

# In silico Analysis, Molecular Docking, Molecular Dynamic, Cloning, Expression and Purification of Chimeric Protein in Colorectal Cancer Treatment

This article was published in the following Dove Press journal:  
*Drug Design, Development and Therapy*

Hassan Dana<sup>1,2,\*</sup>  
Ghanbar  
Mahmoodi Chalbatani<sup>1,\*</sup>  
Elahe Gharagouzloo<sup>1</sup>  
Seyed Rouhollah Miri<sup>1</sup>  
Fereidoon Memari<sup>1</sup>  
Reza Rasoolzadeh<sup>3</sup>  
Mohammad Reza Zinatizadeh<sup>3</sup>  
Peyman Kheirandish Zarandi<sup>3</sup>  
Vahid Marmari<sup>2</sup>

<sup>1</sup>Cancer Research Center, Cancer Institute of Iran, Tehran University of Medical Science, Tehran, Iran;

<sup>2</sup>Department of Biology, Damghan Branch, Islamic Azad University, Damghan, Iran; <sup>3</sup>Department of Biology, Science and Research Branch, Islamic Azad University, Tehran, Iran

\*These authors contributed equally to this work

**Introduction:** Colorectal cancer (CRC) is a type of cancer in humans that leads to high mortality and morbidity. CD166 and CD326 are immunoglobulins that are associated with cell migration. These molecules are included in tumorigenesis of CRC and serve a great marker of CRC stem cells. In the present study, we devised a novel chimeric protein including the V<sub>1</sub>-domain of the CD166 and two epitopes of CD326 to use in diagnostic or therapeutic applications.

**Methods:** In silico techniques were launched to characterize the properties and structure of the protein. We have predicted physicochemical properties, structures, stability, MHC class I binding properties and ligand-receptor interaction of this chimeric protein by means of computational bioinformatics tools and servers. The sequence of chimeric gene was optimized for expression in prokaryotic host using online tools and cloned into pET-28a plasmid. The recombinant pET28a was transformed into the *E. coli* BL21DE3. Expression of recombinant protein was examined by SDS-PAGE and Western blotting.

**Results:** The designed chimeric protein retained high stability and the same immunogenicity as of the original proteins. Bioinformatics data indicated that the epitopes of the synthetic chimeric protein might induce B-cell- and T-cell-mediated immune responses. Furthermore, a gene was synthesized using the codon bias of a prokaryotic expression system. This synthetic gene expressed a bacterial expression system. The recombinant protein with molecular weights of 27kDa was expressed and confirmed by anti-his Western blot analysis.

**Conclusion:** The designed recombinant protein may be useful as a CRC diagnostic tool and for developing a protective vaccine against CRC.

**Keywords:** chimeric protein, molecular docking, molecular dynamic, vaccine

## Introduction

Colorectal cancer (CRC) is a prevalent type of cancer that leads to high mortality and morbidity worldwide.<sup>1</sup> In the United States, CRC is the fourth most frequent types of cancer, leading to 8.0% of the total new diagnosed cancer cases.<sup>2</sup> Furthermore, according to the statistics from the National Cancer Institute, about 35.6% of the people suffering from CRC died between 2009–2015 (<https://seer.cancer.gov/statfacts/html/colorect.html>). CRC is treatable if detected/diagnosed early. However, only 20–25% of patients with metastatic disease are diagnosed at early stages and are treated palliatively.<sup>3</sup> Hence, a rapid diagnostic of CRC is urgently required.

CD166 and CD326 have been frequently found to be overexpressed in tumor cells. Both of these molecules have been projected as the potential targets for

Correspondence: Fereidoon Memari;  
Sayed Rohollah Miri  
Cancer Research Center, Cancer Institute of Iran, Tehran University of Medical Science, Tehran, Iran  
Tel +98-910-5040-602; +98-912-6543-321  
Email [memarife@tums.ac.ir](mailto:memarife@tums.ac.ir);  
[Srmiri@sina.tums.ac.ir](mailto:Srmiri@sina.tums.ac.ir)

diagnostics and therapy of CRC.<sup>4–6</sup> CD166, the Activator Leukocyte Cell Adhesion Molecule (ALCAM), is a member of the immunoglobulin superfamily. This molecule is a vital factor not only for cell survival, motility, and cell growth but also for invasion during tumor progression and metastases.<sup>7,8</sup> CD166 is a glycoprotein that was initially discovered as an MHC-I for the cell surface receptor of T lymphocyte (CD6).<sup>9</sup> It contributes to 1) heterotypic adhesion to the lymphocyte cell-surface receptor (CD6)<sup>10</sup> and 2) homotypic adhesion. This involves five extracellular immunoglobulin domains (VVC2C2C2). Based on the functional mapping analyses, the V domain in the extracellular region of CD166 is essential for both types of cell-cell adhesion.<sup>11</sup> The V domain is identified as the main MHC-I binding domain at the N-terminal immunoglobulin domain. It is comprised of two parts as V1 and V2 with 93 and 110-amino acids length, respectively.<sup>12,13</sup> Also, CD166 has been reported as an appropriate marker for CRC.<sup>14</sup>

CD326 is a member of a subgroup of transmembrane glycoproteins in the immunoglobulin superfamily and is also known as the epithelial cell adhesion molecule (EpCAM).<sup>15</sup> CD326 is expressed at low levels in the healthy epithelial cells but highly expressed in the cancerous epithelial cells such as CRC cells, where it performs essential functions like an epithelial-specific intracellular cell-adhesion activity.<sup>16</sup> However, recently, it has been shown that its role is not only limited to cell adhesion but is also involved in cell migration, proliferation, differentiation, and cellular signaling.<sup>17</sup> Furthermore, CD326 expression has been associated with CRC carcinogenesis, and its expression might be a beneficial biomarker for the clinical diagnosis of CRC.<sup>6</sup> This suggests that CD326 could be a potential target for the immunotherapeutic treatment of CRC.<sup>18</sup>

Furthermore, CD326 has three antibody-binding sites, two of which are linked to the second CD326 extracellular domain and the other can detect the cytoplasmic tail of the protein for binding. Monoclonal antibody MOC31 identifies extracellular motifs between amino acids 27 to 59. In comparison, the 311-1K2 monoclonal antibody detects the extracellular region between amino acids 143 to 164.<sup>16</sup> Recently, there has been a focus on CD166 and CD326 as target molecules for the treatment of CRC that incorporate vaccine candidates including monoclonal antibodies, antibody fragment-targeted tumor necrosis factor-related apoptosis-inducing MHC1 (TRAIL) fusion protein, and toxin-conjugated antibody fragments.

The advancement in immune-informatics has made possible the “in silico” design of new molecules and predicting its functionality. This strategy helps in shortlisting better molecules before testing in vitro or in vivo conditions. Various research groups have designed diverse in silico molecules using the immune-informatics approach.<sup>18–20</sup>

In the present study, we use in silico techniques for designing, optimizing the expression in a suitable host, and predicting physical, chemical, structural properties, and stability. Next, we identified MHC class I binding and T cell epitopes to allow the acceleration of the strong antigenic and immune responses. Finally, we investigated the ligand-receptor interaction.

An unguided experiment, which had searched for antigenic and immunogenic regions, was laborious and resource-intensive. The computational approaches could speed up the process and simplify the evaluation process to a great extent. In the end, we synthesized the codon-optimized gene and cloned and expressed this engineered chimeric protein. Hence, the novel chimeric protein was determined as a candidate for cancer immunotherapy. It could rapidly be identified in silico and subjected to in vitro and in vivo confirmatory reports.

## Materials and Methods

### Schematic Representation of the Workflow

We generated a systematic workflow (Figure 1) for the design of the chimeric protein of the V<sub>1</sub>-domain of the CD166 and two epitopes of CD326.

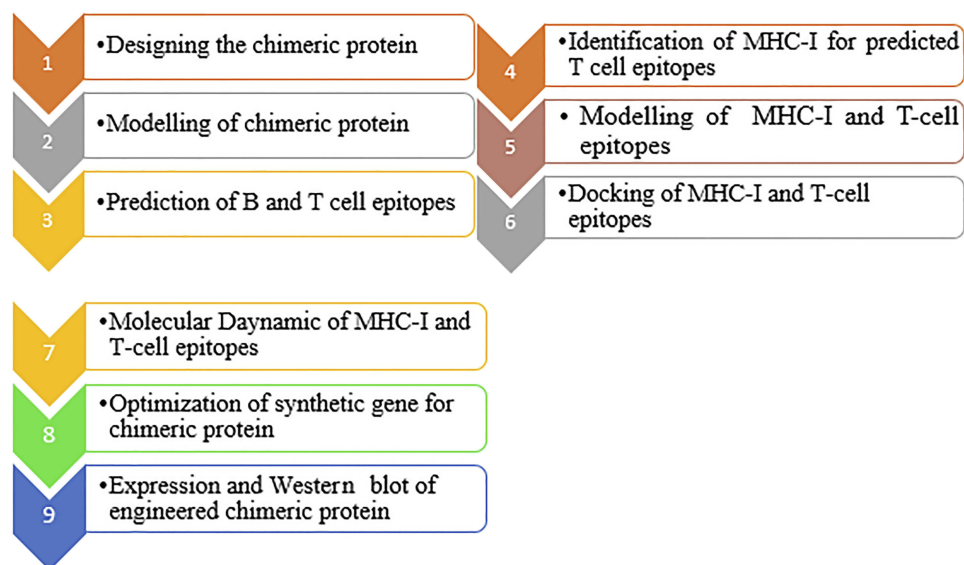
### Protein Selection and Design

We regained the sequence of CD326 (Gene ID: 4072) and the V<sub>1</sub> domain of CD166 (Gene ID: 214) from the National Center for Biotechnology Information (NCBI: [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)).

To design a chimeric protein containing the V<sub>1</sub> domain of CD166 and epitopes from CD326, we selected the amino acids from 23 to 74 (detected by antibody MOC 31) and the amino acids from 117 to 167 (identified by antibody 311-1K2) from the CD326 protein; and the amino acids 28 to 120 of the CD166 protein in the V<sub>1</sub> domain were selected. These fragments were linked together via a general rigid linker (EAAAK).

### The Secondary and Tertiary Structures of the Engineered Chimeric Protein

The tertiary structure of the chimeric protein was built using homology modeling. The modeling of the chimeric



**Figure 1** A systematic workflow of the design of a chimeric protein.

protein was done using advanced modeling based on multiple templates with Modeller9.20.<sup>21</sup> Since the chimeric protein possesses domains from two proteins (CD326 and CD166), the templates 4MZV A and 5A2F A were utilized to build the complete assembly of the chimeric protein using Modeller9.20. Finally, to evaluate the stereochemistry of the modeled protein, the PDB structure of the chimeric protein was submitted to the Procheck\_NT, followed by using the Ramachandran plot for structural stability.<sup>22</sup> The solvent accessibility, alpha-helix, random coil, and beta sheets structures were analyzed using the Predict Protein server.<sup>23</sup> We used the Protparam server for predicting the stability and other physiochemical properties of the chimeric protein.<sup>24</sup>

### Prediction of the Allergenic Protein

To predicting the allergenicity of the chimeric protein, the amino acid sequence of the chimeric protein was imported to the AlgPred software and SDAP database.<sup>25</sup>

### Prediction of Antigenicity

We predicted the antigenic nature of the chimeric protein using the VaxiJen sever, which is based on the proteins' physical and chemical properties.<sup>26</sup> To differentiate between antigenic and non-antigenic proteins, a threshold of 0.5 was used.

### Prediction of B-Cell Epitopes

The B cell epitope, both continuous and discontinuous, was predicted using the EliPro server using the modeled

chimeric protein PDB structure as an input.<sup>27</sup> The scoring of the EliPro server ranges from 0 to 1 with a cut-off of 0.5. A score of less than 0.5 is considered as a non-epitope sequence, and a score of greater than 0.5 is considered as an epitope sequence.

### Prediction of T-Cell Epitopes

CTLpred was utilized to predict cytotoxic T cell epitopes in the chimeric protein. Furthermore, the NetMHC 1 server was also used to predict the affinity for cytotoxic T cell epitopes with the predicted MHC I molecule. The CTLpred database predicts a maximum score of 1. A score of greater than 0.5 was considered as a CTL epitope.

