



## Designing a Multi-Epitope Antigen for Serodiagnosis of *Strongyloides stercoralis* Based on L3Nie.O1 and IgG Immunoreactive Epitopes

Ahmad Movahedpour<sup>1,2</sup>, Zohreh Mostafavi-Pour<sup>3,4\*</sup>, Bahador Sarkari<sup>5,6\*</sup>, Mortaza Taheri-Anganeh<sup>7</sup>, Navid Nezafat<sup>8</sup>, Amir Savardashtaki<sup>1,9</sup>, and Younes Ghasemi<sup>8</sup>

1. Department of Medical Biotechnology, School of Advanced Medical Sciences and Technologies, Shiraz University of Medical Sciences Shiraz, Iran
2. Student Research Committee, Shiraz University of Medical Sciences, Shiraz, Iran
3. Recombinant Protein Laboratory, Department of Biochemistry, School of Medicine, Shiraz University of Medical Sciences, Shiraz, Iran
4. Autophagy Research Center, Shiraz University of Medical Sciences, Shiraz, Iran
5. Department of Parasitology and Mycology, School of Medicine, Shiraz University of Medical Sciences, Shiraz, Iran
6. Basic Sciences in Infectious Diseases Research Center, Shiraz University of Medical Sciences, Shiraz, Iran
7. Cellular and Molecular Research Center, Cellular and Molecular Medicine Institute, Urmia University of Medical Sciences, Urmia, Iran
8. Pharmaceutical Sciences Research Center, Shiraz University of Medical Sciences, Shiraz, Iran
9. Infertility Research Center, Shiraz University of Medical Sciences, Shiraz, Iran

### Abstract

**Background:** Serological diagnosis of *Strongyloides stercoralis* (*S. stercoralis*) is frequently challenging because of cross-reactivity with other parasitic nematodes. Therefore, it is necessary to introduce novel serological tests with high performance to properly diagnose this neglected parasitic infection. The purpose of the current study was to design a multi-epitope construct for the diagnosis of *S. stercoralis*.

**Methods:** For the purpose of this study, first, highly antigenic segments and potential immunodominant epitopes of *S. stercoralis* were identified from two antigenic proteins, and then all of the selected parts were linked by an appropriate linker. Next, the physico-chemical features of the designed construct were analyzed. Then, tertiary structures of the construct were built and evaluated to find out the best one. Lastly, the amino acid sequence was reverse-translated and optimized for over-expression in *Escherichia coli* (*E. coli*).

**Results:** The bioinformatic evaluation indicated that the designed protein construct could be hydrophilic, thermostable, and acidic and the estimated half-life was more than 10 hr in *E. coli*.

**Conclusion:** According to the results of the study, the designed construct could be used as an efficient antigen in the ELISA system for serological diagnosis of human strongyloidiasis.

*Avicenna J Med Biotech* 2022; 14(2): 114-124

**Keywords:** Antigens, Multi-epitopes, Serological diagnosis, *Strongyloides stercoralis*

### Introduction

*Strongyloides stercoralis* (*S. stercoralis*) is a pathogenic neglected parasitic roundworm that causes human strongyloidiasis. This soil-transmitted helminth has spread throughout the world, especially in tropical and subtropical regions<sup>1</sup>. Currently, *S. stercoralis* infects about 370 million people in the world<sup>1,2</sup>. The mortality rate in patients with hospitalization is 16.7% and in immunocompromised patients is as high as 60-85%<sup>3</sup>. The

*S. stercoralis* life cycle is complex and includes direct, indirect, and auto-infective ones<sup>4</sup>. Regarding the property of the auto-infective cycle, the parasite remains undiagnosed for years and causes a chronic disease that resists treatment<sup>5</sup>. Nearly, half of the patients with strongyloidiasis are asymptomatic, which results in the development of a chronic disease due to the unique life cycle of the parasite<sup>2</sup>.

Hyperinfection and distribution can be seen in high-risk groups, including patients undergoing glucocorticoid therapy, patients with hematologic malignancy, co-infected patients with human T Lymphotropic Virus type I (HTLV-1) and HIV, and those with chronic alcohol abuse<sup>6</sup>. Diagnostic methods for strongyloidiasis can fall into two main categories of serological and non-serological approaches, and often a mixture of both methods is required for accurate and proper diagnosis. Non-serological systems are based on intestinal parasite recognition, molecular methods, and antigen detection, while the serological approaches are based on the finding of anti-*S. stercoralis* antibodies<sup>5,7</sup>.

Due to the parasitic load and low severity of infection, it is difficult to definitively diagnose strongyloidiasis by stool examination, so the sensitivity of this method is very low. Although molecular methods are more sensitive, there are differences in the accuracy of PCR reports<sup>8,9</sup>. A few serological tests are presently commercially accessible to diagnose strongyloidiasis, but possible cross-reactions with other parasites have been considered an obstacle in this case. Therefore, the application of recombinant and synthetic antigens has been suggested as suitable options for diagnosing this disease<sup>10</sup>.

The advent of *in silico* methods has provided a platform for predicting the molecular structures and investigating the receptors and ligands interactions<sup>11,12</sup>. Today, computer science and information technology are utilized to analyze biological information such as DNA decoding and predicting the structure and function of proteins, and accordingly, a new field called science bioinformatics is being developed<sup>11,13-16</sup>. Studies show that bioinformatics can help biological and medical studies by saving time, decreasing the number of experiments, and designing novel biological molecules with desirable features<sup>17-21</sup>. Computational immunology is a newly introduced field, using computer science in immunology. Today, computational immunology can be used in various fields, including identification of potential antigens, estimation of B and T cell epitopes, and designing various forms of vaccine. The use of computational tools can decrease the time and cost in experimental procedures<sup>22</sup>. Considering the advantages mentioned above, the purpose of the present study was computational designing of a multi-epitope antigen originated from two main antigens of *S. stercoralis* to diagnose human strongyloidiasis.

## Materials and Methods

### Study design

In the current study, firstly, the proteins with antigenic properties were selected. Then, using the bioinformatics softwares, the epitopic areas of B cells were selected. The designated epitopes were linked using appropriate amino acid linkers and the 3D model of the ultimate structure was constructed and validated by *in silico* tools. Lastly, for the expression of the designed

structure in the *Escherichia coli* (*E. coli*) host, the amino acid sequence was reverse-translated to the DNA and optimized.

### Estimating antigenicity

EMBOSS server v 6.6.0.0 (available at <http://www.bioinformatics.nl/cgi-bin/emboss/antigenic>) was employed to predict the antigenicity features of the proteins<sup>23</sup>. EMBOSS is an algorithm that contains several programs for various analyses, for instance, alignment of sequences, searching databases, recognition of motifs in proteins, prediction of patterns in nucleotide sequences, and analysis of codon usage<sup>24,25</sup>. Also, it identifies antigenic parts of proteins according to physico-chemical properties, amino acid composition, and epitope redundancy.

In the present work, default parameters (minimum length of the antigenic region=6, output report format=EMBOSS motif) were set up to define the antigenic segments of the selected proteins<sup>23</sup>.

