CD171 Multi-epitope peptide design based on immuno-informatics approach as a cancer vaccine candidate for glioblastoma

Seyed Amir Hossein Mohammadzadeh Hosseini Moghri\textsuperscript{a}, Ghanbar Mahmoodi Chalbatani\textsuperscript{b}, Mojtaba Ranjbar\textsuperscript{a}\textsuperscript{,}\textsuperscript{c}, Catarina Raposo\textsuperscript{d} and Arefeh Abbasi\textsuperscript{a}\textsuperscript{,}\textsuperscript{d}

\textsuperscript{a}Faculty of Biotechnology, Department of Microbial Biotechnology, Amol University of Special Modern Technologies, Amol, Iran; \textsuperscript{b}Department of Medical Immunology, School of Medicine, Tehran University of Medical Sciences, Tehran, Iran; \textsuperscript{c}Faculdade de Ciências Farmacêuticas, Universidade Estadual de Campinas (UNICAMP), Campinas, Brazil; \textsuperscript{d}Faculty of Basic Sciences, Department of Biology, Semnan University, Semnan, Iran

Communicated by Ramaswamy H. Sarma

\textbf{ABSTRACT}

Glioblastoma (GB) is a common primary malignancy of the central nervous system, and one of the highly lethal brain tumors. GB cells can promote therapeutic resistance and tumor angiogenesis. The CD171 is an adhesion molecule in neuronal cells that is expressed in glioma cells as a regulator of brain development during the embryonic period. CD171 is one of the immunoglobulin-like CAMs (cell adhesion molecules) families that can be associated with prognosis in a variety of human tumors. The multi-epitope peptide vaccines are based on synthetic peptides with a combination of both B-cell epitopes and T-cell epitopes, which can induce specific humoral or cellular immune responses. Moreover, Cholera toxin subunit B (CTB), a novel TLR agonist was utilized in the final construct to polarize CD4+ T cells toward T-helper 1 to induce strong cytotoxic T lymphocytes (CTL) responses. In the present study, several immune-informatics tools were used for analyzing the CD171 sequence and studying the important characteristics of a designed vaccine. The results included molecular docking, molecular dynamics simulation, immune response simulation, prediction and validation of the secondary and tertiary structure, physicochemical properties, solubility, conservancy, toxicity as well as antigenicity and allergenicity of the promising candidate for a vaccine against CD171. The immuno-informatic analyze suggested 12 predicted multi-epitope peptides, whose construction consists of 582 residues long. Therewith, cloning adaptation of the designed vaccine was performed, and eventually sequence was inserted into pET30a (+) vector for the application of the anti-glioblastoma vaccine development.

\textbf{1. Introduction}

Glioblastoma (GB) is the most aggressive type of glioma and corresponds to the majority of primary central nervous system (CNS) malignancy forms in adults. GB is supposed to respond for more than 50% of all intracranial malignancies (Davis et al., 2001). CD171 plays a regulating role in neural cell development during the embryonic period, tumor cell survival and migration (Maness & Schachner, 2007). Researchers have proven the overexpression of CD17 in solid tumors, such as gliomas and colorectal cancer, as a prognostic factor (Boo et al., 2007). Moreover, CD171 is only expressed on the surface of cancer cells, not their normal tissues (Huszar et al., 2006) also it has a significant role in embryonic and fetal morphogenesis and it is an absent factor (Boo et al., 2007). Moreover, CD171 is only expressed on the surface of cancer cells, not their normal tissues (Huszar et al., 2006) also it has a significant role in embryonic and fetal morphogenesis and it is an absent factor (Boo et al., 2007). In the present study, several immune-informatics tools were used for analyzing the CD171 sequence and studying the important characteristics of a designed vaccine. The results included molecular docking, molecular dynamics simulation, immune response simulation, prediction and validation of the secondary and tertiary structure, physicochemical properties, solubility, conservancy, toxicity as well as antigenicity and allergenicity of the promising candidate for a vaccine against CD171. The immuno-informatic analyze suggested 12 predicted multi-epitope peptides, whose construction consists of 582 residues long. Therewith, cloning adaptation of the designed vaccine was performed, and eventually sequence was inserted into pET30a (+) vector for the application of the anti-glioblastoma vaccine development.

This glycoprotein has a regulatory function in cell adhesion, development, survival, and metastasis of tumor cells (Schäfer & Altevogt, 2010; Siesser & Maness, 2009). CD171 ectodomain is abnormal in tumor cells, due to cleavage by the ADAM10 protease and consequent auto stimulation, resulting in cellular motility and proliferation (Gavert et al., 2008; Riedle et al., 2009; Yang et al., 2011). Currently, considering the various advantages of multi-epitope peptide-based vaccine, which consists of high specificity, good safety, stability, ease of production and storage, it has become an area of increasing interest in the field of vaccine research; this is even more promising in light of advances in immunoinformatics and vaccinology (Skwarczynski & Toth, 2014). Multi-epitope peptide vaccines are based on synthetic peptides with a combination of many B-cell and T-cell epitopes that can induce specific humoral and/or cellular immune responses. Prediction of B- and T-cell epitopes have been the focus of computational vaccinology and, given the potential translational implications, several bioinformatics tools have been developed (Doytchinova & Flower, 2002; Flower et al., 2001).
Applying the cholera toxin subunit B (CTB) in designed multi-epitope as TLR Ligands agonists is a novel TLR4 agonist that significantly robust immunostimulatory effects in cancer immunotherapy (Nezafat et al., 2014; Phongsisay et al., 2015). The aim of the present study is to analyze, using computational methods, the sequence and structure of CD171 and to predict potential linear epitopes of CD171 that may be targets of B and T-cells also molecular docking and molecular dynamics simulation studies revealed strong interactions of the vaccine with MHC class I to elicit the cytotoxic immune response. This multi-epitope design will provide information for a promising peptide vaccine based on the epitope for cancer therapy.

