chimera protein with EGFR was docked using HADDOCK 2.4 (https://wenmr.science.uu.nl/haddock2.4/), and ClusPro 2.0 (https://cluspro.bu.edu/login.php) servers.

ClusPro web server performs rigid docking using a Fast Fourier Transform (FFT) correlation method and then, finds the most likely models of ligand-receptor interactions that are finally refined by energy minimization. This server yields the top ten largest docked structures with the lowest docking energy clustered by Weighted Score based on interface RMSD between the number of neighbors within 9-Å radii and three parameters, including native complex, the average number of docked structures within less than 10-Å radius and the average of lowest server archived IRMSD. Therefore, the best cluster has the highest Weighted Score (Kozakov et al., 2013, 2017; Vajda et al., 2017). The focused docking approach was conducted by HADDOCK 2.4 to obtain more accurate results followed by three stages of energy minimizing and final refinements. The top largest clusters were sorted based on haddock score which is calculated by the weighted sum of individual energy terms, including electrostatic, van der Waals, desolvation, AIP energies, and a buried surface area (BSA) (van Zundert et al., 2016). The active residues of ligand and receptor were defined based on the crystal structure of Egr-TGF alpha complex while the other surface residues with 6.5 Å distance from active residues were automatically mentioned as passive residues by the HADDOCK protocol (Garrett et al., 2002; van Zundert et al., 2016). Then, the binding affinity of x-ray crystallography complex of TGFα and EGFR was evaluated.

Afterward, a model of the ligand with the highest score of binding to the receptor was selected to be optimized via MD simulation. Next, the average structures relevant to the stable part of the trajectory were subject to MD simulation of ligand-receptor complex.

2.6. MD simulations of ligand and ligand-receptor complex

The GROMACS® 2019.6 software was used to perform MD simulations with the amber 99SB + ILDN force field. The protein was solvated in TIP3P explicit water model in a cubic box and the distance between the solute and each side of the box was kept at 0.8 nm. The sodium (30 Na+) counter ions were added to maintain the neutrality of the system. Energy minimization was performed for 50,000 steps, using the steepest descent method, to prevent inappropriate geometry and steric clashes. Then, the system was equilibrated under NVT ensemble at 310K for 200 ps using V-rescale method followed by NPT ensemble at 1 bar for 200 ps using Parrinello–Rahman barostat (Bussi et al., 2007; Parrinello & Rahman, 1981). In both phases of system equilibration, the positions of all heavy atoms were restrained by a force constant of 1000 kJ/mol/nm. Ultimately, the full dynamic production was performed for 100 ns under NPT condition (T= 310 K, p = 1 bar). The LINCS algorithm was used to constrain all covalent bond lengths allowing for a 2.0 fs time step (Hess et al., 1997).

The output of the MD simulation was analyzed using tools that were included in the Gromacs package. The changes in the ligand secondary structure during the simulation were estimated by DSSP software (Kabsch & Sander, 1983). To assess the stability of the fusion protein (ligand) and ligand-receptor complex, multiple parameters, including root-mean-square deviation (RMSD), root-mean-square fluctuation (RMSF), and radius of gyration (Rg) were determined. The number of hydrogen bonds at the ligand-receptor interface was analyzed. Furthermore, every 10 ps of the stable time of simulation was extracted to calculate the end-point binding free energy and perform the energy decomposition analysis using Molecular Mechanic/Poisson–Boltzmann Surface Area (MM-PBSA) and Molecular Mechanics Generalized Born Surface Area (MM-GBSA) methods, respectively (Miller et al., 2012; Tresanco, 2021). Based on the MM-PBSA or MM-GBSA
methodology, the binding free energy was calculated using the following equations (Chen et al., 2016; Massova & Kollman, 1999):

$$\Delta G_{\text{bind}} = G_{\text{complex}} - (G_{\text{protein}} + G_{\text{ligand}})$$

$$= \Delta E_{\text{MM}} + \Delta G_{\text{sol}} + T\Delta S$$

The binding free energies ($\Delta G_{\text{bind}}$) was determined by the total free energy of complex ($\Delta G_{\text{complex}}$), the energy of unbound ligand ($\Delta G_{\text{ligand}}$), and receptor ($\Delta G_{\text{receptor}}$). The Molecular Mechanics Energies in the gas-phase ($\Delta E_{\text{MM}}$) contain electrostatic ($\Delta E_{\text{elec}}$) and the van der Waals energies ($\Delta E_{\text{vdw}}$). The solvation free energies ($\Delta G_{\text{sol}}$) include polar solvation ($\Delta G_{\text{polar}}$) and nonpolar solvation ($\Delta G_{\text{nonpolar}}$) energies. The contribution of conformational entropy ($-T\Delta S$) was not considered in the present study (Chen et al., 2016; Massova & Kollman, 1999).

The Xmgrace tool was used to plot all graphs and the UCSF Chimera program was employed to visualize the results. Furthermore, the LigPlus (https://www.ebi.ac.uk/thornton-srv/software/LigPlus/) was employed to depict the H-bonding and hydrophobic interactions (Laskowski & Swindells, 2011).

### 2.7. B cell and T cell epitope and allergenic sites prediction

For linear and discontinuous B cell epitope prediction BepiPred-2.0 (http://www.cbs.dtu.dk/services/BepiPred/), IEDB server (including Kolaskar and Tongaonkar (http://tools.iedb.org/bcell/), and ElliPro (http://tools.iedb.org/ellipro/) methods, and DiscoTope web server (http://www.cbs.dtu.dk/services/DiscoTope/) were employed (Jespersen et al., 2017; Kolaskar & Tongaonkar, 1990; Kringelum et al., 2012; Ponomarenko et al., 2008). To predict epitopes by BepiPred-2.0, the threshold of 0.5 was set. To analyze B cell epitopes, the window size was set to seven using Kolaskar and Tongaonkar method, and the sequences with an average threshold of 1.008 were defined as epitopes.

