Journal of Medical Microbiology and Infectious Diseases

JoMMID

ISSN: 2345-5349
eISSN: 2345-5330

Structure Evaluation of IroN for Designing a Vaccine against *Escherichia Coli*, an *In Silico* Approach

Fateme Sefid[1, 2], Roghayyeh Baghban[3], Zahra Payandeh[4]*, Bahman Khalesi[5], Mohammad Mahmoudi Gomari[6]

[1]Department of Medical Genetics, Shahid Sadoughi University of Medical Science, Yazd, Iran; [2]Department of Biology, Science and Art University, Yazd, Iran; [3]Department of Medical Biotechnology, Faculty of Advanced Medical Science, Tabriz University of Medical Science, Tabriz, Iran; [4]Immunology Research Center, Tabriz University of Medical Sciences, Tabriz, Iran; [5]Department of Research and Production of Poultry Viral Vaccine, Razi Vaccine and Serum Research Institute, Agricultural Research Education and Extention Organization (AREEO), Karaj, Iran; [6]Department of Medical Biotechnology, Faculty of Paramedical, Iran University of Medical Science, Tehran, Iran

ARTICLE INFO

Original Article

Keywords: Urinary Tract Infections, Vaccine, Iron Receptor, Bioinformatics, OMP

Received: Jul. 21, 2018
Received in revised form: Sep. 20, 2019
Accepted: Sep. 24, 2019
DOI: 10.29252/JoMMID.7.4.93

***Correspondence**
Email: zpayandeh58@yahoo.com
Tel: +98 41 33371328
Fax: +98 41 33371311

ABSTRACT

Introduction: Some strains of *Escherichia Coli*, including intestinal pathogenic strains, commensal strains, and extra intestinal pathogenic *E. coli* (ExPEC) have a significant impact on human health status. A standard vaccine designed based on conserved epitopes can stimulate a protective immune response against these pathogens. Additionally, enhanced expression at the infection site as a pathogenesis factor in disease is crucial for an ideal vaccine candidate. The IroN protein plays a role in severe infections of *E. coli*. Hence, this protein will assist in developing the novel and more efficient treatments for *E. coli* related infections. A better understanding of protein tertiary structure can help to percept their functions and also their interactions with other molecules. There is a growing interest in using bioinformatics tool to make accurate predictions about the functional, immunological, and biochemical features of target antigens. **Method:** Herein, we aimed to predict the structure of the IroN protein upon its folding and determine their immunological properties. **Results**: In the present study, using bioinformatics analyses, we identified the highly antigenic regions of IroN protein. Our designed vaccine candidate had the highest immunological properties and folded into a typical beta-barrel structure. **Conclusion**: The approach of assigning structural and immunological properties of the target antigen to design the vaccine candidate could be deployed as an efficient strategy to circumvent the challenges ahead of empirical methods without dealing with ethical concerns of animal usage and human participants. Although the obtained results are promising, further experimental studies could bring about more insights on the efficiency of the designed vaccine.

INTRODUCTION

Some strains of *Escherichia Coli* that are biologically significant to humans health include intestinal pathogenic strains, commensal strains, and extraintestinal pathogenic *E. coli* (ExPEC) [1]. The ExPEC bacteria as the normal intestinal flora do not cause gastroenteritis in humans. Nevertheless, a wide variety of infections occur after their entry into extraintestinal sites [2]. The ExPEC strains cause a wide range of urinary tract infections. Hence, prevention of infection associated with ExPEC, as an essential cause of severe sepsis, is needed from both economic and medical point of view. This demand has become more critical due to antibiotic resistance, especially to trimethoprim-sulfamethoxazole, which has led to an increase in *E. coli* urinary tract infections in the United States [3]. Moreover, the occurrence rate of severe extraintestinal infection related to *E. coli* increases with age. Therefore, a combination of increased antimicrobial resistance and infection rates will probably make the control of extraintestinal *E. coli* infections more challenging in the future [3].

A standard vaccine candidate designed based on conserved epitopes can induce a protective immune reaction but requires to be displayed on the surface of bacteria. Also, enhanced expression at the infection site is crucial for an ideal vaccine candidate [4].

IroN is a novel gene that is reported to have enhanced expression in human blood, urine, and ascites [5]. Expression of *IroN* gene was revealed to be regulated by iron ions (Fe^{2+}); a homology search based on possible peptide sequence proposed that Iron was a siderophore receptor.

[DOI: 10.29252/JoMMID.7.4.93]

Epidemiological data from some studies have shown an increased incidence of IroN expression among urinary tract infections. This evidence proposes that IroN, especially in urinary tract infection, operates as a receptor of siderophore and is a critical factor [5].

Good knowledge of the role of IroN in severe *E. coli* infections and also the nature of IroN will assist in developing novel and more effective medications for *E. coli* infections. A better understanding of the tertiary structure of proteins might help to predict their functions and interactions with other molecules. Furthermore, rational engineering of proteins depends on the comprehension of their 3D structures. The 3D protein structure information can be useful in the design of vaccines, drugs, and predictions of conformational epitopes. The enormous number of identified protein sequences against the small number of structural annotations emphasizes the requirement for identification of tertiary protein structure. Protein structure determination by the experimental method remains a significant challenge due to its high failure rates [6].

The determination of 3D protein structures using experimental approaches is expensive and time-consuming. Besides, purification and crystallization of outer membrane proteins are challenging. Today, bioinformatics tools can assist biologists in different disciplines, *e.g.*, prediction of 3D protein structure. Some algorithms and approaches are accessible for the prediction of protein structures. The most accurate approach for protein structure prediction is homology modeling. This approach relies on the idea that the protein with similar amino acid sequences could fold into similar structures. Crystal structures of some IroN homologous proteins, including FepA, FptA, FpvA, and FhuA in some pathogens, are available in the PDB server [7].

Nevertheless, the 3D structure of IroN has not been identified so far. Herein, we aimed to predict the structure of the IroN protein upon its folding. The 3D structure of these proteins could pave the way for researchers to determine their immunological properties. Knowledge of the immunological properties of these proteins could be used to design vaccine candidates containing the regions with the highest density of immune-dominant epitopes. Different regions could be connected using linkers with high flexibility.

The bacteria utilize several strategies for Fe acquisition. A new approach capable of blocking all probable iron uptake mechanisms relies on antibodies that target all antigens involved in its metabolism. [8].

MATERIAL AND METHODS

Sequence and template search. Searching for the IroN protein sequence within the NCBI database (http://www.ncbi.nlm.nih.gov/protein) revealed a sequence of this protein under the accession number AAS80269.1. The BLAST search was performed against the sequences available in a non-redundant protein database (http://blast.ncbi.nlm.nih.gov/Blast.cgi). The BLAST result provided information about the putative conserved domains of the query. The experimentally resolved structures with homology to the sequence of the query protein were searched against protein data bank (PDB) [9, 10].

Primary sequence analysis and Subcellular localization. The properties of the IroN protein, such as theoretical pI, molecular weight, number of positive and negative residues, amino acid composition, instability, and aliphatic index were calculated by Protparam [11] software (http://expasy.org/tools/protparam.html). The subcellular positions of the vaccine candidate were determined by CELLO [13] subCELlular LOcalization predictor (http://cello.life.nctu.edu.tw/) and PSLpred (http://crdd.-osdd.net/raghava/pslpred/).