2. Materials and methods

2.1. CD171 sequence retrieval and structural prediction

The amino acid sequences of CD171 with accession number: NP_001265045.1 were retrieved from the NCBI database in FASTA format (Geer et al., 2010). TMHMM tool was used to demonstrate if the peptides are in transmembrane regions or not. TMHMM supported the hidden Markov model (HMM) method, which specializes in the modeling of globular domains, helix caps and other various regions of cell membrane proteins (http://www.cbs.dtu.dk/services/TMHMM/) (Krogh et al., 2001).

2.2. B-cell epitope prediction

The objective is the prediction of the B cell epitope to find a potential antigen that would interact with B lymphocytes and initiate an immune response (Nair et al., 2002). The linear B cell epitopes have variable peptide lengths, from 2 to 85. The BepiPred-2.0 webserver was used for prediction of linear B cell epitopes (http://www.cbs.dtu.dk/services/BepiPred/). This method is based on a random forest algorithm trained on epitopes annotated from antibody-antigen protein structures (Jespersen et al., 2017).

2.3. T-cell epitope prediction

The Immune Epitope Database (IEDB), HLA allele frequencies and reference sets with maximum population coverage for the selected epitopes were used. The prediction of the most probable epitopes interacting with different MHC class I, II alleles were chosen based on the percentile cut-off was set at 0.5 and 1 for MHC class I and II, respectively. In both cases, to find a good binding affinity, the cut-off value IC50 was set at lower 50 nm and 150 nm to obtain a better level of confidence in the prediction of epitopes for MHC class I and II, respectively, and the antigenicity score. The candidate epitopes for MHC class I were determined by both IEDB and NetMHCpan EL 4.0 methods, and for MHC class II by IEDB 2.22 recommended method binding prediction tool (http://tools.iedb.org/mhci/), (http://tools.iedb.org/mhcii/) (Zhang et al., 2008).

2.4. Antigenicity, conservancy, allergenicity and toxicity extrapolation

The antigenicity characteristics of the peptide were screened using an online antigen prediction server, VaxiJen v2.0 (http://www.ddgpharmfac.net/vaxijen/). The Threshold for tumor models was set at ≥ 0.4 (Doytchinova & Flower, 2007) and the conservancy of the selected epitopes was examined using the epitope conservancy tool (http://tools.iedb.org/conservancy/) (Bui et al., 2007). The web-based servers AllerTOP v.2.0 (http://www.ddgpharmfac.net/AllerTOP/) (Dimitrov et al., 2014) and AllergenFP v.1.0 (http://www.ddg-pharmfac.net/AllergenFP/) (Dimitrov et al., 2014) were used to prediction of the epitopes allergenicity. The ToxinPred webserver was applied for the prediction of toxic peptides (https://webs.iiitd.edu.in/raghava/toxinpred). This method was developed based on the machine learning technique and quantitative matrix using different properties of peptides (Gupta et al., 2013).

2.5. Construction of multi-epitope vaccine candidate sequence

The candidate vaccine sequence was generated based on the overlapping of the predicted B-cell and T-cell epitopes of the predicted peptide containing linear B- and T-cell epitopes were fused using AAY linkers (Khatoon et al., 2017). Using a DPRVPSS linker, the cholera toxin subunit B (CTB) with accession no. CAA53976.1 was chosen as an adjuvant, which was constructed within the final vaccine model at the N-terminal to potentiate the immunogenic capacity of our peptide by stimulating the innate immunity. The adjuvant potential of CTB has been reported in several animal models, indicating that the adjuvant potential would be scalable to complex species (Baldauf et al., 2015). The CTB sequence was retrieved from the UniProt server (http://www.uniprot.org/).

2.6. Physicochemical properties and solubility prediction

Various physicochemical features of candidate peptides which included theoretical pI, aliphatic index, instability index, estimated half-life in the mammalian reticulocytes in vitro, extinction coefficient, molecular weight, and grand average of hydropathicity (GRAVY) were determined using the online web server ProtParam (http://web.expasy.org/protparam/) (Gasteiger et al., 2005). Estimated solubility in water of the multi-epitope vaccine peptide was evaluated using the Pepscale (http://pepscale.com/) (Lear & Cobb, 2016).

2.7. Secondary structure prediction

The secondary structure of final peptide was predicted with Position-specific iterated BLAST (PSI-blast) based on secondary structure PREDiction (PSIPRED), (http://bioinf.cs.ucl.ac.uk/psipred/) (Buchan et al., 2013). Also, the RaptorX Property web server (http://raptorx.uchicago.edu/StructurePropertyPred/predict/) was employed to predict the secondary structure properties of the chimeric protein (Wang et al., 2016).
2.8. Tertiary structure prediction and validation
The homology modeling tool, I-TASSER (Iterative Threading ASSEMBly Refinement) server (https://zhanglab.ccmb.med.umich.edu/I-TASSER/), was applied for three-dimensional (3D) structure prediction of the final multi-epitope vaccine peptide. Through multiple threading alignments and iterative structural assembly simulations, the I-TASSER first generates three-dimensional (3D) atomic models from an amino acid sequence (Roy et al., 2010). Furthermore, we perform analyze and visualization by generating a 3D model in PDB format using PyMOL (https://pymol.org/2/) (DeLano, 2002). In addition, the validation of the built model was performed as a vital step to the detection of potential errors in predicted 3D models (Khatoo et al., 2017). Subsequently, a Ramachandran plot was generated using the RAMPAGE server (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php). The Ramachandran plot energetically displays a visualization of allowed and disallowed dihedral angles psi (\(\psi\)) and phi (\(\phi\)) of the amino acid and is calculated based on the van der Waal radius of the side chain. The RAMPAGE results demonstrate the percentage of residues in allowed and disallowed regions that define the quality of the modeled structure (Lovell et al., 2003). The entire quality of the final vaccine model was defined by ProSA-web (https://prosa.services.came.sbg.ac.at/prosa.php). ProSA-web provides a protein tertiary structure validation and calculates an overall 3D structure quality score; if the calculated score is outside the characteristic range of native proteins the structure is likely to have errors (Wiederstein & Sippl, 2007).