### Linear B cell epitopes

Epitopes are regions of the antigen that are detected by the paratope region of the immunoglobulin. The linear epitope is a continuous amino acid sequence. The conformational epitope consists of amino acids that may not be in the same sequence but reside near each other in a 3D structure. A critical step in designing a multi-epitope protein for vaccine production or building a serological test is the prediction of B cell epitopes. Thus, bioinformatics methods are desirable for prediction of linear B cell epitopes in proteins. In addition, bioinformatics approaches can provide cost and time-saving methods for predicting probable B cell epitopes in a target protein<sup>26</sup>.

The linear antigenic B cell epitopes were identified via BCPred v2.0 (available at <http://ailab.cs.iastate.edu/bcpreds/predict.html>), an innovative way for linear B cell epitopes estimation by the subsequence kernel. The BCPred resource was applied for identification of the B cell epitopes with a size of 20 amino acids, set to 75% specificity. Furthermore, the server uses an innovative approach of a subsequence kernel with 74.57% accuracy. Then, the chosen variants of each antigen are submitted in plain form, and approximately 4 to 7 segments are predicted as the final epitopes<sup>27</sup>.

The linear B cell epitopes are also estimated by ABCPred tools of immune epitope database (available at <http://crdd.osdd.net/raghava/abcpred/>) with default settings. The best predictions by ABCPred occur when 16-mer peptide (ABCP16) is set up<sup>28</sup>. ABCPred server can define B cell epitopes in the proteins through artificial neural networks. Furthermore, ABCPred is the primary online software established according to the recurrent neural network by fixed-length patterns with 65.93% accuracy<sup>27</sup>.

### Designing the construct

The B cells epitopes that were selected by the aforementioned servers were fused *via* appropriate amino

acid linkers. Finally, an amino acid construct was built based on a multi-epitope approach.

#### **Assessment of the physico-chemical features**

Stability and efficiency of a protein structure in a biological system are related to their physico-chemical properties. Numerous physico-chemical features including amino acid sequence, extinction coefficient, instability index, grand average of hydropathicity index (GRAVY), aliphatic index, isoelectric point (pI), and molecular weight can help predicting the stability, activity, and overall properties of the protein<sup>29</sup>. There are various online softwares that can determine the physico-chemical features of a given protein.

ProtParam tool (available at <http://web.expasy.org/ProtParam/>) is an online software at ExPASy server that was utilized in our study to analyze different physico-chemical properties of our designed construct, based on its amino acid sequence<sup>30</sup>.

#### **In silico cloning**

The OPTIMIZER (available at <http://genomes.urv.es/OPTIMIZER/>) was utilized for reverse translation of amino acid to DNA and codons to prepare the optimized DNA sequence of the construct for cloning and expression in *E. coli* host. Moreover, critical parameters of the final optimized nucleotide sequence, including GC content, Codon Adaptation Index (CAI), and Codon Frequency Distribution (CFD) were analyzed by Gen Script rare codon analysis (available at <https://www.genscript.com/tools/rare-codon-analysis>)<sup>13</sup>.

#### **Building 3D model for the construct**

The 3D structures of proteins can help to predict the protein functions and provide useful information for designing the regulator molecules that can regulate the functions of proteins. For 3D structure of our construct, the GalaxyTBM (Template-based modeling, available at <http://galaxy.seoklab.org/>) was applied<sup>31</sup>.

TBM, also termed homology modeling or comparative modeling, usually consists of the following steps<sup>32,33</sup>.

1- adopting the homologous proteins with experimentally defined structures as templates; 2- alignment of the target sequences and templates; 3- building the 3D model, based on the alignment; and 4- refinement of the models. Current methods regularly treat each step distinctly, and the full TBM process can then be set up by merging methods for each of the above steps<sup>31</sup>.

Primarily, templates are chosen by rescoring HH-search results which assign more weights for the score of secondary structure. Among the re-ranked top 20 homologs, various templates are chosen *via* eliminating structural deviations according to mutual TM scores for kernel alignment<sup>34,35</sup>.

PROMALS3D is utilized for multiple sequence alignment of kernel segments removing unaligned termini<sup>29</sup>. Alignments of terminal sequences are attached afterward. Primary model structures are then constructed based on the templates and the alignment *via*

CSA (conformational space annealing) for total optimization of the regions, originated from templates<sup>36,37</sup>.

#### **Tertiary structure validation**

A significant step for discovering the biological procedures at a molecular level is the availability of a structural model of the protein. After processing, preparing, and reviewing the 3D model using online softwares such as ProSA, RAMPAGE, and ERRAT, the constructed model's quality and reliability were assessed<sup>38,39</sup>. A critical online software, used to examine the 3D models of protein for possible errors, is ProSA<sup>40</sup>. This tool is utilized in the diagnosis of empirically identified structures<sup>34,35</sup>, theoretical models<sup>36</sup>, and protein engineering<sup>41</sup>. ProSa z-score can determine the 3D models' quality (<http://prosa.services.came.sbg.ac.at/prosa.php>)<sup>42</sup>. The total quality score assessed by ProSA for a definite input structure is shown in a plot that indicates the scores of all empirically defined protein chains, available in the Protein Data Bank (PDB)<sup>43</sup>. This property connects the score of a specific model to the scores calculated from all defined structures that are saved in PDB. Hard segments of a 3D structure are known by a plot of local quality scores and the same scores are mapped on a display of the tertiary structure, using color codes. ProSA can help for validating 3D protein structures before submitting to PDB<sup>44</sup>.

Moreover, the quality of generated models was evaluated with PROCHECK by Ramachandran plot (available at <http://mordred.bioc.cam.ac.uk/rapper/rampage.php>)<sup>45,46</sup>. Ramachandran Plot (RP) indicates each pair of ( $\phi$ ,  $\psi$ ) angles in 3D structure of a specified protein in a simple logical way<sup>13</sup>. Ramachandran plot has long been employed in every step of crystallography and protein modelling to identify the main chain torsion angles ( $\phi$ ,  $\psi$ )<sup>46</sup>.

Also, the package calculates the Ramachandran angles at the central residue in the stretch of three amino acids which have definite types of flanking residues. The package shows the Ramachandran plots with a thorough analysis of output. This package is accommodated with all the 3D structures accessible in the Protein Data Bank (PDB)<sup>47</sup>.

The ERRAT (available at <http://services.mbi.ucla.edu/ERRAT/>) is a tertiary structure validation algorithm for analyzing the procedure of crystallographic model construction and refining preserved by UC Global Health Institute, University of California, USA<sup>48</sup>. The ERRAT explains the statistics of non-bonded interaction between different atoms, and a score of 50 is commonly appropriate for the tertiary structure assessment<sup>49</sup>. In the current research, six types of noncovalently bonded atom-atom interactions (CC, CN, CO, NN, NO, and OO) in 3D models were measured<sup>48</sup>.

#### **Molecular dynamics simulations**

Molecular Dynamics (MD) simulations were carried out by GROMACS 2019.6 package<sup>50</sup>, using the Amber99sb force field<sup>51</sup>. At first, multiepitope antigen was

placed in a triclinic simulation box (6.20700 nm×18.54500 nm×11.35100 nm) and filled with water molecules using TIP3P model<sup>52</sup>.