Antigenicity properties of the IroN protein. The prediction of antigenicity for IroN performed using Vaxijen server 2.0 version [12] (http://www.ddg-pharmfac.net/vaxijen/VaxiJen-/VaxiJen.html).

Secondary structure prediction. The secondary structure of the antigen was determined by phyre2 (http://www.sbg.bio.ic.ac.uk/phyre2) and the_SOPMA[14] (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=n-psa_sopma.html). These servers are the highly accurate predictors of protein secondary structure which make consensus predictions from multiple alignments.

Topological Predictions. Topology prediction for integral transmembrane (TM) proteins would yield important clues about the function of a protein and guide further experimental studies. Various methods with high accuracies are available for topology prediction, *i.e.*, a specification of the membrane-spanning regions and in/out orientation concerning the membrane.

The TopCons software [15] (http://topcons.cbr.su.se/) predicts the membrane topology of the target protein and determines its signal peptides. This server contains SPOCTOPUS (an algorithm for combined prediction of SPs and membrane protein topology), RHYllus, OCTOPUS (Using a novel combination of hidden Markov models and artificial neural networks). OCTOPUS predicts the correct topology for 94% of a dataset of 124 sequences with known structures.

Transmembrane helices of target proteins were predicted by TMHMM [16] (http://www.cbs.dtu.dk/servic-es/TMHMM/) and SPOCTOPUS (http://octopus.cbr.su.se/).

The location of transmembrane β-strands of the protein sequence in Gram-negative bacteria along with the topology of the loops was predicted by PRED-TMBB [17] at (http://biophysics.biol.uoa.gr/PRED-TMBB/).

The SignalP 4.1 [18] server (http://www.cbs.dtu.dk/-services/SignalP/) was used to predict the presence and location of signal peptide cleavage sites in amino acid sequence from different organisms. Based on several approaches, *i.e.*, artificial neural networks, the cleavage sites, and the signal peptide were detected.

Further scrutiny. ToxinPred (http://crdd.osdd.net/-raghava/toxinpred/protein.php) was used to determine the toxicity of the protein. AlgPred (http://crdd.osdd.net/-raghava/algpred/submission.html) was used to determine the possible allergic feature of the protein.

Modeling the 3D protein structure. PS^2v^2, SWISS-MODEL, LOMETS, and Phyre2 were used to model the 3D protein structure.

[DOI: 10.29252/JoMMID.7.4.93]

The PS^2v^2 [19] (http://ps2.life.nctu.edu.tw) is an automated homology modeling server. This server finds the best model for a query, based on a combination of PSI-BLAST, IMPALA, and T-Coffee and performing an effective template alignment.

The fully automated SWISS-MODEL software [20], available in the ExPASy web server (https://swissmodel-.expasy.org/) and the program DeepView [21] (Swiss Pdb-Viewer) was used for template-based modeling of the protein 3D structure.

The LOMETS [22] is a Local Meta-Threading-server (http://zhanglab.ccmb.med.umich.edu/LOMETS/). It is an online web service for protein structure prediction. Ten locally-installed threading programs (FUGUE, HHsearch, MUSTER, PPA, PROSPECT2, SAM-T02, SPARKSX, SP3, FFAS, and PRC) are used by LOMETS to generate 3D models by collecting high-scoring target-to-template alignments.

The Protein Homology/analogy Recognition Engine V 2.0 (Phyre2) [23] (http://www.sbg.bio.ic.ac.uk/phyre2/html-/page.cgi?id=index) is a significant update to the original Phyre server. In Phyre2, comparing to Phyre, some features such as accuracy of modeling and user interface have been improved. In contrast to Phyre that uses the profile-profile alignment method, Phyre2, by using hidden Markov models algorithm via HHsearch1, significantly improves the accuracy of alignment and modeling. In addition, Phyre2, using a new ab initio folding simulation called Poing, can model regions of proteins with no detectable homology to known structures. Poing promotes Phyre2 efficiency by doing combine multiple templates. Constrain of distance from available models are treated as linear elastic springs. Poing then synthesizes the entire protein model in the presence of these springs and at the same time models unconstrained regions using its physics simulation.

Models evaluation and refinement. The structural quality of all predicted 3D models was assessed by Rampage software (http://mordred.bioc.cam.ac.ukrapper/rampag-e.php). This server provides the Ramachandran plots for each analyzed model [24].

Mod Refiner is an atomic-level, high-resolution protein structure refinement server (http://zhanglab.ccmb.med.-umich.edu/ModRefiner), which is used for improving the quality of the predicted models [25].

Protein Structure and Membrane Alignment. The PPM server (http://opm.phar.umich.edu/server.php) was used to calculate the rotational and translational positions of transmembrane and peripheral proteins in membranes. The recently identified proteins are available at this site. Many membrane-associated proteins at the PDB database have already been pre-calculated and can be found in the OPM server [26].

Ligand binding site prediction. COFACTOR (http://zhanglab.ccmb.med.umich.edu/COFACTOR) is a structure, sequence, and protein-protein interaction (PPI) based method to annotate protein molecules for their biological functions. COFACTOR was used to identify functional sites and homologies. Functional properties such as Gene Ontology (GO), Enzyme Commission (EC), and ligand-binding sites, will be explored using the best functional homology templates. Further insights about the GO of the target protein can be achieved by UniProt-GOA (sequence and sequence-profile alignments) and STRING (protein-protein interaction inferrals) [27] servers.

Functionally and Structurally Important Residues. The conserved functional and structural amino acids of the protein sequence were determined by Conseq software (http://conseq.tau.ac.il/). The software parameters were set as follows: PSI-BLAST for five iterations against Uniprot database with an E-value of 0.01 and the maximum likelihood (ML) score as a method of calculating amino acid conservation score [28].

Serving the protein 3D structure as an input file, the functional sites on the protein surface were predicted by InterProSurf [29] (http://curie.utmb.edu/pattest9.html) using patch analysis. Consurf program (http://consurf-.tau.ac.il/) was used for color-coded illustration of the functional residues on the protein structure.

Surface accessible pockets and Clefts analyses. The GHECOM (Grid-based HECOMi finder) server (http://strcomp.protein.osaka-u.ac.jp/ghecom/) was emplo-yed to find the multi-scale pockets on protein surfaces using mathematical morphology. Binding sites and active sites of proteins and DNAs are often associated with structural pockets and cavities [30]. The castP server (http://sts.bioe-.uic.edu/c_astp/) is a pocket finding server that uses the weighted Delaunay triangulation and the alpha complex for shape measurements. It was used for identification and measurements of surface accessible pockets as well as interior inaccessible cavities, for proteins and other molecules [31].

The depth of the cavity, the sizes of the cavity, the ligand-binding sites, and PK of the target protein were analyzed by Depth program (http://mspc.bii.a-star.edu.sg/tankp/help.html). This server calculates the closest distance of a residue/atom to bulk solvent [32]. For further evaluation, the discovery studio surface accessibility analyzer along with GHECOM, castP, and Depth program was used to compute surface accessibility for vaccine candidate epitopes.