The distinct discontinuous epitope defined by ElliPro is three repeated clustering steps based on the protrusion index $S$ and the distance $R$ between each residue’s centers of mass. The minimum score and maximum distance (Angstrom) are set to 0.5 and 6 Å, respectively (Ponomarenko et al., 2008). To get the highest sensitivity and specificity values for epitope identification via DiscoTope the threshold was set to $-3.7$. Surface accessibility of epitopes was evaluated by the Emini surface accessibility scale server (Emini et al., 1995). The flexibility and hydrophilicity were evaluated by Karplus and Schulz and Parker methods, respectively (http://tools.iedb.org/bcell/) (Karplus & Schulz, 1985; Parker et al., 1986).

The sequence of chimera protein was loaded to the IEDB server to predict the T cell epitope by PickPocket and SMM methods (http://tools.iedb.org/mhci/) (Peters & Sette, 2005; Zhang et al., 2009). Also, SYFPEITHI (http://www.syfpeithi.de/bin/MHCServer.dll/EpitopePrediction.htm) and Netmhcpan 4.1 (http://www.cbs.dtu.dk/services/NetMHCPan/) servers were used to find every sequence that can bind with all known MHC I and MHC II alleles. Furthermore, NetMHCIpan servers (http://www.cbs.dtu.dk/services/NetMHCIpan/) were applied to MHCII prediction (Rammensee et al., 1999; Reynisson et al., 2020). The allergenicity of chimera protein was evaluated by using SDAP (https://fermi.utmb.edu/) and algpred (http://www.imtech.res.in/raghava/algpred/) servers (Ivanciuc et al., 2003; Saha & Raghava, 2006). The antigenicity of the fusion protein was evaluated through the Vaxijen web server (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html) which uses an alignment-independent algorithm to predict protective antigens (Doytchinova & Flower, 2007). For all tools, the default setting was selected.

### 3. Results

#### 3.1. Structure analysis

The structure of arazyme is composed of N-terminal zinc-dependent metalloprotease and C-terminal Peptidase_M10_C domains. The two designed fusion protein constructs have the molecular weight of 59,955.38 Da containing the 504 amino acid fragments of arazyme, a fragment of 17 amino acid fragments of TGF$\alpha$ (third loop), and a rigid linker (A(EAAAK)4ALEA(EAAAK)4A). The physicochemical properties are presented in Table 1.

#### 3.2. 3D structure prediction and qualification

Five models were built by each of the I-TASSER and Robetta servers. Among templates used by Robetta and I-TASSER to construct the models, the crystal structure of metalloprotease from S. Marcescens (PDB id: 1af0) showed the greatest similarity based on the comparison of RMSD, Q and SDM values (Table S1). The multiple sequence and structural alignment between Robetta models and the metalloprotease of *Serratia* are shown in Figure 2. To find the number of residues in favored, allowed, and disallowed regions, the Ramachandran plot analysis was performed by Procheck and MolProbity. For each chimera, the model with the better score and the most residues in the favored region was subjected to Ligand-receptor docking. The three-dimensional structure of Arazine-linker-TGF$\alpha$L3, and TGF$\alpha$L3-linker-Arazyme, predicted by Robetta was selected as the best model because of the highest residues percentage in the favored region according to PROCHECK and MolProbity Ramachandran distributions (Table 2). Also, the models were qualified by Qmean, ProSA, and ERRAT. The Robetta models show the highest ERRAT percentage quality factor. Based on ERRAT, a quality factor higher than 95% has a good high resolution (not higher than 3 Å). Also, a higher ProSA and Qmean score revealed that all Robetta models are closer to X-ray structures. The alignment made by the IPBA webserver between the arazyme predicted model and each one of the fusion proteins shows that Robetta models obtain lower RMSD compared to I-TASSER models for the two fusion proteins (Table S2).

#### 3.3. Ligand-receptor docking

The interaction energy between different regions of TGF$\alpha$ with the receptor was calculated by autodock vina software
Table 1. Some physicochemical properties of Arazyme-linker-TGFαL3, TGFαL3-linker-Arazyme, TGFαL3, and arazyme.

| Construct                  | Theoretical pI | Estimated half-life * (per hour) | Instability index | Aliphatic index: Grand average of hydropathicity (GRAVY) |
|----------------------------|----------------|----------------------------------|-------------------|-------------------------------------------------------|
| Arazyme-linker-TGFαL3      | 4.73           | 30                               | 24.90             | 69.17                                                 | −0.366                                      |
| TGFαL3-linker-Arazyme      | 4.73           | 100                              | 24.90             | 69.17                                                 | −0.366                                      |
| TGFαL3                     | 5.99           | 100                              | 4.18              | 91.76                                                 | 0.224                                      |
| Arazyme                    | 4.59           | 30                               | 26.94             | 68.39                                                 | −0.403                                      |

*In mammalian reticulocytes, in vitro.

Figure 2. Sequence alignment of (a) Arazyme-linker-TGFαL3 and (b) TGFαL3-linker-Arazyme with the sequence of the metalloprotease from S. marcescens. The dark blue indicates the identical residues and light blue shows similar residues. The structural alignment between the crystal structure of the metalloprotease from S. Marcescens (red) and Robetta predicted models (blue), (c) Arazyme-linker-TGFαL3, and (d) TGFαL3-linker-Arazyme. The linker and TGFαL3 fragments are represented in green and pink, respectively.
cates fluctuations in compactness during the initial 70,000 ps, the protein. The radius of the Gyration value of protein indicated have higher flexibility compared to other regions of

The change in the number of hydrogen bonds was measured during MD simulation. As shown in Figure 5d, four H-bond was formed by the complex at initial of MD simulation which was increased to about 11 hydrogen bond at 2000 ps and then decreased to about 3 at 4000 ps. Afterward, the number of formed H-bond fluctuated between about 3 to 13 which remained up to 45,000 ps and then increased to about 15 at 60,000 ps. After this time, the hydrogen bond decreased to reach 10 at end of simulation time (Figure 6).

The contribution of van der Waals, electrostatic, and the polar and non-polar solvations to the MMPBSA binding free energies was presented in Table 4. Also, Figure S1 shows the MMGBSA binding free energy decomposition analysis.

