



Designing a Chimeric Vaccine Against Colorectal Cancer

Ziba Veisi Malekshahi,¹ Masoumeh Rajabibazl,^{2,3} Walead Ebrahimizadeh,⁴ Jafar Amani,^{5,*} and Babak Negahdari^{1,**}

¹Department of Medical Biotechnology, School of Advanced Technologies in Medicine, Tehran University of Medical Sciences, Tehran, IR Iran

²Department of Clinical Biochemistry, Faculty of Medicine, Shahid Beheshti University of Medical Sciences, Tehran, IR Iran

³School of Advanced Technologies in Medicine, Shahid Beheshti University of Medical sciences, Tehran, IR Iran

⁴Department of Medical Biotechnology, School of Advanced Technologies in Medicine, Shiraz University of Medical Sciences, Shiraz, IR Iran

⁵Applied Microbiology Research Center, Systems Biology and Poisonings Institute, Baqiyatallah University of Medical Sciences, Tehran, IR Iran

*Corresponding author: Dr. Jafar Amani, Applied Microbiology Research Center, Systems Biology and Poisonings Institute, Baqiyatallah University of Medical Sciences, Vanak Sq. Molasadra St. P.O. Box: 19395-5487, Tehran, IR Iran. Tel: +98-2182482568, Fax: +98-2188068924, E-mail: jafar.amani@gmail.com

**Corresponding author: Dr. Babak Negahdari, School of Advanced Technologies in Medicine, Tehran University of Medical Sciences, Enghelab Sq. Italia St. Tehran, IR Iran. Tel: +98-2143052000, Fax: +98-2188991117, E-mail: b-negahdari@sina.tums.ac.ir

Received 2016 October 22; Revised 2016 December 10; Accepted 2017 December 12.

Abstract

Background: Colorectal cancer is the major cause of death worldwide. Recently, cancer immunotherapy has been used in cancer prevention and treatment which has led to immune system response to cancerous cells and their elimination. Recent studies were conducted to specifically target tumor marker to design vaccine.

Objectives: Here, we designed a vaccine chimeric including CEA and CA19-9 against colorectal cancer.

Methods: The construct was analyzed, using bioinformatics tools and servers. The physicochemical properties, structures, antigenicity, stability, MHC binding properties, and ligand-receptor docking of chimeric protein (CE-CA) were predicted.

Results: The results showed the structure of CE-CA has suitable form and separates each domain; furthermore, the epitope mapping did not change after combination.

Conclusions: The results showed that the construct can be appropriate vaccine against colorectal cancer and could generate potent immune response against this cancer. In silico tools are useful to design vaccine.

Keywords: Cancer Anti-Gens, Bioinformatics, CEA, CA19-9, B-Cell Epitope

1. Background

Colorectal cancer is considered a major gastrointestinal. Colorectal cancer is the second cancer-related cause of death after lung cancer around the world (1). Based on the epidemiologic studies, each year 1 million new cases are diagnosed with colorectal cancer and 500 000 patient with cancer lose their life to the disease (2). The highest incidence of colorectal cancer is reported in North America, Australia, New Zealand, Europe, and Japan and the lowest incidence is in South America, Asia, and Africa (3). Colorectal cancer has the third highest number of patients and is the second reason cause of death due to cancer (4). Both genders have equal chance of developing colorectal cancer and the disease is mostly seen in elderly people. Approximately, 90% of cases are diagnosed after the age of 50; thus, the target group for diagnostic assayed is this group (5).

Common treatment includes surgery, chemotherapy, and radiotherapy, which in most cases are accompanied by the regression of cancer and metastasis. Furthermore,

these treatment methods are considered invasive and hazardous and are ineffective in prevention of metastasis. Studies show that the immune system is capable against cancer and can prevent metastasis if properly guided. Thus, enhancing the immune system and introduction of tumor associated antigens to the immune cells can greatly affect the outcome of treatment (6). One of the methods in boosting the immune system toward battling cancer is the use of cancer vaccines, which has significantly gained the interest of cancer researchers in the past few years. The main goal of cancer immunotherapy is immune reorganization and removal of cancer cells. Several research studies showed that antigens expressed by tumor cells can elicit specific cellular and humoral immune responses (7). Antigen cancer vaccines are important in immunotherapy consisting of cytotoxic T lymphocyte (CTL) epitopes from tumor-specific antigens (TSAs) or tumor-associated antigens (TAA) (8).

One of the major drawbacks in developing an effective diagnosis method or treatment in colorectal cancer is lack

of specific tumor marker. Most tumor markers in colorectal cancer are genetic mutation that are detected through the analysis of multiple genes, such as P53, Ras or microsatellite instability and assaying for loss of heterozygosity long arm of chromosome 18 (9). The most commonly used protein markers are carcinoembryonic antigen (CEA) and CA19-9. Using each of these markers alone does not provide sufficient evidence to conform or rule out cancer. Usually both of these markers are used to provide reasonable evidence from the status of cancer (10, 11). If a cancer vaccine is to be developed against colorectal cancer, these 2 antigens need to be used. The proper choose of domains of the antigens and constructing a new cancer antigen that can stimulate the immune system against colorectal cancer cells have vital impact on the quality of the vaccine (12). Thus, bioinformatics plays an important role in design and development of a cancer vaccine. There are several cancer vaccines that have been developed so far, among those, regarding colorectal cancer, CEA had the most application (13). However CEA alone is not recognized as a specific colorectal cancer tumor associated antigen hence, using CEA in combination with another vaccine candidate could increase the specificity of the vaccine. In that regard, CA19-9 antigen comprises the appropriate characteristics of a vaccine candidate. CA19-9 is a cell-surface antigen that has a significant expression increase in colorectal cancer, and it has already been approved as a marker of colorectal cancer. Thus, based on available data, these 2 antigens in combination can provide specificity for production of colorectal cancer vaccine. In this study, using bioinformatics study, we aim at designing a colorectal cancer vaccine based of immune-dominant domains of CEA and CA19-9 proteins.

2. Methods

2.1. Domains Selection and Construct Design

CEA is a large antigen with a molecular weight of 180 kDa and it is significantly glycosylated; thus, approximately 60% of the weight of complete protein with PTM belongs to carbohydrates (14). Since most expression systems including simple Eukaryotes are unable to correctly mimic the glycosylation pattern of mammals, chosen domains need to have minimum glycosylation sites. Study on CEA glycosylation pattern revealed that domains 1, 4, 5, and 7 with respectively 2, 4, 3, and 3 glycosylation sites have less glycosylation among the proteins 7 domains. Domain 7 is located in the c-terminus of the proteins and, therefore, it is less likely to be recognized and interacted with the immune system components. Thus, domains 1, 4, and 5 are chosen for production of final construct. These domains contain the immunogenic sites and can fold independently to mimic the conformation of complete CEA.

Unlike CEA, CA19-9 is a much smaller protein with only 1 glycosylation site. Hence, after the removal of first 64 amino acids that are either signal peptide or cytoplasmic domain, the complete sequence was included in the final chimeric construct. This chimeric construct includes domain 1 of CEA attached to domains 4 and 5 via flexible Serine-Glycine linker, which are, then, linked to the CA19-9 sequence by a rigid linker to provide independent folding for each section of protein. The construct was named CE-CA. The related sequences of major antigens colorectal cancer (CEA, CA19-9) were obtained from Uniprot/Swiss-port and NCBI (15). The sequences were submitted to the basic local alignment search tool (BLAST) to confidence that the selected sequences were conserved (16).

2.2. Antigenicity and Allergenicity Evaluation

APPLE and ALGPred servers were performed to analyze the allergenicity of the construct from sequence derived structural and physicochemical properties of the whole protein. The accuracy of ALGPred could be over than 80% by combined approaches (17). VaxiJen server was used to the prediction of protective antigens, tumor antigens, and subunit vaccines (18).

2.3. 3D structural Model Prediction

Since proper folding of the recombinant chimeric construct have vital impact on the B-cell and T-cell immune response, the construct needs to have similar folding compared to the template proteins, namely CEA and CA19-9. Analysis of the second and third structures were carried out by GOR (19) and I-TASSER servers were, then, aligned with both CEA and CA19-9 (20).

2.4. Homology Modeling

Homology modeling of chimeric protein was performed by *phyre2* (protein Homology/analogy Recognition Engine V 2.0) server at (21).

2.5. Evaluation of Model Stability and Validation

3D structural stability of chimeric protein was analyzed by Swiss-pdbViewer software for energy minimization and *RAMPAGE* server (22).

For Tertiary structure validation, *ProSA* server (23), *PROCHECK* server (24), and *ERRAT* server (25) were used. *ProSA-web* is a server to check 3D models of protein structure for potential errors. The *proSA-web* z-score is shown in a plot that contains the z-score of experimentally determined protein structure in PDB. The residue-by-residue stereo chemical qualities of protein structure were validated, using Ramachandran plot obtained from

PROCHECK. The ERRAT server is a protein structure verification algorithm evaluating the statistics of non-bounded interactions between different atom types compared to a database of reliable high-resolution crystallography structures.