To confirm the total charge neutrality of the simulated systems, sufficient amounts of counter-ions were added instead of water molecules keeping physiological salt concentration at 0.15 M. Periodic boundary conditions were applied in three spatial dimensions. An energy minimization process was carried out using the steepest descent method to remove large forces and relax the systems<sup>53</sup>.

The first phase of equilibration involved 500 ps stimulation under the canonical ensemble (NVT) followed by a 500 ps NPT equilibration. The temperature was maintained at 277 °K (refrigerator temperature) and room temperature (298 °K) and the pressure was maintained at 1.0 bar with the Berendsen thermostat and barostat<sup>54</sup>. In total, after two phases of equilibrations, 150 ns molecular dynamics simulations were done to record trajectories using the leap-frog algorithm and solve the equations of motion with a time step of 0.002 ps. The Particle Mesh Ewald (PME) method, Lennard-Jones potential, and LINCS algorithm were used for computation of long range electrostatics, van der Waals interactions, and covalent bond constraints, respectively<sup>55-57</sup>.

## Results

### Sequence retrieval

The amino acid sequences of L3Nie.01 and IgG immunoreactive antigens in FASTA format were obtained from the NCBI database (Table 1).

### Immunoinformatic analyses

The EMBOSS server showed the most antigenic sites, based on the score of each residue (Table 2).

### Defining linear B cell epitopes

According to BCPred and ABCPred, the linear epitopes of each antigen were predicted. Thus, the epitope sequences for each of the antigens, along with their primary and terminal amino acids, the number of amino acids, and the score for each sequence are shown in table 3. Then, using Microsoft Excel software, the epitopes mentioned above were chosen. The final selected segments were again compared with the segments in the CLC Genomics program for the two antigenic proteins; both the conserved regions and the best immunogenic proteins were chosen. In the last step, the multi-epitope protein was designed, using the appropriate linkers between selected regions. The finally selected regions of L3Nie.01 and IgG immunoreactive proteins are presented in table 4.

### Multi-epitope construct

The selected B cell epitope regions of L3Nie.01 and IgG immunoreactive proteins, shown in table 4, were fused via the EAAAK linkers, and the multi-epitope construct was designed.

### Evaluation of the physico-chemical features

The physico-chemical features of the designed structure were calculated by ProtParam. The Molecular Weight (MW) and pI value of the designed protein were 26.8 kDa and 5.05. The pI value showed the acidic nature of the protein. Also, the construct is composed of

Table 1. Prediction of antigenic regions of L3Nie.01 and IgG immunoreactive proteins via EMBOSS server

| Source protein            | Sequences                 | Start and end of aa | Max score |
|---------------------------|---------------------------|---------------------|-----------|
| <b>L3Nie.01</b>           |                           |                     |           |
|                           | KKPIKKPIKKPGPKPIRPIVKPKPK | 55-79               | 71        |
|                           | KNQLKN                    | 29-34               | 29        |
|                           | KNQLKN                    | 42-47               | 42        |
| <b>IgG immunoreactive</b> |                           |                     |           |
|                           | GFGVVEKGDRVYVVKYS         | 194-210             | 207       |
|                           | IGHLAVKGW                 | 153-161             | 155       |
|                           | KAQAYAEVIARLGR            | 117-131             | 123       |
|                           | ATGHFTQLVWKG              | 178-189             | 184       |
|                           | RQFANVLQ                  | 217-224             | 221       |
|                           | EIALYNF                   | 164-170             | 165       |
|                           | YRIAHGAKKLIKSK            | 99-112              | 111       |

Table 2. Linear B cells analysis of L3Nie.01 and IgG immunoreactive by BCPRED and ABCPred

| Protein            | Server  | Number of linear B cell epitopes | Maximum score |
|--------------------|---------|----------------------------------|---------------|
| L3Nie.01           | BCPred  | 33                               | 1             |
| L3Nie.01           | ABCPred | 206                              | >.95          |
| IgG immunoreactive | BCPred  | 111                              | Up to 0.999   |
| IgG immunoreactive | ABCPred | 136                              | Up to 0.86    |

Table 3. Selected B cell epitopes regions of L3Nie.01 and IgG immunoreactive proteins

| Protein                   | Start-end position | Sequence  | Length of sequence |
|---------------------------|--------------------|---|--------------------|
| L3Nie.01                  | 6-65               | ENQDQKQLENQDQKQLENQDQKQNLKNQSENQDQKQNLKNQSENQDQKPKIKPKIKP | 60                 |
|                           | 124-156            | PEEPEGPEEPEGPAGPEEPRDDDDGVDEEDERD                         | 33                 |
| <b>IgG immunoreactive</b> |                    |   |                    |
|                           | 7-40               | GKLIYTYNGNDYDTKEAMEDAIQRDYPDKIFTFG                        | 34                 |
|                           | 91-131             | LFNEQNKYRIAHGAKKLIKSKDLEKKAQAYAEVIARLGRLE                 | 41                 |
|                           | 178-224            | TGHFTQLVWKGTTAGFGVVEKGDRVYVVAKYSPPGNYPRQFAANVLQR          | 47                 |

Table 4. The evaluation of five 3D structures via Ramachandran, ERRAT, and ProSA-Web servers

| 3D models | Ramachandran plot results   | ERRAT overall quality factor | ProSA-Web (Z scores) |
|-----------|---|------------------------------|----------------------|
| Model 1   | Favored region: 93.2%<br>Allowed Region: 6.3%<br>Outlier region: 0.5% | 89.720                       | - 3.45               |
| Model 2   | Favored region: 94.7%<br>Allowed Region: 4.9%<br>Outlier region: 0.5% | 89.163                       | - 3.53               |
| Model 3   | Favored region: 92.7%<br>Allowed Region: 6.8%<br>Outlier region: 0.5% | 95.755                       | - 3.32               |
| Model 4   | Favored region: 92.7%<br>Allowed Region: 6.8%<br>Outlier region: 0.5% | 91.262                       | - 3.55               |
| Model 5   | Favored region: 93.7%<br>Allowed Region: 5.9%<br>Outlier region: 0.5% | 87.745                       | - 3.41               |

237 amino acids. The predictable half-life of the designed construct was 1 hr (mammalian reticulocytes *in vitro*), 30 min (yeast *in vivo*), and >10 hr (*E. coli in vivo*). The Instability Index (II) was found to be 28.04, which showed that the construct could be considered as a stable protein. Aliphatic index and GRAVY value were calculated as 54.47 and -1.378, which indicated the hydrophilic nature of the construct and showed that the protein can interact with water molecules.

**Optimization of codons**

Reverse-translation and codon optimization were done by the online software for over-expression in *E. coli*. The final optimized sequence was analyzed via the GenScript. The results indicated that the CAI of the sequence was 0.94. The CAI of 1.0 is ideal and CAI >0.8

could be considered optimal for expression in the bacterial host (Figure 1A). The GC content of the gene was 53.65% (Figure 1B). A GC content in range of 30% to 70% is a suitable value. All peaks outside this range would have an adverse effect on transcriptional and translational efficiency. Codon Frequency Distribution (CFD) of 91-100, 81-90, 71-80, and 31-40 in the gene was 81, 5, 10, and 1%, respectively (Figure 1C).