3.6. T cell epitope prediction

Netmhcpan 4.1 presents a rank to choose the best epitopes by the comparison of random natural peptides. The threshold of less than 0.5 was expressed as a strong binder. The epitope sequence, which has the highest affinity and the

Table 2. The Ramachandran distributions of predicted models of arazyme, TGF-xL3, and fusion proteins analyzed by PROCHECK and MolProbity webservers.

|                  | Arazyme | TGF-xL3 | Arazyme-linker-TGF-xL3 | TGF-xL3-Linker-Arazyme |
|------------------|---------|---------|------------------------|------------------------|
| PROCHECK Ramachandran distributions | I-TASSER | Robetta | I-TASSER | Robetta | I-TASSER | Robetta | I-TASSER | Robetta |
| (most favored, additional allowed, generously allowed and disallowed regions, respectively) | 75.3% | 90.9 | 69.2 | 67.6 | 91.2 | 67.2 | 91.2 |
| MolProbity rotamers analysis | 7.85 | 0 | 0 | 8.45 | 0.00 | 8.69 | 0.00 |
| (poor and favored respectively) | 83.54 | 99.49 | 100 | 77.23 | 100.00 | 81.69 | 99.77 |
| (outlier and favored respectively) | 81.67 | 98.21 | 80 | 76.46 | 98.41 | 10.27 | 98.23 |

(Table S3). Meanwhile, the two fusion proteins were docked to the EGFR by using ClusPro and haddock servers. Haddock and ClusPro results showed a higher affinity of the fusion protein, Arazyme-linker-TGF-xL3 to EGFR (Table 3).

3.4. MD trajectory analysis of Arazyme-linker-TGF-xL3

The trajectories generated from the 100 ns MD were used to evaluate the stability of Arazyme-linker-TGF-xL3. As shown in Figure 3a, the DSSP analysis of Arazyme-linker-TGF-xL3 revealed that the secondary structure of amino acids in most areas of protein remains conserved during the simulation period and β-sheets are the most stable structures, although changes have occurred as follows: The residues of 21–32 showed a conformational change from α-helix to bend and coil conformation; the residues of 438–443 undergoes a conformational change from α-helix to the coil; and the residues of 520–540 exhibited a conformational change from α-helix to bend conformation. Totally, the content of α-helix declined from 24.6% to 11.1% while the content of coil and bend increased from 41.9% to 52%.

The RMSD plot (Figure 3b) for protein revealed a quick increase in RMSD value up to 0.4 nm at the initial 4 ps of simulation. Afterward, a gradual increase in RMSD value was observed that reached up to 0.5 nm after 35,000 ps, and a sudden increase was seen up to 0.7 nm at 48,000 ps. Then, the value showed a decreasing trend to reach 0.55 nm at 65,000 ps which fluctuating around this value during simulation time. Then, from 85,000 ps to the end, the value slightly increased to reach 0.6 nm. Therefore, the stable state was considered to be between 70,000 and 100,000 ps.

The RMSF plot of the Cα atoms against residues number of protein was generated from the stable state of RMSD. As shown in Figure 3c, the terminal residues in N- and C- terminal have higher flexibility compared to other regions of the protein. The radius of the Gyration value of protein indicates fluctuations in compactness during the initial 70,000 ps, which then remain relatively constant till the end of the simulation (Figure 3d).

3.5. MD simulation of ligand-receptor interactions

The average MD structure of Arazyme-linker-TGF-xL3 obtained from 100 ns simulation was docked against EGFR using HADDOCK 2.4. The best cluster with the largest cluster size (62 structures with a similar binding site) and the highest HADDOCK score (−111.2) was subjected to MD simulation (Table S4). To investigate the stability and dynamic, MD simulation for 100 ns was performed. Figure 4 shows the superimposition of the binding pose of Arazyme-linker-TGF-xL3 in docking with EGFR at 10 ns intervals.

The changes of RMSD value of ligand and receptor and their interaction were shown in Figure 5a. The complex revealed a sharp increase in RMSD value at the first 25,000 ps to about 0.6 nm which fluctuated slightly around this value until the end of the simulation. The stable state of the ligand-receptor complex was considered 70,000 ps, according to the RMSD plot.

The RMSF of the ligand with the time before and after docking with the receptor was measured based on 70 ns of trajectory data. As shown in Figure 5b, the docked ligand showed less fluctuations in residues 510–567 compared to ligand alone while the RMSF value of other residues is similar between docked and unbounded ligand. The compactness of the complex was measured using the Radius of gyration (Figure 5c). At first 4000 ps, the Rg increased to 3.725 nm, thereafter it dropped to 3.57 nm at about 6000 ps which then continuously fluctuated around 3.6 nm till the end of the simulation.

The average MD structure of Arazyme-linker-TGF-xL3 obtained from 100 ns simulation was docked against EGFR using HADDOCK 2.4. The best cluster with the largest cluster size (62 structures with a similar binding site) and the highest HADDOCK score (−111.2) was subjected to MD simulation (Table S4). To investigate the stability and dynamic, MD simulation for 100 ns was performed. Figure 4 shows the superimposition of the binding pose of Arazyme-linker-TGF-xL3 in docking with EGFR at 10 ns intervals.

The changes of RMSD value of ligand and receptor and their interaction were shown in Figure 5a. The complex revealed a sharp increase in RMSD value at the first 25,000 ps to about 0.6 nm which fluctuated slightly around this value until the end of the simulation. The stable state of the ligand-receptor complex was considered 70,000 ps, according to the RMSD plot.

The RMSF of the ligand with the time before and after docking with the receptor was measured based on 70 ns of trajectory data. As shown in Figure 5b, the docked ligand showed less fluctuations in residues 510–567 compared to ligand alone while the RMSF value of other residues is similar between docked and unbounded ligand. The compactness of the complex was measured using the Radius of gyration (Figure 5c). At first 4000 ps, the Rg increased to 3.725 nm, thereafter it dropped to 3.57 nm at about 6000 ps which then continuously fluctuated around 3.6 nm till the end of the simulation.