Single-scale amino acid properties assay. IEDB and BCEPred were employed in single-scale amino acid properties assay. The IEDB database [33] (http://tools.immuneepitope.org/tools/bcell/iedb) uses various input parameters to predict the location of B-cell epitopes. It uses certain features of the protein sequence for the detection of B-cell epitope position. BCEPred server [34] (http://webs.iiitd.edu.in/raghava/bcepred/) provides various physicochemical properties or a combination of them to predict B-cell epitopes.

Prediction of linear and conformational epitopes along with the immunogenic regions. The identification of possible B-cell epitopes is crucial for vaccine design, immunological assays, and antibody production. Therefore, bioinformatics tools capable of reliable detection of linear B-cell epitopes are highly desirable. BepiPredv2.0 [35]

[DOI: 10.29252/JoMMID.7.4.93]

(http://www.cbs.dtu.dk/services/BepiPred/) predicted the location of linear B-cell epitopes using a combination of a hidden Markov model and a propensity scale method with the threshold of 0.5. Svmtrip was another server used for antigenic epitope prediction (http://sysbio.unl.edu/-SVMTriP/prediction.php). Such prediction methods are based upon the amino acid properties including hydrophilicity, solvent accessibility, secondary structure, flexibility, and antigenicity. In addition, based on the known linear epitope databases such as Bcipep and FIMM, there also exist some methods using machine learning algorithms such as Hidden Markov Model (HMM), Artificial Neural Network (ANN), and Support Vector Machine (SVM) to locate linear epitopes [36].

Conformational epitopes of the antigen were determined using the predicted 3D model as input structure. Discotope v1.1 (http://www.cbs.dtu.dk/services/DiscoTope/) predicted the discontinuous B-cell epitopes using an algorithm that calculates a novel score for epitope propensity of amino acids and their surface accessibility. The method utilizes calculation of surface accessibility (estimated in terms of contact numbers) and a novel epitope propensity amino acid score. The final scores are calculated by combining the propensity scores of residues in spatial proximity and the contact numbers. Novel definition of the spatial neighborhood used to sum propensity scores and half-sphere exposure as a surface measure [37]. The Ellipro server is a new tool that was used to locate the conformational epitopes from the regions of the protein which are protruding from the protein's globular surface. This server also predicts the antibody epitopes (http://tools.iedb.org/ellipro/). The ElliPro implements three algorithms for epitope prediction process; it approximates the shape of the protein as an ellipsoid, it computes a protrusion index (PI) for each residue, and it clusters the neighboring residues using calculated PI values [38]. "Minimum score" and "Maximum distance" for epitope prediction specify 0.5 and 6 respectively.

Determining the Best Immunogenic Regions. The regions of the protein with the highest density of both linear and conformational epitopes have the priority for the selection process. The selected regions should also meet other criteria, including favorable single scale amino acid properties, physicochemical properties, and antigenicity. Regarding these criteria, two regions of the antigen were qualified to be included in the final vaccine candidate. Further analyses were performed on the selected regions including the antigenicity by vaxijen (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html), B-cell epitopes by IEDB server (http://tools.immuneepitope.org/-tools/bcell/iedb_input) and physicochemical properties by Protparam server (http://expasy.org/tools/protparam.html). The performed analyses were also repeated on the IroN protein to compare with the selected regions.

RESULTS

Sequence and Template Search. The IroN protein, containing 725 amino acids, was stored as a FASTA file format. The search for similar protein sequences by the BLASTP tool returned numerous homologous sequences. The PSI-BLAST search against PDB also returned several hits with experimentally resolved structures. The first hit of the search with the highest score (PDB code of 5FP2_A with 60% identity) was selected as the template structure for the homology modeling. The outer membrane receptor superfamily was the putative conserved domain predicted for IroN protein.

Physicochemical Properties and Subcellular localization. The analyses on the physicochemical properties of the IroN protein revealed 78 negatively charged (Asp + Glu) amino acids and 72 positively charged (Arg + Lys) residues. The protein is expected to be stable according to the predicted instability index. A summary of the computed physicochemical properties is available in Table 1. The obtained results were in line with the results from the Discovery studio software. The subcellular location of the protein is predicted to be the outer membrane with a highly reliable CELLO score (4.635) and a high accuracy of PSLpred (90.2%). The computed parameters included the molecular weight, theoretical pI (isoelectric point), instability index, aliphatic index and grand average of hydropathicity (indicates the solubility of the proteins: positive GRAVY (hydrophobic), negative GRAVY (hydrophilic)) and Vaxijen score are summarized in Table 1.

Antigenicity properties of IroN protein. Vaxijen antigenicity score calculated for IroN about 0.7889.

Secondary structure prediction. The secondary structure of the proteins was composed of Coil, helix, and beta-sheet strands. The accuracy of the 3D structure prediction could be validated using the secondary structure of the protein. The predicted secondary structure composition of the IroN protein is as follows: alpha helix (16.41%), extended strand (27.17%), beta-turn (10.21%), and random coil (46.21%). The results of the SOPMA were consistent with the results of the phyre2. Figure 1 shows the graphical representation of the predicted secondary structure.

Topological Predictions. The 2D topology model of IroN indicated that the protein was inserted in the membrane by 22 transmembrane anti-parallel β-strands (Fig. 2). These transmembrane regions could put the IroN protein in the category of β-barrel membrane proteins. The strands of the β-barrel structure were connected with turns at the inside or loops at the outside of the membrane bilayer. No TMHs (transmembrane helixes) were predicted to exist within the sequence of the IroN protein. The first 24 residues of the IroN protein were predicted to be the signal peptide.

Table 1. Physicochemical properties of the IroN protein

Protein name	VaxiJen Score	Number of amino acids	Molecular weight	Theoretical pI	Instability index	Aliphatic index	GRAVY
IroN	**0.7889**	**725**	**79134.54**	**5.78**	**33.97 (stable)**	**76.79**	**-0.552**

[DOI: 10.29252/JoMMID.7.4.93]

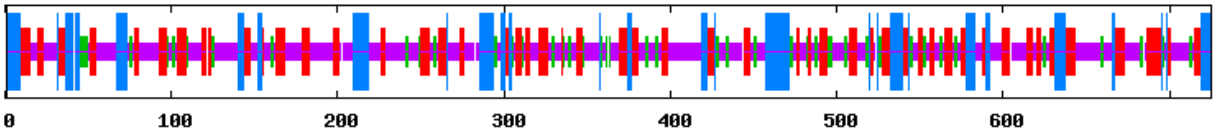


Fig. 1. Graphical representation of SOPMA results. Red, Extended strand; blue, Alpha-helix; green, Beta turn, and yellow, Random coil