**3D modeling, refinement, and validation of the tertiary structure**

The GalaxyTBM created five 3D structures. Then, ERRAT, ProSA, and Ramachandran plot servers were utilized for analyzing the quality of tertiary models and choosing the best one (Figures 2-4). The probable errors of the five models were determined by ProSA with a z-

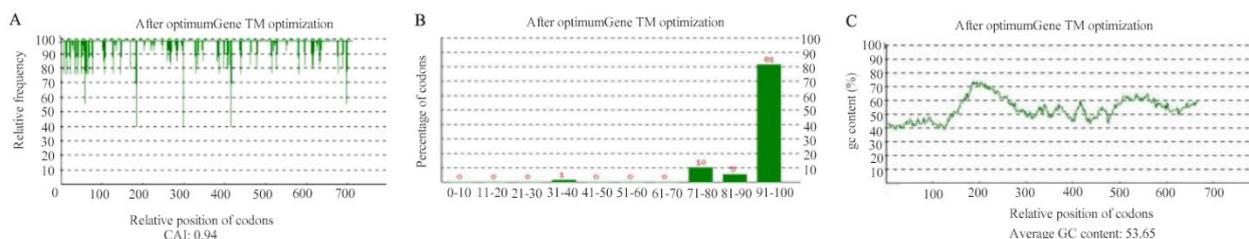


Figure 1. Analysis of three important features of the designed sequence for high-level protein expression in *E. coli* host. A. Codon Adaptation Index (CAI), B. GC content, C. Codon with the Frequency Distribution (CFD).

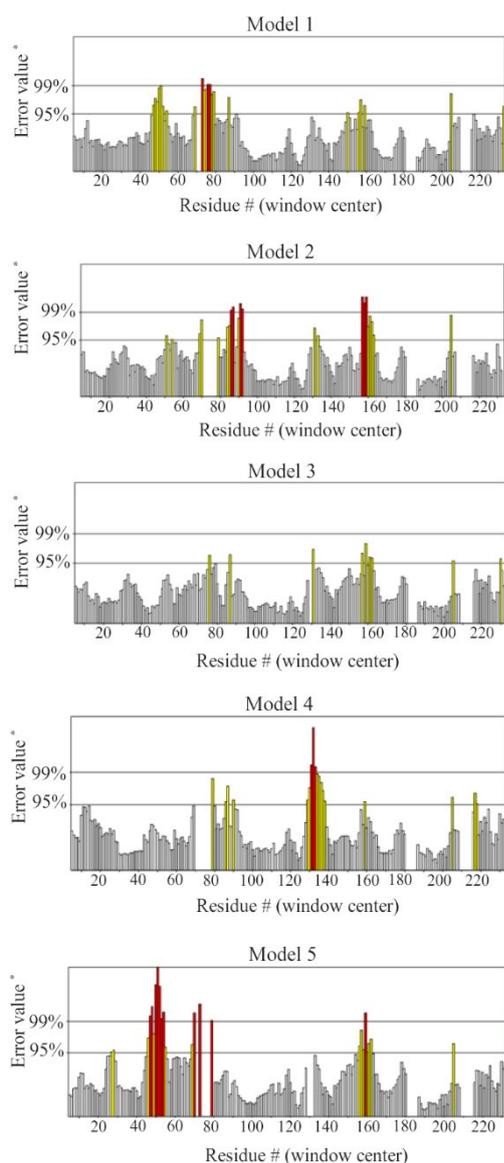


Figure 2. The ERRAT analysis results of models 1, 2, 3, 4, 5.

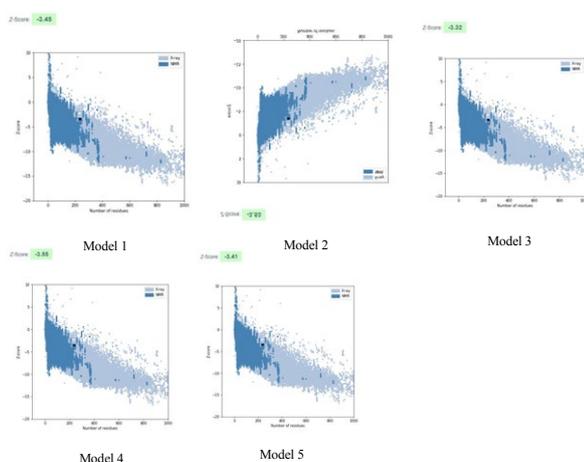


Figure 3. The ProSA scores of models 1, 2, 3, 4, 5.

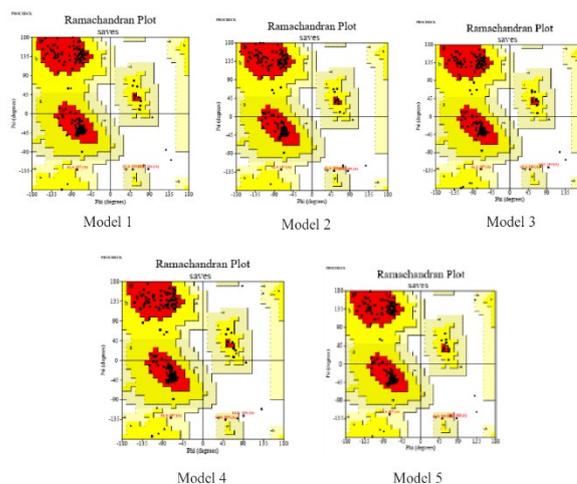


Figure 4. The Ramachandran plot analysis of five models.

score, which indicates similarity of the target 3D structure and experimentally defined structures (Table 5). ProSA computes the total quality score (z-score) of the protein and compares it with experimentally defined structures of proteins in the same size according to PDB results. Positive values show the problem in the created 3D models and the highest negative z-score indicates the highest quality of tertiary structure of the designed construct. The results of the Ramachandran plot and ERRAT analysis for five models are illustrated in table 3. After comparing tertiary models, based on the results of the aforementioned servers, model 3 was considered as the most appropriate model. The ERRAT quality and ProSA z-score were 95.755 and -3.32 for model 3. Furthermore, the Ramachandran plot indicated that 92.7, 6.8, and 0.5% of the amino acid residues were situated in the favored region, allowed region, and outlier region, respectively.

**The results of molecular dynamics simulations**

MD simulation study assists us to understand structure and dynamics of multi-epitope antigen with extreme details. For this purpose, GROMACS analysis tools were used to analyze the trajectories in terms of Root Mean Square Deviation (RMSD), Root Mean Square Fluctuation (RMSF), Radius of gyration (Rg), and secondary structure.

**Root mean square deviation (RMSD)**

To confirm the stability of the simulations, the RMSDs of multi-epitope antigen at room and refrigerator temperatures were analyzed during 150 ns simulation time. As seen in figure 5, RMSD plot of multi-epitope at 298 K increased gradually and reached equilibrium around 75 ns and after that remained stable with low RMSD fluctuations. High value of RMSD fluctuations were seen at 277 °K but finally it reached equilibrium around 120 ns and thereafter remained stable with low RMSD fluctuations.