The change in the number of hydrogen bonds was measured during MD simulation. As shown in Figure 5d, four H-bond was formed by the complex at initial of MD simulation which was increased to about 11 hydrogen bond at 2000 ps and then dropped to about 3 at 4000 ps. Afterward, the number of formed H-bond fluctuated between about 3 to 13 which remained up to 45,000 ps and then increased to about 15 at 60,000 ps. After this time, the hydrogen bond decreased to reach 10 at end of simulation time (Figure 6).

The contribution of van der Waals, electrostatic, and the polar and non-polar solvations to the MMPBSA binding free energies was presented in Table 4. Also, Figure S1 shows the MMGBSA binding free energy decomposition analysis.

3.6. T cell epitope prediction

Netmhcpan 4.1 presents a rank to choose the best epitopes by the comparison of random natural peptides. The threshold of less than 0.5 was expressed as a strong binder. The epitope sequence, which has the highest affinity and the

Table 3. The docking of TGF-x, Arazyme-linker-TGF-xL3, and TGF-xL3-linker-Arazyme with EGFR using Haddock and Cluspro servers.

| Docking server | TGF-x | Arazyme-linker-TGF-xL3 | TGF-xL3-Linker-Arazyme |
|----------------|-------|------------------------|------------------------|
| Haddock score  | −166.4 | −71.0 | −42.0 |
| Cluspro score  | −975.3 | −678.5 | −668 |
lowest rank in each allele serotype was selected based on Netmhcpan prediction. The IC50 value less than 50 nm was mentioned as a strong binding affinity by PickPocket and SMM servers but SYFPEITHI gives a score to each residue of epitope oligomer based on the frequency of amino acids in natural ligands. For MHCI, the best five predicted epitopes from different alleles serotypes were shown in Table S5 and the related sequences were highlighted in Figure 7. For MHCII, the five epitopes identified by NetMHCIIpan and SYFPEITHI were displayed in Table S6 and Figure 7.

3.7. B cell epitope prediction and allergenicity

The five high-ranked linear epitopes are predicted by ellipro, as well as Kolaskar and Tongaonkar method whose physicochemical properties are shown in Table 5. Based on the ElliPro, Kolaskar, and Tongaonkar method, eight linear B cell epitopes with the length ranged from 11 to 17 residues and 17 epitopes with the length ranged from 6 to 20 residues were identified, respectively. All the above three tools identify the residues 476–481 (EALLSY) and 487–494 (VSDLALNI). The probability average of EALLSY and VSDLALNI obtained from BepiPred-2.0 were 0.561 and 0.509, respectively. Furthermore, the VSDLALNI sequence obtains the highest Vaxijen score (1.1432) among all the best-predicted B cell epitopes.

Also, 89 and 326 residues were defined as conformational B-cell epitopes by DiscoTope and ellipro, respectively (Figure 8).
SDAP and algpred servers showed no potential allergen similar to any allergens region in their library.

4. Discussion

Ligand-based therapy by targeting the epidermal growth factor receptor (EGFR) is an approach to design anti-cancer drugs. Increased level of EGFR in various tumor cells makes it a candidate in the subject of designing anticancer drugs (Lim et al., 2017; Yousefi et al., 2016). In the current study, we designed two fusion proteins Arazyme-linker-TGFαL3 and TGFαL3-linker-Arazyme using arazyme as an effective anti-cancer protein and non-mitogenic fragment (loop three) of TGFα as a ligand of EGFR (Coleman & Tsongalis, 2016). A rigid 46 amino acid linker (AEAAAK)4ALEA(EAAAK)4A which includes eight helix-forming motifs was selected to increase stability as well as improving the biological activity of fusion protein (Chen et al., 2013). The increased expression level of the fusion protein with helical linkers is due to its secondary structure that is resistant to enzymatic cleavage. Also, the true folding of fusion proteins is enhanced by helical linkers resulting in higher stability and expression level (Amet et al., 2009).

The physicochemical analysis of two chimera proteins is similar to that of arazyme. Since different residues of the N-terminal of protein determine the stability of protein, ProtParam estimates the stability of 30 h for arazyme and Arazyme-linker-TGFαL3 based on the N-end rule. But TGFαL3-linker-Arazyme showed stability of 100 h owing to Amino-terminal valine which is responsible for the highest stability among other amino acids (Gonda et al., 1989). Moreover, the instability index of all proteins is <40, and the high aliphatic index, which is the relative volume of aliphatic side chains of protein, indicates that the stability of designed chimera proteins has not changed. The negative value of GRAVY (Grand Average of Hydropathy) implies that proteins are hydrophilic and have a better reaction with water.

The tertiary structure of arazyme, TGFαL3, and fusion proteins were modeled by I-TASSER and Robetta web servers. Then, all predicted models were evaluated by ERRAT, PROCHECK Ramachandran, MolProbity, and QMEAN value. The quality assessment analysis by all web servers showed that predicted models by Robetta obtained a better score than the I-TASSER server. In addition, the folding of the Robetta predicted models showed higher similarity to the arazyme territory structure due to their lower RMSD calculated by IPBA.
Molecular docking of two fusion proteins to calculate binding energy and strength affinity between ligand and receptor was conducted by blind, rigid, and focused methods. By considering β-sheets position, the binding affinity of the different sequence lengths of TGFα was evaluated. Interestingly, loop three of TGFα, which includes half of one large beta-sheet, has an energy that is more negative and therefore was selected to make fusion protein. Then the affinity of the two constructs was compared by HADDOCK 2.4, and Cluspro 2.0 and the results showed that the construct, Arazyme-linker-TGFαL3 was the best because of more negative scores, although the entire sequence of TGFα was the best. Therefore, Arazyme-linker-TGFαL3 was subjected to MD simulations and epitope investigations. The radius of gyration and RMSD of the optimized structure of Arazyme-linker-TGFαL3 revealed the relative conformational equilibrium of fusion protein, although RMSF analysis showed that both C- and N-terminal are the most flexible region. Meanwhile, the secondary structure of fusion proteins remains relatively stable after simulation. Afterward, the interactions of the average structure of Arazyme-linker-TGFαL3 with EGFR were evaluated by 100 ns of MD simulation. The RMSD values indicated the overall stability of the complex. Also, the changes in the compactness of the complex were not significant during simulation time. The comparison of RMSF values between bound and unbound ligand indicated that the region comprising linker and TGFαL3 has the lower flexibility in the ligand-receptor complex which may be due to interaction with the receptor.