2.9. In silico cloning adaptation of designed vaccine
The Java Codon (JCat) adaptation tool server (http://www.prodoric.de/JCat/) was performed for reverse translation and codon optimization, in order to construct the multi-epitope vaccine in a selected expression vector (Grote et al., 2005). Codon optimization leads to a higher expression rate of the final vaccine in E. coli K12 as an expression host because the use of the condon of the human and the selected host differs from each other. Three additional options were applied to avoid the transcription of rho-independent termination, prokaryote ribosome binding site and cleavage site of restriction enzymes to increase the efficiency of a translation process. The JCat result includes the codon adaptation index (CAI) and the percentage GC content, which can be used to evaluate protein expression levels. CAI provides information on codon usage biases; the ideal CAI score is 1.0, however, greater than 0.8 is considered a good score (Morla et al., 2016). The GC content of a sequence should range between 30–70% and outside this range is considered unfavorable translational and transcriptional efficiencies (Ali et al., 2017). To clone the optimized, the final vaccine sequence was reversed and Nde I and Xho I restriction sites was added at the N- and C-terminal sites of the final construct, respectively. Finally, the optimized sequence (with restriction sites) was inserted into the pET-30a (+) vector using the SnapGene restriction cloning module to ensure vaccine expression.

2.10. Molecular docking of vaccine candidate with MHC I
Interaction between the MHC class 1 binding and vaccine is highly considerable to elicit the cytotoxic immune response therefore we performed molecular docking between the vaccine model and MHC class 1 (PDB ID- 1AKJ) which was obtained from the PDB database (https://rcsb.org) through ClusPro 2.0 (https://cluspro.bu.edu/login.php) as an online protein-protein docking server. This software incorporates three techniques; firstly, a fast Fourier transform (FFT) correlation approach, secondly, root-mean-square deviation (RMSD) based clustering of best energy conformations and lastly is an assessment of the stability of clusters (Kozakov et al., 2017; Vajda et al., 2017).

2.11. Molecular dynamics simulation of the vaccine-MHC I
Eventually, molecular dynamics simulation was studied through the i-MOD server (i-MODS) (http://imods.chaconlab.org). Via i-MODS the protein binding stability and minimal deformation of the docked complex were examined through normal mode analyze (NMA) which shows the inherent motion of the complex coordinates. The stability of the protein is represented in terms of its main-chain deformability plot that depicts the binding stability and potency of the ligand to the receptor according to the atomic fluctuations, atomic motions or B-factor values, covariance matrix, elastic network model, also eigen score that demonstrates the rigidity of the motion of the complex and is comparable to the deformation profiles. As demonstrated using independent component algorithms, a low Eigen score shows less stability and easy deformation of the atoms coordinates (López-Blanco et al., 2014).

2.12. Immune response simulation
To interpret the immune response profile of the designed multi-epitope vaccine, immune simulations were carried out via the C-ImmSim server available at (http://kraken.iac.rm.cnr.it/C-IMMSIM/) (Moghri et al., 2021). The C-ImmSim is an agent-based computational immune response simulator that utilizes position-specific score matrix (PSSM) and machine learning methods for predicting epitope and immune interactions, respectively (Rapin et al., 2010). The simulation was performed with default parameters. The simulation volume was 1,000, simulation steps were 1,000, the random seed was 12,345, and the vaccine injection with no LPS (Abraham Peele et al., 2021; Samad et al., 2020). Other parameters were kept as default.

3. Results
3.1. CD171 sequence analyze and structural prediction
The amino acid sequence of CD171 contains 1257 amino acids, which ectodomain residues ranging from 19 to 1120
comparison, the multi-epitope vaccine were selected, containing both B- and T-cell epitopes. By candidates for multi-epitope vaccine overlapping peptides (CTL) and 81 of Helper T Lymphocytes (HTL)), a total of 12 adjacent T-cell epitopes (95 of Cytotoxic T Lymphocytes). Taken together, based on the 29 B-cell epitopes and the 176 linear B cell epitopes, 4 epitopes, including “EASGKPEV”, “WREGSQRKH”, “PLDEGGKQ” and “VPKEGQ”, had the VaxiJen score of more than one, which indicates the high antigenicity nature of these epitopes and that they can be considered the most potential antigenic B cell epitopes, as shown in Supplementary Table S1.

3.2. Prediction of B-cell epitopes

Results using the ABCpred server and based on VaxiJen scores showed 29 linear B cell epitopes, varying in peptide lengths from 6 to 37. However, among the 29 predicted linear B cell epitopes, 4 epitopes, including “LQANDTGRY”, “RGYNVTYWR”, “LTDLSPHLRY”, “LQANDTGRFY”, “RLMAEGAPK”, “VTMQGQNLNY”, “ETARLDCQV”, “FFTSTPEGV”, “QVKGHLRGY”, “RFQGYRRCF”, “SLGSAHAY”, “TRFVTAINK”, “TYWREGSOR”, “KTNGTRVR”, “MAVTNGTGR”, “GTAMSHEIR”, “GSQRKHSKR” and “ETWRIINGI”; and among the 81 predicted MHC Class II epitope, 5 epitopes, including “SFTITGNSNSFAQRF”, “KILHIKQDERVTMQGQ”, “ILHIKQDERVTMQGQ”, “NSKLHIKQDERVTM” and “LHIKQDERVTMQQGQ” had the VaxiJen score of more than one, which demonstrates the high antigenicity nature of these epitopes, as shown in Supplementary Tables S2 and S3 respectively.