Table 5. The evaluation of five 3D structures via Ramachandran, ERRAT, and ProSA-Web servers

| 3D models      | Ramachandran plot results   | ERRAT overall quality factor | ProSA-web (Z scores) |
|----------------|---|------------------------------|----------------------|
| <b>Model 1</b> | Favored region: 93.2%<br>Allowed Region: 6.3%<br>Outlier region: 0.5% | 89.720                       | - 3.45               |
| <b>Model 2</b> | Favored region: 94.7%<br>Allowed Region: 4.9%<br>Outlier region: 0.5% | 89.163                       | - 3.53               |
| <b>Model 3</b> | Favored region: 92.7%<br>Allowed Region: 6.8%<br>Outlier region: 0.5% | 95.755                       | - 3.32               |
| <b>Model 4</b> | Favored region: 92.7%<br>Allowed Region: 6.8%<br>Outlier region: 0.5% | 91.262                       | - 3.55               |
| <b>Model 5</b> | Favored region: 93.7%<br>Allowed Region: 5.9%<br>Outlier region: 0.5% | 87.745                       | - 3.41               |

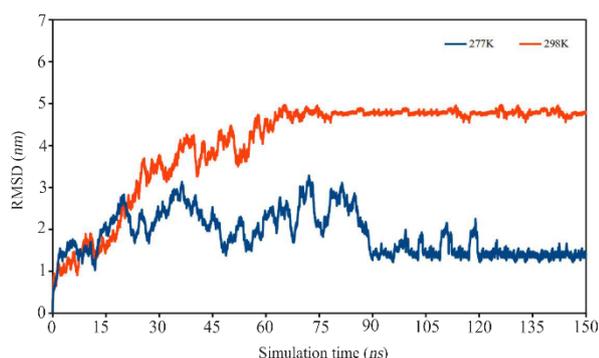


Figure 5. RMSD of multi-epitope antigen at 277 K and 298 K during total simulation time.

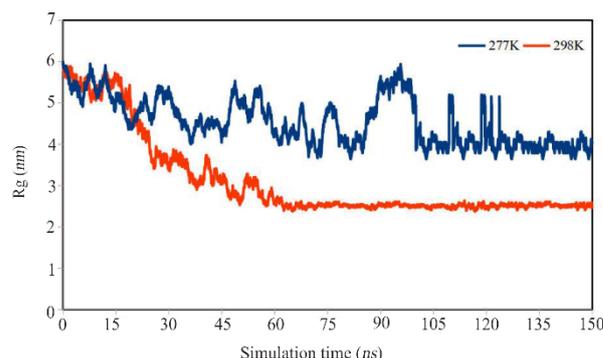


Figure 6. Time dependence changes of  $R_g$  of multi-epitope antigen during the simulation at 277 °K and 298 °K.

**Radius of gyration ( $R_g$ )**

The compactness of multi-epitope antigen structure during the MD simulation time was assessed by analysis of the  $R_g$ . As seen in figure 6, the radius of gyration of multi-epitope antigen at 298 °K decreased gradually which represents the compact structure of multi-epitope antigen. Finally, it reached to equilibrium around 75 ns and after that remained stable with low fluctuations. High value of  $R_g$  fluctuations was seen at 277 °K showing the somewhat unfolded multi-epitope antigen at 277 °K but it reached equilibrium around 120 ns and thereafter its fluctuation was unnoticeable.

**Root mean square fluctuation (RMSF)**

The analysis of RMSFs of multi-epitope antigen can be used as a reference to estimate flexibility of the residues (Figure 7). The results of the RMSF analysis of multi-epitope antigen showed that the highly fluctuating areas appeared in the random coil structures. Moreover,

some residues have a higher RMSF value at 277 °K than 298 °K, showing the partially unfolded multi-epitope antigen at 277 °K.

**Secondary structure**

The secondary structure of the multi-epitope antigen was calculated with the DSSP code of GROMACS. Figure 8 shows the  $\alpha$ -helix,  $\beta$ -sheet, and other secondary structures of the multi-epitope antigen in two simulations. As seen in this figure, at two temperatures, the main secondary structures were rather stable and major conformational change of multi-epitope antigen was a simple motion during the 150 ns simulation time.

**Discussion**

Strongyloidiasis is usually difficult to diagnose as the larval output is irregular and the parasite load is low. Serological methods, based on the detection of anti-*S. stercoralis* antibodies, are believed to be more applicable.

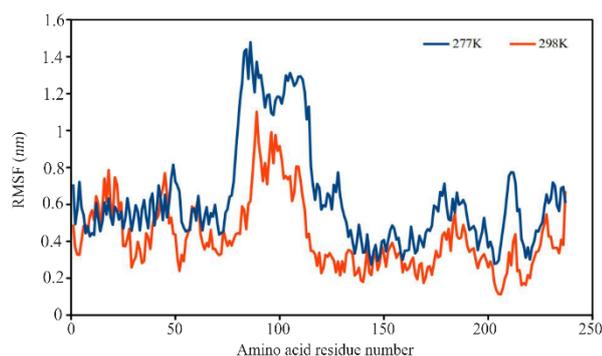


Figure 7. RMSF of residues of multi-epitope antigen from time-averaged positions during the last 30 ns at 277 °K and 298 °K.

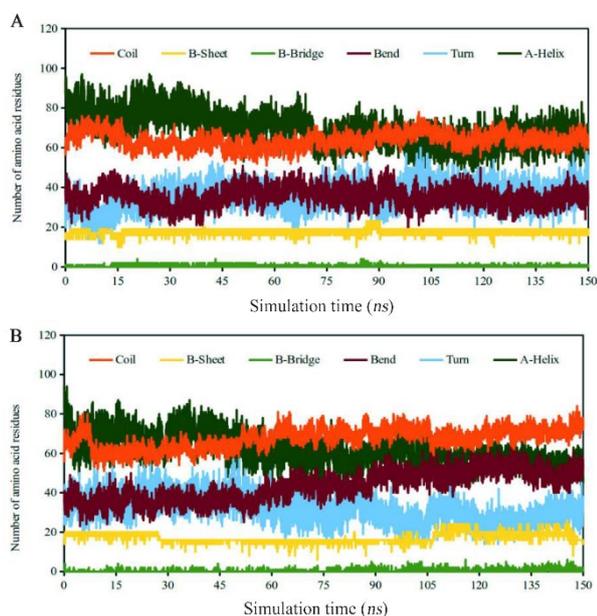


Figure 8. Variation of the secondary structure versus time for the multi-epitope antigen at a) 277 °K and b) 298 °K.

However, the cross-reactivity with other parasites antigens decreases the performance of serological tests for the diagnosis of strongyloidiasis<sup>58</sup>. Today, ELISA and related methods are commonly used serological tests due to low costs, easy use, less variable results, and ease of automation<sup>59</sup>. Using recombinant DNA technology, all protein antigens could be easily produced and applied for designing an ELISA kit<sup>60-62</sup>. Diagnostic ELISA kits based on selected epitopes of B cells are novel and efficient approaches for serological diagnosis of infectious diseases<sup>63,64</sup>.