Figure 6. Hydrogen bonds (dash line) and hydrophobic bonds (spoked arcs) interactions between Arazyme-linker-TGFαL3 and EGFR after 100 ns of MD simulation.
Table 5. The physicochemical properties and Vaxijen score of the best high-ranked linear epitope predicted by Ellipro server as well as Kolaskar, and Tongaonkar’s method.

| Web servers | The sequence of predicted epitopes | Hydrophilicity (average) | Flexibility (average) | Accessibility (average) | Vaxijen score |
|-------------|-----------------------------------|--------------------------|----------------------|------------------------|---------------|
| Ellipro     | MSCLJNNQLMSIEPQMSTKKAITESSLA      | 1.730                    | 1.015                | 1.02                   | 0.2362        |
|             | AGQGDQHVDHFSGGAEGALSYNASNVDLALNG  | 1.357                    | 0.984                | 0.588                  | 0.4722        |
|             | GHQAPDFLVQGVDVATDF                 | 3.218                    | 0.970                | 0.845                  | 1.0011        |
|             | VAEAAKEAAKEAAKEAEEAAKAAEAKAAEAAAKAEEAAKEAAKEAAKEAAKEAAKAA 
|             | AKEAAAKAVCHGSVGARCEHADLL          |                         |                      |            |               |
|             | RENQTWNGYKVFGQPVKLT5FPDQYKFSSTNVAGDTL5SKFSAEQQQ               | 1.836                    | 1.015                | 1.14                   | 0.6127        |
|             | ANITTFEVGQAKANITFGNYSDRPGHYDYTQ   | 3.021                    | 1.011                | 1.44                   | 0.7693        |
| Kolaskar and Tongaonkar | SSLAAAAAYNVIDOLLHYH          | 2.096                    | 0.973                | 0.677                  | 0.2094        |
|             | YKVFGQPVLTSFDPDYK                 | -0.051                   | 1.001                | 0.866                  | 0.594         |
|             | QAKLSLQSWS                       | 0.346                    | 0.990                | 0.677                  | 0.8053        |
|             | QAYAFLPNT                        | -0.090                   | 0.938                | 0.539                  | 0.1514        |
|             | EIGHALGALS                      | -0.043                   | 0.943                | 0.253                  | 0.8727        |

Figure 7. The sequence of MHCI and MHCII epitopes of Arazyme-linker-TGFαL3 predicted by Netmhcpan 4.1 (bolded), PickPocket (highlighted) SMM (double underline), SYFPEITH (wavy underline for MHCI epitopes and in a box for MHCII epitopes), and NetMHCLipan (in dotted box).

MM/PBSA and MM/GBSA are the most popular methods to estimate the free binding energy of ligand to receptor although the physicochemical properties of molecules such as entropy and atomic charges as well as simulation time and sampling protocols can affect performance, accuracy, and differences in the calculation of binding energy (Genheden & Ryde, 2015). In the current study, the negative values of MM/PBSA indicated the contribution of nonbonded interactions in the binding affinity of Arazyme-linker-TGFαL3 to EGFR. In addition, the highest favorable contribution was related to van der Waals forces. Also, non-polar solvation energy and electrostatic interactions are favorable for ligand-receptor binding. However, the polar contributions were unfavorable for binding to EGFR.

Based on the crystal structure analysis of the complex of EGFR external domain and intact TGFα, LEU 566, HID 563, CYS 561, and VAL 551 from TGFαL3 region.

Different approaches including BepiPred-2.0, Kolaskar, and Tongaonkar, ElliPro, and DiscoTope were adopted to identify linear and conformational B-cell epitopes. BepiPred-2.0 is the server, which uses a random forest algorithm to derive epitope data from crystal structures and obtain higher quality results (Jespersen et al., 2017). By utilizing the physicochemical properties of the known epitopes, the Kolaskar, and Tongaonkar method showed 75% accuracy in epitope prediction (Kolaskar & Tongaonkar, 1990). Two algorithms for continuous epitope prediction by ElliPro are the approximation of protein surface as an ellipsoid and the protrusion index (PI) which is the percentage of outside residues in an ellipsoid (Ponomarenko et al., 2008). Two common predicted B cell epitopes, namely EALLSY and VSDLALNI were positioned near C-terminal peptidase of arazyme.

If a residue appears to be greater than two nm of the water-accessible surface, it is considered a surface residue. By considering this scale, the fractional surface probability of amino acids in a hexapeptide sequence was defined by Emini server formulae (Emini et al., 1985). The epitope, ANITTFEVGQAKANITFGNYSDRPGHYDYTQ which is near the N-terminal of the Zinc-dependent metalloprotease sequence of arazyme obtains the highest Hydrophilicity, flexibility, and accessibility scores.

The DiscoTope combines log-odds ratio score and surface measurement calculated from half-sphere exposure to
improve the performance of epitope prediction (Kringelum et al., 2012). Based on the DiscoTope score, the residues with a score of higher than 1.9 have 95% Specificity. In this manner, all predicted epitope residues of 154–165 (SQDRPGHYDYDTQAYAFLPNTI) show the highest specificity which was included in the discontinuous epitopes obtaining the highest score from ElliPro.

Figure 8. Conformational B cell epitopes of Arazyme-linker-TGFαL3 predicted by DiscoTope and ellipro. Green balls are predicted epitopes. DiscoTope predicted residues, including (a) 15, 16, 53–62, 81–83, 97–99, 101–113, and 115–117; (b) 152–168, 175, 177–179, 187–198, and 202; (c) 222–238, and 253–256. ElliPro predicted residues, including (d) 75–117, 119, 133, 134, 136, 138–149, 151–168, 171–183, 186, 188–197, 216, 277, 280 and 281; (e) 3–36, 386, 396, 404, 405, 413–422, 428–433, 438, 442–502, 504–514, and 516–552; (f) 51–64, 66, 67,70,71,73,74, 218–240, 250, 254, 427, and 553–567.