2.6. Analysis of mRNA

The secondary mRNA structure was predicted by GeneBee (26) and mRNA resulted from both these sequences analyzed by mFold server (27).

2.7. Codon Optimization

The nucleotide sequence of the CE-CA construct was optimized by Genscript Optimization Gene TM algorithm (www.genscript.com, Piscataway, New Jersey USA) based on codon bias of *E. coli* (28).

2.8. Analysis of Physical and Chemical Properties of the CE-CA Chimeric Protein

Physical and chemical properties were obtained using the ExPasy ProtParam (29). Physical and chemical properties included amino acid composition, molecular weight, pI, Grand average of hydropathicity (GRAVY), solubility in natural pH and pH ranging from 4 to 12, half-life and total number of positive and negative residues in prokaryotic and eukaryotic systems.

2.9. Solvent Accessibility Prediction

The prediction of protein solubility were performed by Proso server (30) and SOLpro server (31).

2.10. Analysis of Conserved Domains and Protein Localization

Transmembrane and conserved domain were performed, using DAS server (32). The amino acid conservation was determined by PRALINE server (33). Also, conserved functional and structural amino acid were performed by ConSurf server (34). The protein localization was estimated, using CELLO server in eukaryotic and prokaryotic system (35). The prediction of membrane protein topology and signal peptides were analyzed by OC-TOPUS server (36). In addition, SignalP 4.1 server (37) predicted presence and location of signal peptide cleavage site in amino acid sequence.

2.11. Prediction of B-Cell Epitopes

The linear B-cell epitopes were carried out by Bcepred server, ABCpred server, and Bepipred servers. Discotope 1.2 (38-40) and SEPPA servers were used to predict discontinuous B-cell epitopes (41). In addition, Ellipro server (42) was used to predict both linear and discontinuous B-cell epitopes. Ellipro is a method that predicts epitopes based on solvent accessibility and flexibility.

2.12. Prediction of Cleavage Sites

Proteasome cleavage sites of the chimeric protein were predicted by Netchop 3.1 (43), MAPPP (44), and PCPS (45). Peptides binding affinity to TAP protein was computed by TAPPred (46).

2.13. Prediction of MHC Binding Peptides Affinity

The chimeric protein was analyzed for MHC binding peptides. NetCTL (47), SYFPEITHI (48), and CTLpred (49) were used for MHC-presented epitopes and MHC-specific anchor and peptide motifs. NetMHC server (50) was used to product a neural network prediction of binding affinities for MHC.

2.14. Prediction of T-Cell Epitopes

For prediction of peptides from the antigenic sequence binding with MHC class I, Propred-I (51) and nHLAPred servers were used (52).

For prediction of peptides from the chimeric protein binding with MHC classII HLApred (52), MHC2Pred (53) and Propred were used (54).

2.15. Prediction of Post-Translational Modification

The prediction of N-glycosylation sites in chimeric protein was estimated by NetGlycate 1.0 (55) and NetNGlyc 1.0 servers (56). The prediction of O-glycosylation sites in this protein was performed by YinOYang 1.2 (57) and NetOGlycate servers (58). Myristoylator server (59) was used for prediction of N-terminal myristoylation site and 6 was estimated for prediction of phosphorylation sites. Prediction of potential c-terminal GPI-modification site was done by GPI server (60).

3. Results

3.1. The Design and Construction of Chimeric Gene

Two fragments of proteins, 551 amino acids from CEA, and CA19-9 of major colorectal cancer antigens were selected. These antigens were selected as a chimeric structure. Seine-Glycine linker was designed to separate the domains of CEA. Linkers consisting of EAAAK repeats were used to separate the domains of CEA and CA19-9. It was shown that helix formation can be stabilized by these linkers between different domains. Four repeated EAAAK sequences were introduced between 2 domains for more flexibility and efficient separation. The EcoRI and HindIII restriction sites for cloning in prokaryotic vectors were successfully introduced at the N- and C-terminal of sequences, respectively. The arrangement of fragment junction and linkers sites are shown in Figure 1A.

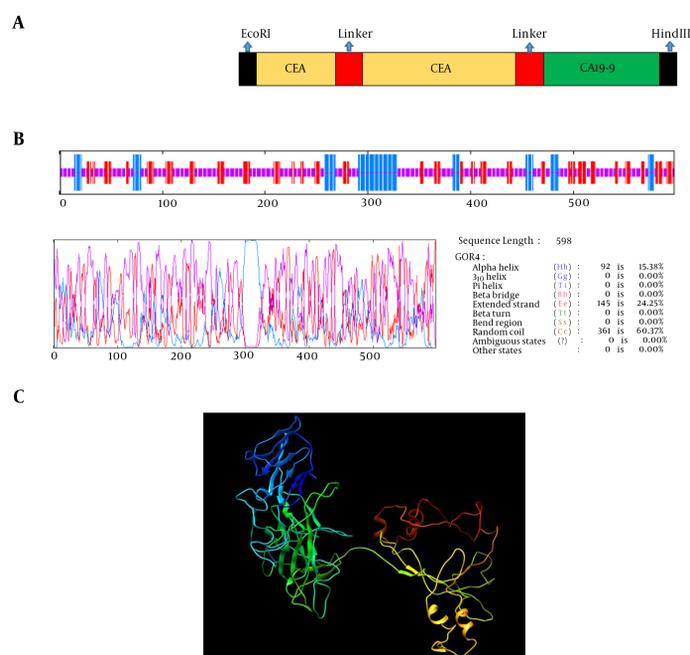


Figure 1. A, Schematic Model of Construction; B, analysis of chimeric CE-CA protein secondary structure; C, prediction of 3D structure of chimeric protein (CE-CA) by I-TASSER server. The result was viewed by Accelrys Discovery studio Visualizer 1.7 software.

3.2. Antigenicity and Allergenicity Evaluation

Antigen index by vaxijen server for CEA, CA19-9, and chimeric protein was 0.45, 0.40, and 0.48, respectively. The allergenicity analyses by APPLE and ALGPred servers showed the antigen as non-allergen.

3.3. Secondary Structure Prediction

The prediction of the secondary structure of the chimeric protein was performed by GOR server. The results showed that total residue is 598, of which alpha helix (15.38%), extended strand (24.25%), and random coil (60.37%) are structural constituents of the chimeric protein. The secondary structure prediction of the chimeric protein is shown in Figure 1B.

3.4. Homology Modeling

Phyre 2 employs the alignment of hidden Markov models via HH search to improve alignment accuracy and detection rate. This model also incorporates Poing, a new ab initio folding simulation to model regions of protein with no detectable homology with known structures.

3.5. Tertiary Structure Prediction

I-TASSER server was used for the prediction of the tertiary structure of protein. Tertiary prediction results of

the chimeric protein construction using I-TASSER showed a protein with 4 domains attached together with a linker (Figure 1C). The confidence score (C-score) for estimating the quality of the predicted models is typically in the range of -5 to 2. The C-score of models predicted by I-TASSER was -1.48. The z-score of the input structure was within the range of scores typically found for native protein of the similar size. Also, the template modeling (TM)-score for this model was 0.53 ± 0.15 and root-mean-square deviation (RMSD) was 11.2 ± 4 . The tertiary structures of the chimeric protein construct are shown in Figure 1C.

3.6. Evaluation of Model Stability and Validation

The quality and potential errors of 3D structure were investigated by ERRAT server and ProSA-web. The z-score of structure was -3.15 and the overall quality factor plot (ERRAT) of structure was 64.07%. The quality and potential errors of 3D structure showed in Figures 2A, 2B, and 2C.