Nowadays, bioinformatics tools and approaches help to design protein molecules with desirable and appropriate features<sup>65-68</sup>. *In silico* methods have revolutionized the field of medical biotechnology by presenting a platform *via* predicting molecular structures and the molecular interactions between the designed therapeutics and

their potential target receptors<sup>69</sup>. Thus, online bioinformatics tools were used to design a multi-epitope protein for the diagnosis of strongyloidiasis, based on B cell linear epitopes of L3Nie.01 and IgG immunoreactive antigens.

Bisoffi *et al* showed that the L3Nie.01 antigen has no cross-reactivity with other parasite antigens<sup>70</sup>. Furthermore, it has been shown that the diagnostic sensitivity of recombinant L3Nie.01 antigen is in the range of 72.3 to 78.9%, and its specificity ranges from 85.1 to 93.6%<sup>58</sup>. Therefore, L3Nie.01-ELISA has been suggested as a serological method for assessing and determining treatment outcomes in endemic areas and has a significant role in monitoring public health interventions against strongyloidiasis<sup>71</sup>.

Linker design is a very important and fundamental principle in the construction of *de novo* designed multi-domain proteins. Controlling the distance and orientation of domains in order to maximize the function of the desired protein is a difficult task in protein engineering. In 2001, Arai *et al* engineered and designed a suitable linker that aimed to provide logical consistency, flexibility, and spacing between the domains of a multifunctional fusion protein. It is shown that EAAAK helix binders can very effectively separate fusion protein domains, and that distances between domains can be controlled by changing EAAAK binding motifs. Therefore, the high potential of helical binders was indicated to be useful in protein engineering and design of multifunctional fusion proteins<sup>72</sup>. In the present study, two servers were utilized for enhancing the prediction accuracy of linear epitopes with 20- and 16-mer B cell epitopes. Then, Microsoft Excel software was used to detect the most common and overlapping regions of each chosen protein. The selected regions of linear B cell epitopes were fused *via* EAAAK linker and a novel multi-epitope construct was built. To fix the distance between functional domains and intensify the expression of the final product in the bacterial host, EAAAK helical linker with the support of Glu-Lys salt bridge created a coherent structure<sup>73,74</sup>. The absence of the linkers between the epitopes leads to the production of connected epitopes (neo-epitopes) and disturbance of protein function<sup>75</sup>. Farhani *et al* utilized EAAAK as a rigid linker for binding HTL epitopes in their construct and designed a stable multi-epitope vaccine against pathogenic *Shigella* spp<sup>76</sup>. Furthermore, Vakili *et al* used EAAAK peptide linker for binding multi-epitope construct of *Leishmania infantum* antigens to peptide adjuvant in N- and C- terminals which results in making a stable structure<sup>74</sup>.

The DNA sequence of the construct was optimized based on codon usage of *E. coli* host. Analysis of optimization parameters including CAI, CFD, and GC contents by GenScript server showed that the optimized nucleotide sequence was suitable for overexpression in *E. coli* host.

The physico-chemical features of the designed construct were also predicted *via* the bioinformatics tool.

The construct was hydrophilic based on the GRAVY score -1.378, which demonstrates its easy interaction with water molecules. The pI value 5.05 showed the acidic feature of the designed protein. Besides, the high aliphatic index 54.47 showed that the designed protein could be more thermostable, and also, the value of the instability index 28.04 revealed that the designed construct could be stable. The 3D model building and validation can open a horizon to identify the structure of the proteins. For this reason, five 3D models of the construct were built by GalaxyTBM. Moreover, ProSA and ER-RAT were utilized to evaluate and select the best model.

### Conclusion

In the present study, immunodominant linear B cell epitopes from two antigens of *S. stercoralis*, were predicted and linked together *via* proper linkers. The study of physico-chemical structures of designed protein construct showed that the structure is thermostable, hydrophilic, and acidic with a half-life of more than 10 hr in *E. coli*. Finally, it was found that the designed chimeric protein can express greatly in *E. coli*. The designed protein can be used in the construction of a high-performance serological test for the diagnosis of human strongyloidiasis.

### Acknowledgement

This study was financially supported by the office of vice-chancellor for research of Shiraz University of Medical Sciences (Grant No. 18780). The results described in this paper were part of PhD thesis of Ahmad Movahedpour.

### Conflict of Interest

The authors declare that they have no conflict of interest.

### References

- Casavechia MTG, Lonardoni MVC, Venazzi EAS, Campanerut-Sá PAZ, da Costa Benalia HR, Mattiello MF, et al. Prevalence and predictors associated with intestinal infections by protozoa and helminths in southern Brazil. *Parasitol Res* 2016;115(6):2321-9.
- Krolewiecki A, Nutman TB. Strongyloidiasis: a neglected tropical disease. *Infect Dis Clin North Am* 2019;33(1):135-51.
- Iriemenam NC, Sanyaolu AO, Oyibo WA, Fagbenro-Beyioku AF. Strongyloides stercoralis and the immune response. *Parasitol Int* 2010;59(1):9-14.
- Lundberg L, Pinkham C, de la Fuente C, Brahms A, Shafagati N, Wagstaff KM, et al. Selective inhibitor of nuclear export (SINE) compounds alter New World alphavirus capsid localization and reduce viral replication in mammalian cells. *PLoS Negl Trop Dis* 2016;10(11):e0005122.
- Balachandra D, Ahmad H, Arifin N, Noordin R. Direct detection of Strongyloides infection via molecular and antigen detection methods. *Eur J Clin Microbiol Infect Dis* 2021;40(1):27-37.
- Mishra S, Patnayak R, Panda S, Jena A. Strongyloides stercoralis hyperinfection in a patient with acute lymphoblastic leukemia. *J Postgrad Med* 2021;67(2):113-4.
- Arifin N, Hanafiah KM, Ahmad H, Noordin R. Serodiagnosis of Strongyloides stercoralis infection. *J Microbiol Immunol Infect* 2019;52(3):371-8.
- Repetto SA, Ruybal P, Solana ME, López C, Berini CA, Soto CDA, et al. Comparison between PCR and larvae visualization methods for diagnosis of Strongyloides stercoralis out of endemic area: a proposed algorithm. *Acta Trop* 2016;157:169-77.
- Olsen A, van Lieshout L, Marti H, Polderman T, Polman K, Steinmann P, et al. Strongyloidiasis—the most neglected of the neglected tropical diseases? *Trans R Soc Trop Med Hyg* 2009;103(10):967-72.
- Krolewiecki AJ, Ramanathan R, Fink V, McAuliffe I, Cajal SP, Won K, et al. Improved diagnosis of Strongyloides stercoralis using recombinant antigen-based serologies in a community-wide study in northern Argentina. *Clin Vaccine Immunol* 2010;17(10):1624-30.
- Mohabatkar H, Behbahani M, Moradi M. A consie in silico prediction report of a potential prion like domain in SARS-COV-2 polyprotein. *J Microbiol Biotechnol Food Sci* 2021:e4813-e.
- Haghighi O. In silico study of the structure and ligand preference of pyruvate kinases from Cyanobacterium synechocystis sp. PCC 6803. *Appl Biochem Biotechnol* 2021;193(11):3651-71.
- Mehrpour K, Mirzaei SA, Savardashtaki A, Nezafat N, Ghasemi Y. Designing an HCV diagnostic kit for common genotypes of the virus in Iran based on conserved regions of core, NS3-protease, NS4A/B, and NS5A/B antigens: an in silico approach. *Biologia* 2021;76(1):281-96.
- Taheri-Anganeh M, Khatami SH, Jamali Z, Savardashtaki A, Ghasemi Y, Mostafavi-Pour Z. In silico analysis of suitable signal peptides for secretion of a recombinant alcohol dehydrogenase with a key role in atorvastatin enzymatic synthesis. *Mol Biol Res Commun* 2019;8(1):17-26.
- Behbahani M, Moradi M, Mohabatkar H. In silico design of a multi-epitope peptide construct as a potential vaccine candidate for Influenza A based on neuraminidase protein. *In silico Pharmacol* 2021;9(1):36.
- Sadeghi M, Moradi M, Madanchi H, Johari B. In silico study of garlic (*Allium sativum* L.)-derived compounds molecular interactions with  $\alpha$ -glucosidase. *In Silico Pharmacol* 2021;9(1):11.
- Mohammadi S, Mostafavi-Pour Z, Ghasemi Y, Barazesh M, Pour SK, Atapour A, et al. In silico analysis of different signal peptides for the excretory production of recombinant NS3-GP96 fusion protein in Escherichia coli. *Int J Peptide Research Therapeutics* 2019;25(4):1279-90.
- Mousavi P, Mostafavi-Pour Z, Morowvat MH, Nezafat N, Zamani M, Berenjian A, et al. In silico analysis of several signal peptides for the excretory production of reteplase in Escherichia coli. *Current Proteomics* 2017;14(4):326-35.
- Jamali Z, Taheri-Anganeh M, Entezam M. Prediction of potential deleterious nonsynonymous single nucleotide polymorphisms of HIF1A gene: A computational approach. *Comput Biol Chem* 2020;88:107354.