### Molecular Docking of Selected Epitope with MHC-I

In this study, two servers, namely the GalaxyPepDock and HPEPDOCK, were used for docking between the peptide and MHC class 1.<sup>28</sup> The GalaxyPepDock webserver is available without restrictions at <http://galaxy.seoklab.org/pepdock>. It works based on similarity-based docking by uncovering templates from the database of experimentally verified structures and building models operating energy-based optimization that permits structural flexibility. The HPEPDOCK is an innovative online server for investigating protein-peptide docking based on the hierarchical algorithm. The HPEPDOCK web-server is accessible at <http://huanglab.phys.hust.edu.cn/hpepdock/>.

## Molecular Dynamics Simulation of Selected Epitopes with MHC-I

We used RMSD, RMSF, SASA, and gyration analysis in the molecular dynamics study. The simulated time was 100 ns. MD simulations were calculated operating the GROMACS ver. 5.1.2 software with the GROMOS 96, 54A7 force field. These simulations were achieved at constant pressure (1 atm) and temperature (300 K). The system was situated in the box (cubic box) with the SPC water model and dimensions of 90 \* 90 \* 90 Å<sup>3</sup>.

## Protein Solubility Prediction

The solubility of the chimeric protein and polarity of various residues were estimated by DSSP and further online programs such as VADAR ([redpoll.pharmacy.ualberta.ca/vadar/](http://redpoll.pharmacy.ualberta.ca/vadar/)) and the PROSO II server ([mips.helmholtz-muenchen.de/prosoII](http://mips.helmholtz-muenchen.de/prosoII)).<sup>29</sup> Moreover, the estimation of the mean residue accessible surface area (ASA) in the chimeric protein was done using the NetSurfP server ([www.cbs.dtu.dk/services/NetSurfP/](http://www.cbs.dtu.dk/services/NetSurfP/)).<sup>30</sup>

## Optimization of the Synthetic Gene for Expression of the Recombinant Protein

The nucleotide sequence for the chimeric protein was codon-optimized for the bacterial expression system (Genscript, USA). Parameters such as frequency of optimal codon (FOP), codon adaptation index (CAI), and GC content were also analyzed.

The secondary structure of mRNA was predicted for the chimeric gene by the Mfold Web-based server ([mfold.rna.albany.edu](http://mfold.rna.albany.edu))<sup>31</sup> and by the CentroidFold webserver (Sato et al 2009), both before and after codon optimization.

## Hosts and Plasmids

Individually, we utilized Top10 and BL21DE3 *E. coli* strains as hosts for cloning and utilized Pet-28a expression for recombinant gene expression.

## Sub-Cloning Recombinant Gene and Transformation of Recombinant Vector

Ultimately, we synthesized the optimized sequence and inserted it into the pBSK (+) vector. For amplifying the multi-epitope, *E. coli* TOP 10 was transformed via the pBSK (+) vector. The double digestion was done with NcoI and XhoI (Fermentas, Lithuania) after the plasmid elicitation (Fermentas, Lithuania). Also, the pET28a vector was digested with similar enzymes. Double digestion was

completed at 37°C. We studied the digested fragments by agarose gel electrophoresis, and the pET28a vector and the recombinant gene parts were purified (Fermentas, Lithuania). Lastly, the ligation of the V<sub>1</sub> domain to the pET28a was done using the T4 DNA ligase (Fermentas, Lithuania). The capability of *E. coli* BL21 (DE3) host cells was assessed based on the calcium chloride approach. The XhoI and NcoI enzymes were utilized to validate the efficiency of the transformation via the double digestion of plasmid.

## Analysis of Recombinant Gene Expression

About 20 µL of overnight pre-cultured transformed bacterial cells were inoculated to a 2 mL fresh LB medium (containing kanamycin (100 µg.mL<sup>-1</sup>)) and incubated for 2 h in an incubator shaker at 150 rpm (37°C) until reaching the optimized optical density. Next, 20 µL IPTG (100 µg.mL<sup>-1</sup>) was added to the medium to initiate protein expression, and incubation was performed in the shaker incubator for 6 h at 150 rpm and 37°C. The blend was centrifuged for 5 min at 5000 rpm and 4°C. The supernatant was disposed, and 60 µL urea (8 M) was added to the precipitate. The SDS-PAGES was used for validating the expression of the V1-domain. The mix of urea and the precipitated sample was solved in a 1x SDS-PAGE sample buffer. Eventually, the samples and protein molecules were loaded on the 15% SDS-PAGE and run with the 100v.

## Chimeric Protein Purification

Due to the presence of histidine sequences in the selected protein, the Ni-NTA column was used for the purification of the chimeric protein. *E. coli* BL21 (DE3) containing the recombinant vector (pET28a:: V1- Moc31-311\_1k2) was cultured in a volume of 1 liter and after reaching an OD<sub>600</sub> = 0.5, was induced with IPTG to a final concentration of 0.5 mM overnight. After this, the bacteria were collected by centrifugation and their precipitate was lysed with 5 mL of lysis buffer using an ultrasound pulse of 2–4 for 15 seconds with 5/3 power. After centrifugation, the supernatant was applied to the Ni-NTA column and after washing with buffers containing 20 mM and 50 mM imidazole, the soluble chimeric protein was separated from the column in response to the buffer containing 300 mM imidazole, furthermore placed in a suspension buffer, and finally, dialysis occurred in response to PBS buffer. After SDS-PAGE, the chimeric protein was observed on the polyacrylamide gel.

## Western Blot Analysis

Western blotting using polyclonal antibody anti histidine was used to confirm the expression of the recombinant protein.

## Results

### Engineering of a Chimeric Gene

The sequences of the CD166 and CD326 were obtained from NCBI and used to create a chimeric gene. The sequence of the CD166 V<sub>1</sub>-domain and two extracellular epitopes of CD326 were used to design a chimeric construct. These three fragments were linked together by a general linker (EAAAK). This linker was used to separate and provide stability to all three domains. To purify the recombinant protein, 6xHis-tag was added at the C-terminal of the gene. Thus, a chimeric gene with 759 nucleotides and protein-coding with a length of 253 amino acids was engineered (Figure 2).

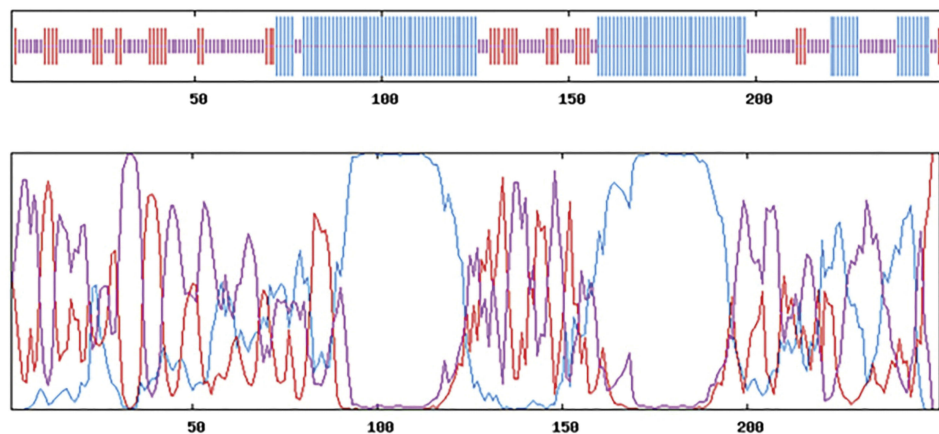
### Antigenicity and Allergenicity Evaluation

To predict the allergenicity of the chimeric protein, we submitted the sequence to the ALGPred and SDAP database and found that the chimeric protein is non-allergen.



**Figure 2** Schematic diagram of the final chimeric proteins designed.

**Notes:** Sequence length: 253. Alpha helix (Hh): 109 is 43.25%.  $3_{10}$  helix (Gg): 0 is 0.00%. Pi helix (Ii): 0 is 0.00%. Beta bridge (Bb): 0 is 0.00%. Extended strand (Ee): 41 is 16.27%. Beta turn (Tt): 0 is 0.00%. Bend region (Ss): 0 is 0.00%. Random coil (Cc): 102 is 40.48%.



**Figure 3** Prediction of the secondary structure of the chimeric protein.

## The Secondary and Tertiary Structure of the Engineered Chimeric Protein

The PORTER and SOPMA online web-based servers predicted the secondary structure of the chimeric protein. This structure consists of 252 amino acids that are made up of an alpha helix (43.25%) and random coil (40.48%) (Figure 3). The tertiary structure of the chimeric protein was prepared by operating homology modeling. The template for homology modeling was attained by pBLAST against the PDB database. Two templates were obtained, such as 5A2F “A” (1–127) and 4MZV “A” (121–246) with identities of 74.63% and 72.73%, respectively. Considering both the templates, advanced modeling was carried out using both the templates to build a full-length model of a chimeric protein (Figure 4).

### Evaluation of Model Stability

The modeled tertiary structure of the chimeric protein contained 74.8% residue in the core region, 16.4% in the allowed region, 8.8% residue in the generously permitted region, and 0.4% residue in the disallowed region in the Ramachandran plot, suggesting that the protein is predicted to be stable. The percentage of the residue was 74.8% in the favored region (core beta), 16.4% in the allowed (core-alpha), and 8.8% in the outlier (core left-handed alpha). The backbone dihedral angles were termed  $\phi$  and  $\psi$ , which were rationally accurate (Figure 5).

## The Engineered Chimeric Protein Is Antigenic and Stable

The chimeric protein was further analyzed for its immunogenicity using an independent alignment server; ie, Vaxijen -

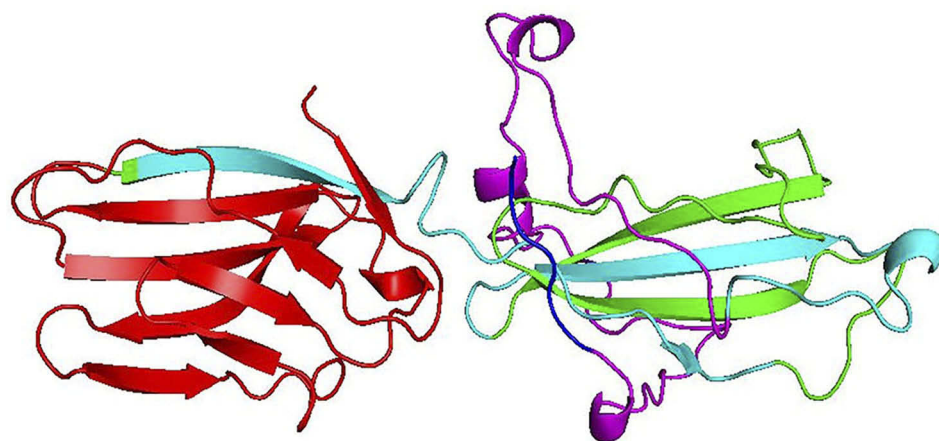


Figure 4 The tertiary structure of the chimeric protein.