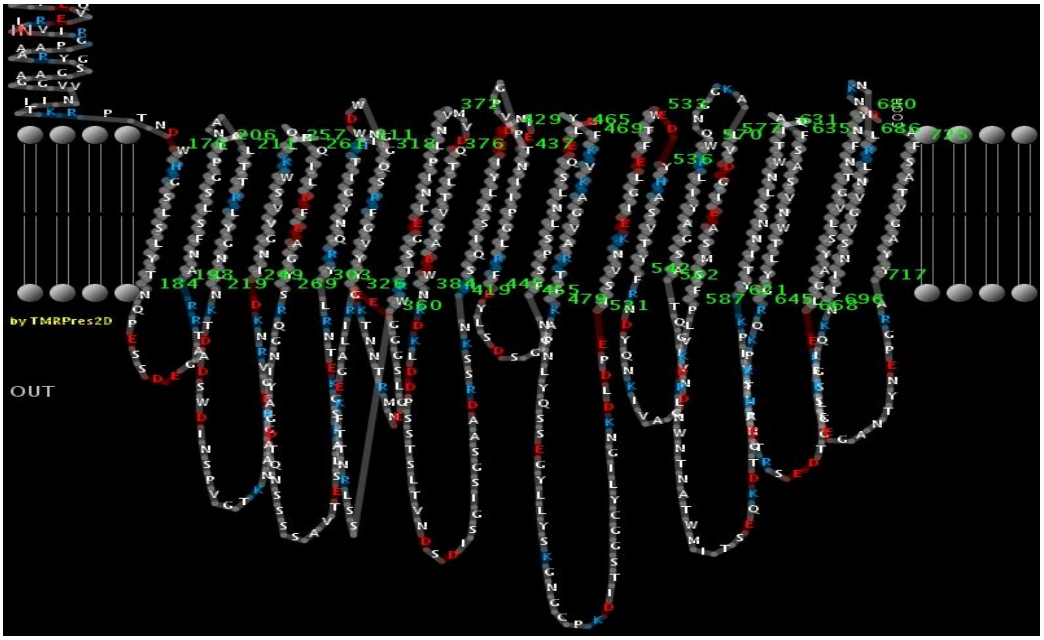


Fig. 2. A 2D topology model of IroN. A cork domain at N terminus and a beta-barrel domain at the C terminus are the main predicted substructures of the protein. The beta-barrel domain consists of 22 transmembrane beta-sheets, 11 large extracellular loops, and 10 short periplasmic turns.

Further scrutiny. ToxinPred predicts no toxic peptide along the protein sequence. The results for allergenicity prediction demonstrated that the protein does not contain significant allergen potency.

Structure Prediction by Homology Modeling and Threading Method. All of the employed servers have managed to make structural models of the IroN protein. Amongst them, two models were predicted by the Swiss model, one model by Ps^2v^2, one model by Phyre2 and ten models by the Lomets Meta server. The obtained models were all taken for further quality assessment.

Models Quality Assessment. The Ramachandran plots of the predicted models show the percentage of residues located in favored, allowed, and outlier regions. This plot gives useful information about the protein structure and could be used to validate the predicted 3D model. The φ–ψ set angles have distinct regions, and each spot in the Ramachandran plot corresponds to a particular secondary structure. According to the results of the Ramachandran plots, the model generated by LOMETS was way more qualified compared to the other models. About 97.4% of the residues from the selected model were located at favored regions, while only 0.8% of the residues were at outlier regions. All of the data for model validations are summarized in Table 2.

Improving Model Quality. The model generated by LOMETS was selected as the best-predicted model based on the quality scores. Using multiple templates and the combinatorial methods employed to produce the models could be the reason behind the accurate predictions of the LOMETS. Performing the model refinement process, the quality of the selected model was improved. The improvement in the model quality is shown by comparing the results of the Ramachandran plot analysis both before and after the refinement step (Fig. 3). The percentage of the residues in the favored region increased from 98.3% to 98.7%, while the percentage of residues in the outlier region decreased from 0.35% to 0.27%.

Protein and Membrane Alignment. The orientation of the protein was predicted by the PPM server (Fig. 4). The results indicated that the predicted model for the IroN protein could be accommodated within the lipid bilayer. The server employs a computational approach for optimizing the spatial arrangement of protein structures in lipid bilayers. This calculation has already been performed for all unique structures of transmembrane proteins and a significant number of peripheral proteins stored in the PDB database. For calculations, each protein is considered as a rigid body that freely floats in a hydrophobic slab of adjustable thickness. The calculation could be carried out up to a precision of ~1 Å and 2° for hydrophobic thicknesses and tilt angles (Table 3).

[DOI: 10.29252/JoMMID.7.4.93]

Table 2. The results for the Ramachandran plot of the predicted structures

Program	Favored region	Allowed region	Outlier region
FFAS-3D	97.4%	1.7%	1.0%
PRC	95.4%	2.5%	2.1%
SP3	97.4%	1.8%	0.8%
PROSPECT2	87.3%	7.7%	5.0%
pGenTHREADER	81.1%	13.0%	5.9%
FFAS03	96.7%	2.2%	1.1%
Neff-PPAS	92.7%	4.4%	2.9%
SPARKS-X	94.6%	4.6%	0.8%
wdPPAS	95.7%	2.6%	1.7%
MUSTER	97.2%	1.9%	0.8%
Swiss model	96.2%	3.5%	0.3%
Swiss model	90.4%	8.3%	1.3%
PS2V2	93.4%	5.1%	1.5%
Phyre2	95.6%	3.5%	0.9%

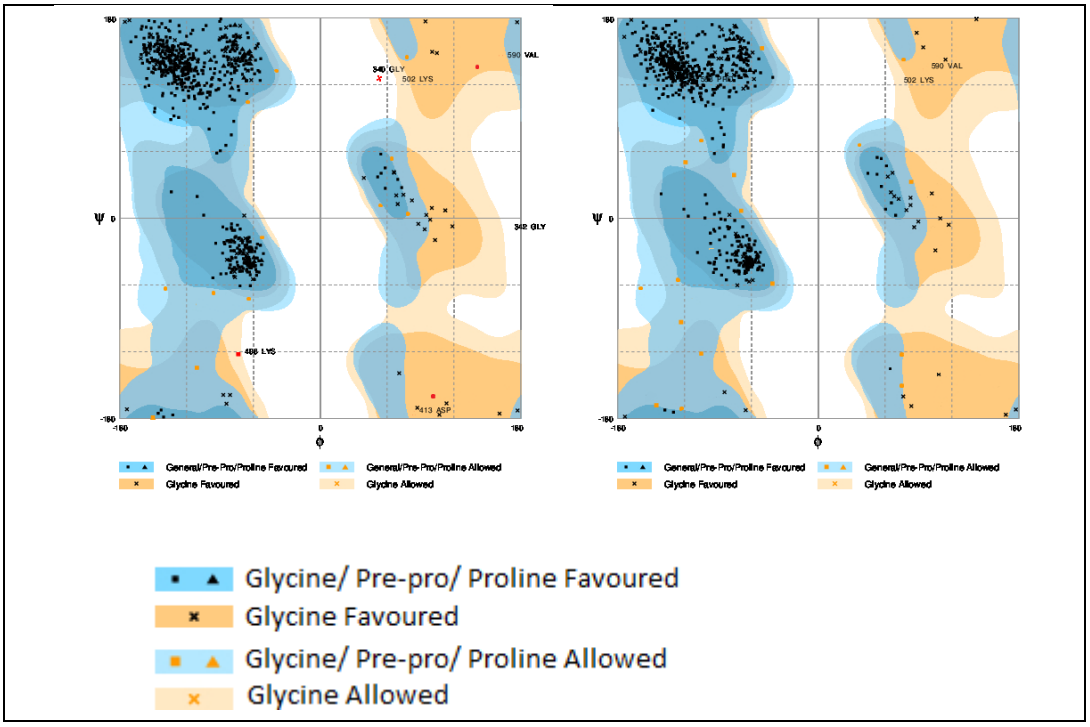


Fig. 3. Model quality assessment using Ramachandran plot before and after refinement for IroN models

Protein-Ligand Binding Site Prediction. The interaction site between the ligand and the receptor protein was determined by the COFACTOR software. The results indicated that the conserved residues between the crock domain and the large extracellular loops of the barrel in the iron-binding site had the highest C-score about 0.7. The high C-score suggests that these residues are involved in the formation of the predicted ligand-binding site (Fig. 5). Cscore is the confidence score of the predicted binding site. Cscore values range in between [0-1]; where a higher score indicates a more reliable ligand-binding site prediction.