20. Haghghi O, Moradi M. In silico study of the structure and ligand interactions of alcohol dehydrogenase from *Cyanobacterium Synechocystis* sp. PCC 6803 as a key enzyme for biofuel production. *Appl Biochem Biotechnol* 2020; 192(4):1346-67.
21. Nabati F, Moradi M, Mohabatkar H. In silico analyzing the molecular interactions of plant-derived inhibitors against E6AP, p53, and c-Myc binding sites of HPV type 16 E6 oncoprotein. *Mol Biol Res Commun* 2020;9(2):71-82.
22. Rapin N, Lund O, Bernaschi M, Castiglione F. Computational immunology meets bioinformatics: the use of prediction tools for molecular binding in the simulation of the immune system. *PloS One* 2010;5(4):e9862.
23. Rice P, Longden I, Bleasby A. EMBOSS: the European molecular biology open software suite. *Trends Genet* 2000;16(6):276-7.
24. Lamprecht A-L, Naujokat S, Margaria T, Steffen B. Semantics-based composition of EMBOSS services. *J Biomed Semantics* 2011;2(Suppl 1):S5.
25. Goldenberg D, Pasmanik-Chor M, Pirak M, Kass N, Lublin A, Yehekel A, et al. Genetic and antigenic characterization of sigma C protein from avian reovirus. *Avian Pathol* 2010;39(3):189-99.
26. Tehrani SS, Jahangiri A, Taheri-Anganeh M, Maghsoudi H, Khalili S, Fana SE, et al. Designing an outer membrane protein (Omp-W) based vaccine for immunization against *Vibrio* and *Salmonella*: an in silico approach. *Recent Patents on Biotechnology*. 2020;14(4):312-24.
27. Saha S, Raghava GPS, editors. BcePred: prediction of continuous B-cell epitopes in antigenic sequences using physico-chemical properties. *International Conference on Artificial Immune Systems*; 2004: Springer.
28. Saha S, Raghava GPS. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins* 2006;65(1):40-8.
29. Garg VK, Avashthi H, Tiwari A, Jain PA, Ramkete PW, Kayastha AM, et al. MFPPi-multi FASTA ProtParam interface. *Bioinformation* 2016;12(2):74.
30. Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD, Bairoch A. Protein identification and analysis tools on the ExPASy server. *The proteomics protocols handbook*. 2005:571-607.
31. Ko J, Park H, Seok C. GalaxyTBM: template-based modeling by building a reliable core and refining unreliable local regions. *BMC Bioinform* 2012;13(1):198.
32. Pearce R, Zhang Y. Toward the solution of the protein-structure prediction problem. *J Biol Chem* 2021;297(1):100870.
33. Sánchez R, Pieper U, Melo F, Eswar N, Martí-Renom MA, Madhusudhan M, et al. Protein structure modeling for structural genomics. *Nat Struct Biol* 2000;7(Suppl): 986-90.
34. Söding J. Protein homology detection by HMM-HMM comparison. *Bioinformatics* 2005;21(7):951-60.
35. Zhang Y, Skolnick J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res* 2005;33(7):2302-9.
36. Joo K, Lee J, Seo JH, Lee K, Kim BG, Lee J. All-atom chain-building by optimizing MODELLER energy function using conformational space annealing. *Protein* 2009; 75(4):1010-23.
37. Ko J, Park H, Heo L, Seok C. GalaxyWEB server for protein structure prediction and refinement. *Nucleic Acids Res* 2012;40(Web Server issue):W294-W7.
38. Haghghi O, Davaeifar S, Zahiri HS, Maleki H, Noghabi KA. Homology modeling and molecular docking studies of glutamate dehydrogenase (GDH) from *Cyanobacterium Synechocystis* sp. PCC 6803. *Int J Peptide Research Therapeutics* 2020;26(2):783-93.
39. Gharbavi M, Johari B, Rismani E, Mousazadeh N, Taromchi AH, Sharafi A. NANOG decoy oligodeoxynucleotide-encapsulated niosomes nanocarriers: a promising approach to suppress the metastatic properties of U87 human glioblastoma multiforme cells. *ACS Chem Neurosci* 2020;11(24):4499-515.
40. Sippl MJ. Recognition of errors in three-dimensional structures of proteins. *Proteins* 1993;17(4):355-62.
41. Beissenhirtz MK, Scheller FW, Viezzoli MS, Lisdat F. Engineered superoxide dismutase monomers for superoxide biosensor applications. *Anal Chem* 2006;78(3):928-35.
42. Randhawa V, Jamwal R. The synthetic community as they present a variety of synthetic challenges. *Int Res J Biochem Bioinform* 2011;1(4):95-102.
43. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The protein data bank. *Nucleic Acids Res* 2000;28(1):235-42.
44. 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-W10.
45. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *J Applied Crystallography* 1993;26(2):283-91.
46. Lovell SC, Davis IW, Arendall III WB, De Bakker PI, Word JM, Prisant MG, et al. Structure validation by Ca geometry:  $\phi$ ,  $\psi$  and C $\beta$  deviation. *Proteins* 2003;50(3): 437-50.
47. Sheik S, Sundararajan P, Hussain A, Sekar K. Ramachandran plot on the web. *Bioinformatics* 2002;18(11): 1548-9.
48. Colovos C, Yeates TO. Verification of protein structures: patterns of nonbonded atomic interactions. *Prot Sci* 1993;2(9):1511-9.
49. Zafar M, Khan H, Rauf A, Khan A, Lodhi MA. In silico study of alkaloids as  $\alpha$ -glucosidase inhibitors: Hope for the discovery of effective lead compounds. *Front Endocrinol (Lausanne)* 2016;7:153.
50. Junghans C, Phillips JL, Wall ME. Gromacs IC Tutorial. Los Alamos National Lab.(LANL), Los Alamos, NM (United States); 2014.
51. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J Chem Theory Comput* 2015;11(8):3696-713.