3.7. mRNA Structure Prediction

The secondary structure of mRNA was predicted using mfold. The 5' terminus of the gene was folded typically as in all bacterial gene structures. The minimum free energy for secondary structure formed by RNA molecules was predicted. All 42 structural elements obtained in this analysis

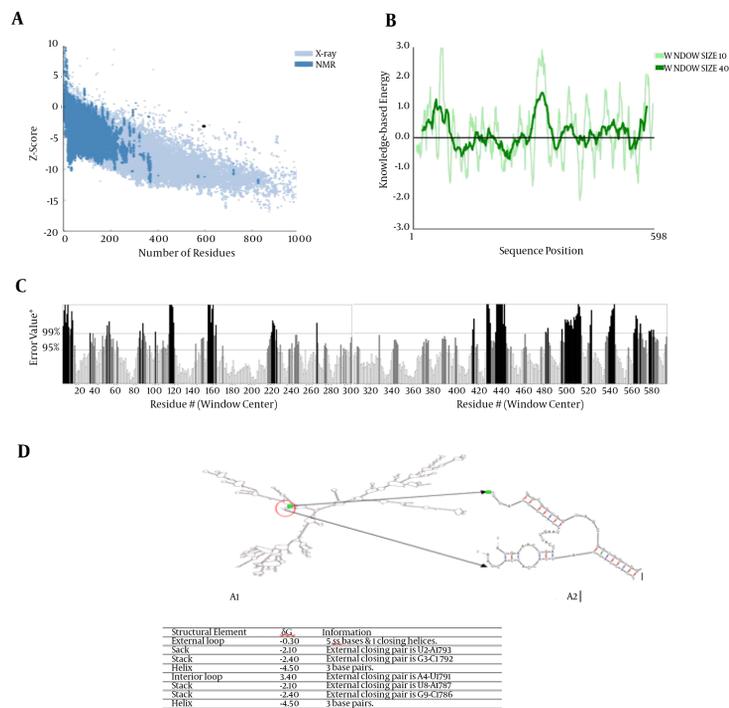


Figure 2. A, Z-score Plot for 3D Structure of Chimeric Protein Displayed Using NMR Spectroscopy (Dark Blue) and X-Ray Crystallography (Light Blue); B, the plot showed local model quality using plotting energies as a function of amino acids sequence position; C, The overall quality factor plot (ERRAT) of structure is 64.07%. The result of ERRAT plot showed the region of the 3D structure that can be disproved at the 95% confidence level in gray lines and region of the 3D structure that can be disproved at the 99% level displayed in black lines; D, Analysis of mRNA stability and start codon position in the structure and free energy details for mRNA structure by mfold server.

showed RNA Folding. The mRNA structure had a free energy of -527.74 kcal/mol and the first nucleotide at 5' did not have a long stable hairpin or pseudoknot (Figure 2D). The data have shown that the mRNA was stable enough for efficient translation in the new Host (Figure 2D).

3.8. Codon Optimization Analysis

Life technologies "Gene Optimizer" service is a gene optimization technology that can modify both recombinant and naturally gene sequences to gain the highest conceivable level of expression in any expression system. Both the wild type and construct were analyzed for their codon bias and GC content. The analysis of the sequence encoding, the optimized chimeric construct, and wild type gene are shown in Figures 3 - 5. The codon adaptation index (CAI) of chimeric construct was 0.77, while that of wild type gene was 0.66 (Figure 3). The percentage of codon having a frequency distribution of 90 to 100 in wild chimeric gene was 45% and 85% for *E. coli* and mus, respectively, which was significantly improved to 95% for *E. coli* and 85% for mus in the optimized gene sequence (Figure 3). The overall GC content was reduced from 56.78 to 51.54, which should increase the overall stability of mRNA from the synthetic

gene. Within the recombinant chimeric construct, splice sites, polyadenylation signal, instability elements, and all the cis-acting sites that may have a negative influence on the expression rate, were removed. Furthermore, the necessary restriction sites (EcoRI and HindIII) were at the end of the sequence for cloning purpose.

3.9. Evaluation of Model Stability

The profile of energy minimization was calculated by spdbv (Swiss-pdbviewer). The amount -23911.445 kcal/mol indicated that the recombinant protein had acceptable stability compared to that of original structure of each domain. Additionally, the data obtained by Ramachandran plot confirmed the structural stability of the protein (Figure 4).

3.10. Analysis of Physical and Chemical Properties of the CE-CA Protein

The primary structure analysis of a chimeric protein was performed using ProtParam software. The number of amino acids was 598. The molecular weight of chimeric protein was about 66.558 KDa. Isoelectric point (pI) was

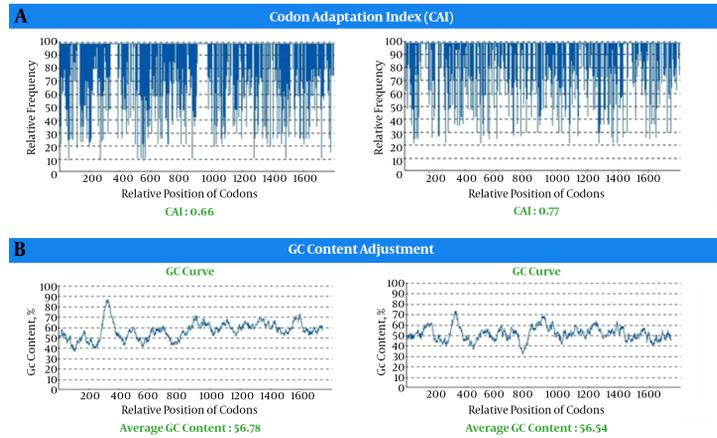


Figure 3. Analysis of the Sequence Encoding the Optimized Chimeric Construct and Wild Type Gene; **A**, codon adaptation index (CAI); **B**, percentage distribution of codons.

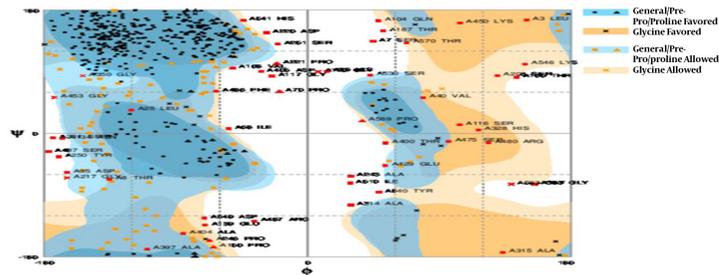


Figure 4. Evaluation of Model Stability Based on Ramachandran Plot Plot

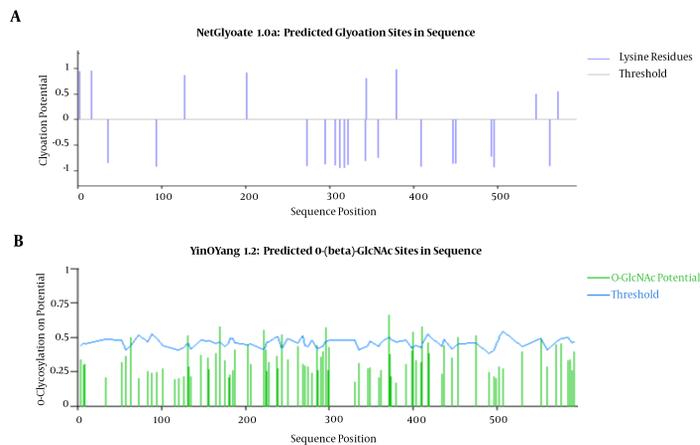


Figure 5. The Results of Glycation Sites on the Chimeric Protein; NetGlycate (A) and YinOYang 1.2 (B).

8.38. The total numbers of negatively (Asp + Glu) and positively (Arg + Lys) charged residues were 53 and 57, respectively. The half-life of this chimeric protein was 30 hours

(mammalian reticulocytes, in vitro), > 20 hours (yeast, in vivo), and > 10 hours (Escherichia coli, in vivo). Instability index was computed to be 43.97, thus the chimeric pro-

tein as unstable. Aliphatic index of chimeric protein was 74.48. Extinction coefficient of chimeric protein at 280 nm was $95855 \text{ M}^{-1}\text{cm}^{-1}$. The grand average of hydrophobicity (GRAVY) was -0.466.

3.11. Solvent Accessibility Prediction

Solvent accessibility prediction was estimated using Proso server. The solvent accessibility distribution was characterized, using the major hydrophobic and polarity properties of residual patterns. These patterns identified that the mean residue accessible surface area (ASA) has given a high solvent accessibility value, approximately 50% (Table 1).

In order to be confident about the lack of protein precipitant in cell during expression and solubility of protein, Proso server was used. According to the algorithm of the server, scores above 0.5 are soluble form. The solubility score of chimeric protein was 0.842. Thus, the result showed that recombinant protein has a high solubility. The study of protein charge at pH (4 - 10) indicated that protein in physiologic pH is stable and has a high solubility. The amount of protein charge was obtained by protein calculator server (Table 1).

3.12. Analysis of Conserved Domains and Protein Localization

Prediction of subcellular localization: subcellular localization of CE-CA in prokaryotic and eukaryotic systems was predicted by CELLO. The result of localization prediction showed that chimeric protein as an extracellular protein in both prokaryotic and eukaryotic systems.

3.13. Prediction of B-Cell Epitopes

Different parameters such as hydrophobicity, flexibility, exterior accessibility, exposed surface, and antigenicity were used to predict the chimeric protein epitopes. The epitopes located on the surface of the protein could interact easily with antibodies. Bcepred software was used in different parameters including hydrophobicity, Antigenicity, flexibility, accessibility, polarity, and exposed surface to determine the continuous B-cell epitope (Table 2). The results of this analysis included peptides and their corresponding threshold scores. The higher the threshold score, the higher the specificity and binding affinity. Discontinuous B-cell epitopes were predicted by Ellipro software (Table 3). The results of Ellipro software showed 6 set of discontinuous B-Cell epitopes. Discotop server was used for the prediction of conformational B-cell Epitopes (Table 4). Also, SEPPA server was used for conformation B-cell epitope Prediction.