3.3. Prediction of T-cell epitopes

Among the 95 predicted MHC Class I epitope, 18 epitopes, including “LQANDTGRY”, “RGYNVTYWR”, “LTDLSPHLRY”, “LQANDTGRFY”, “RLMAEGAPK”, “VTMQGQNLNY”, “ETARLDCQV”, “FFTSTPEGV”, “QVKGHLRGY”, “RFQGYRRCF”, “SLGSAHAY”, “TRFVTAINK”, “TYWREGSOR”, “KTNGTRVR”, “MAVTNGTGR”, “GTAMSHEIR”, “GSQRKHSKR” and “ETWRIINGI”; and among the 81 predicted MHC Class II epitope, 5 epitopes, including “SFTITGNSNSFAQRF”, “KILHIKQDERVTMQGQ”, “ILHIKQDERVTMQGQ”, “NSKLHIKQDERVTM” and “LHIKQDERVTMQQGQ” had the VaxiJen score of more than one, which indicates the high antigenicity nature of these epitopes, as shown in Supplementary Tables S2 and S3 respectively.

3.4. Multi-epitope vaccine

Taken together, based on the 29 B-cell epitopes and the 176 adjacent T-cell epitopes (95 of Cytotoxic T Lymphocytes (CTL) and 81 of Helper T Lymphocytes (HTL)), a total of 12 candidates for multi-epitope vaccine overlapping peptides were selected, containing both B- and T-cell epitopes. By comparison, the multi-epitope vaccine “QTYMAVQGST AYLLCKAFGAPVP5QWLDDEGT1VLQDER” contained eight (highest count) T-cell epitopes. These data and the antigenicity and allergenicity of the vaccine candidates were shown in Table 1.

3.5. Prediction of the physicochemical properties and solubility of the vaccine candidate

The multi-epitope vaccine had an estimated molecular weight of 64.5 kDa, a theoretical isoelectric point (pI) of 6.23, which is below seven, demonstrating a high proportion of negative charge against positively charged residues in the protein. The large average of the hydropathicity index (GRAVY) was −0.429, indicating that it was a hydrophilic protein; therefore, it can interact in aqueous solutions (Ali et al., 2017). The aliphatic index was 75.14, indicating that the vaccine can be stable over a wide temperature range. The estimated half-life in mammalian reticulocytes in vitro was 30 hours, and >20 hours in yeast and >10 hours in E. coli. The extinction coefficient has been calculated from 94660 to 95035 M-1 cm-1, which provides support for a quantitative study of protein-ligands and protein-protein interaction in solution, and indicates poor water solubility. The final vaccine demonstrates appropriate physicochemical properties such as high stability which can improve immunogenicity, bioavailability and decrease the probable side effects without any allergenicity (Mahdevar et al., 2021; Safavi et al., 2020). The results of the physicochemical properties and also the allergenicity of the candidate peptide are shown in Table 2.

3.6. Multi-epitope vaccine design

After the prediction of B- and T-cell epitopes, we fused them in order to generate a multi-epitope peptide by using suitable linkers. In order to produce sequences with minimized junctional immunogenicity, AAY linkers were combined between the predicted epitopes. Cholera toxin subunit B (CTB) sequence as an adjuvant was constructed within the final vaccine model at the N-terminal, using a DPRVPSS linker to improve the immunogenic capacity of the peptides by stimulating the innate immune response. Moreover, to aid in protein purification and identification, a 6xHis tag was added at the C-terminal. The final vaccine peptide was constructed with 582 residues derived from 12 merged peptide sequences. A schematic diagram of the vaccine is displayed in Figure 2A.
| NO. | Sequence | Position | Count of T Epitopes | Count of B Epitopes | Ab (Vaxijen score) | AllergenFP | AllerTOP | Toxicity |
|-----|----------|----------|---------------------|---------------------|-------------------|------------|----------|----------|
| 1   | YQSPHGSF1ITGNNSFAQ | 68-87 | 1 | 1 | 1 | 0.9340 | PROBABLE ALLERGEN | PROBABLE ALLERGEN | Non-Toxin |
| 2   | VLPCNPPSAELIRYWMNSKHLIQDERVTM | 136-168 | 3 | 1 | 1 | 0.6857 | PROBABLE ALLERGEN | PROBABLE NON-ALLERGEN | Non-Toxin |
| 3   | HLVALQGQPLVLEIAGFPTPTIKLWRPSGMPADRVITYQN | 232-273 | 1 | 1 | 1 | 0.4400 | PROBABLE NON-ALLERGEN | PROBABLE NON-ALLERGEN | Non-Toxin |
| 4   | VQGRPQPTEVTRINGPVVEALQDKYRIQGALIL | 337-372 | 1 | 1 | 1 | 0.3826 | PROBABLE NON-ALLERGEN | PROBABLE NON-ALLERGEN | Non-Toxin |
| 5   | VQPSDITQCEARNRHGLLLANAYIVYV | 375-402 | 4 | 1 | 1 | 0.4472 | PROBABLE NON-ALLERGEN | PROBABLE NON-ALLERGEN | Non-Toxin |
| 6   | QTYMAVQGSTAYLLCKAFGAPVPSQWLDEGTVLQDER | 415-456 | 5 | 3 | 1 | 0.2626 | PROBABLE NON-ALLERGEN | PROBABLE NON-ALLERGEN | Non-Toxin |
| 7   | PYVHYTFRTVANKYGPGEPSVEVTVPQEAPEKNPVVKGEGNETT | 662-710 | 2 | 4 | 1 | 0.6377 | PROBABLE NON-ALLERGEN | PROBABLE NON-ALLERGEN | Non-Toxin |
| 8   | QWRPOQTRGFWOEIVSDFRIVS | 734-757 | 2 | 1 | 1 | 0.5074 | PROBABLE ALLERGEN | PROBABLE ALLERGEN | Non-Toxin |
| 9   | TSTFVPYEINTQAVNSOQKGPEQVTGISGEDYPQAIPELEGILNSS | 759-808 | 3 | 2 | 1 | 0.7410 | PROBABLE NON-ALLERGEN | PROBABLE NON-ALLERGEN | Non-Toxin |
| 10  | FNGRGGPASEFTFSPVGPVHPE | 879-903 | 1 | 4 | 1 | 0.6699 | PROBABLE NON-ALLERGEN | PROBABLE NON-ALLERGEN | Non-Toxin |
| 11  | SPRLRYRFLQATTKENGPAIVR | 965-988 | 1 | 4 | 1 | 0.6276 | PROBABLE NON-ALLERGEN | PROBABLE NON-ALLERGEN | Non-Toxin |
| 12  | FRFHIFKALGEEKGASLSPQYVSYQSSYQWDLQPD | 1026-1066 | 4 | 2 | 1 | 0.4705 | PROBABLE ALLERGEN | PROBABLE NON-ALLERGEN | Non-Toxin |
### Table 2. Physicochemical properties of the vaccine candidate (MW: Molecular weight, Ai: Aliphatic index and Ec: Extinction coefficient).