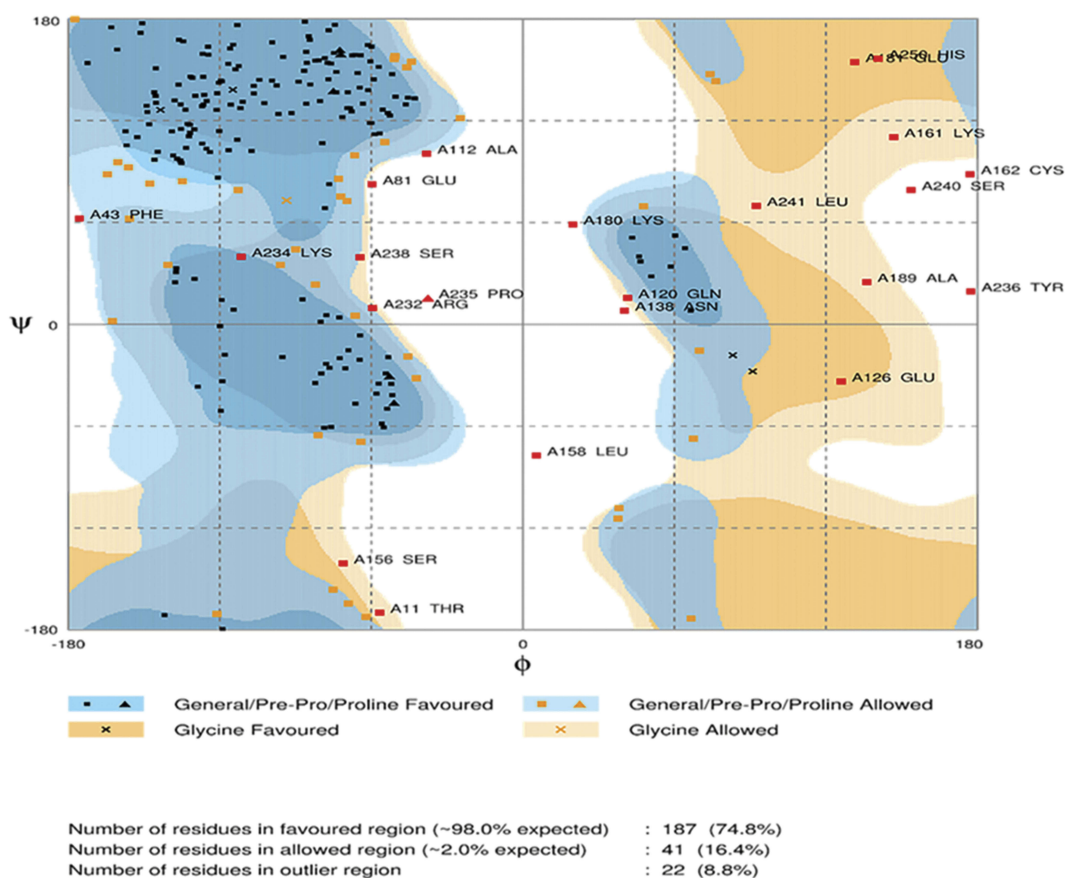


Figure 5 Evaluating model quality based on the Ramachandran plot.

v2. The threshold was kept at 0.5 to foretell possible antigenic and non-antigenic proteins. The chimeric protein was predicted as immunogenic with a score of 0.5709. ProtParam was applied to predict the physicochemical properties of the chimeric protein. The molecular weight and theoretical PI of

the chimeric protein were predicted as 28 kDa and 8.41, respectively. There are a total of 34 negatively charged residues and 38 positively charged residues in the chimeric protein. The instability index was computed to be 31.61, which classifies the molecule as stable.

## Identify Interfering Areas

Raptor X software was used to identify the interfering and suppressor regions of the structure in the recombinant protein. In terms of the amino acid sequence, this software predicts that the interfering areas in the correct folding accounts only for 2% of the total amino acids. The results of this assessment are shown in [Figure 6](#).

## The Chimeric Protein Possesses B Cell Epitopes

Since the B cell epitope has a critical role in antibody recognition, we evaluated the predicted tertiary configuration of the chimeric protein for the presence of continuous and discontinuous B cell epitopes using the EliPro server. The results indicate that the chimeric protein contains both continuous B cell epitopes ([Table 1A](#)) and discontinuous epitopes ([Table 1B](#)).

## The Chimeric Protein Possesses CTL Cell Epitope with a Specific MHC-I Restriction Element

The CTLpred server was used for predicting the cytotoxic T cell epitopes. The score predicted by the CTLpred is based on the specificity and sensitivity of the interaction between CTL and MHC I. Three epitopes were predicted by the CTLpred with the peptide scores of 1, 0.99, and 0.99 ([Table 2](#)). The binding affinity of the CTL epitope and MHC I molecule were verified using the NetMHC server. Finally, we selected the highest-scoring peptide and MHC showing the highest binding affinity for further analysis.

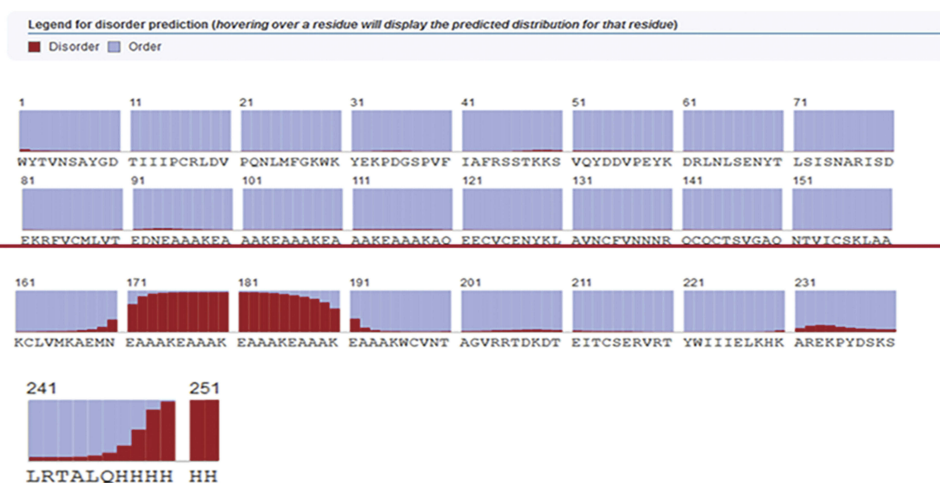
## The CTL Epitope Forms a Stable Complex with MHC I

The docking based on the minimum energy of the peptide and MHC I analysis with the HPEPDOCK showed the 10 best docking scores. We selected peptide number 2, which has a docking score of  $-209.839$  ([Table 3](#), [Figure 7A](#)). Furthermore, we used the Galaxy Webserver for confirming the output of the HPEPDOCK server. The result of the galaxy showed us that peptide number 2 has the best interaction with MHC I ([Table 3](#), [Figure 7B](#)).

In the RMSD analysis, the protein produced more oscillating properties over 100 ns than when the peptide was set up with MHC1 binding. During the simulation, the average RMSD for the protein is more than the average RMSD for the complex of peptide-MHC1. This graph shows that the MHC1 connecting to the protein occurs after the primary modifications. Moreover, the stable RMSD and roughly constant fluctuation were stimulated in 100 ns ([Figure 8](#)).

Due to the presence of the B-sheet receptor, relatively high stability was observed when the MHC1 was connected to the receptor during the duration of the simulation. In the graph of C-alpha root-mean-square fluctuation (RMSF) based on residues, the position of every amino acid residue is shown. It is observed that the RMSF value is in accordance with the RMSD graph and the RMSF of the protein is more than the RMSF of the peptide-MHC1 complex. Moreover, the residue range of RMSF for the peptide and the MHC1 complex is much higher than the natural state of the residues ([Figure 9](#)).

In the SASA analysis, it was found that the solvent-accessible surface area for the peptide is significantly less than the peptide-MHC1 complex. This graph shows that the



**Figure 6** The prediction of interfering regions in protein structure and 25 (9%) positions predicted as disordered.

**Table 1** A) the Linear Epitope of Protein Multiple Structures for Lymphocyte B and B) the Discontinuous Epitope of Protein Multiple Structures for Lymphocyte B

A						
No.	Chain	Start	End	Peptide	Number of Residues	Score
1	A	88	99	LVTEDNEAAAKE	12	0.808
2	A	16	26	CRLDVPQNLMF	11	0.763
3	A	33	38	KPDGSP	6	0.742
4	A	179	187	AKEAAAKEA	9	0.735
5	A	149	168	AQNTVICSKLAAKCLVMKAE	20	0.717
6	A	42	57	AFRSTTKKSVQYDDVP	16	0.703
7	A	220	252	TYWIIILKHKAREKPYDSKSLRTALQHIIIIIIII	33	0.699
8	A	64	69	NLSENY	6	0.614
9	A	204	211	RRTDKDTE	8	0.583
10	A	117	122	AKAQEE	6	0.529
B						
No.	Residues				Number of Residues	Score
1	A:K33, A:P34, A:D35, A:G36, A:S37, A:P38				6	0.742
2	A:C16, A:L18, A:D19, A:V20, A:P21, A:Q22, A:N23, A:L24, A:M25, A:F26, A:I41, A:A42, A:F43, A:R44, A:S45, A:S46, A:T47, A:K48, A:K49, A:S50, A:V51, A:Q52, A:Y53, A:D54, A:D55, A:V56, A:P57, A:K60, A:D61, A:S66, A:E67, A:N68, A:Y69, A:T70, A:L88, A:V89, A:T90, A:E91, A:D92, A:N93, A:E94, A:A95, A:A96, A:A97, A:K98, A:E99				46	0.735
3	A:A179, A:K180, A:E181, A:A182, A:A183, A:A184, A:K185, A:E186, A:A187, A:K190				10	0.696
4	A:A119, A:E121, A:E122, A:V124, A:C125, A:E126, A:N127, A:Y128, A:A149, A:Q150, A:N151, A:T152, A:V153, A:I154, A:C155, A:S156, A:K157, A:L158, A:A159, A:A160, A:K161, A:C162, A:L163, A:V164, A:M165, A:K166, A:A167, A:E168, A:M169, A:N170, A:E171, A:A172, A:G202, A:R204, A:R205, A:T206, A:D207, A:K208, A:D209, A:T210, A:E211, A:I223, A:I224, A:I225, A:E226, A:L227, A:K228, A:H229, A:K230, A:A231, A:R232, A:E233, A:K234, A:P235, A:Y236, A:D237, A:S238, A:K239, A:S240, A:L241, A:R242, A:T243, A:A244, A:L245, A:Q246, A:H247, A:H248, A:H249, A:H250, A:H251, A:H252				71	0.651
5	A:Y2, A:T3, A:R17				3	0.542

**Table 2** Epitopes Detectable by T Cell Recombinant Multi-Topo Structure Using the CTLpred Database

Peptide Rank	Start Position	Sequence	Score	Prediction
1	50	SVQYDDVPE	1	Epitope
2	14	IPCRLDVPQ	0.99	Epitope
3	68	NYTLLSISNA	0.99	Epitope

protein has been cracked connecting to MHC1, and the SASA is higher than the protein alone. Therefore, after MHC1 coupling, we observe that the surface of the receptor structure expands (Figure 10).

Rg is utilized to compute the compactness and stability of the protein. This parameter is sensitive to the degree of folding/unfolding of the protein. Rg is shown as:

$$Rg = ((1/N(\sum (ri - rcm)^2 Ni = 1))^{1/2}$$

From the gyration graph, it is seen that the SASA results are verified. Moreover, an increase in the radius of gyration for the peptide and the MHC1 complex is observed with an average of about 1.7 nm if the average gyration of the peptide alone is 1.2 nm (Figure 11).