3.14. Prediction of Cleavage Sites

The cleavage site on the construct protein was analyzed by Net Chop server. The Net Chop server produced neural network predictions for cleavage sites of the human proteasome. Number of cleavage site was 64 (data not shown). The prediction of binding affinity of TAP binder in chimeric protein was performed using TAPPred server. The result of TAPPred showed 41 peptides have high binding affinity and 171 peptides have intermediate binding affinity to TAP protein.

3.15. Prediction of T-Cell Epitopes

CTLpred is a direct method for prediction of CTL epitopes. The score of CTLpred- Predicted epitopes are shown in (Table 5).

NetCTL 1.2 is a server for prediction of CTL epitopes in the chimeric protein sequence. Based on the prediction methods, the scores were defined and thresholds were explained by using sensitivity and specificity of integrated peptides value (Table 6).

3.16. Prediction of MHC Binding Peptide

The conserved peptide sequence with the highest binding score to MHC class I and II was predicted, using propepI and propepII servers, respectively (Tables 7 and 8). The result of this servers showed that 14 MHC class I alleles and 19 MHC class II alleles were found to identify the common T-cell epitopes.

3.17. Prediction of Post-Translational Modification

The 8 glycation sites on the chimeric protein have been found (Figure 5). NetNGly predicted 10 asparagine amino acids at position (71,82,133,154,163,178,235,269,357,408) to be N-glycosylated. YinOYang 1.2 predicted 12 glycation sites on this protein. For N-terminal, myristoylation using myristoylator found no site. The result of NetPhos server showed that 34 sites have phosphorylation. The result of GPI showed no GPI lipid anchor site found in the sequence.

4. Discussion

Colorectal cancer is a leading cause of cancer-related deaths all over the world (61). The success of any cancer vaccine depends on the selection of a suitable target antigen and presentation pathway (62). No vaccine currently exists to colorectal cancer. So, it is urgently needed to search for finding an effective vaccine for colorectal cancer. Vaccination can stimulate the immune system and increase adaptive to a disease. Cellular immunity has an important role in cancer vaccines (63, 64). Vaccine efficacy can be assessed

Table 1. Accessible Surface Area (ASA) Calculation for CE-CA Protein Complex (A); the chart of Protein Charge Based on pH (B)

Probe radius	POLAR Area/Energy	APOLAR Area/Energy	Total Area/Energy	Number of Surface Atoms	Number of Buried Atoms
1.400	10330.50	23621.67	33952.17	3012	1683

Table 2. Continuous B-Cell Epitopes Predicted in Chimeric Protein by Bcepred Software

Prediction Parameters	Epitope Positions
Hydrophobicity	11-17, 36-45, 52-59, 97-104, 112-128, 131-145, 173-179, 203-209, 216-222, 279-293, 307-318, 395-411, 487-499, 543-551, 556-566.
Flexibility	33-42, 109-126, 128-135, 172-178, 268-275, 280-293, 327-339, 367-374, 429-435, 446-452, 472-478, 484-497, 526-532, 541-548, 553-563.
Accessibility	11-17, 32-45, 52-71, 81-87, 92-110, 131-143, 149-157, 170-182, 187-195, 199-205, 216-222, 224-235, 243-251, 257-265, 268-278, 287-296, 307-318, 331-349, 357-363, 383-389, 393-399, 405-415, 421-438, 443-455, 475-484, 486-500, 525-551, 553-577.
Turns	129-139, 160-166, 173-181, 203-210, 280-288, 370-376, 562-568, 590-598.
Exposed surface	333-344, 444-453, 489-499, 530-538, 540-550, 556-567.
Polarity	11-17, 33-44, 136-145, 289-295, 307-318, 322-328, 331-344, 381-393, 420-431, 478-501, 513-519, 531-538, 540-550, 555-577, 588-598.
Antigenic propensity	15-31, 44-52, 73-80, 85-97, 105-111, 164-174, 180-187, 190-197, 204-218, 232-241, 275-281, 294-300, 326-332, 349-357, 359-377, 417-426, 452-461, 512-530, 548-557.

with ability to induce CD8+ or CD4+ T cell. MHC I restriction depends on CD8+ cytotoxic T-cells (CTL) and MHC II restriction with CD4+ helper T-cells (TH). Thus, B-cell and T-cell epitopes mapping play a vital role in designing vaccines (65). For over a century, the role of immune system in controlling cancer was ambiguous. The vaccine strategies used against cancer depend on how well the target antigens are defined (66). As recent advances in colorectal cancer, tumor antigen identified specific molecular target in colorectal cancer cells (67). These finding indicated how immune responses are generated in patients with cancer. The data help the development of new vaccine strategies. Tumor-associated antigens (TAA) are proteins expressed using cancer cells that can be defined based on recognition by T cells (68). CEA and CA19-9 are 2 TAAs extensively studied in colorectal cancer. Thus, we selected CEA and CA19-9 proteins that play an important role in colorectal cancer. In colorectal cancer, several antigens have been found over-expressed, but not mutated. The most studied are CEA and CA19-9 (69). CEA is a member of the immunoglobulin superfamily and a useful target for vaccine (70). Several studies have shown that well-differentiated colorectal cancers produced more CEA per gram of total protein. Recent studies showed that the CEA overexpressed in > 90% of colorec-

tal cancers and this antigen is weakly recognized by the immune system (71). CA19-9 level was an important prognostic factor for the recurrence of colorectal cancer. The expression of CA19-9 has been described in colorectal cancer and increased in advanced stages of colorectal cancer (72). The epitope is a part of the antigen that was identified by the immune system. T-cell epitopes on the surface of an antigen present cell (APC) and bound to major histocompatibility (MHC) molecules to induce immune response. The identification of epitopes by T-cells and, then, the induction of immune response have a main role in individual's immune system. While the prediction of epitopes, investigation of the binding affinity of antigenic peptides to the MHC molecules is the main aim (73).

The chimeric construct contains CEA- CA19-9 peptide for expression in *E. coli* designed (Figure 1A). In order to design chimeric protein, we selected epitopes from residues 500 to 700 of CEA and amino acid residues 130 to 400 of CA19-9. The constructed chimeric protein requires appropriate linkers to bind protein domains. Linkers play a critical role in displaying different domains of chimeric protein; based on the linker containing EAAAk, repeats were designed (Figure 1A). To improve the transcription efficiency and transcript stability and enhance recombinant protein production, codon optimization was performed. Codon adaptation index (CAI) was the major factor used for a gene optimization, with a range of 0 to 1. An ideally biased gene would be a CAI of 1.0, although no natural bacterial gene reaches this theoretical value. CAI index increased from 0.66 in the wild type gene to 0.77 in chimeric optimized gene sequence, indicating that the optimized gene sequence could be expressed well (Figure 3). The prediction of allergenic protein is important for modification of proteins in therapeutics. This result showed that the chimeric protein was not allergen. By using VaxiJen server, immunogenicity of chimeric protein was predicted. Models derived include bacterial, viral, tumor, parasite, and fungal kingdoms. The accuracy rate of server was between 70% to 97%. The solubility score of chimeric protein was 0.842 showing this protein can be purified under normal condition when expressed in *E. coli*. Messenger RNA secondary has a major role in the protein expression. mfold is the software used for prediction of RNA secondary structure. The characterization of low ΔG and energy of the start codon could help ribosome binding and translation initiation. All 42 structural have folding of the

Table 3. Discontinuous B-Cell Epitopes Predicted in Chimeric Protein by Ellipro Software