| Sequence | LENGTH | Ab(Vaxijen score) | AllergenFP | AllerTOP | GRAVY | PI | MW (Da) | Solubility | Ai | Ec (M<sup>-1</sup> cm<sup>-1</sup>) |
|----------|--------|-------------------|------------|----------|--------|----|---------|------------|----|------------------|
| MIKLKFGVFFTTLSSAYAHTGP | S82 | 0.4913 | Probable | PROBABLY PROBABLY | -0.365 | 6.23 | 64576.83 | Poor water solubility | 75.14 | 94660 - 95035 |
| QNTIDLCAEYHNTQYTLNDKIF | | | | | | | | | | |
| SYTELAGKREMAITIFKNGAIF | | | | | | | | | |
| QVEPVGSQHIDSKKAIERMK | | | | | | | | | |
| DTTRAYITKEKVLWNN | | | | | | | | | |
| KTPHAIAISMANDPRVPSYQ | | | | | | | | | |
| SPHGSGITITGNNSNFQAAAY | | | | | | | | | |
| VLPCCNPSSAEPMRIYWMNSK | | | | | | | | | |
| ILHIKQDRTMAAYHVLVAL | | | | | | | | | |
| QGQPLVCEIAEGFPTIK | | | | | | | | | |
| WLRPSGMPADTVQNYA | | | | | | | | | |
| AYVQGRQPQEVWRINGIP | | | | | | | | | |
| VEELAKQDRYIRQGLIILA | | | | | | | | | |
| AYVQPQSDTMTQCEARNR | | | | | | | | | |
| GLLLNNAYYYVAAYQTYMA | | | | | | | | | |
| VQGSTAYLCLKAFQAPVPSV | | | | | | | | | |
| QWLDGQTVLQGDERAAPPY | | | | | | | | | |
| VHYTFRVATINKYGPGPSVSE | | | | | | | | | |
| TVTPEAEPKNVDKGE | | | | | | | | | |
| NETTAAAYQWRQGRPGQWE | | | | | | | | | |
| QIVSIGFVLSAATTSTFVPEIK | | | | | | | | | |
| VQAVNSQGKIEPQVITIGSSE | | | | | | | | | |
| DYPQRLEPEGIELNSAAAYFNG | | | | | | | | | |
| RGSGASEFTSTPTEGPGPEA | | | | | | | | | |
| AYSFHLRFFQLOATTKEGGEA | | | | | | | | | |
| IVRAAYRFHMLKALGEGEKCGGLSL | | | | | | | | | |
| SPQVSYNQSSSTQWDLQPD | | | | | | | | | |
| TDHHHHHH | | | | | | | | | |

**Figure 2.** Schematic presentation and structure prediction of the final multi-epitope vaccine construct. (A) vaccine construct (S82 residues long) containing an adjuvant (green) in N-terminal linked with 12 epitopes through a DPRVPS linker (red). 12 epitopes fused together using AAY linkers (Light cyan). A 6x-His tag is added at the carboxy terminus for purification and identification purposes. (CTB; cholera toxin subunit B), (B) PSIPRED predicted secondary structure of final chimeric peptide and (C) Predicted 3D model homology structure of final vaccine, helix, sheet and loop respectively are represented by red, yellow and green.
3.7. Secondary structure prediction

The PSIPRED prediction method was used to predict the secondary structure of the final chimeric peptide, which accomplished output analyze obtained from the PSI-BLAST and was submitted in FASTA format. The obtained secondary structure prediction revealed that the protein had to contain 2% alpha helix, 47% beta strand, and 49% coil (Figure 2B). As well, considering the accessibility of the amino acids to solvents, it was predicted that 50% would be exposed, 32% exposed to the medium exposed, and 17% would be buried. The RaptorX Property server predicted a total of 39 residues (6%) to be located in disordered domains.