PickPocket, Netmhcpan, SMM, and SYFPEITHI servers were used to predict MHCI and MHCII epitopes. The prediction of the MHC-I epitope was made through methods, including an artificial neural network (ANN), a stabilized matrix method (SMM), and PickPocket separately. Netmhcpan and PickPocket use artificial neural networks and computational methods to make accurate predictions for different MHC
with no allele data (Reynisson et al., 2020; Zhang et al., 2009). SMM and SYFPEITHI rely on experimental data of different MHC molecules to make a prediction (Peters & Sette, 2005; Rammensee et al., 1999). The best-identified mhcl and mhcll epitopes by the above servers were different, although there was some overlap between the suggested epitopes. SDAP web server employs an algorithm to find allergens based on the conserved region of amino acid side chains or its database of IgE epitopes (Ivanciuc et al., 2003). Algpred employs different approaches, including SVM, IgE epitope, ARPs BLAST, and MAST to predict allergens (Saha & Raghava, 2006). Based on the results obtained from these servers, there is no allergen-specific IgE for chimera protein.

Our results indicate that the fusion protein, Arazyme-linker-TGF-αL3 shows affinity to the EGFR receptor which is expressed at a high level in the tumor cell. Subsequently, blocking of EGFR and elimination of proliferative signals are expected. Also, because of numerous predicted B cell and T cell epitopes, the immune responses to cancer cells can be elevated.

Acknowledgement

This manuscript is an extract of a Ph.D. thesis by Abdolamir Ghadaksaz submitted to Islamic Azad University, Science and Research Branch, Tehran, Iran.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Abdolamir Ghadaksaz http://orcid.org/0000-0003-4105-4403
Abbas Ali Imani Fooladi http://orcid.org/0000-0001-7339-8257
Hamideh Mahmooodzadeh Hosseini http://orcid.org/0000-0002-3987-0164

References

Amet, N., Lee, H.-F., & Shen, W.-C. (2009). Insertion of the designed helical linker led to increased expression of TF-based fusion proteins. Pharmaceutical Research, 26(3), 523–528. https://doi.org/10.1007/s11095-008-9767-0

Amjadi, G., Parivar, K., Fazollah Mousavi, S., & Imani Fooladi, A. (2020). Anti-cancer effects of recombinant arzyme from Seratia Proteomaculans. Journal of B.U.ON: Official Journal of the Balkan Union of Microbiology, 25(1), 531–542.

Appert-Collin, A., Hubert, P., Crémel, G., & Bennassroune, A. (2015). Role of ErbB receptors in cancer cell migration and invasion. Frontiers in Pharmacology, 6, 283–283. https://doi.org/10.3389/fphar.2015.00283

Benkert, P., and Biasini, M., & Schwede, T. (2011). Toward the estimation of the absolute quality of individual protein structure models. Bioinformatics (Oxford, England), 27(3), 343–350. https://doi.org/10.1093/bioinformatics/bttq662

Bersanetti, P. A., Park, H.-Y., Bae, K. S., Son, K.-H., Shin, D.-H., Hirata, I. Y., Juliano, M. A., Carmona, A. K., & Juliano, L. (2005). Characterization of an arzyme, an exocellular metalloprotease isolated from Seratia proteomaculans culture medium. Enzyme and Microbial Technology, 37(6), 574–581. https://doi.org/10.1016/j.enzmictec.2005.01.041

Bryson, C. J., and Jones, T. D., & Baker, M. P. (2010). Prediction of immunogenicity of therapeutic proteins: Validity of computational tools. BioDrugs: Clinical Immunotherapeutics, Biopharmaceuticals and Gene Therapy, 24(1), 1–8. https://doi.org/10.2165/11318560-000000000-00000

Bussi, G., and Donadio, D., & Parrinello, M. (2007). Canonical sampling through velocity rescaling. Journal of Chemical Physics, 126(1), 014101. https://doi.org/10.1063/1.2408420

Chen, X., and Zaro, J. L., & Shen, W.-C. (2013). Fusion protein linkers: Property, design and functionality, Advanced Drug Delivery Reviews, 65(10), 1357–1369. https://doi.org/10.1016/j.addr.2012.09.039

Chen, F., Liu, H., Sun, H., Pan, P., Li, Y., Li, D., & Hou, T. (2016). Assessing the performance of the MM/PBSA and MM/GBSA methods. 6. Capability to predict protein-protein binding free energies and re- rank binding poses generated by protein-protein docking . Physical Chemistry Chemical Physics: PCCP, 18(32), 22129–22139. https://doi.org/10.1039/c6cp03670h

Coleman, W. B., & Tsongalis, G. J. (2016). The Molecular Basis of Human Cancer. Springer. https://doi.org/10.1007/978-1-9992-125-1

Colovos, C., & Yeates, T. O. (1993). Verification of protein structures: Patterns of nonbonded atomic interactions. Protein Science: A Publication of the Protein Society, 2(9), 1511–1519. https://doi.org/10.1002/pro.556020916

Das, M., Mohanty, C., & Saooh, S. K. (2009). Ligand-based targeted therapy for cancer tissue. Expert Opinion on Drug Delivery, 6(3), 285–304. https://doi.org/10.1517/17452409078016

De Vries, S. J., and Van Dijk, M., & Bonvin, A. M. J. M. J. (2010). The HADDOCK web server for data-driven biomolecular docking. Nature Protocols, 5(5), 883–897. https://doi.org/10.1038/nprot.2010.32

Doytchinova, I. A., & Flower, D. R. (2007). VaxiJen: A server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinformatics, 8(1), 4. https://doi.org/10.1186/1471-2105-8-4

Emini, E. A., Hughes, J. V., Perlow, D. S., & Boger, J. (1985). Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. Journal of Virology, 55(3), 836–839. https://doi.org/10.1128/JVI.55.3.836-839.1985