No	Residues	Number of Residues	Score
1	A:D405, A:V406, A:G407, A:N408, A:K409, A:T410, A:T411, A:F441, A:W442, A:G443, A:P444, A:P445, A:S446, A:K447, A:M448, A:Q449, A:K450, A:P451, A:V474, A:P476, A:G477, A:R478, A:M479, A:R480, A:F482, A:D483, A:D484, A:L485, A:F486, A:R487, A:G488, A:E489, A:T490, A:G491, A:K492, A:D493, A:E495, A:K496, A:S497, A:H498, A:S499, A:W500, A:L501, A:S502, A:T503, A:G504, A:W505, A:F506, A:T507, A:M508, A:V509, A:I510, A:A511, A:V512, A:E513, A:L514, A:C515, A:D516, A:H517, A:V518, A:H519, A:M523, A:V524, A:P525, A:P526, A:N527, A:C529, A:S530, A:Q531, A:R532, A:P533, A:R534, A:L535, A:Q536, A:R537, A:M538, A:P539, A:Y540, A:H541, A:Y542, A:Y543, A:E544, A:P545, A:K546, A:G547, A:P548, A:D549, A:E550, A:D555, A:H556, A:H557, A:F558, A:I559, A:T570, A:E571, A:K572, A:R573, A:V574, A:F575, A:S576, A:S577, A:W578, A:A579, A:Q580, A:L581, A:Y582, A:G583, A:H584, A:T585, A:F586, A:S587, A:H588, A:P589, A:S590, A:W591, A:H593, A:H594, A:H596, A:H597, A:H598	121	0.724
2	A:M1, A:K2, A:L3, A:T4, A:I5, A:E6, A:S7, A:T8, A:P9, A:F10, A:N11, A:V12, A:A13, A:E14, A:G15, A:K16, A:E17, A:L21, A:V22, A:H23, A:N24, A:L25, A:P26, A:Q27, A:H28, A:L29, A:F30, A:G31, A:Y32, A:S33, A:W34, A:Y35, A:K36, A:G37, A:E38, A:R39, A:V40, A:D41, A:G42, A:N43, A:R44, A:Q45, A:I46, A:I47, A:G48, A:Y49, A:V50, A:I51, A:G52, A:T53, A:Q54, A:Q55, A:A56, A:T57, A:P58, A:G59, A:P60, A:A61, A:Y62, A:S63, A:G64, A:R65, A:E66, A:I67, A:I68, A:Y69, A:P70, A:N71, A:A72, A:S73, A:L74, A:L75, A:I76, A:Q77, A:N78, A:I79, A:I80, A:Q81, A:N82, A:D83, A:T84, A:G85, A:F86, A:Y87, A:T88, A:L89, A:H90, A:Y91, A:I92, A:K93, A:S94, A:D95, A:L96, A:V97, A:N98, A:E99, A:E100, A:A101, A:T102, A:G103, A:Q104, A:F105, A:R106, A:V107, A:Y108, A:P109, A:E110, A:L111, A:G112, A:G113, A:G114, A:G115, A:S116, A:G117, A:G118, A:G119, A:G120, A:S121, A:G122, A:G123	120	0.724
3	A:E139, A:D140, A:E141, A:A143, A:W159, A:W160, A:V161, A:N162, A:N163, A:Q164, A:S165, A:L166, A:P167, A:V168, A:S169, A:P170, A:R171, A:L172, A:Q173, A:T182, A:L183, A:L184, A:S185, A:V186, A:T187, A:R188, A:N189, A:D190, A:V191, A:G192, A:P193, A:G197, A:G217, A:P218	34	0.691
4	A:A319, A:R338, A:P339, A:V340, A:N341, A:L342, A:L355, A:G356, A:N357, A:K358, A:T359, A:L360, A:P361, A:S362, A:R363, A:E38	16	0.657
5	A:L174, A:S175, A:N176, A:D177	4	0.634
6	A:F430, A:V431, A:N432, A:R433, A:T434, A:P435, A:V438, A:F439, A:I440	9	0.599

Table 4. The Prediction of Discontinuous Epitopes of Chimeric Protein by Discotop Server

| Start and End Position |
|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
| 27-9 | 69-10 | 232-10 | 344-12 | 441-13 | 480-12 | 503-8 | 537-18 | 587-14 |
| 38-21 | 81-16 | 233-10 | 345-16 | 442-9 | 481-14 | 504-8 | 547-23 | 588-12 |
| 39-13 | 82-14 | 243-14 | 346-13 | 443-8 | 482-18 | 516-10 | 548-21 | 589-10 |
| 40-11 | 94-10 | 244-16 | 347-11 | 444-10 | 483-14 | 517-10 | 549-20 | 590-15 |
| 41-13 | 95-9 | 255-12 | 360-10 | 445-12 | 484-13 | 520-17 | 550-17 | 591-14 |
| 42-15 | 133-14 | 256-12 | 399-11 | 446-20 | 485-15 | 521-23 | 551-22 | 592-12 |
| 43-20 | 134-14 | 257-11 | 400-12 | 447-16 | 486-15 | 522-15 | 552-17 | 593-9 |
| 44-15 | 135-16 | 258-12 | 405-19 | 448-17 | 487-15 | 523-20 | 553-23 | 594-8 |
| 50-12 | 151-15 | 259-13 | 406-13 | 449-18 | 488-14 | 524-16 | 555-19 | 595-8 |
| 51-8 | 152-15 | 260-13 | 407-10 | 450-19 | 489-10 | 525-16 | 556-14 | 596-8 |
| 52-6 | 154-16 | 262-14 | 408-16 | 451-20 | 490-19 | 526-9 | 557-13 | 597-11 |
| 53-8 | 163-12 | 285-15 | 409-18 | 452-16 | 491-22 | 527-8 | 558-20 | 598-12 |
| 54-13 | 164-10 | 286-16 | 410-20 | 453-12 | 492-17 | 528-8 | 559-16 | |
| 55-14 | 165-10 | 287-19 | 411-18 | 454-12 | 493-18 | 529-9 | 560-21 | |
| 56-16 | 166-11 | 288-11 | 427-17 | 470-11 | 494-15 | 530-9 | 561-21 | |
| 57-14 | 167-11 | 289-18 | 428-15 | 473-20 | 495-14 | 531-15 | 562-20 | |
| 58-10 | 175-12 | 338-13 | 429-11 | 474-21 | 496-13 | 532-10 | 563-18 | |
| 59-15 | 176-12 | 339-16 | 430-7 | 476-26 | 497-13 | 533-13 | 564-18 | |
| 60-12 | 189-13 | 340-18 | 431-10 | 477-25 | 498-15 | 534-14 | 565-18 | |
| 67-10 | 205-14 | 341-13 | 432-14 | 478-24 | 499-17 | 535-20 | 566-21 | |
| 68-7 | 218-15 | 342-13 | 433-16 | 479-18 | 500-13 | 536-16 | 567-27 | |

Table 5. MHC Restriction of CTL Epitope Prediction by CTLpred Based on Artificial Neural Network in CE-CA

Peptide Rank	Start Position	Sequence	Score	MHC Restriction
1	30	FGYSWYKGE	1.000	HLA-B*2705, HLA-B*5301, HLA-Cw*0401, HLA-B*2703
2	164	QSLPVSPRL	1.000	HLA-Cw*0401, HLA-G
3	180	TLTLLSVTR	1.000	HLA-A24, HLA-Cw*0401, HLA-G

RNA construct at 37°C and the best structure had ΔG = -527.74 kcal/mol. The data from mRNA structure prediction showed that the mRNA was stable enough for efficient translation in *E. coli* (Figure 2D).

The physicochemical parameters of chimeric protein were analyzed by ProtParam software. The protein pI value (8.38) showed that the protein has an acidic nature. Extinction coefficient of CE-CA at 280nm was high (95855

M¹cm⁻¹). On the basis of instability index, expasy ProtParam classifies the chimeric protein as unstable (instability index, 43.97). For chimeric protein, the grand average of hydropathicity (GRAVY) was -0.466. The low GRAVY index of this chimeric protein infers that CE-CA could result in a better interaction with water. GORIV program was used for secondary structural analysis. The very high coil structural content of CE-CA (60.37%) was due to the rich content of

Table 6. NetCTL-1.2 Predictions Using MHC Super Type A1.Threshold 0.750000; CE-CA Chimeric Protein, Number of MHC Ligands 16 Identified; Number of Peptides 590

Position	Sequence	aff	aff-Rescale	Cle	Tap	COMB
270	ITEKNSGLY	0.7720	3.2778	0.9000	2.9230	3.5590
206	HSDPVILNV	0.4673	1.9841	0.9763	-0.0510	2.1280
222	TISPSYTY	0.4306	1.8283	0.9683	2.8830	2.1177
186	VTRNDVGPY	0.2312	0.9815	0.9291	3.0680	1.2743
394	MNDAPTTGY	0.2323	0.9863	0.9560	2.7750	1.2684
464	LVFPNMEAY	0.2112	0.8967	0.9660	3.1490	1.1991
513	ELCDHVHVY	0.2114	0.8975	0.9478	2.7960	1.1795
221	PTISPSYTY	0.1985	0.8428	0.9769	2.3530	1.1070
574	VFSSWAQLY	0.1730	0.7344	0.9341	3.2820	1.0386
242	AASNPPAQY	0.1604	0.6809	0.9684	3.0940	0.9809
347	ITDGVVPL	0.1699	0.7215	0.9633	0.8340	0.9077
534	RLQRMPYHY	0.1365	0.5796	0.9757	3.0110	0.8765
321	AKANVFHY	0.1168	0.4960	0.9382	3.2890	0.8011
83	DTGFYTLHV	0.1573	0.6680	0.8121	-0.0450	0.7876
54	QQATPGPAY	0.1118	0.4747	0.9468	3.0400	0.7687

Table 7. The Result of Prediction for MHC I Epitopes by Propred Server

No	Epitope Sequence	Position
1	LVHNLPOHL	21-29
2	IYPNASLL	67-75
3	IQNDTGFYT	80-88
4	GQFRVYPEL	103-111
5	TCEPEIQNT	147-155
6	EPEIQNTTY	149-157
7	QNTTYLWVV	153-161
8	LLSVTRNDV	183-191
9	HSDPVILNV	206-214
10	ITEKNSGLY	270-278
11	PTTGYADV	398-406
12	SLVRVIQRA	454-462
13	WLSTGWFTM	500-508
14	RVFSSWAQL	573-581