## The Codon-Optimized Engineered Gene Is Stable

The chimeric gene was assembled based on the *E. coli* codon optimization. The codon practice was biased for *E. coli* (BL21) by increasing the codon adaptation index (CAI) from 0.42 to 0.96. When CAI is zero, it means that all of the synonym codons have been used to a degree in one gene, and the value of one equals the maximum of the codon inclination rate. In this case, only the optimal codons are used. Consequently, it represents a more permanent expression of a gene,

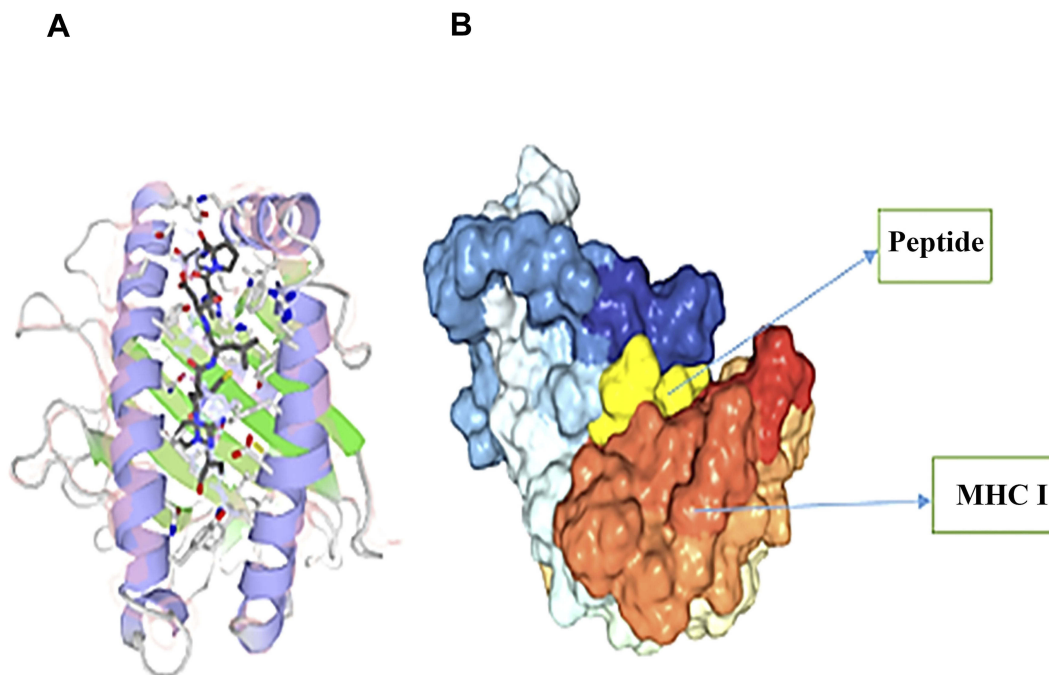
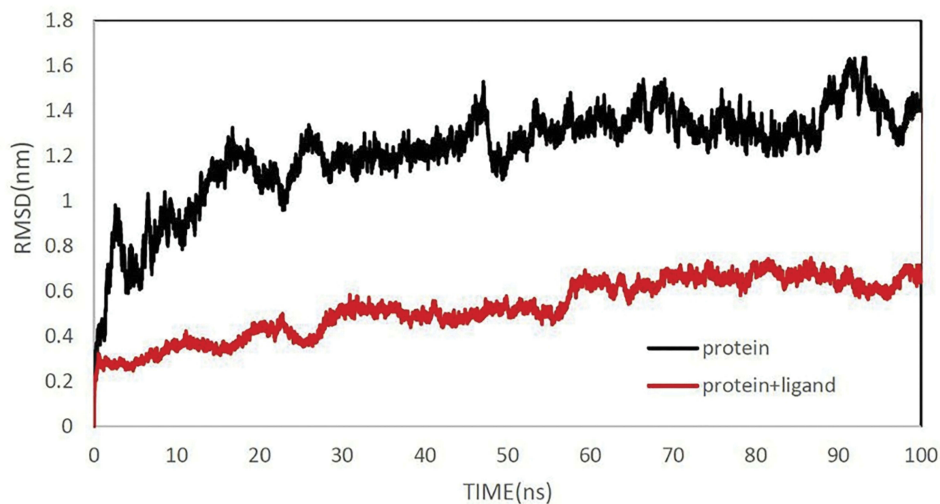


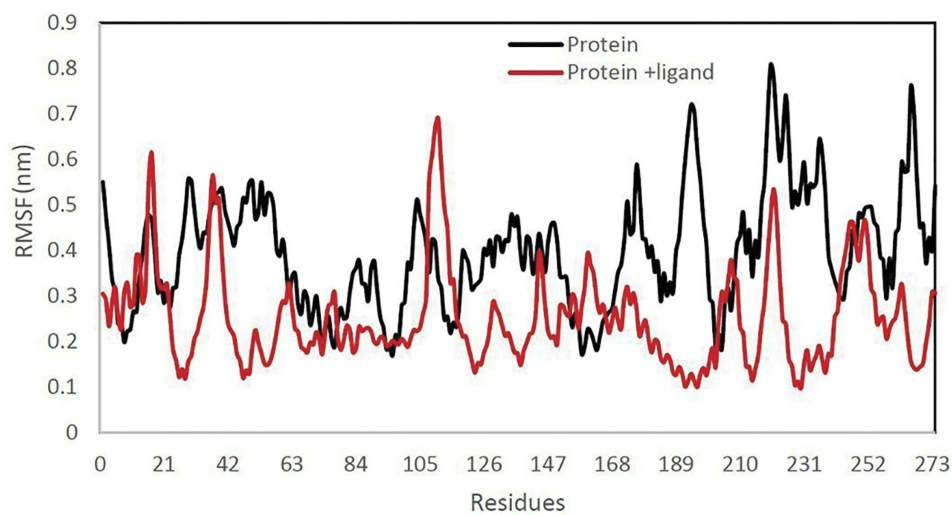
**Table 3** The Interaction Between MHC I and Peptide Based on the HPEPDOCK and Galaxy Webserver

Row	Peptide	Protein Structure Similarity (TM-Score)	Interaction Similarity Score	Estimated Accuracy	Docking Score	RMSD Score
1	SVQYDDVPE	0.967	125	0.873	-203.23	174.23
2	IPCRLDVPO	0.967	171	0.967	-209.839	-
3	NYTLSISNA	0.951	144	0.9	-206.876	-

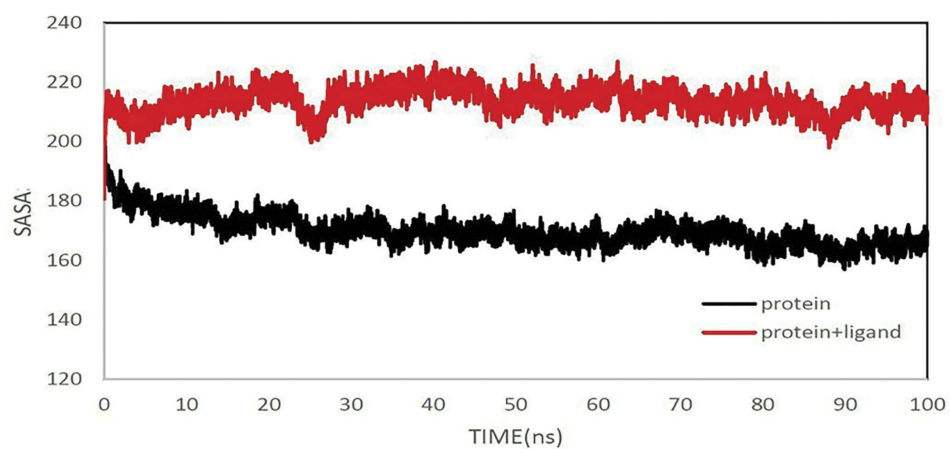
determining the level of its expression. Given this parameter, CAI of 1 is the best possible expression, more than 0.8 is a good state of expression, and 0.6 has an

average expressive state (Figure 12). In addition, the frequency of optimal codon (FOP) was improved from 42 to 78 in by optimizing the gene sequence. The FOP

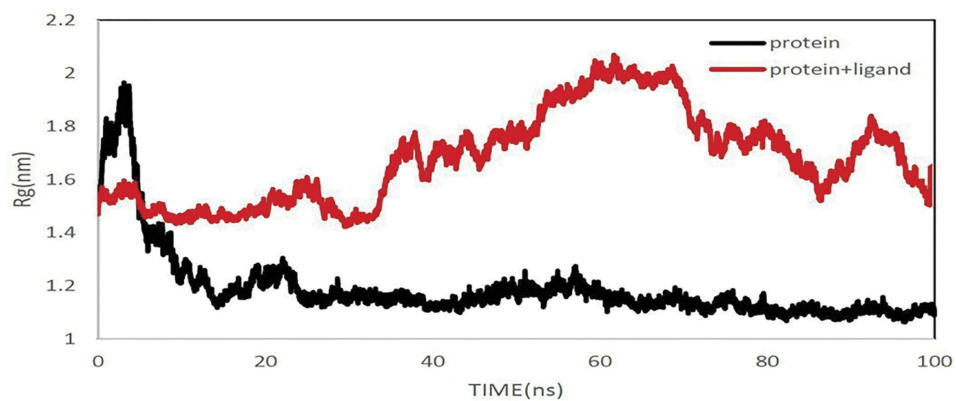
**Figure 7** (A) The predicted protein-peptide complex structures by the galaxy webserver and (B) The docking scores predicted with HPEPDOCK.**Figure 8** Root mean square deviations (RMSD) of the protein and C $\alpha$  RMSD a function of residue number for peptide and peptide+MHC class I, respectively, over 100 ns.



**Figure 9** Root mean square fluctuation of protein and C $\alpha$  atoms as a function of residue number for peptide and peptide+MHC I, respectively.



**Figure 10** The solvent-accessible surface area of protein. SASA for peptide and peptide+MHC I over 100 ns.



**Figure 11** The radius of gyration (Rg) of protein. The radius of gyration of peptide and peptide+MHC I over 100 ns.

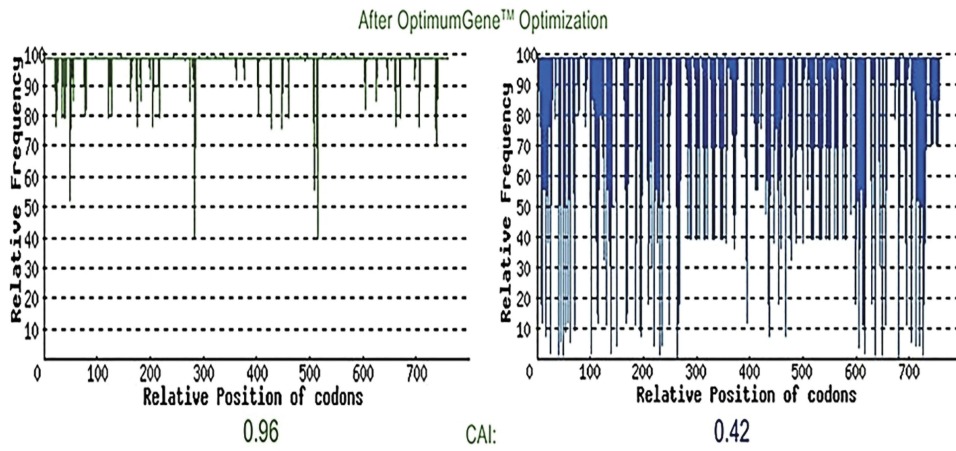


Figure 12 Codon Adaptation Index (CAI) (Right: before optimization and Left: after optimization).

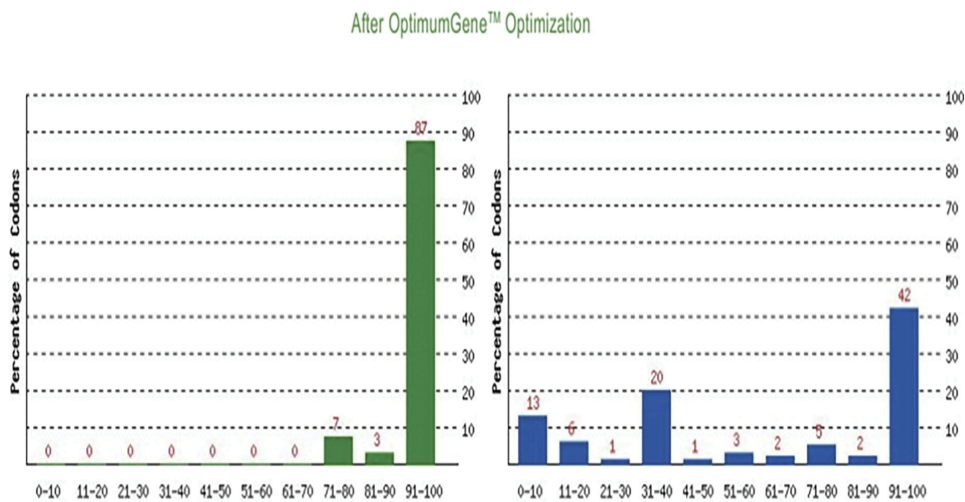


Figure 13 Frequency of Optimal Codons (FOP) (Right: before optimization and Left: after optimization).

of the optimum represents the exact expression of the codons encoding amino acids. It means that given the degree of changes that can occur throughout the

sequence, for a particular amino acid, the codon at the host can have the highest expression efficiency. The percentage distribution of codons in computed codon