3.8. Tertiary structure modeling and validation

In total, five models of tertiary structure of the designed chimeric protein were predicted based on 10 threading templates by the I-TASSER server. The Z-score values of the 10 chosen templates ranged from 0.72 to 12.98 and showed good alignment. The C-score values of five predicted models ranged from −3.11 to −1.16. The range of C-score values is typically between −5 and 2, with higher values showing higher confidence. The model with the highest C-score was selected from the homology modeling (Figure 2C). This model had an estimated TM-score of 0.53 ± 0.15 with an estimated RMSD of 11.2 ± 4.6 Å. The TM-score has been suggested as a scale for measuring the structural similarity between two structures (Ali et al., 2017) and to overcome the problem of RMSD, which is sensitive to local error. A TM-score of more than 0.5 indicates a correct topology model. These cut-off values are independent of protein length. RAMPAGE, a determinative tool, was assigned to evaluate the reliability model, generate a Ramachandran plot and determine the energy of the stable conformation of the psi (ψ) and phi (Φ) twisting or dihedral angles for each amino acid. The results of the tertiary structure validation of the Ramachandran plot analyze show that the number of residues from favorable regions was 69.5% and, additionally, the allowed region residues were 20.2%, and only 10.3% of the residues were found in the outlier region. The total percentage of favoured and allowed region residues was 89.7%, while more than 90% is an ideal result to make the mode believable and convincing (Figure 3A). The quality and potential errors in the crude 3D model were verified by ProSA-web, which gave a Z-score of −2.25 for the chosen model of the input vaccine protein (Figure 3B). Both the Ramachandran plot and the ProSA-web score authenticated the quality of the CD171 3D model.

3.9. In silico cloning adaptation of designed vaccine

The Java Codon adaptation tool (JCat) was utilized to optimize the use of codons in the vaccine constructed on strains of E. coli K12, for maximum protein expression and differentiate the human and E. coli expression system. The optimized codon sequence has 1746 nucleotides in length. The Codon Adaptation Index (CAI) of the optimized codon sequence was found to be 0.96. The optimal range of the CAI index was between zero to one, showing the probable success of the target gene expression. The GC sequence content was 53.3%, which is also satisfactory because it shows the possibility of good expression of the vaccine candidate in the host E. coli. The ideal percentage range of GC content is between 30% and 70%. Eventually, using SnapGene software for restriction cloning, the recombinant plasmid sequence was constructed by adding the adapted codon sequences into the pET30a (+) vector between Nde I and XhoI restriction sites (Figure 4).

3.10. Molecular docking of the final vaccine constructs with MHC I

The results for molecular docking between the final vaccine and MHC I demonstrated 30 clusters for each docked complex ranked (0–29) based on cluster members and furthermore the weighted scores (energies) of each cluster were...
Presented. Among all the generated docking models, cluster 3 of the MHC I–vaccine docked complex with 35 members having the lowest energy of $-1388.7$ was selected for further analyze as the best-docked complex, suggesting that the vaccine model occupies the receptor properly and indicating good binding affinity. The weighted scores of the selected clusters demonstrate the docked structure is more stable than each structure alone thus confirming the rational design of the peptide vaccine with respect to the MHC I and affirm that it is able to identify the correct epitopes in the MHC I structure. The interaction surface residues of docked structures are illustrated in Figure 5.

3.11. Molecular dynamics simulation of the final vaccine-MHC I complex

Molecular dynamics simulation through the iMOD server was used to evaluation of the binding stability and physical movements of the vaccine- MHC I docked complex. The range of main-chain deformability was 0.1 to 0.9 Å through the chain hinge (regions as high deformability) distortion evaluates and showed low deformation and with the steady binding of the complex (Figure 6A). The B-factor values are proportional to the root mean square and demonstrated that there were few atomic fluctuations in the complex (Figure 6B). Eigen score of the complex was calculated as $8.367 \times 10^{-7}$ which is closely associated with the energy needed to structure deformation (Figure 6C). The results of the covariance matrix between the pairs of residues indicate their correlations, correlated (red color), non-correlated (blue color), and uncorrelated (white color) (Figure 6D). The elastic network model results showed the pairing of atomic coordinates through distance-dependent spring analyze (Figure 6E). Single dot in the elastic network plot indicates one spring and is colored according to the stiffness of the complex with respective atomic pairs. The obtained spring models that are dark grey in color show the stability and compactness of the binding complex. These significant molecular dynamic simulation results suggest the complex rigidity and stable binding of the vaccine complex with few atomic fluctuations and a low deformation scores.

3.12. Immune response simulation

The results of the simulated immunogenicity and immune response profile of the multiepitope-designed vaccine show...
the secondary and tertiary responses are characterized by higher levels of IgM, IgG + IgM, and IgG1 + IgG2 and B-cell populations (Figure 7A–C). This profile indicates the development of immune memory and subsequent clearance of the antigen. A similar pattern is also seen in TH (helper) and TC (cytotoxic) cell populations with corresponding memory development (Figure 7D and E). During the exposure time, production levels of cytokine profile were additionally observed (Figure 7F).