Residues of Functional and Structural Significance. The results obtained from the Interprosurf software suggest that the outer membrane loops and the cork domain are the most functionally important regions of the protein (Fig. 6). These residues were mostly located at the protein surface. These results were in line with the results obtained from the ligand-binding site prediction.

Table 3. Trance-membrane segment positions predicted by PPM Orientation and evaluation server

Server	TM segment positions
PPM	171-178|183-191|199-207|223-230|235-242|287-295|299-306|345-352|360-367|420-427|433-439|464-471|475-482|510-517|521-528|560-568|572-579|606-613|621-628|649-656|664-671|695-702

[DOI: 10.29252/JoMMID.7.4.93]

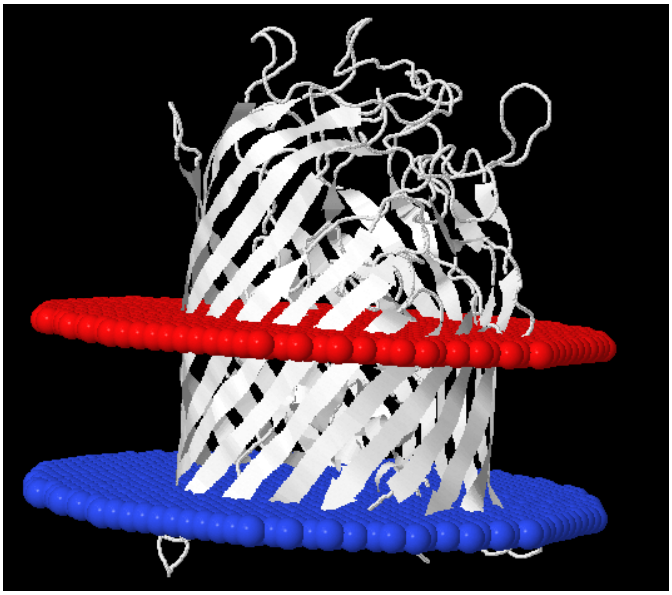


Fig. 4. The protein accommodation within the lipid bilayer. The beta-barrel composition of the protein is located within the membrane

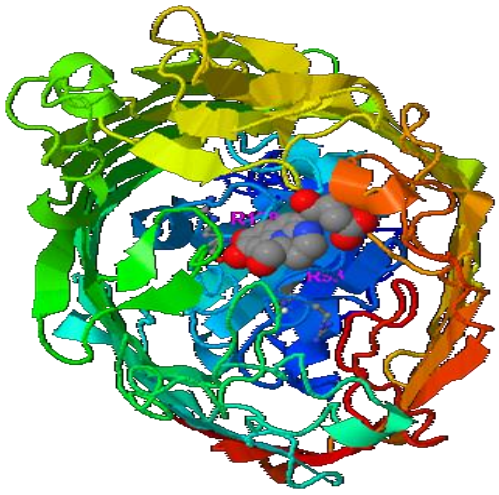
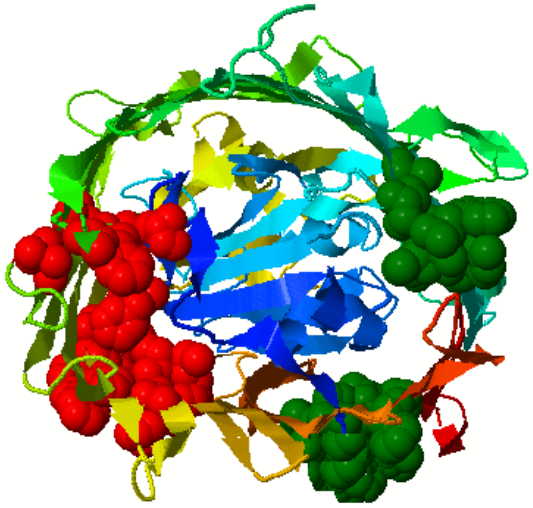


Fig. 5. Lateral and top view of the IroN structure in interaction with the ligand molecule. The ribbon representation shows the protein 3D structure, and the space-filling representation shows the ligand (blue: Fe, gray: Hydrogen, red: Oxygen).



The most significant functional residues:
41,44,45,336,376,378,477,478,493,494,495,497,
530,532,533,474,475,499,528,165,167,184,658,
659,660,582,583,584,585,609,610,611,612,632,634,635

Fig. 6. Interpro surf prediction. The clusters of the functional residues are depicted in space fill representation. The Blue cluster is the residue with the highest functional scores, and the green clusters are the residues with less functional significance.

Surface accessibility and Cleft regions. By computing a pocket score for each residue, the GHECOM server found 5 pocket regions on the surface of the IroN protein. The predicted pockets at small molecule binding sites and the active site had even more significant pocket value. Therefore, the results for pocket prediction could be used to confirm binding and active sites. The results obtained from the GHECOM server are shown in Figure 7. The results of the GHECOM software were in concordance with the results of the surface accessibility prediction by discovery studio.

In order to determine the pocket regions by CastP server, the solvent accessible surface (SA, Richards' surface) and molecular surface (MS, Connolly's surface) were analytically considered, and three sources were employed to annotate biologically relevant functional residues. The server predicted 8 pocket areas in protein structure. Figure 8 shows

[DOI: 10.29252/JoMMID.7.4.93]

the atoms of the charge relay system in a functional pocket of the protein. The atoms of residues that lie in the pocket are highlighted for better illustration.

The potential binding sites (PBS) of the protein refer to residues and atoms which bind to ligands directly, and their positions are close to the ligand-binding sites. A binding cavity is defined as a protein sub-structure which has conserved geometrical and chemical properties complementary to its bound ligand. The residue depth, accessible area value pairs are used to parametrize the probability of each amino acid to be a part of the binding cavity. The plot shows both mean and standard deviation of depth values for the residues. Figure 9 shows the probability of the residues to form a binding site, residue depth plot, and a 3D rendition of the cavity prediction.