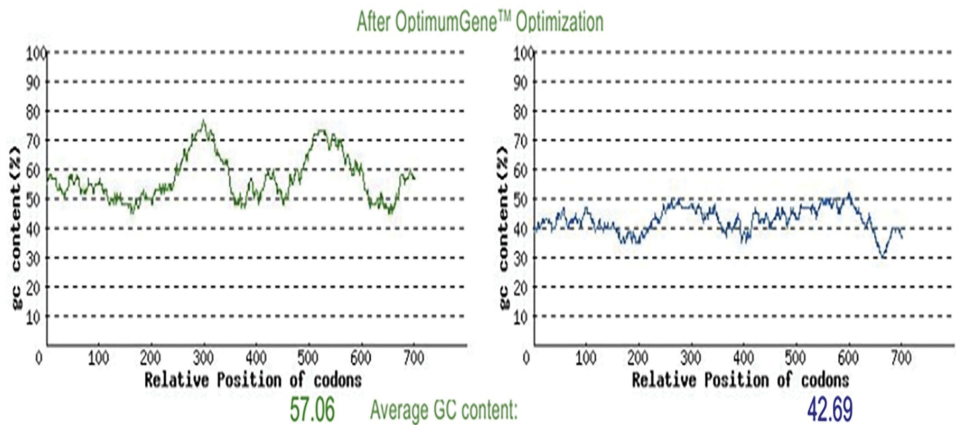


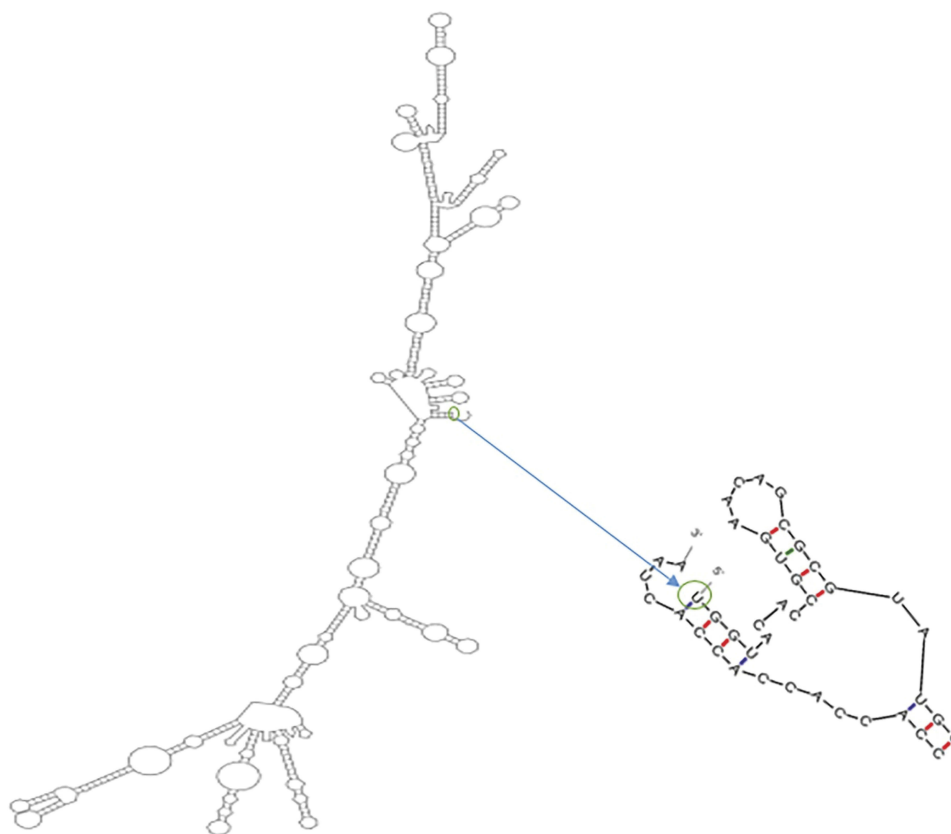
Figure 14 GC Content Adjustment (Right: before optimization and Left: after optimization).

Optimized 1	TGG <b>TACACCGTGAACAGCGCG</b> TAT <b>GGCGAC</b> ACCATT <b>ATTATTCCGTGCCGCGCTGGATGTG</b>
Original 1	TGGTATACTGTAAATTCAGCATATGGAGATACCATTATCATACTTGGCCACTTGACGTA
Optimized 61	<b>CCGCAGAACCTGATGTTTGGCAAGTGG</b> <b>AAGTACGAGAAACCGGACGGTAGCCCGGTGTTCC</b>
Original 61	CCTCAGAATCTCATGTTTGGCAAATGGAAATATGAAAAGCCCGATGGCTCCCCAGTATTT
Optimized 121	ATT <b>GCGTTTTCGTAGCAGCACCA</b> AGAAA <b>AGCGTGCAGTACGACGATGTTCCGGAGTATAAG</b>
Original 121	ATTGCCTTCAGATCCTCTACAAAGAAAAGTGTGCAGTACGACGATGTACCAGAATACAAA
Optimized 181	<b>GATCGTCTGAACCTGAGCG</b> AAAACT <b>TATACCCTGAGCATTAGCAACGCGCGTATCAGCGAC</b>
Original 181	GACAGATTGAACCTCTCAGAAAACTACACTTTGTCTATCAGTAATGCAAGGATCAGTGAT
Optimized 241	<b>GAGAAACGTTTTCGTGTGCATGCTGGTTACC</b> GAGGATAACGAAGCT <b>GCGGCGAAAGAGGGCG</b>
Original 241	GAAAAGAGATTTGTGTGCATGCTAGTAACTGAGGACAACGAAGCTGCTGCTAAAGAAGCT
Optimized 301	<b>GCGGCGAAAGAGGGCGGCGGCG</b> AAA <b>GAGGGCGGCGGCG</b> AAA <b>GAGGGCGGCGGCGAAGGGCGCAG</b>
Original 301	GCTGCTAAAGAAGCTGCTGCTAAAGAAGCTGCTGCTAAAGAAGCTGCTGCTAAAGCACAG
Optimized 361	<b>GAAGAGTGCGTGTGCG</b> AAAACTACAAGCT <b>GCGGGTGA</b> ACTGCTTT <b>GTTAACAACAACCGT</b>
Original 361	GAAGAATGTGTCTGTGAAAACTACAAGCTGGCCGTAAACTGCTTTGTGAATAATAATCCT
Optimized 421	<b>CAGTGCCAATGCACCAGCGTGGTGCG</b> CAA <b>ACACCGTT</b> ATTTGCAGCAAGCT <b>GCGGCGG</b>
Original 421	CAATGCCAGTGTACTTCAGTTGGTGACAAAATACTGTCATTTGCTCAAAGCTGGCTGCC
Optimized 481	AAAT <b>GCCTGGTTATGAAAGCGGAGATGAATGAGGCTGCGGCG</b> AAA <b>GAGGGCGGCGGCG</b> AAA
Original 481	AAATGTTTGGTGATGAAGGCAGAAATGAATGAAGCTGCTGCTAAAAGAAGCTGCTGCTAAA
Optimized 541	<b>GAGGGCGGCGGCG</b> AAA <b>GAGGGCGGCGGCG</b> AAA <b>GAGGGCGGCGGCG</b> AAATGG <b>TGCGTGAACACC</b>
Original 541	GAAGCTGCTGCTAAAGAAGCTGCTGCTAAAGAAGCTGCTGCTAAATGGTGTGTGAACACT
Optimized 601	<b>GCGGGCGTTTCGTCTACC</b> GACAAGG <b>TACCGAGATCACCTGCAGCGAACGTGTTTCGTACC</b>
Original 601	GCTGGGGTCAGAAGAACAGACAAGGACACTGAAATAACCTGCTCTGAGCGAGTGAGAACC
Optimized 661	<b>TATTGGATCATTATCGAGCTGAAGC</b> ACAAA <b>GCGCGT</b> GAAA <b>AGCCGTATGACAGCAAGAGC</b>
Original 661	TACTGGATCATCATTGAACTAAAACACAAAAGCAAGAGAAAAACCTTATGATAGTAAAAGT
Optimized 721	<b>CTGCGTACCGCGCTGCAACATCACCACCACC</b> ACTAA
Original 721	TTGCGGACTGCACTGCAGCATCATCATCATCATTA

**Figure 15** DNA Alignment; comparing nucleotide for peer to peer in the optimized and primary sequences.

quality groups. The value of 100 is set for the codon with the highest usage frequency for a given amino acid in the desired expression organism (Figure 13). In the optimized sequence, the CG content changed from 42.69% to 57.06%. The ideal percentage range of GC

content is between 30–70%. Peaks of % GC content in a 60 bp window have been removed (Figure 14). In Figure 15, the nucleotide sequences of the original and optimized nucleotide sequences are compared peer to peer.



**Figure 16** The analysis of mRNA stability and start codon in the structure.

## mRNA Structure Prediction

The Mfold Web-based server ([mfold.rna.albany.edu](http://mfold.rna.albany.edu)) was used to examine the mRNA secondary structure of the chimeric gene before and after codon usage. The results of the RNAfold and GeneBee ([www.genebee.msu.edu/services/rna2\\_reduced.html](http://www.genebee.msu.edu/services/rna2_reduced.html)) were established by the CentroidFold webserver. As can be seen, the mRNA was steady enough for effective translation in the prokaryotic host.<sup>32</sup>

The “Mfold” server was used to evaluate the original and optimized sequences of the minimum free energy for 33 structures of chimeric mRNA. The calculations showed that the  $\Delta G$  of the best-predicted structure regarding the optimized assembly and the original one was  $-261.70$  kcal/mol and  $-196.6$  kcal/mol, respectively. The primary nucleotides at 5' did not possess a long stable hairpin or pseudoknot. Thus, the attachment of the ribosomes to the translation initiation site and the resulting translation procedure is achieved in the target host. The data gained in the “RNAfold”

webserver match the results (Figure 16). According to the energy dot plot, the minimum energy of the original (Figure 17) is less than the optimized (Figure 18) prediction of the mRNA secondary structure of the chimeric protein.

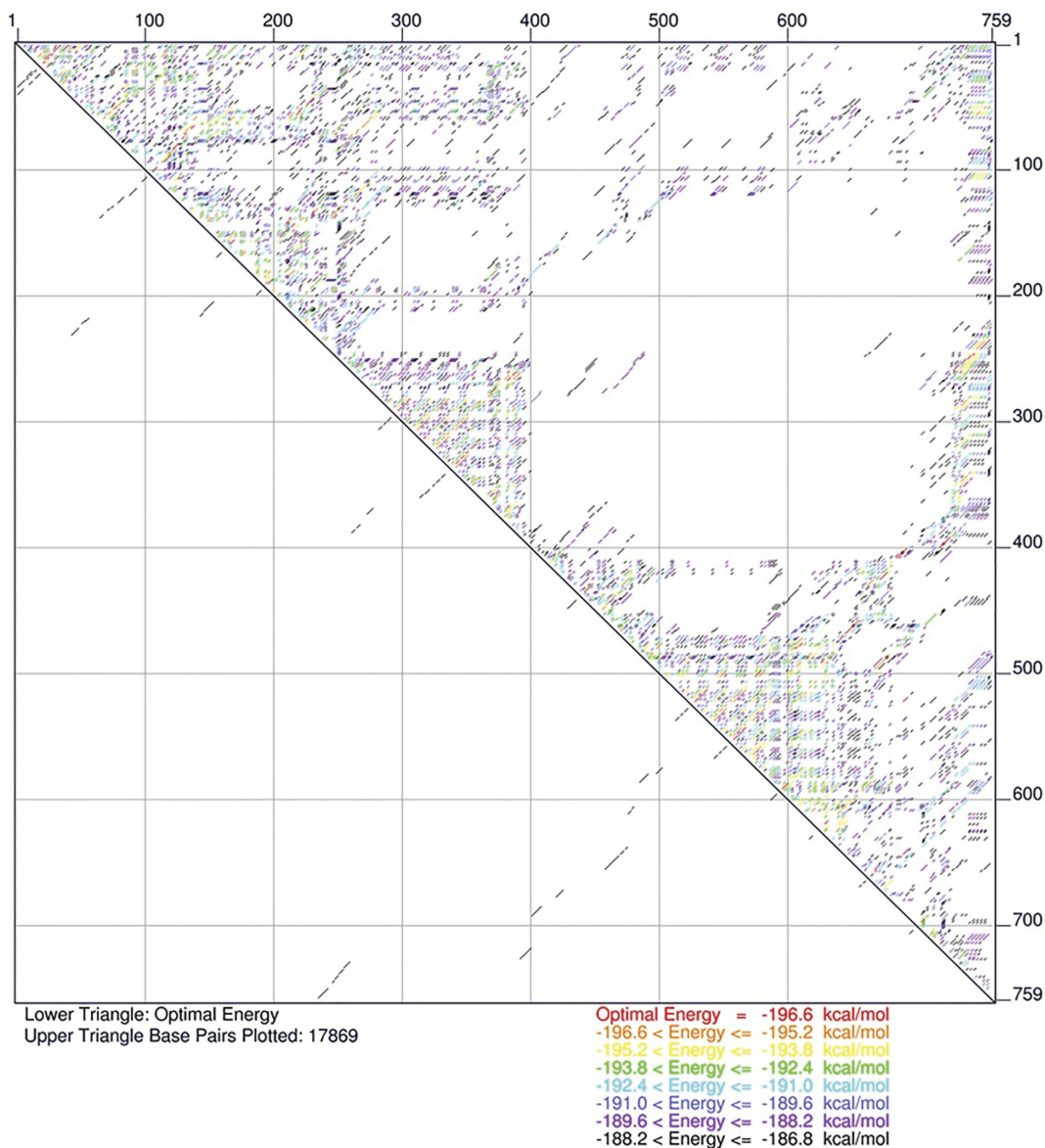
## Cloning and Expression of the Engineered Protein

The synthesized recombinant gene with a size of 759 bp was placed between two *Nco*I and *Xho*I restriction enzyme sites and cloned into the vector of pBSK (Figure 19).