Felgner, S., Kocijancic, D., Frahm, M., & Weiss, S. (2016). Bacteria in cancer therapy: Renaissance of an old concept. International Journal of Microbiology, 2016, 8451728. https://doi.org/10.1155/2016/8451728

Garrett, T. P. J., McKern, N. M., Lou, M., Elleman, T. C., Adams, T. E., Lovrecz, G. O., Zhu, H.-J., Walker, F., Frenkel, M. J., Hoyne, P. A., Jorissen, R. N., Nice, E. C., Burgess, A. W., & Ward, C. W. (2002). Crystal structure of a truncated epidermal growth factor receptor extracellular domain bound to transforming growth factor α. Cell, 110(6), 763–773. https://doi.org/10.1016/S0002-8674(02)00940-6

Gasteiger, E. (2005). Protein identification and analysis tools on the ExPASy server. In The proteomics protocols handbook (pp. 571–607). Springer.

Gelly, J.-C., Joseph, A. P., Srinivasan, N., & de Brevern, A. G. (2011). IPBA: A tool for protein structure comparison using sequence alignment strategies. Nucleic Acids Research, 39(Web Server issue), W18–W23. https://doi.org/10.1093/nar/gkr333

Geheden, S., & Ryde, U. (2015). The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. Expert Opinion on Drug Discovery, 10(5), 449–461. https://doi.org/10.1517/17460441.2015.1032936

Gonda, D. K., Bachmair, A., Wünnen, I., Tobias, J. W., Lane, W. S., & Varshavsky, A. (1989). Universality and structure of the N-end rule. The Journal of Biological Chemistry, 264(28), 16700–16712. https://doi.org/10.1002/jbc.1070021

Hess, B., Bekker, H., Berendsen, H. J. C., & Frenkel, G. J. E. M. (2008). LINCS: A Linear Constraint Solver for molecular simulations. Journal of Computational Chemistry, 29(12), 1463–1472. https://doi.org/10.1002/jcc.20876

Ivanciuc, O., and Schein, C. H., & Braun, W. (2003). SDAP: Database and computational tools for antigens. Nucleic Acids Research, 31(1), 359–362. https://doi.org/10.1093/nar/gkg010

Jespersen, M. C., Peters, B., Nielsen, M., & Marcatili, P. (2017). BepiPred-2.0: Improving sequence-based B-cell epitope prediction using conformational epitopes. Nucleic Acids Research, 45(W1), W24–W29. https://doi.org/10.1093/nar/gkw346

Johnson, M. S., and Sutcliffe, M. J., & Blundell, T. L. (1990). Molecular anatomy: Phyletic relationships derived from three-dimensional
structures of proteins. *Journal of Molecular Evolution*, 30(1), 43–59. https://doi.org/10.1007/BF02102452

Kabsch, W., & Sander, C. (1983). Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22(12), 2577–2637. https://doi.org/10.1002/bip.36022121

Kaplus, P. A., & Schulz, G. E. (1985). Prediction of chain flexibility in proteins. *Naturwissenschaften*, 72(4), 212–213. https://doi.org/10.1007/BF01195768

Kolaskar, A. S., & Tongaonkar, P. C. (1990). A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Letters*, 276(1-2), 172–174. https://doi.org/10.1016/0014-5793(90)80535-Q

Kozakov, D., Beglov, D., Bohnuud, T., Mottarella, S. E., Xia, B., Hall, D. R., & Vaja, S. (2013). How good is automated protein docking? *Proteins*, 81(12), 2159–2166. https://doi.org/10.1002/prot.24403

Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993). PROCHECK: A program to check the stereochemical quality of protein structures. *Acta Crystallographica Section D*, 50(6 Pt 1), 265–266. https://doi.org/10.1107/0001749

Lokhande, K. B., and Nagar, S., & Swamy, K. V. (2019). Molecular interaction studies of Deguelin and its derivatives with Cyclin D1 and Cyclin E in cancer cell signaling pathway: The computational approach. *Scientific Reports*, 9(1), 1778–1713. https://doi.org/10.1038/s41598-018-38332-6

Massova, I., & Kollman, P. A. (1999). Computational Alanine scanning to probe protein – protein interactions: A novel approach to evaluate binding free energies. *Journal of the American Chemical Society*, 121(36), 8133–8143. https://doi.org/10.1021/ja990935k

Miller, B. R., McGee, T. D., Swalls, J. M., Homeyer, N., Gohlke, H., & Roitberg, A. E. (2012). MMPBSA.py: An efficient program for end-state free energy calculations. *Journal of Chemical Theory and Computation*, 8(9), 3314–3321. https://doi.org/10.1021/ct300418h

Naqui, A.A.T., Mohammad, T., Hasan, G.M., & Hassan, M.J. (2018). Advancements in docking and molecular dynamics simulations towards ligand-receptor interactions and structure-function relationships. *Current Topics in Medicinal Chemistry*, 18(20), 1755–1768. https://doi.org/10.2174/15680266186666181025114157

Nestor, J. J., Newman, S. R., DeLustra, B., Todaro, G. J., & Schreiber, A. B. (1985). A synthetic fragment of rat transforming growth factor α with receptor binding and antigenic properties. *Biochemical and Biophysical Research Communications*, 129(1), 226–232. https://doi.org/10.1016/0006-291X(85)91426-3

Parker, J. M., and Guo, D., & Hodges, R. S. (1986). New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: Correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. *Biochemistry*, 25(19), 5425–5432. https://doi.org/10.1021/bi00367a013

Parrinello, M., & Rahman, A. (1981). Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics*, 52(12), 7182–7190. https://doi.org/10.1063/1.328693

Patil, R., Das, S., Stanley, A., Yadav, L., Sudhakar, A., & Varma, A. K. (2010). Optimized hydrophobic interactions and hydrogen bonding at the target-ligand interface leads the pathways of drug-designing. *PloS One*, 5(8), e12029. https://doi.org/10.1371/journal.pone.0012029