52. Mark P, Nilsson L. Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. *J Physical Chemistry A* 2001;105(43):9954-60.
53. Hess B, Kutzner C, Van Der Spoel D, Lindahl E. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory Comput* 2008;4(3):435-47.
54. Berendsen HJ, Postma Jv, van Gunsteren WF, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. *J Chem Phys* 1984;81(8):3684-90.
55. Darden T, York D, Pedersen L. Particle mesh Ewald: An N log (N) method for Ewald sums in large systems. *J Chem Phys* 1993;98(12):10089-92.
56. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. *J Chem Phys* 1995;103(19):8577-93.
57. Hess B, Bekker H, Berendsen HJ, Fraaije JG. LINCS: a linear constraint solver for molecular simulations. *J Computational Chemistry* 1997;18(12):1463-72.
58. Fradejas I, Herrero-Martínez J, Lizasoain M, Rodríguez de las Parras E, Pérez-Ayala A. Comparative study of two commercial tests for *Strongyloides stercoralis* serologic diagnosis. *Trans R Soc Trop Med Hyg* 2018;112(12):561-7.
59. Khatami SH, Taheri-Anganeh M, Movahedpour A, Savardashtaki A, Ramezani A, Sarkari B, et al. Serodiagnosis of human cystic echinococcosis based on recombinant antigens B8/1 and B8/2 of *Echinococcus granulosus*. *J Immunoassay Immunochem* 2020;41(6):1010-20.
60. Savardashtaki A, Mostafavi-Pour Z, Arianfar F, Sarkari B. Comparison of the utility of recombinant B8/2 subunit of the antigen B, native antigen, and a commercial ELISA kit in the diagnosis of human cystic echinococcosis. *Iran Biomed J* 2019;23(4):246-52.
61. Savardashtaki A, Sarkari B, Arianfar F, Mostafavi-Pour Z. Immunodiagnostic value of *Echinococcus granulosus* recombinant B8/1 subunit of antigen B. *Iran J Immunol* 2017;14(2):111-22.
62. Hemmati M, Seghatoleslam A, Rasti M, Ebadat S, Mosavari N, Habibagahi M, et al. Expression and purification of recombinant *Mycobacterium tuberculosis* (TB) antigens, ESAT-6, CFP-10 and ESAT-6/CFP-10 and their diagnosis potential for detection of TB patients. *Iran Red Crescent Med J* 2011;13(8):556.
63. Lu Y, Li Z, Teng H, Xu H, Qi S, Gu D, et al. Chimeric peptide constructs comprising linear B-cell epitopes: application to the serodiagnosis of infectious diseases. *Scientific Reports* 2015;5(1):1-11.
64. Vale DL, Lage DP, Machado AS, Freitas CS, de Oliveira D, Galvani NC, et al. Serodiagnosis of canine leishmaniasis using a novel recombinant chimeric protein constructed with distinct B-cell epitopes from antigenic *Leishmania infantum* proteins. *Vet Parasitol* 2021;296:109513.
65. Tehrani SS, Goodarzi G, Naghizadeh M, Khatami SH, Movahedpour A, Abbasi A, et al. Suitable signal peptides for secretory production of recombinant granulocyte colony stimulating factor in *Escherichia coli*. *Recent Pat Biotechnol* 2020;14(4):269-82.
66. Khatami SH, Taheri-Anganeh M, Arianfar F, Savardashtaki A, Sarkari B, Ghasemi Y, et al. Analyzing signal peptides for secretory production of recombinant diagnostic antigen B8/1 from *Echinococcus granulosus*: an in silico approach. *Mol Biol Res Commun* 2020;9(1):1-10.
67. Vafadar A, Taheri-Anganeh M, Movahedpour A, Jamali Z, Irajie C, Ghasemi Y, et al. In silico design and evaluation of scFv-CdtB as a novel immunotoxin for breast cancer treatment. *Int J Cancer Manag* 2020;13(1):e96094.
68. Taheri-Anganeh M, Amiri A, Movahedpour A, Khatami SH, Ghasemi Y, Savardashtaki A, et al. In silico evaluation of PLAC1-fliC as a chimeric vaccine against breast cancer. *Iran Biomed J* 2020;24(3):173.
69. Rahmati M, Johari B, Kadivar M, Rismani E, Mortazavi Y. Suppressing the metastatic properties of the breast cancer cells using STAT3 decoy oligodeoxynucleotides: a promising approach for eradication of cancer cells by differentiation therapy. *J Cell Physiol* 2020;235(6):5429-44.
70. Bisoffi Z, Buonfrate D, Sequi M, Mejia R, Cimino RO, Krolewiecki AJ, et al. Diagnostic accuracy of five serologic tests for *Strongyloides stercoralis* infection. *PLoS Negl Trop Dis* 2014;8(1):e2640.
71. Vargas P, Krolewiecki AJ, Echazú A, Juárez M, Cajal P, Gil JF, et al. Serologic monitoring of public health interventions against *Strongyloides stercoralis*. *Am J Trop Med Hyg* 2017;97(1):166-72.
72. Arai R, Ueda H, Kitayama A, Kamiya N, Nagamune T. Design of the linkers which effectively separate domains of a bifunctional fusion protein. *Protein Eng* 2001;14(8):529-32.
73. Partovi Nasr M, Motalebi M, Zamani MR, Jourabchi E. In silico analysis and expression of osmotin-EAAAK-LTP fused protein. *J Genetic Resources* 2020;6(1):41-8.
74. Vakili B, Eslami M, Hatam GR, Zare B, Erfani N, Nezafat N, et al. Immunoinformatics-aided design of a potential multi-epitope peptide vaccine against *Leishmania infantum*. *Int J Biol Macromol* 2018;120:1127-39.
75. Khan M, Khan S, Ali A, Akbar H, Sayaf AM, Khan A, et al. Immunoinformatics approaches to explore *Helicobacter pylori* proteome (Virulence Factors) to design B and T cell multi-epitope subunit vaccine. *Scientific Reports* 2019;9(1):1-13.
76. Farhani I, Nezafat N, Mahmoodi S. Designing a novel multi-epitope peptide vaccine against pathogenic *Shigella* spp. based immunoinformatics approaches. *Int J Peptide Res Therap* 2019;25(2):541-53.