Table 8. The Result of Prediction for MHC II Epitopes by Propred Server

No	Epitope Sequence	Position
1	FNVAEGKEV	10-18
2	LVHNLPOHL	21-29
3	VVIGTQAT	49-57
4	IYPNASLL	67-75
5	FYTLHVIKS	86-93
6	WVNNQSLPV	160-168
7	YRPGVNLSL	230-238
8	YGLRGRSR	329-337
9	YVILGNKT	351-359
10	IVSSSHLL	369-377
11	IRMNDAPTT	392-400
12	YRVVAHSSV	412-420
13	VVIFWGPP	438-446
14	VRVQRAGL	456-464
15	FPNMEAYAV	466-474
16	FRGETGKDR	486-494
17	FTMVIAVEL	506-514
18	YGMVPPNYC	521-529
19	WAQLYGITF	578-586

more flexible glycine and hydrophobic proline (Figure 1B). The three-dimensional (3D) structure of proteins was of major importance in functional properties of the protein sequence. The three-dimensional model of the chimeric protein was generated, using I-TASSER online software. Our results showed that I-TASSER software can predict the folds as well as good resolution model for our chimeric protein (Figure 1C). RMSD and TM-score were used to evaluate the predicted models. The best RMSD value was the result of our model on template, which consisted of 598 amino acids. Expected TM-score of 0.53 ± 0.15 confirms the correctness of the model. TM-score more than 0.5 shows an accurate topological model. Its confidence was achieved by Z-score and C-score. The Z-score indicates measures the deviation of the total energy of the structure with respect to an energy distribution derived from random conformations and overall model quality. For native protein, Z-score outside a range characteristic indicate erroneous

structure. The results of ProSA-web showed that synthetic chimeric protein has features, which are the characteristic of native structures. In Ramachandran plot analysis, the residues rate was 72.8% favored region, 18.1% allowed, and 9.1% in outlier region. Thus, based on Ramachandran plot prediction, our chimeric structure showed desirable protein stability. A negligible 7.4% of the residues were in Ramachandran plot analysis to be in outlier region that could probably be due to the presence of chimeric junctions (Figure 4). Identification of B-cell epitopes is a crucial step for satisfactory design of vaccines. B-cell epitopes are the specific region of an antigenic surface protein. On the basis of the structural prediction and solvent

accessibility, B-cell epitopes for the chimeric protein could be predicted. In order to predict B-cell epitopes, several different method such as hydrophobicity method, accessibility method, antigenicity method, flexibility method, and secondary structure analysis have been developed. All methods together were performed to obtain results good enough to predict the B-cell epitopes. The results of the most similar B-cell epitopes of this chimeric protein were indicated in above table.

Glycosylation analysis showed that constructing has high glycosylation sites. Glycosylation may decrease antigenicity and immunogenicity of the vaccine product. The existence of myristoylation signal in N-terminal raises vaccine efficiency (Figure 3). Various methods were used for B-cell epitope prediction. The results of prediction showed that there are 14 consensus MHC class I binding regions and 19 consensus MHCII class binding regions in the chimeric protein sequence (Tables 7 and 8). The prediction of CTL epitopes in chimeric protein structure was done by NETCTL server. This server showed that 16 MHC ligands were identified in CE-CA protein (Table 6). The CTLpred server showed the score of epitopes in chimeric protein. The cutoff score was 0.51 (Table 5). NetMHC 3.4 server predicted peptide binding to different HLA alleles by artificial neural networks (ANNs). Three same peptide sequences with high log score were recognized as strong MHB binder in CE-CA chimeric protein. All MHC binding peptides were used for suitable immune response. Propred server predicted MHCII binding regions in antigenic protein sequences. MHC class-II binding peptide prediction in chimeric protein was performed by Propred server and this server showed 57 alleles query in this protein (Table 8).

4.1. Conclusions

In this study, we designed a novel chimeric vaccine for cancer immunotherapy. Our results showed that epitopes of the chimeric protein could induce B-cell and T-cell mediated immune responses, which are important for a protective vaccine against colorectal cancer.

Acknowledgments

This study was supported by a grant from Tehran University of Medical Sciences. The authors declare no conflict of interests.

Footnotes

Authors' Contribution: None.

Conflict of Interests: None.

Financial Disclosure: None.

References

1. Safaee A, Fatemi SR, Ashtari S, Vahedi M, Moghimi-Dehkordi B, Zali MR. Four years incidence rate of colorectal cancer in Iran: a survey of national cancer registry data - implications for screening. *Asian Pac J Cancer Prev*. 2012;**13**(6):2695–8. [PubMed: 22938443].
2. Ohkado A, Yanabu N, Kimura F, Kamimura K, Sumiyoshi T, Wakuya J, et al. [A case of mesothelioma of the pleura with pneumothorax]. *Kyobu Geka*. 1989;**42**(7):565–8. [PubMed: 2796096].
3. Bretthauer M. Evidence for colorectal cancer screening. *Best Pract Res Clin Gastroenterol*. 2010;**24**(4):417–25. doi: 10.1016/j.bpg.2010.06.005. [PubMed: 20833346].
4. Irvin G3, Horsley J3, Caruana JJ. The morbidity and mortality of emergent operations for colorectal disease. *Ann Surg*. 1984;**199**(5):598–603. [PubMed: 6609686].
5. Moghimi-Dehkordi B, Safaee A, Zali MR. Prognostic factors in 1,138 Iranian colorectal cancer patients. *Int J Colorectal Dis*. 2008;**23**(7):683–8. doi: 10.1007/s00384-008-0463-7. [PubMed: 18330578].
6. Fioretti D, Iurescia S, Fazio VM, Rinaldi M. DNA vaccines: developing new strategies against cancer. *J Biomed Biotechnol*. 2010;**2010**:174378. doi: 10.1155/2010/174378. [PubMed: 20368780].
7. Finn OJ. Cancer vaccines: between the idea and the reality. *Nat Rev Immunol*. 2003;**3**(8):630–41. doi: 10.1038/nri1150. [PubMed: 12974478].
8. Mellgren SI, Johnsen HJ. [Steroid-induced myopathy. Clinical and muscle pathologic aspects with a case report]. *Tidsskr Nor Laegeforen*. 1979;**99**(3):163–4. [PubMed: 419494].
9. Ullenhag E, van Dalen A, Haglund C, Hansson L, Holinski-Feder E, Klapdor R, et al. Tumour markers in colorectal cancer: European Group on Tumour Markers (EGTM) guidelines for clinical use. *Eur J Cancer*. 2007;**43**(9):348–60. doi: 10.1016/j.ejca.2007.03.021. [PubMed: 17512720].
10. Ullenhag GJ, Frodin JE, Jeddi-Tehrani M, Strigard K, Eriksson E, Samanci A, et al. Durable carcinoembryonic antigen (CEA)-specific humoral and cellular immune responses in colorectal carcinoma patients vaccinated with recombinant CEA and granulocyte/macrophage colony-stimulating factor. *Clin Cancer Res*. 2004;**10**(10):3273–81. doi: 10.1158/1078-0432.CCR-03-0706. [PubMed: 15161680].
11. Dahlquist G, Sterky G, Ivarsson JI, Tengvald K, Wall S. Health problems and care in young families-load of illness and patterns of illness behaviour. *Scand J Prim Health Care*. 1987;**5**(2):79–86. [PubMed: 3616277].
12. Paterson Y, Singh R. *Administering a Listeria vaccine with a strain expressing an antigen fused to a truncated LLO protein; cancer vaccines; anticarcinogenic agents*. Google Patents; 2010.
13. Bos R, van Duikeren S, van Hall T, Lauwen MM, Parrington M, Berinstein NL, et al. Characterization of antigen-specific immune responses induced by canarypox virus vaccines. *J Immunol*. 2007;**179**(9):6115–22. [PubMed: 17947686].
14. Hammarstrom S. The carcinoembryonic antigen (CEA) family: structures, suggested functions and expression in normal and malignant tissues. *Semin Cancer Biol*. 1999;**9**(2):67–81. doi: 10.1006/scbi.1998.0119. [PubMed: 10202129].
15. Misener S, Krawetz SA. *Bioinformatics methods and protocols*. **132**. Springer Science & Business Media; 1999.
16. Gish W, States DJ. Identification of protein coding regions by database similarity search. *Nat Genet*. 1993;**3**(3):266–72. doi: 10.1038/ng0393-266. [PubMed: 8485583].
17. Casado V, Lluís C, Canela E, Franco R, Mallol J. The distribution of A1 adenosine receptor and 5'-nucleotidase in pig brain cortex subcellular fractions. *Neurochem Res*. 1992;**17**(2):129–39. [PubMed: 1538830].