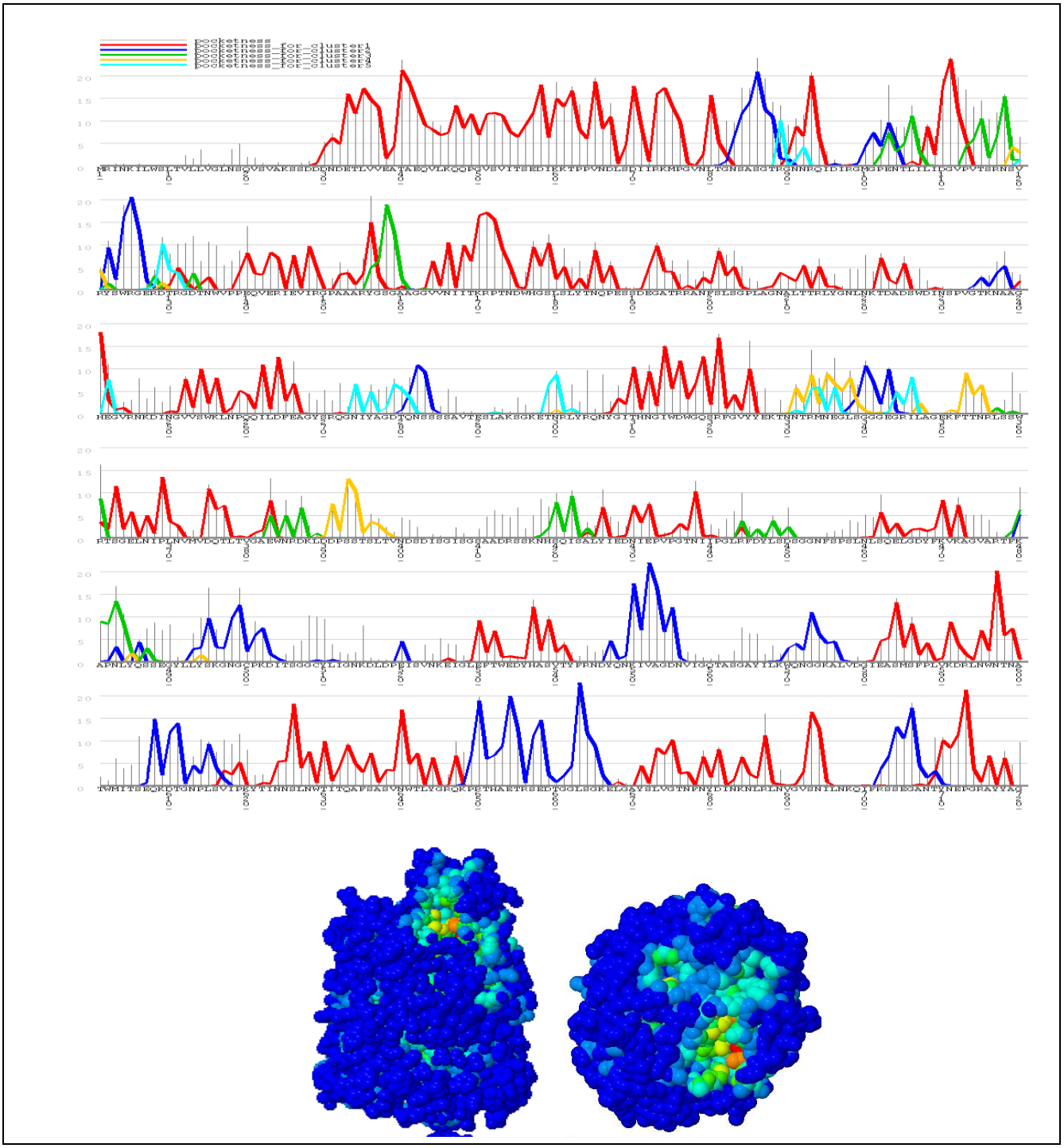
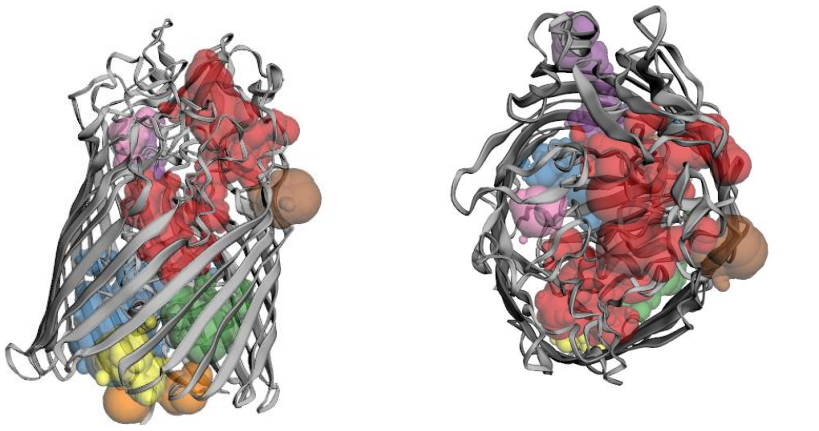


Fig. 7. GHECOM results showing graph residue-based pocketness and Jmol view of pocket structure. Above: graph residue-based pocketness. The height of the bar shows the value of pocketness [%] for each residue. The color of pocketness bar indicates cluster number of pocket (red: cluster 1, blue: cluster 2, green: cluster 3, yellow: cluster 4, cyan: cluster 5). Below: Jmol view of pocket structure based on pocketness color.

[DOI: 10.29252/JoMMID.7.4.93]



Volume(SA)	Area (SA)	PocID
957.760	1836.922	1
889.189	1027.845	2
276.631	433.143	3
223.358	494.937	4
210.832	223.188	5
179.549	232.353	6
137.292	181.333	7
71.415	152.324	8

Fig. 8. CastP results showing the accessible surface pockets as well as interior inaccessible cavities. Residues are colored based on area and volume size. The most important one is illustrated in red and others in blue, green, purple, orange, yellow, brown, and pink.