Pereira, F. V., Melo, A. C. L., de Melo, F. M., Mourao-Sá, D., Silva, P., Berzaghi, R., Herbreteau, C. A., Coelho-Dos-Reis, J., Scutti, J. A., Orriaga, C. S. T., Pereira, R. M., Julioan, L., Marotta, A. C. A., Carmona, A. K., Camara, N. O. S., Tsuji, M., Travassos, L. R., & Rodrigues, E. G. (2016). TL94-mediated immunomodulatory properties of the bacterial metalloprotease araziney in pre-clinical tumor models. *Onc-immunology*, 5(7), e1178420. https://doi.org/10.21626/2016.1178420

Peters, B., & Sette, A. (2005). Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. *BMc Bioinformatics*, 6, 132. https://doi.org/10.1186/1471-2105-6-132

Ponomarenko, J., Bui, H.-H., Li, W., Fusseder, N., Bourne, P. E., Sette, A., & Peters, B. (2008). EliPro: A new structure-based tool for the prediction of antibody epitopes. *BMc Bioinformatics*, 9, 514 https://doi.org/10.1186/1471-2105-9-514

Potocnakova, L., Bhide, M., & Pulzova, L. B. (2016). An introduction to B-cell epitope mapping and in silico epitope prediction. *Journal of Immunology Research*, 2016, 6760830. https://doi.org/10.1155/2016/6760830

Prisant, M. G., Williams, C. J., Chen, V. B., Richardson, J. S., & Richardson, D. C. (2020). New tools in MolProbity validation: CaBLAM for CryoEM backbone, UnDowser to rethink “waters,” and NGL viewer to recaputure online 3D graphics. *Protein Science*, 29(1), 315–329. https://doi.org/10.1002/prot.25250

Ramam, S., Vernon, R., Thompson, J., Tyka, M., Sadreyev, R., Pei, J., Kim, D., Kellogg, E., DiMaio, F., Lange, O., Kinch, L., Sheffler, W., Kim, B-H., Das, R., Grishin, N. V., & Baker, D. (2009). Structure prediction for CASP8 with all-atom refinement using Rosetta. *Proteins*, 77(59), 89–99. https://doi.org/10.1002/prot.22540

Rammensee, H., Bachmann, J., Emmerich, N. P., Bachor, O. A., & Stevanovic, S. (1999). SYFPEITHI: Database for MHC ligands and peptide motifs. *Immunogenetics*, 50(3-4), 213–219. https://doi.org/10.1007/s002510050095

Reynisson, B., Alvarez, B., Paul, S., Peters, B., & Nielsen, M. (2020). NetMHCpan-4.1 and NetMHCIIpan-4.0: Improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Research*, 48(W1), W449–W454. https://doi.org/10.1093/nar/gkaa379

Roy, A., and Kucukural, A., & Zhang, Y. (2010). I-TASSER: A unified platform for automated protein structure and function prediction. *Nature Protocols*, 5(4), 725–738. https://doi.org/10.1038/nprot.2010.5

Saha, S., & Raghava, G. P. S. (2006). AlgPred: Prediction of allergenic proteins and mapping of IgE epitopes. *Nucleic Acids Research*, 34(Web Server issue), W202–W209. (Web Server issue), https://doi.org/10.1093/nar/gkl343
Song, Y., DiMaio, F., Wang, R. Y.-R., Kim, D., Miles, C., Brunette, T., Thompson, J., & Baker, D. (2013). High-resolution comparative modeling with RosettaCM. *Structure (London, England: 1993), 21*(10), 1735–1742. https://doi.org/10.1016/j.str.2013.08.005

Tresanco, M. S. V. (2021). gmx_MMPBSA. https://doi.org/10.5281/ZENODO.4732765

Trott, O., & Olson, A. J. (2010). AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry, 31*(2), 455–461. https://doi.org/10.1002/jcc.21334

Vajda, S., Yueh, C., Beglov, D., Bohnuuc, T., Mottarella, S. E., Xia, B., Hall, D. R., & Kozakov, D. (2017). New additions to the ClusPro server motivated by CAPRI. *Proteins, 85*(3), 435–444. https://doi.org/10.1002/prot.25219

van Zundert, G. C. P., Rodrigues, J. P. G. L. M., Trellet, M., Schmitz, C., Kastritis, P. L., Karaca, E., Melquiond, A. S. J., van Dijk, M., de Vries, S. J., & Bonvin, A. M. J. J. (2016). The HADDOCK2.2 web server: User-friendly integrative modeling of biomolecular complexes. *Journal of Molecular Biology, 428*(4), 720–725. https://doi.org/10.1016/j.jmb.2015.09.014

Wiederstein, M., & Sippl, M. J. (2007). ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research, 35*(Web Server issue), W407–W410. https://doi.org/10.1093/nar/gkm290

Wykosky, J., Fenton, T., Furnari, F., & Cavenee, W. K. (2011). Therapeutic targeting of epidermal growth factor receptor in human cancer: Successes and limitations. *Chinese Journal of Cancer, 30*(1), 5–12. https://doi.org/10.5732/cjc.010.10542

Yotsumoto, F., Sanui, A., Fukami, T., Shirotori, K., Horiuchi, S., Tsujioka, H., Yoshizato, T., Kuroki, M., & Miyamoto, S. (2009). Efficacy of ligand-based targeting for the EGF system in cancer. *Anticancer Research, 29*(11), 4879–4885. https://doi.org/10.1517/17425240902780166

Yousefi, F., Mousavi, S. F., Siadat, S. D., Aslani, M. M., Amani, J., Rad, H. S., & Fooladi, A. A. I. (2016). Preparation and in vitro evaluation of antitumor activity of TGF-$\alpha$L3-SEB as a ligand-targeted superantigen. *Technology in Cancer Research & Treatment, 15*(2), 215–226. https://doi.org/10.1177/1533034614568753

Zhang, H., and Lund, O., & Nielsen, M. (2009). The PickPocket method for predicting binding specificities for receptors based on receptor pocket similarities: Application to MHC-peptide binding. *Bioinformatics (Oxford, England), 25*(10), 1293–1299. https://doi.org/10.1093/bioinformatics/btp137