4. Discussion

Malignant gliomas are rare and indicate an incidence of 2.5% of the leading cause of cancer death (Hanif et al., 2017). Many vaccines against glioblastoma have been examined in cell cultures and animal models, but none have been used for therapeutic application in humans so far (Kindy et al., 2016; Oh et al., 2014). Epitope-based vaccines as a promising approach have been considered to generate a specific immune response and avoid responses against other unfavorable epitopes on the complete antigen (Moise et al., 2015). Various advantages of this approach include increased safety, the opportunity to rationally engineer the epitopes for increased potency and breadth, and the ability to focus immune responses on conserved epitopes (Zhou et al., 2009) but also achieved phase-I clinical trials (Jiang et al., 2017; Lennerz et al., 2014; Romeli et al., 2020; Slingluff et al., 2013; Toledo et al., 2001). Recently in the same approach, the immunoinformatics designed vaccine has been used for designing multi-epitope vaccines against melanoma (Safavi et al., 2019), breast cancer (Mahdevar et al., 2021), Hendra virus (Kamthania et al., 2019), Leishmania (Ropón-Palacios et al., 2019), and also SARS-CoV-2 (Moghi et al., 2021). Reports demonstrate CD171 is only expressed on the surface of cancer cells, not their normal tissues therefore the immune responses generated against the CD171 will affect cells expressing high levels of CD171 (Debiec et al., 1998; Gavert et al., 2008; Huszar et al., 2006; Kowitz et al., 1993). Moreover, cancer cells reexpress CD171 which has a significant role in embryonic and fetal morphogenesis and it is absent in healthy human adult tissue (Pechriggl et al., 2017). In other hand we can incorporate the personal medicine and tumor profiling before administration. Although it still requires validation by biological or experimental approaches as a cancer vaccine candidate for glioblastoma. In this research, several precise bioinformatics tools were applied to gather information about the candidate vaccine; the first step was the prediction of B- and T-cell epitopes in the protein sequence and, in order to design a multi-epitope peptide vaccine, the selected epitopes were linked using suitable linker sequences (Meza et al., 2017). To the production of sequences with minimized junctional immunogenicity, AAY linkers (Khatoon et al., 2017) were inserted between the predicted epitopes, thus leading to the rational multi-epitope vaccine design construction (Meza et al., 2017). Also, in order to attain a high level of expression
and improved bioactivity of the fusion peptide the DPRVPSS linker was inserted between the adjuvant protein sequence and the designed epitopes. The constructed vaccine candidate indicated that it contains 28 and 25 numbers of high-affinity MHC class I, MHC class II respectively, and 12 numbers of linear B-cell epitopes by the analyzing of immuno-informatics results. The absence of allergenic properties of the designed peptide chimera indicates could be promising potential as a vaccine candidate. The designed multi-epitope vaccine has been shown Vaxijen scores 0.49 and 0.50, with and without the addition of an adjuvant sequence respectively. Similar antigenicity scores with and without an adjuvant sequence can suggest that it may be valuable to express the chimeric peptide without the adjuvant. As a result, the vaccine candidate molecular weight is 64.5 kDa and its water solubility was predicted to be poor. In order to do many biochemical and functional research, the solubility of the overexpressed recombinant protein in the E. coli host is one of the needed agents (Khatoon et al., 2017). The candidate vaccine theoretical pI is predicted to be 6.23 showing that the vaccine is acidic in nature. The aliphatic index (75.14) indicated that the vaccine has aliphatic side chains, showing potential hydrophobicity. All these above parameters demonstrate the vaccine is thermally stable.

Considering the importance of secondary and tertiary structures knowledge in target protein as an essential factor in vaccine design (Meza et al., 2017) the secondary structure results demonstrated that the protein contained mainly of coils (49%), with only 6% of residues disordered. The 3D structure of the vaccine candidate demonstrated desirable properties based on Ramachandran plot predictions. The Ramachandran plot indicates that most of the residues are found in the favoured and allowed regions (89.7%) with few residues in the outlier region which shows satisfaction with the quality of the overall model. The vaccine was evaluated via docked and dynamic with MHC I that can trigger cytotoxic T lymphocytes. Molecular docking assessments showed an appropriate and stable interaction between the vaccine and the MHC I. Furthermore, significant molecular dynamic simulation results suggest the complex rigidity and stable binding of the vaccine with few atomic fluctuations and a low deformation index. These observations of immune response simulation suggest that the designed vaccine likely induces immune reactions as evidenced by a marked increase in the generation of secondary responses. The codon optimization was performed to attain the high-level expression of our recombinant vaccine in E. coli (strain K12). Both the codon adaptability index (0.96) and the GC content (53.3%) were desirable for high-level expression of the vaccine in bacteria. Eventually to the validation of our immuno-informatics results, the peptide expression in a bacterial system and fulfilling the various immunological assays is essential.

5. Conclusions

In the formulation of novel drugs or vaccines, immunoinformatics approaches provide new insights and appear as
interdisciplinary strings that take different materials to overcome the difficulties related to a long time and high expense of therapy development. Herein, several immuno-informatics tools were applied to design a promising vaccine peptide to successfully generate a functional vaccine and the suggested vaccine candidate could potentially be used to glioblastoma neoplasm cell elimination.

**Disclosure statement**

No potential conflict of interest was reported by the authors.

**ORCID**

Mojtaba Ranjbar [http://orcid.org/0000-0002-3884-3447](http://orcid.org/0000-0002-3884-3447)

**Authors contribution**

Work design and conceptualization was done by MR and GMC. All the Authors contribution

**References**

Abraham Peele, K., Srihansa, T., Krupanidhi, S., Ayyagari, V. S., & Venkateswarulu, T. C. (2021). Design of multi-epitope vaccine candidate against SARS-CoV-2: A in-silico study. Journal of Biomolecular Structure and Dynamics, 39(10), 3793–3801. https://doi.org/10.1080/07391102.2020.1770127

Ali, M., Pandey, R. K., Khatoon, N., Narula, A., Mishra, A., & Prajapati, V. K. (2017). Exploring dengue genome to construct a multi-epitope based subunit vaccine by utilizing immunoinformatics approach to battle against dengue infection. *Scientific Reports*, 7(1), 1–3. Aug 23 https://doi.org/10.1038/s41598-017-09199-w

Baldauf, K. J., Royal, J. M., Hamorsky, K. T., & Matoba, N. (2015). *Cholera* vaccine candidate could potentially be used to glioblastoma multiforme: A review of its epidemiology and pathogenesis. *Glioblastoma multiforme*, 1, 1–20. Dec https://doi.org/10.1089/gbi.2014.0138

DeLano, W. L. (2002). *PyMol*: An open-source molecular graphics tool. *CCP4 Newsletter on Protein Crystallography*, 40(1), 82–92.