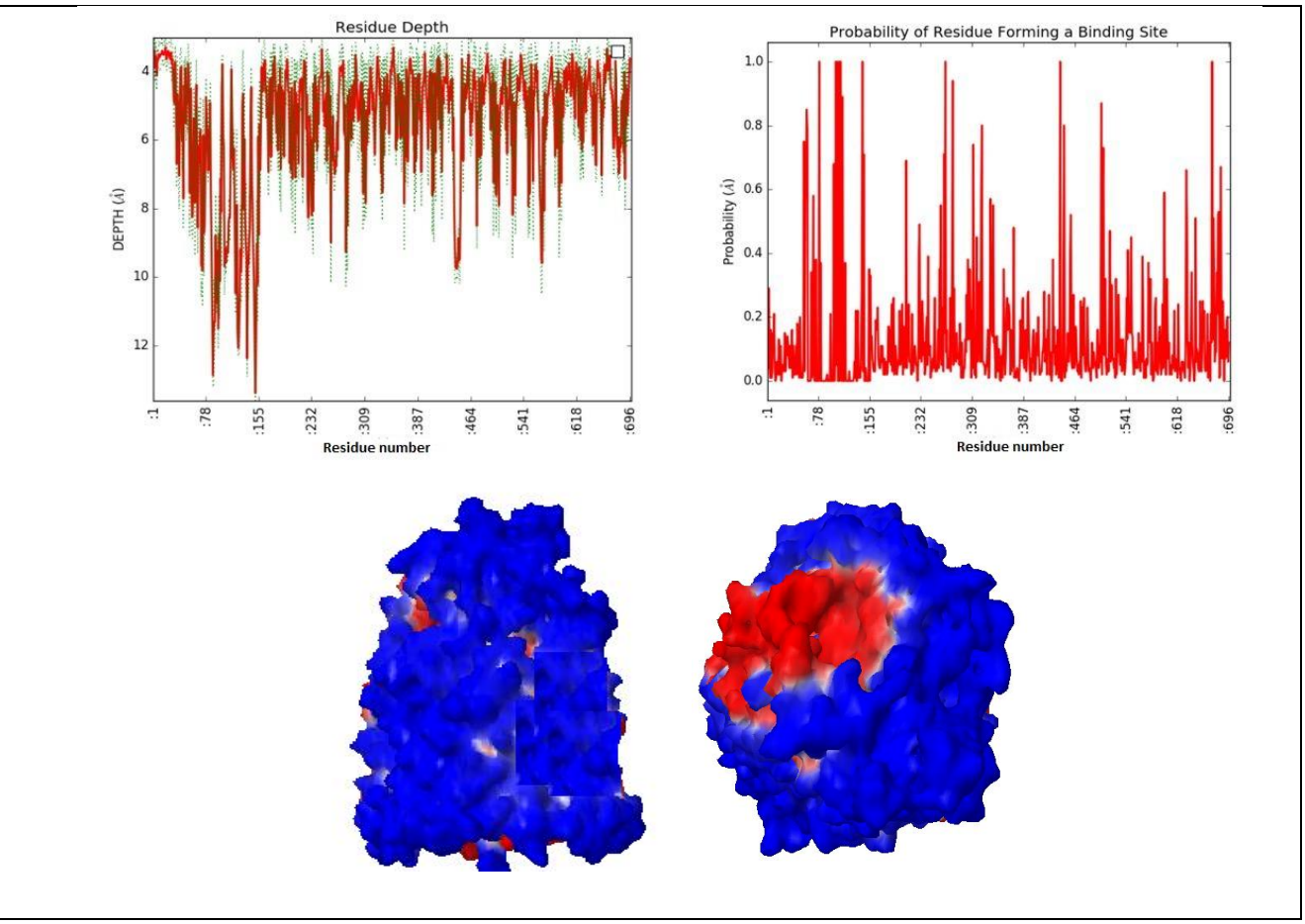


Fig. 9. Probability of residue forming a binding site and residue depth plot and a 3D rendition of the cavity prediction. Top: Probability of residue forming a binding site and residue depth plot. Below: A 3D rendition of the cavity prediction is shown using Jmol. Residues of the predicted binding cavity are colored red while the rest of the protein is colored blue.

Single-scale amino acid properties assay. The assay for single-scale amino acid properties revealed that these properties were detectable throughout the whole antigen. However, a region spanning the amino acids from 542 to 670 indicated to possess the most salient properties. This region was a segment of the beta-barrel domain of the protein. All of the predicted properties from the IEDB and BCEpred servers were taken into consideration to determine the selected region

Epitope Prediction. The linear B-cell epitopes were predicted to be of higher density in the region spanning the 557-667 amino acids. The highest score for a predicted linear B-cell epitope was related to the

[DOI: 10.29252/JoMMID.7.4.93]

"RQKPRTHAETRSEDTGGLSGKE" sequence (located at 646-667 region).

The linear epitopes predicted by Svmtrip were ranked based on their scores. The predicted epitope with the highest score was a sequence spanning the 628-647 amino acids (NWTITQAFSASVNWTLYGRQ).

ElliPro software returned 19 linear epitopes and 4 conformational B-cell epitopes. Figure 10 shows the best predicted conformational and linear epitopes of the antigen with the highest PI (protrusion index). The sequence "AGDNVIGQTASGAYILKWQ" which spans the 553-571 amino acids, was predicted to be the best linear epitope. The outer membrane loops of the antigen were predicted to be involved in the formation of conformational epitopes using the Discotope program (Figure 11).

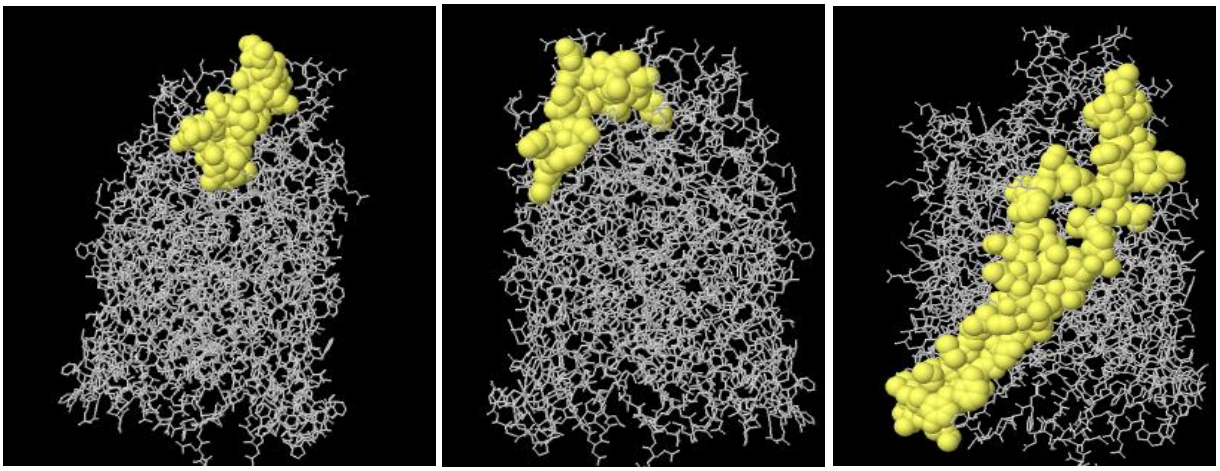


Fig. 10. Discontinuous epitopes predicted by the Ellipro server. The epitopic regions are in yellow. Epitopes were mapped on 3D models using Discovery Studio Visualizer 2.5.5 software.