The recombinant gene that was synthesized by transformed pBSK (+) was amplified into *E. coli* top10. It was digested and subcloned into the pET-28a expression vector successfully (Figure 20).

In the next step, the recombinant vector (pET-28a + gene) was transfected into *E. coli* BL21 (DE3). We verified the accuracy of transformation by double digestion of the plasmid with the *Nco*I and *Xho*I enzymes (Figure 21).

## Fold of original at 37 C.

 $\Delta G$  in Plot File = 9.8 kcal/mol

**Figure 17** The energy dot plot for the original chimeric mRNA, the optimal energy is  $-196.6$  kcal/mol.

The recombinant colonies were chosen and cultured in an LB medium and induced by IPTG. The SDS-PAGE technique was applied to investigate the expression of the recombinant protein. Given that the molecular weight of the recombinant protein is 27 kDa, 15% SDS-PAGE gel was utilized (Figure 22).

## Purification and Expression of the Chimeric Protein

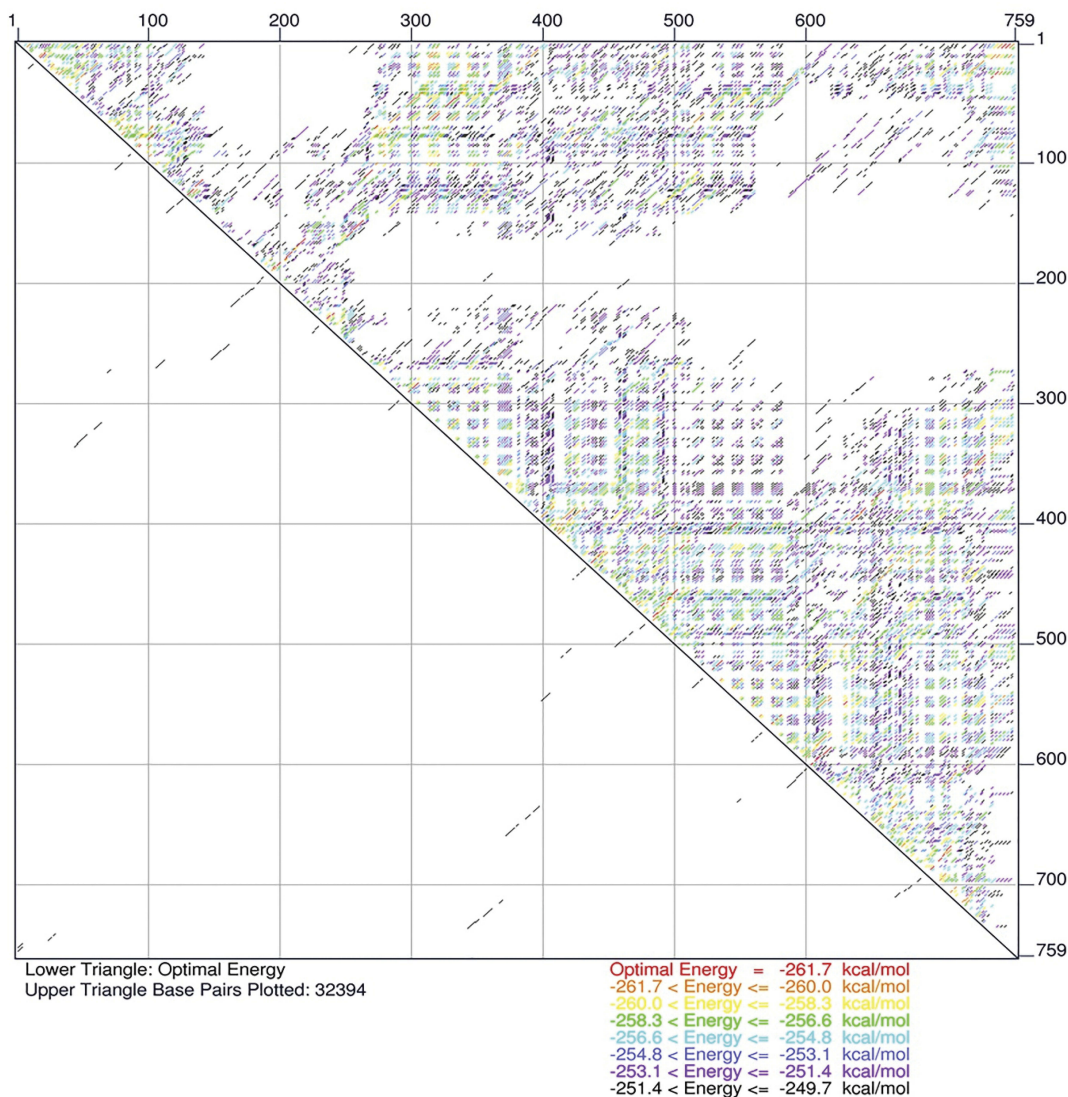
As the protein of choice was expressed with a histidine sequence, the Ni-NTA column was used for the chimeric protein purification. The chimeric protein was suspended

Output of boxplot\_ng (©)  
mfold\_util 4.7

Created Mon Feb 18 13:34:03 2019

## Fold of optimized at 37 C.

$\delta G$  in Plot File = 12.0 kcal/mol



**Figure 18** The energy dot plot for the optimized chimeric mRNA, the optimal energy is  $-261.7$  kcal/mol.

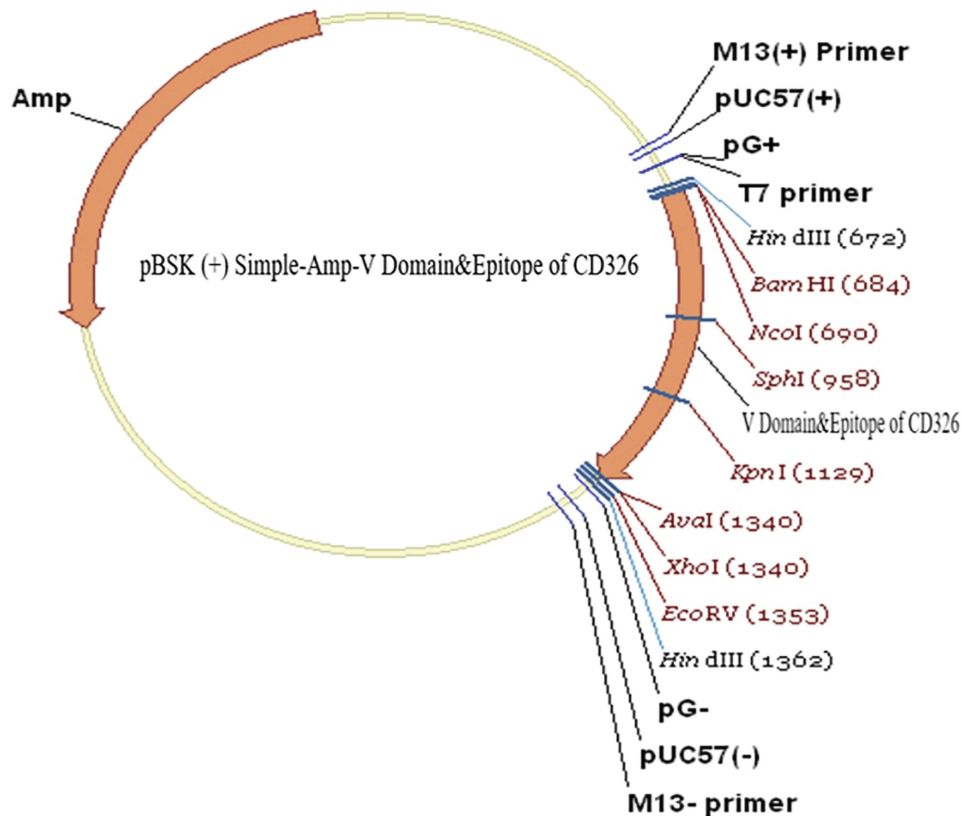
in a buffer containing 300 mM imidazole. After SDS-PAGE, the chimeric protein was observed on the polyacrylamide gel (Figure 23).

Due to the presence of the 6-histidine sequence at the amino end of the recombinant protein, Western blot analysis was performed using a polyclonal anti-histidine antibody to confirm the expression. In addition, the 27-kDa band of the chimeric protein was observed upon detection with diaminobenzidine substrate (DAB) (Figure 24).

## Discussion

Chemotherapy and radiation therapy are developed for the treatment of cancer. However, these methods often cause undesirable side effects. Currently, antibody-based targeted therapy is widely used against tumor antigens for treating patients with cancer.

The bioinformatics approaches can lead to a significant reduction in time, expense, and failure in experimental attempts. However, bioinformatics predictions may not



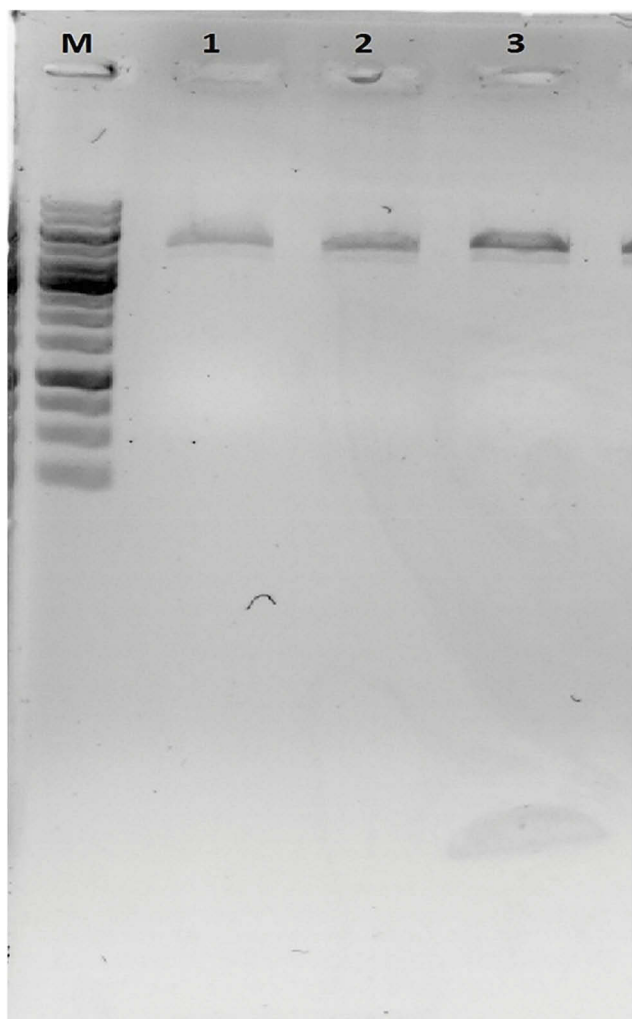
**Figure 19** Schematic figure of construct map pBSK (+) Simple-Amp-V<sub>1</sub>-domain of the CD166 and two extracellular immunogenic epitopes of CD326.

always be in concordance with experimental results. In this regard, with the advancement of software and continuous information regarding the relationship between the structure and function of the protein, they could play a crucial responsibility in vaccine design, development of a protein suitable for antigen preparation used in immunoassay, structural studies, drug-protein, and protein-protein interaction analysis.