18. Doytchinova IA, Flower DR. Vaxijen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics*. 2007;**8**:4. doi: [10.1186/1471-2105-8-4](https://doi.org/10.1186/1471-2105-8-4). [PubMed: [17207271](https://pubmed.ncbi.nlm.nih.gov/17207271/)].
19. Sen TZ, Jernigan RL, Garnier J, Kloczkowski A. GOR V server for protein secondary structure prediction. *Bioinformatics*. 2005;**21**(11):2787-8. doi: [10.1093/bioinformatics/bti408](https://doi.org/10.1093/bioinformatics/bti408). [PubMed: [15797907](https://pubmed.ncbi.nlm.nih.gov/15797907/)].
20. Lissoni P, Galli MA, Barni S, Colombo V, Pelizzoni F, Fumagalli R, et al. [The cardiovascular toxicity of interleukin-2: the pathogenic mechanisms and treatment]. *G Ital Cardiol*. 1990;**20**(7):631-5. [PubMed: [2245901](https://pubmed.ncbi.nlm.nih.gov/2245901/)].
21. Kelley LA, Sternberg MJ. Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc*. 2009;**4**(3):363-71. doi: [10.1038/nprot.2009.2](https://doi.org/10.1038/nprot.2009.2). [PubMed: [19247286](https://pubmed.ncbi.nlm.nih.gov/19247286/)].
22. Lovell SC, Davis IW, Arendall W3, de Bakker PI, Word JM, Prisant MG, et al. Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins*. 2003;**50**(3):437-50. doi: [10.1002/prot.10286](https://doi.org/10.1002/prot.10286). [PubMed: [12557186](https://pubmed.ncbi.nlm.nih.gov/12557186/)].
23. Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res*. 2007;**35**(Web Server issue):W407-10. doi: [10.1093/nar/gkm290](https://doi.org/10.1093/nar/gkm290). [PubMed: [17517781](https://pubmed.ncbi.nlm.nih.gov/17517781/)].
24. Chambers JE, Yarbrough JD. Parathion and methyl parathion toxicity to insecticide-resistant and susceptible mosquitofish (*Gambusia affinis*). *Bull Environ Contam Toxicol*. 1974;**11**(4):315-20. [PubMed: [4433817](https://pubmed.ncbi.nlm.nih.gov/4433817/)].
25. Flickinger KS, Culp LA. Dermal fibroblasts from Down's syndrome patients share a cycloheximide-induced deficiency in collagen adhesion responses with normal aging cells. *Exp Cell Res*. 1990;**189**(2):189-201. [PubMed: [2142462](https://pubmed.ncbi.nlm.nih.gov/2142462/)].
26. Tochilin VI. [Nodular goiter and cancer of the thyroid]. *Vestn Khir Im I I Grek*. 1989;**143**(11):18-20. [PubMed: [2633425](https://pubmed.ncbi.nlm.nih.gov/2633425/)].
27. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res*. 2003;**31**(13):3406-15. [PubMed: [12824337](https://pubmed.ncbi.nlm.nih.gov/12824337/)].
28. Graf M, Deml L, Wagner R. Codon-optimized genes that enable increased heterologous expression in mammalian cells and elicit efficient immune responses in mice after vaccination of naked DNA. *Mol Diagn Infect Dis*. 2004;197-210.
29. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, et al. *Protein identification and analysis tools on the EXPASY server*. Springer; 2005.
30. Smialowski P, Martin-Galiano AJ, Mikolajka A, Girschick T, Holak TA, Frishman D. Protein solubility: sequence based prediction and experimental verification. *Bioinformatics*. 2007;**23**(19):2536-42. doi: [10.1093/bioinformatics/btl623](https://doi.org/10.1093/bioinformatics/btl623). [PubMed: [17150993](https://pubmed.ncbi.nlm.nih.gov/17150993/)].
31. Magnan CN, Randall A, Baldi P. SOLpro: accurate sequence-based prediction of protein solubility. *Bioinformatics*. 2009;**25**(17):2200-7. doi: [10.1093/bioinformatics/btp386](https://doi.org/10.1093/bioinformatics/btp386). [PubMed: [19549632](https://pubmed.ncbi.nlm.nih.gov/19549632/)].
32. Cserzo M, Wallin E, Simon I, von Heijne G, Elofsson A. Prediction of transmembrane alpha-helices in prokaryotic membrane proteins: the dense alignment surface method. *Protein Eng*. 1997;**10**(6):673-6. [PubMed: [9278280](https://pubmed.ncbi.nlm.nih.gov/9278280/)].
33. Heringa J, Frishman D, Argos P. *Chapter 4 Computational methods relating protein sequence and structure*. 1997.
34. Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res*. 2010;**38**(Web Server issue):W529-33. doi: [10.1093/nar/gkq399](https://doi.org/10.1093/nar/gkq399). [PubMed: [20478830](https://pubmed.ncbi.nlm.nih.gov/20478830/)].
35. Yu CS, Chen YC, Lu CH, Hwang JK. Prediction of protein subcellular localization. *Proteins*. 2006;**64**(3):643-51. doi: [10.1002/prot.21018](https://doi.org/10.1002/prot.21018). [PubMed: [16752418](https://pubmed.ncbi.nlm.nih.gov/16752418/)].
36. Viklund H, Elofsson A. OCTOPUS: improving topology prediction by two-track ANN-based preference scores and an extended topological grammar. *Bioinformatics*. 2008;**24**(15):1662-8. doi: [10.1093/bioinformatics/btn221](https://doi.org/10.1093/bioinformatics/btn221). [PubMed: [18474507](https://pubmed.ncbi.nlm.nih.gov/18474507/)].
37. Nielsen H, Engelbrecht J, Brunak S, von Heijne G. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng*. 1997;**10**(1):1-6. [PubMed: [9051728](https://pubmed.ncbi.nlm.nih.gov/9051728/)].
38. El-Manzalawy Y, Dobbs D, Honavar V. Predicting linear B-cell epitopes using string kernels. *J Mol Recognit*. 2008;**21**(4):243-55. doi: [10.1002/jmr.893](https://doi.org/10.1002/jmr.893). [PubMed: [18496882](https://pubmed.ncbi.nlm.nih.gov/18496882/)].
39. Saha S, Raghava GP. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins*. 2006;**65**(1):40-8. doi: [10.1002/prot.21078](https://doi.org/10.1002/prot.21078). [PubMed: [16894596](https://pubmed.ncbi.nlm.nih.gov/16894596/)].
40. Larsen JE, Lund O, Nielsen M. Improved method for predicting linear B-cell epitopes. *Immune Res*. 2006;**2**:2. doi: [10.1186/1745-7580-2-2](https://doi.org/10.1186/1745-7580-2-2). [PubMed: [16635264](https://pubmed.ncbi.nlm.nih.gov/16635264/)].
41. Sun J, Wu D, Xu T, Wang X, Xu X, Tao L, et al. SEPPA: a computational server for spatial epitope prediction of protein antigens. *Nucleic Acids Res*. 2009;**37**(Web Server issue):W612-6. doi: [10.1093/nar/gkp417](https://doi.org/10.1093/nar/gkp417). [PubMed: [19465377](https://pubmed.ncbi.nlm.nih.gov/19465377/)].
42. Ponomarenko J, Bui HH, Li W, Füsseder N, Bourne PE, Sette A, et al. ElliPro: a new structure-based tool for the prediction of antibody epitopes. *BMC Bioinformatics*. 2008;**9**:514. doi: [10.1186/1471-2105-9-514](https://doi.org/10.1186/1471-2105-9-514). [PubMed: [19055730](https://pubmed.ncbi.nlm.nih.gov/19055730/)].
43. Kesmir C, Nussbaum AK, Schild H, Detours V, Brunak S. Prediction of proteasome cleavage motifs by neural networks. *Protein Eng*. 2002;**15**(4):287-96. [PubMed: [11983929](https://pubmed.ncbi.nlm.nih.gov/11983929/)].
44. Holzhtutter HG, Kloetzel PM. A kinetic model of vertebrate 20S proteasome accounting for the generation of major proteolytic fragments from oligomeric peptide substrates. *Biophys J*. 2000;**79**(3):1196-205. doi: [10.1016/S0006-3495\(00\)76374-0](https://doi.org/10.1016/S0006-3495(00)76374-0). [PubMed: [10968984](https://pubmed.ncbi.nlm.nih.gov/10968984/)].
45. Diez-Rivero CM, Lafuente EM, Reche PA. Computational analysis and modeling of cleavage by the immunoproteasome and the constitutive proteasome. *BMC Bioinformatics*. 2010;**11**:479. doi: [10.1186/1471-2105-11-479](https://doi.org/10.1186/1471-2105-11-479). [PubMed: [20863374](https://pubmed.ncbi.nlm.nih.gov/20863374/)].
46. Bhasin M, Raghava GP. Analysis and prediction of affinity of TAP binding peptides using cascade SVM. *Protein Sci*. 2004;**13**(3):596-607. doi: [10.1110/ps.03373104](https://doi.org/10.1110/ps.03373104). [PubMed: [14978300](https://pubmed.ncbi.nlm.nih.gov/14978300/)].
47. Larsen MV, Lundegaard C, Lambert K, Buus S, Lund O, Nielsen M. Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC Bioinformatics*. 2007;**8**:424. doi: [10.1186/1471-2105-8-424](https://doi.org/10.1186/1471-2105-8-424). [PubMed: [17973982](https://pubmed.ncbi.nlm.nih.gov/17973982/)].
48. Rammensee H, Bachmann J, Emmerich NP, Bachor OA, Stevanovic S. SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics*. 1999;**50**(3-4):213-9. [PubMed: [10602881](https://pubmed.ncbi.nlm.nih.gov/10602881/)].
49. Bhasin M, Raghava GP. Prediction of CTL epitopes using QM, SVM and ANN techniques. *Vaccine*. 2004;**22**(23-24):3195-204. doi: [10.1016/j.vaccine.2004.02.005](https://doi.org/10.1016/j.vaccine.2004.02.005). [PubMed: [15297074](https://pubmed.ncbi.nlm.nih.gov/15297074/)].
50. Lundegaard C, Lund O, Nielsen M. Accurate approximation method for prediction of class I MHC affinities for peptides of length 8, 10 and 11 using prediction tools trained on 9mers. *Bioinformatics*. 2008;**24**(11):1397-8. doi: [10.1093/bioinformatics/btn128](https://doi.org/10.1093/bioinformatics/btn128). [PubMed: [18413329](https://pubmed.ncbi.nlm.nih.gov/18413329/)].
51. Singh H, Raghava GP. ProPred1: prediction of promiscuous MHC Class-I binding sites. *Bioinformatics*. 2003;**19**(8):1009-14. [PubMed: [12761064](https://pubmed.ncbi.nlm.nih.gov/12761064/)].
52. Adams HP, Koziol JA. Prediction of binding to MHC class I molecules. *J Immunol Methods*. 1995;**185**(2):181-90. [PubMed: [7561128](https://pubmed.ncbi.nlm.nih.gov/7561128/)].
53. Lata S, Bhasin M, Raghava GP. Application of machine learning techniques in predicting MHC binders. *Methods Mol Biol*. 2007;**409**:201-15. doi: [10.1007/978-1-60327-118-9_14](https://doi.org/10.1007/978-1-60327-118-9_14). [PubMed: [18450002](https://pubmed.ncbi.nlm.nih.gov/18450002/)].
54. Singh H, Raghava GP. ProPred: prediction of HLA-DR binding sites. *Bioinformatics*. 2001;**17**(12):1236-7. [PubMed: [11751237](https://pubmed.ncbi.nlm.nih.gov/11751237/)].
55. Johansen MB, Kierner L, Brunak S. Analysis and prediction of mammalian protein glycation. *Glycobiology*. 2006;**16**(9):844-53. doi: [10.1093/glycob/cwl009](https://doi.org/10.1093/glycob/cwl009). [PubMed: [16762979](https://pubmed.ncbi.nlm.nih.gov/16762979/)].
56. Gupta R, Jung E, Brunak S. *Prediction of N-glycosylation sites in human proteins*. 2016.
57. Gupta R, Brunak S. Prediction of glycosylation across the human proteome and the correlation to protein function. *Pac Symp Biocomput*. 2002;310-22. [PubMed: [11928486](https://pubmed.ncbi.nlm.nih.gov/11928486/)].
58. Steentoft C, Vakhrushev SY, Joshi HJ, Kong Y, Vester-Christensen MB,