Dimitrov, I., Bangov, I., Flower, D. R., & Doytchinova, I. (2014). AllerTOP v.2—a server for in silico prediction of allergens. *Journal of Molecular Modeling*, 20(6), 2278–2276. Jun https://doi.org/10.1007/s00894-014-2278-5

Doytchinova, I. A., & Flower, D. R. (2002). Quantitative approaches to computational vaccinology. *Immunology and Cell Biology*, 80(3), 270–279. Jun https://doi.org/10.1046/j.1440-1711.2002.01076.x

Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server. In *The proteomics protocols handbook* (pp. 571–607). https://doi.org/10.1385/1-59259-800-0:571

Gatvert, N., Ben-Shmul, A., Ravel, S., & Ben-Zeev, A. (2008). L1-CAM in cancerous tissues. *Expert Opinion on Biological Therapy*, 8(11), 1749–1757. Nov 17 https://doi.org/10.1517/14712598.8.11.1749

Geer, L. Y., Marchler-Bauer, A., Geer, R. C., Han, L., He, J., He, S., Liu, C., Shi, W., & Bryant, S. H. (2010). The NCBI biosystems database. *Nucleic Acids Research*, 38(Database issue), D492–6. Jan https://doi.org/10.1093/nar/gkq318

Grote, A., Hiller, K., Scheer, M., Münch, R., Nörtemann, B., Hempel, D. C., & Jahn, D. (2005). JCat: A novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Research*, 33(Web Server issue), W526–W531. https://doi.org/10.1093/nar/gkq367

Gupta, S., Kapoor, P., Chaudhary, K., Gautam, A., Kumar, R., & Raghava, G. P. (2013). In silico approach for predicting toxicity of peptides and proteins. *PloS One*, 8(9), e73957. https://doi.org/10.1371/journal.pone.0073957

Hafid, F., Muzaffar, K., Perveen, K., Malhi, S. M., & Simjee, S. U. (2017). Glioblastoma multiforme: A review of its epidemiology and pathogenesis through clinical presentation and treatment. *Asian Pacific Journal of Cancer Prevention : APJCP*, 18(1), 3–9. https://doi.org/10.22034/ APJCP.2017.18.1.3

Huszar, M., Moldenhauer, G., Gschwend, V., Ben-Arie, A., Altevogt, P., & Fogel, M. (2006). Expression profile analysis in multiple human tumors identifies L1 (CD171) as a molecular marker for differential diagnosis and targeted therapy. *Human Pathology*, 37(8), 1000–1008. https://doi.org/10.1016/j.humpath.2006.03.014

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), W24–9. Jul 3https://doi.org/10.1093/nar/gkw346

Jiang, P., Cai, Y., Chen, J., Ye, X., Mao, S., Zhu, S., Xue, X., Chen, S., & Zhang, L. (2017). Evaluation of tandem Chlamydia trachomatis MOMP multi-epitopes vaccine in BALB/c mice model. *Vaccine*, 35(23), 3096–3103. May 25https://doi.org/10.1016/j.vaccine.2017.04.031

Kamthania, M., Srivastava, S., Desai, M., Jain, A., Shrivastav, A., & Sharma, D. K. (2019). ImmunoInformatics Approach to design T-cell epitope-
evaluation of a novel multiepitope peptide vaccine consisted of immunodominant epitopes of SYCP1 and ACRBP antigens as a prophylactic melanoma vaccine. *International Immunopharmacology*, 76, 105827Nov 1 https://doi.org/10.1016/j.intimp.2019.105872

Samad, A., Ahammad, F., Nain, Z., Alam, R., Imon, R. R., Hasan, M., & Rahman, M. S. (2020). Designing a multi-epitope vaccine against SARS-Cov-2: An immunoinformatics approach. *Journal of Biomolecular Structure and Dynamics*, 1–7. Jul 17: https://doi.org/10.1080/07391102.2020.1792347

Schäfer, M. K., & Altevogt, P. (2010). L1CAM malfunction in the nervous system and human carcinomas. *Cellular and Molecular Life Sciences : CMLS*, 67(14), 2425–2437. Jul https://doi.org/10.1007/s00018-010-0339-1

Siesser, P. F., & Maness, P. F. (2009). L1 cell adhesion molecules as regulators of tumor cell invasiveness. *Cell Adhesion & Migration*, 3(3), 275–277. Jul 1 https://doi.org/10.4161/cam.3.3.8689

Skwarczynski, M., & Toth, I. (2014). Recent advances in peptide-based subunit nanovaccines. *Nanomedicine (London, England)*, 9(17), 2657–2669. Dechttps://doi.org/10.2217/nmm.14.187

Wang, S., Peng, J., Ma, J., & Xu, J. (2016). Protein secondary structure prediction using deep convolutional neural fields. *Scientific Reports*, 6(1), 18962–18961. Jan 11 https://doi.org/10.1038/srep18962

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), W407–10. Jul 1 https://doi.org/10.1093/nar/gkm290

Yang, M., Li, Y., Chilukuri, K., Brady, O. A., Boulos, M. I., Kappes, J. C., & Galileo, D. S. (2011). L1 stimulation of human glioma cell motility correlates with FAK activation. *Journal of Neuro-Oncology*, 105(1), 27–44. Oct https://doi.org/10.1007/s11060-011-0557-x

Zhang, Q., Wang, P., Kim, Y., Haste-Andersen, P., Beaver, J., Bourne, P. E., Bui, H.-H., Buus, S., Frankild, S., Greenbaum, J., Lund, O., Lundegaard, C., Nielsen, M., Ponomarenko, J., Sette, A., Zhu, Z., & Peters, B. (2008). Immune epitope database analysis resource (IEDB-AR). *Nucleic Acids Research*, 36(Web Server issue), W513–8. May 31 https://doi.org/10.1093/nar/gkn254

Zhou, W. Y., Shi, Y., Wu, C., Zhang, W. J., Mao, X. H., Guo, G., Li, H. X., & Zou, Q. M. (2009). Therapeutic efficacy of a multi-epitope vaccine against Helicobacter pylori infection in BALB/c mice model. *Vaccine*, 27(36), 5013–5019. Aug 6 https://doi.org/10.1016/j.vaccine.2009.05.009