CD326 was first identified in 1979 as an antigen that induces the production of specific antibodies in mice immunized with human CRC cells.<sup>33</sup> The adhesive function and intercellular binding property of the CD326 were recently discovered; however, the potential signaling role of this molecule in cancer has just been considered. Immunohistochemical studies have shown that the CD326 gene is highly expressed in malignant epithelial tumors such as CRC. This increase in expression at various stages of cancer development has made the CD326 protein one of the molecular markers for the target therapeutic options for patients with CRC.<sup>6</sup> The association of the CD326 protein with proliferation, adhesion, tissue consolidation, stimulation of tumor growth, and metastasis also led to the

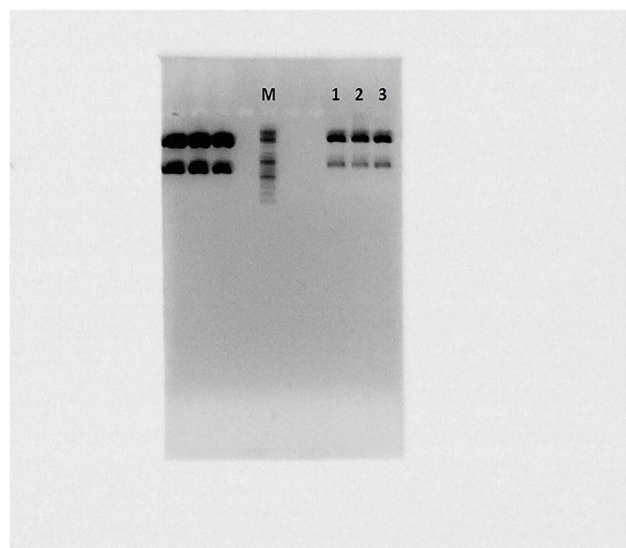
selection of this protein as a marker for cancer treatment. Recent studies have shown that CD166 is frequently upregulated in CRC. Hence, underscoring the importance of CD166 in tumor progression in this disease<sup>34</sup> has made CD166 a new independent prognostic marker. The low expression of CD326 and CD166 in healthy cells compared with tumor cells in CRC is one of the reasons why the effects of CD326<sup>35</sup> and CD166-based therapies are unlikely to have any side effects.<sup>36</sup> In addition, the CD326 protein appears in the natural tissues as a complex on the cell surface and CD9, CD44, and Claudine-7 surrounding it. While the CD326 protein in the tumor cell appears abundantly on the surface of the tumor cell,<sup>37</sup> the anti-chimeric protein antibodies can more likely be attached to healthy cells at low levels. Wiiger et al characterized a novel fully human scFv antibody (scFv anti-CD166 antibody) recognizing CD166 on cancer cells and in tumor tissues that reduce cancer cell invasion and tumor growth.<sup>38</sup> In current years, widespread efforts have been made to produce antibodies against the CD326 protein. The first monoclonal antibody used against CD326 was Eredrelomab, also known as Panorex. In this study, patients who received Edrecolomab





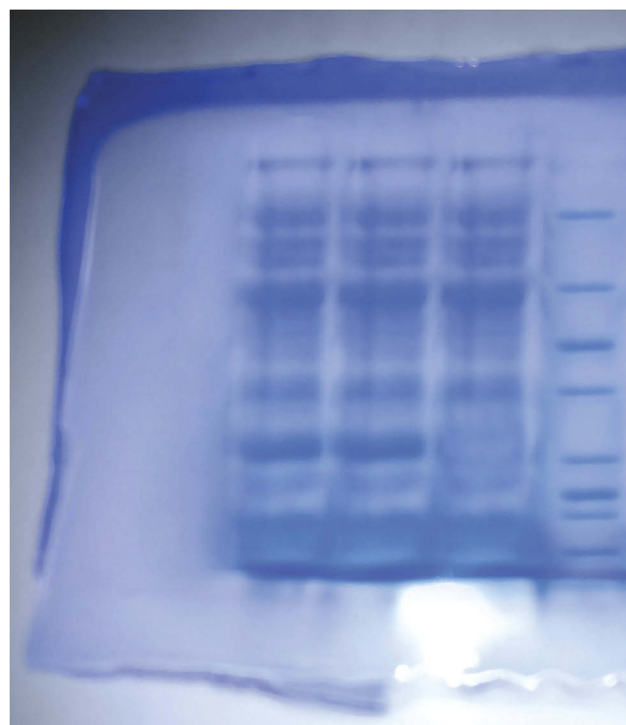
**Figure 20** Extraction of the transformed plasmid (pET-28a Recombinant Vector (containing V<sub>1</sub>-domain of the CD166 and two epitopes of CD326) M, Gene Ruler™ 1 kb ladder; Lane 1, 2, and 3: Extraction of the transformed plasmid.

had a significant reduction in tumor recurrence and death rates. The concept of tumor-initiating cells (TICs) or cancer stem cells (CSCs) has been extended more recently to solid tumors treatment, although it was previously well known in several hematologic cancers. This concept suggests that a tiny subpopulation of cells (CSCs) keeps cancer tissues active by sustaining phenotypically distinct cancer cells. The unique characteristics of these cells enable self-renew and generate progeny cells that lead to cellular heterogeneity of the tumor. So, this model may explain why many modern therapeutic methodologies eventually cause disease relapse. Adhesion molecules might be in tumor cell-endothelial cell adhesion, tumor cell-matrix adhesion, or tumor cell adhesion; all these adhesions are necessary at various times during primary tumor formation or metastasis. The focus of this study is on CD166 and CD326, which

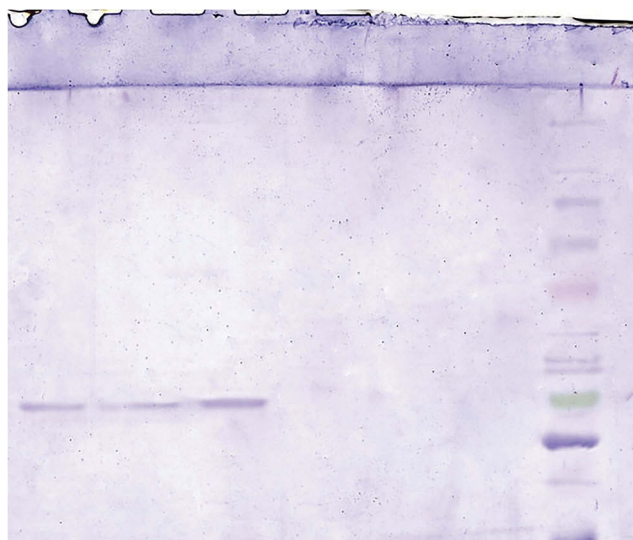


**Figure 21** The double digestion of the pET-28a recombinant vector: The double digestion of pET-28a recombinant vector; M, Gene Ruler™ 1 kb ladder (Lanes 1, 2 and, 3 double digestion).

have been introduced as CRC stem cell markers at various stages of development and recognized as a potential therapeutic and attunement target for CRC. The design of the chimeric antigen of the CD166 protein V domain and the immunogenic epitope of the membrane protein CD326,



**Figure 22** Expression analysis of chimeric protein by SDS-PAGE. column 1, protein marker (column, 3 and 4: Induction of E. coli BL21 (DE3) with IPTG; column 2: Non-induced E. coli BL21 (DE3)).

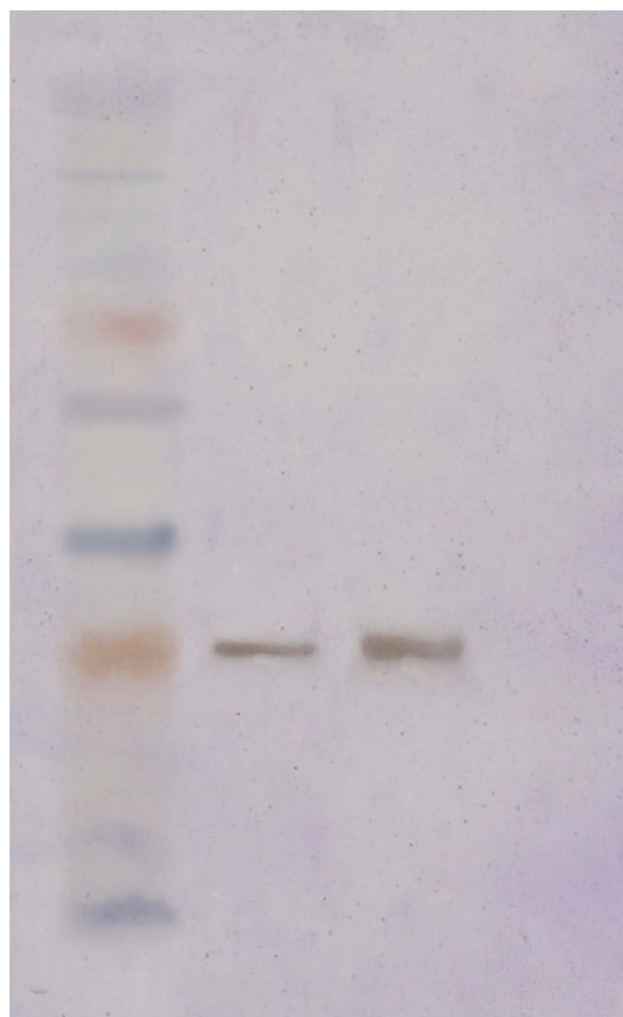


**Figure 23** Isolation and purification of the chimeric protein. (A) Chimeric protein purification by Ni-NTA column, respectively, column 2, 3 and 4 are a chimeric protein with 27K.Da. Column 1 is an MW marker.

bioinformatics analysis, cloning, and ultimately the desired protein suggests that the results of this study can improve the diagnostic methods. This improvement can lead to producing diagnostic kits for early diagnosis of pre-malignant CRC cells and producing a promising vaccine for this cancer.

Prior to the construction of recombinant protein vaccines, as opposed to CRC for clinical function, extensive testing is necessary. Recombinant protein vaccines must be harmless, stable in various conditions, have high immunogenicity, and high specificity in addition to affordable production. In this regard, it is recommended considering the current developments in the subject of vaccine advancement using bioinformatics, immune-informatic tactics and biophysics software. Thus, herein, we made a chimeric protein, including the V1-domain of the CD166 and two epitopes of CD326 with a proper linker (EAAAK) in between. In this way, we fabricated a multi-epitope vaccine with 490 residues in length. We eliminated the instability elements, restriction sites, and every cis-acting site from the construct, which considerably interfered with the cloning.

In silico studies approved active transcription and translation, in addition to the quality expression of the proposed construction in host expression vectors. In quantitative terms, the codons tend to be measured using the Codon Adaptation Index (CAI), which varies between 0 and 1. When the index is equal to 0, it means that all synonymous codons are used to the same extent in the same gene, and a value equal to 1 means the



**Figure 24** Western blot analysis. Chimeric protein with anti- polyclonal Histidine, Column 1, 2 after chimeric protein fusion, 3 hosts of unfused BL21, Column 4 is an MW marker.

maximum codon affinity, in which only the optimal codons are used. Therefore, it will represent a more constant and permanent expression of a gene and thus determines its expression level. In our gene, the CAI index has increased from 0.42 in the wild type sequences to 0.96 in the chimeric optimized gene. Moreover, the overall GC content declined from 42.69 to 57.06%, which should increase the overall stability of mRNA from the synthetic gene. Also, we added the required restriction enzyme sites to the ends of the designate gene for future assays. Codon optimization ensured that the synthetic construct expressed well in the wanted host vector.