- Schjoldager KT, et al. Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *EMBOJ*. 2013;**32**(10):1478–88. doi: [10.1038/emboj.2013.79](https://doi.org/10.1038/emboj.2013.79). [PubMed: [23584533](https://pubmed.ncbi.nlm.nih.gov/23584533/)].
59. Bologna G, Yvon C, Duvaud S, Veuthey AL. N-Terminal myristoylation predictions by ensembles of neural networks. *Proteomics*. 2004;**4**(6):1626–32. doi: [10.1002/pmic.200300783](https://doi.org/10.1002/pmic.200300783). [PubMed: [15174132](https://pubmed.ncbi.nlm.nih.gov/15174132/)].
 60. Eisenhaber B, Bork P, Eisenhaber F. Sequence properties of GPI-anchored proteins near the omega-site: constraints for the polypeptide binding site of the putative transamidase. *Protein Eng*. 1998;**11**(12):1155–61. [PubMed: [9930665](https://pubmed.ncbi.nlm.nih.gov/9930665/)].
 61. Hagggar FA, Boushey RP. Colorectal cancer epidemiology: incidence, mortality, survival, and risk factors. *Clin Colon Rectal Surg*. 2009;**22**(4):191–7. doi: [10.1055/s-0029-1242458](https://doi.org/10.1055/s-0029-1242458). [PubMed: [21037809](https://pubmed.ncbi.nlm.nih.gov/21037809/)].
 62. Vigneron N. Human Tumor Antigens and Cancer Immunotherapy. *Biomed Res Int*. 2015;**2015**:948501. doi: [10.1155/2015/948501](https://doi.org/10.1155/2015/948501). [PubMed: [26161423](https://pubmed.ncbi.nlm.nih.gov/26161423/)].
 63. Dzivenu OK, O'Donnell-Tormey J, O'Donnell-Tormey J. *Cancer and the immune system: the vital connection*. Cancer Research Institute; 2003.
 64. Palucka K, Banchereau J. Cancer immunotherapy via dendritic cells. *Nat Rev Cancer*. 2012;**12**(4):265–77. doi: [10.1038/nrc3258](https://doi.org/10.1038/nrc3258). [PubMed: [22437871](https://pubmed.ncbi.nlm.nih.gov/22437871/)].
 65. Koup RA, Douek DC. Vaccine design for CD8 T lymphocyte responses. *Cold Spring Harb Perspect Med*. 2011;**1**(1). a007252. doi: [10.1101/cshperspect.a007252](https://doi.org/10.1101/cshperspect.a007252). [PubMed: [22229122](https://pubmed.ncbi.nlm.nih.gov/22229122/)].
 66. Dermime S, Armstrong A, Hawkins RE, Stern PL. Cancer vaccines and immunotherapy. *Br Med Bull*. 2002;**62**:149–62. [PubMed: [12176857](https://pubmed.ncbi.nlm.nih.gov/12176857/)].
 67. Somers VA, Brandwijk RJ, Joosten B, Moerkerk PT, Arends JW, Menheere P, et al. A panel of candidate tumor antigens in colorectal cancer revealed by the serological selection of a phage displayed cDNA expression library. *J Immunol*. 2002;**169**(5):2772–80. [PubMed: [12193752](https://pubmed.ncbi.nlm.nih.gov/12193752/)].
 68. Sarcevic B, Spagnoli GC, Terracciano L, Schultz-Thater E, Heberer M, Gamulin M, et al. Expression of cancer/testis tumor associated antigens in cervical squamous cell carcinoma. *Oncology*. 2003;**64**(4):443–9. doi: [10.1159/000070305](https://doi.org/10.1159/000070305). [PubMed: [12759544](https://pubmed.ncbi.nlm.nih.gov/12759544/)].
 69. Perkins GL, Slater ED, Sanders GK, Prichard JG. Serum tumor markers. *Am Fam Physician*. 2003;**68**(6):1075–82. [PubMed: [14524394](https://pubmed.ncbi.nlm.nih.gov/14524394/)].
 70. Turriziani M, Fantini M, Benvenuto M, Izzi V, Masuelli L, Sacchetti P, et al. Carcinoembryonic antigen (CEA)-based cancer vaccines: recent patents and antitumor effects from experimental models to clinical trials. *Recent Pat Anticancer Drug Discov*. 2012;**7**(3):265–96. [PubMed: [22630596](https://pubmed.ncbi.nlm.nih.gov/22630596/)].
 71. Ananenko AA, Malinovskaia VV, Pertseva NG, Bezrukov K, Sultanov AT. [State of lipid peroxidation system and antioxidant defense in acute respiratory viral infection in children and the principles of pathogenetic therapy]. *Pediatriia*. 1989;**1**:27–30. [PubMed: [2710600](https://pubmed.ncbi.nlm.nih.gov/2710600/)].
 72. Park IJ, Choi GS, Jun SH. Prognostic value of serum tumor antigen CA19-9 after curative resection of colorectal cancer. *Anticancer Res*. 2009;**29**(10):4303–8. [PubMed: [19846991](https://pubmed.ncbi.nlm.nih.gov/19846991/)].
 73. Ribas A, Butterfield LH, Economou JS. Genetic immunotherapy for cancer. *Oncologist*. 2000;**5**(2):87–98. [PubMed: [10794799](https://pubmed.ncbi.nlm.nih.gov/10794799/)].