The mRNA secondary configuration is an important factor in the expression of proteins. The outcomes from mRNA prediction by the Mfold server revealed that the

mRNA had enough stability to translate in the host effectively. Subsequently, higher stability brings about a higher expression rate. The mRNA structure was adjusted based on low  $\Delta G$  and energy of the start codon. This character might help ribosome binding and translation initiation. To predict the RNA secondary structure, a genetic algorithm-based RNA secondary structure prediction was merged with comparative sequence analysis to determine the potential folding of the chimeric gene. We also predicted the minimum free energy for the secondary structure of RNA molecules. The program mfold was applied to investigate the mRNA secondary structure of the chimeric gene. The finest structure had  $\Delta G = -261.70$  kcal/mol and  $-196.6$  Kcal/mol. The data displayed that the mRNA was stable enough for efficient translation in the novel host.

The ProtParam software was used to analyze the physio-chemical parameters. Similarly, ExPASy ProtParam categorizes the chimeric protein as a steady protein with an instability index of 31.61.

The analysis of every continuous and discontinuous B-cell epitope demonstrated that the epitopes found on the surface of the protein might interact easily with antibodies, and they were commonly adaptable.

However, overall, the computational methodologies are vital stages and instruments for assessing the vaccines earlier than starting the experimental analysis. The assembled vaccine that utilizes the mentioned immunoinformatics techniques must be experimentally estimated to simplify efficiency and have the achievement of the ultimate fabricated vaccine. Hence, to bring our research on epitope-based vaccine advancement against leptospirosis to an end, in vivo experimental studies can be suggested for measuring the various chemical and physical aspects of this chimeric protein.

## Conclusion

In this study, we designed a novel chimeric vaccine for cancer immunotherapy. Multiple different approaches have been used to activate the immune system against CRC. Here we have evaluated the ability of chimeric protein composing of V1-domain of the CD166 and epitopes of CD326 as a new antitumor candidate. Since it was important to establish the structure-function relation of chimeric protein before starting experimental studies, the chimeric protein has analyzed by various tools and softwares.

Our data showed that the possibility of successful production of a large chimeric protein composing of V1-domain of the CD166 and epitopes of CD326 in the prokaryotic host. In addition, our data signify that the V<sub>1</sub>-domain of the CD166 and epitopes of CD326 of the synthetic chimeric protein could induce both T-cell and B-cell-mediated immune responses. These findings will intensify efforts to develop a vaccine against CRC and may also suggest this synthetic chimeric protein could help to diagnosis of CRC. Thus, the chimeric protein can possibly be utilized to produce CRC diagnostic kits based on the ELISA technique and develop a protective vaccine against CRC. Further studies are required to establish these notions which are the theme of our future researches.

## Acknowledgment

Many thanks to Dr. Davood Afshar for scientific comments.

## Disclosure

The authors report no conflicts of interest in this work.

## References

1. Favoriti P, Carbone G, Greco M, Pirozzi F, Pirozzi RE, Corcione F. Worldwide burden of colorectal cancer: a review. *Updates Surg.* 2016;68(1):7–11.
2. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA Cancer J Clin.* 2015;65(2):87–108. doi:10.3322/caac.21262
3. Nakayama G, Tanaka C, Kodera Y. Current options for the diagnosis, staging and therapeutic management of colorectal cancer. *Gastrointest Tumors.* 2014;1:25–32. doi:10.1159/000354995
4. Fujiwara K, Ohuchida K, Sada M, et al. CD166/ALCAM expression is characteristic of tumorigenicity and invasive and migratory activities of pancreatic cancer cells. *PLoS One.* 2014;9(9):e107247. doi:10.1371/journal.pone.0107247
5. Lieto E, Galizia G, Orditura M, et al. CD26-positive/CD326-negative circulating cancer cells as prognostic markers for colorectal cancer recurrence. *Oncol Lett.* 2015;9(2):542–550. doi:10.3892/ol.2014.2749
6. Han S, Zong S, Shi Q, et al. Is Ep-CAM expression a diagnostic and prognostic biomarker for colorectal cancer? A systematic meta-analysis. *EBioMedicine.* 2017;20:61–69. doi:10.1016/j.ebiom.2017.05.025
7. Smith NR, Davies PS, Levin TG, et al. Cell adhesion molecule CD166/ALCAM functions within the crypt to orchestrate murine intestinal stem cell homeostasis. *Cell Mol Gastroenterol Hepatol.* 2017;3(3):389–409. doi:10.1016/j.jcmgh.2016.12.010
8. Inaguma S, Lasota J, Wang Z, et al. Expression of ALCAM (CD166) and PD-L1 (CD274) independently predicts shorter survival in malignant pleural mesothelioma. *Hum Pathol.* 2018;71:1–7. doi:10.1016/j.humpath.2017.04.032
9. Bowen MA, Patel DD, Li X, et al. Cloning, mapping, and characterization of activated leukocyte-cell adhesion molecule (ALCAM), a CD6 ligand. *J Exp Med.* 1995;181(6):2213–2220. doi:10.1084/jem.181.6.2213

10. Consuegra-Fernández M, Lin F, Fox DA, Lozano F. Clinical and experimental evidence for targeting CD6 in immune-based disorders. *Autoimmun Rev*. 2018;17(5):493–503. doi:10.1016/j.autrev.2017.12.004
11. Lehmann JM, Riethmüller G, Johnson JP. MUC18, a marker of tumor progression in human melanoma, shows sequence similarity to the neural cell adhesion molecules of the immunoglobulin superfamily. *Proc Natl Acad Sci U S A*. 1989;86(24):9891–9895. doi:10.1073/pnas.86.24.9891
12. Swart GW. Activated leukocyte cell adhesion molecule (CD166/ALCAM): developmental and mechanistic aspects of cell clustering and cell migration. *Eur J Cell Biol*. 2002;81(6):313–321. doi:10.1078/0171-9335-00256
13. Fanali C, Lucchetti D, Farina M, et al. Cancer stem cells in colorectal cancer from pathogenesis to therapy: controversies and perspectives. *World J Gastroenterol*. 2014;20(4):923–942. doi:10.3748/wjg.v20.i4.923
14. Wahab SMR, Islam F, Gopalan V, Lam AK. The identifications and clinical implications of cancer stem cells in colorectal cancer. *Clin Colorectal Cancer*. 2017;16(2):93–102. doi:10.1016/j.clcc.2017.01.011
15. Patriarca C, Macchi RM, Marschner AK, Mellstedt H. Epithelial cell adhesion molecule expression (CD326) in cancer: a short review. *Cancer Treat Rev*. 2012;38(1):68–75. doi:10.1016/j.ctrv.2011.04.002
16. Schnell U, Cirulli V, Giepmans BN. EpCAM: structure and function in health and disease. *Biochim Biophys Acta*. 2013;1828(8):1989–2001. doi:10.1016/j.bbamem.2013.04.018
17. Herreros-Pomares A, Aguilar-Gallardo C, Calabuig-Fariñas S, Sirera R, Jantus-Lewintre E, Camps C. EpCAM duality becomes this molecule in a new Dr. Jekyll and Mr. Hyde tale. *Crit Rev Oncol Hematol*. 2018;126:52–63. doi:10.1016/j.critrevonc.2018.03.006
18. Baeuerle PA, Gires O. EpCAM (CD326) finding its role in cancer [published correction appears in *Br J Cancer*. 2007; 7; 96(9):1491]. *Br J Cancer*. 2007;96(3):417–423. doi:10.1038/sj.bjc.6603494
19. Kalamurthi S, Selvaraj G, Junaid M, Khan A, Gu K, Wei DQ. Cancer Immunoinformatics: a promising era in the development of peptide vaccines for human papillomavirus-induced cervical cancer. *Curr Pharm Des*. 2018;24(32):3791–3817. doi:10.2174/1381612824666181106094133
20. Kar PP, Srivastava A. Immuno-informatics analysis to identify novel vaccine candidates and design of a multi-epitope based vaccine candidate against theileria parasites. *Front Immunol*. 2018;9:2213. doi:10.3389/fimmu.2018.02213
21. Webb B, Sali A. Comparative protein structure modeling using MODELLER. *Curr Protoc Protein Sci*. 2016;86:2.9.1–2.9.37. doi:10.1002/cpps.20
22. Hollingsworth SA, Karplus PA. A fresh look at the Ramachandran plot and the occurrence of standard structures in proteins. *Biomol Concepts*. 2010;1(3–4):271–283. doi:10.1515/bmc.2010.022
23. Fioramonte M, Dos Santos AM, McIlwain S, Noble WS, Franchini KG, Gozzo FC. Analysis of secondary structure in proteins by chemical cross-linking coupled to MS. *Proteomics*. 2012;12(17):2746–2752. doi:10.1002/pmic.201200040
24. Grasso EJ, Sottile AE, Coronel CE. Structural prediction and in silico physicochemical characterization for mouse Caltrin I and bovine caltrin proteins. *Bioinform Biol Insights*. 2016;30(10):225–236.
25. Saha S, Raghava GP. AllgPred: prediction of allergenic proteins and mapping of IgE epitopes. *Nucleic Acids Res*. 2006;34(Web Server issue):W202–W209. doi:10.1093/nar/gkl343
26. 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
27. Ponomarenko J, Bui HH, Li W, 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
28. Zhou P, Jin B, Li H, Huang SY. HPEPDOCK: a web server for blind peptide-protein docking based on a hierarchical algorithm. *Nucleic Acids Res*. 2018;46(W1):W443–W450. doi:10.1093/nar/gky357
29. Smialowski P, Doose G, Torkler P, Kaufmann S, Frishman D. PROSO II—a new method for protein solubility prediction. *FEBS J*. 2012;279(12):2192–2200. doi:10.1111/j.1742-4658.2012.08603.x
30. Petersen B, Petersen TN, Andersen P, Nielsen M, Lundegaard C. A generic method for assignment of reliability scores applied to solvent accessibility predictions. *BMC Struct Biol*. 2009;9:51. doi:10.1186/1472-6807-9-51
31. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res*. 2003;31(13):3406–3415. doi:10.1093/nar/gkg595
32. Sato K, Hamada M, Asai K, Mituyama T. CENTROIDFOLD: a web server for RNA secondary structure prediction. *Nucleic Acids Res*. 2009;37(Web Server issue):W277–W280. doi:10.1093/nar/gkp367
33. Herlyn M, Steplewski Z, Herlyn D, Koprowski H. Colorectal carcinoma-specific antigen: detection by means of monoclonal antibodies. *Proc Natl Acad Sci U S A*. 1979;76(3):1438–1442. doi:10.1073/pnas.76.3.1438
34. Weichert W, Knösel T, Bellach J, Dietel M, Kristiansen G. ALCAM/CD166 is overexpressed in colorectal carcinoma and correlates with shortened patient survival. *J Clin Pathol*. 2004;57(11):1160–1164. doi:10.1136/jcp.2004.016238
35. Chaudry MA, Sales K, Ruf P, Lindhofer H, Winslet MC. EpCAM an immunotherapeutic target for gastrointestinal malignancy: current experience and future challenges. *Br J Cancer*. 2007;96(7):1013–1019. doi:10.1038/sj.bjc.6603505
36. Levin TG, Powell AE, Davies PS, et al. Characterization of the intestinal cancer stem cell marker CD166 in the human and mouse gastrointestinal tract. *Gastroenterology*. 2010;139(6):2072–2082.e5. doi:10.1053/j.gastro.2010.08.053
37. Stoecklein NH, Siegmund A, Scheunemann P, et al. EpCAM expression in squamous cell carcinoma of the esophagus: a potential therapeutic target and prognostic marker. *BMC Cancer*. 2006;6:165. doi:10.1186/1471-2407-6-165
38. Wiiger MT, Gehrken HB, Fodstad Ø, Maeldandsmo GM, Andersson Y. A novel human recombinant single-chain antibody targeting CD166/ALCAM inhibits cancer cell invasion in vitro and in vivo tumour growth. *Cancer Immunol Immunother*. 2010;59(11):1665–1674. doi:10.1007/s00262-010-0892-3

## Drug Design, Development and Therapy

Dovepress

### Publish your work in this journal

Drug Design, Development and Therapy is an international, peer-reviewed open-access journal that spans the spectrum of drug design and development through to clinical applications. Clinical outcomes, patient safety, and programs for the development and effective, safe, and sustained use of medicines are a feature of the journal, which has also

been accepted for indexing on PubMed Central. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/drug-design-development-and-therapy-journal>