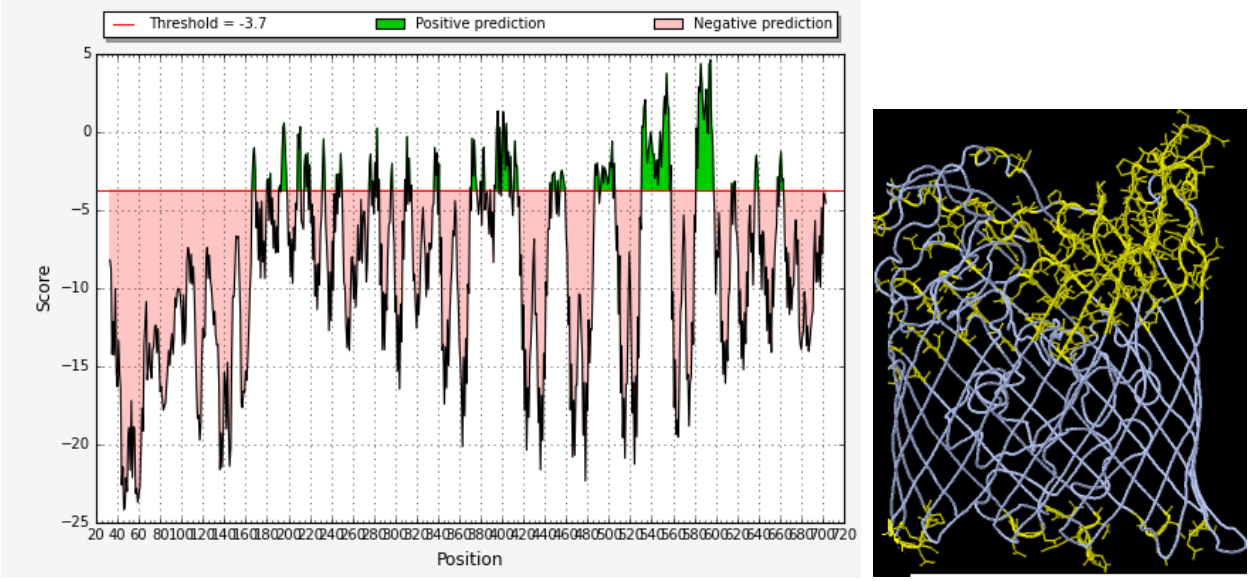


Fig 11. Discontinuous B-cell epitopes predicted from the 3D structure of the protein by Discotope. In the chart, predictions above the threshold (red line) are positive predictions (displayed in green), and predictions below the threshold are negative predictions (displayed in orange). Protein 3D structure is showing in blue. Discontinuous B-cell epitopes at the protein structure differentiated with yellow.

Immunogenic region selection. All of the previously obtained results were considered to determine the most favorable regions for vaccine design. Hence, a region spanning the residues 542 to 693 (including a part of the barrel) was selected to be a suitable region for vaccine design. For confirmation of the properties of this region, the antigenicity score, PI, instability index, solubility, accessibility, flexibility, and secondary structure properties were recalculated for the selected region. A comparison of the properties for the selected region and the same properties for the IroN protein is presented in Table 4.

Flowchart in Figure 12 shows the work done in this research and graphical display of linear and conformational epitopes predicted by various software shows in Figure 13.

[DOI: 10.29252/JoMMID.7.4.93]

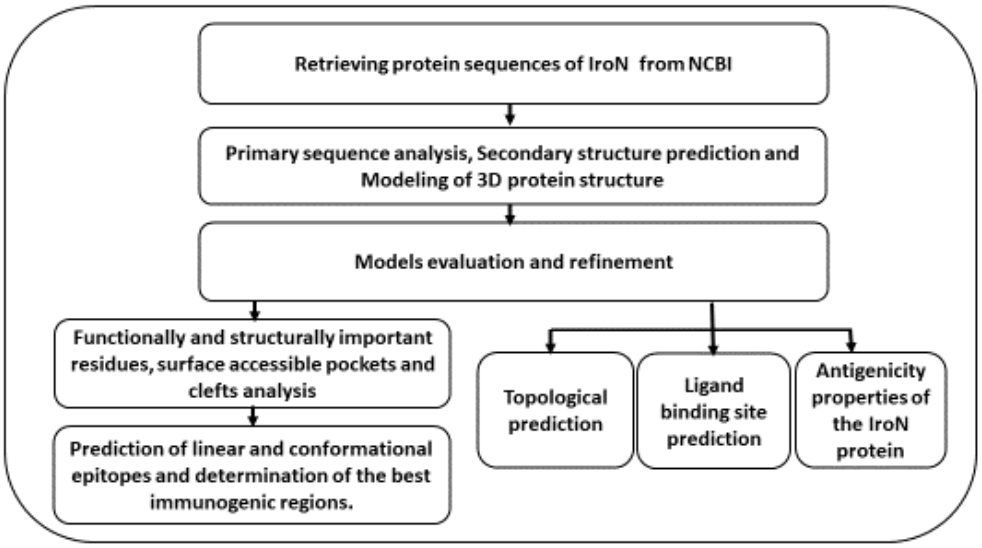


Fig. 12. The flowchart of performed analyses

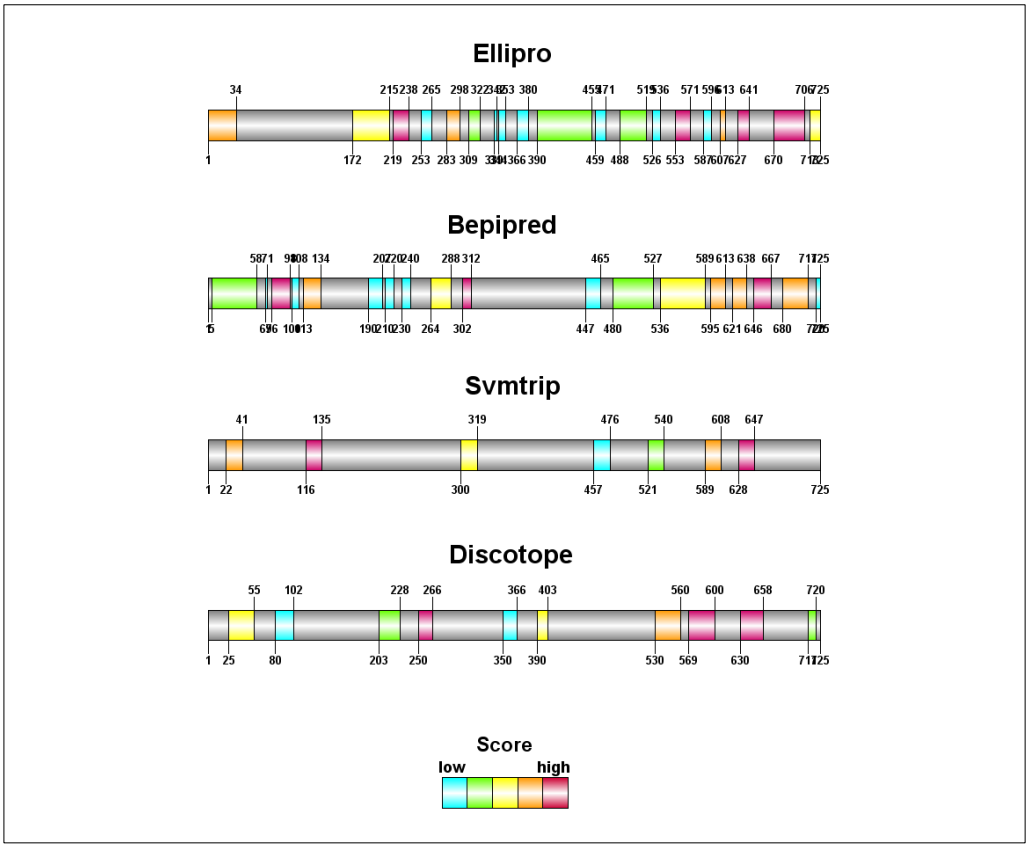


Fig. 13.Graphical display of linear and conformational epitopes predicted by various software. The length and position of epitopes are graphically illustrated on a gray rod. The epitopes with the highest score are shown in red to blue respectively.

Table 4. Average physicochemical properties (such as flexibility, accessibility, turns, exposed surface, polarity, and the antigenic propensity) of the vaccine candidate region and IroN

	Residue number	Weight	Vaxijen score	Instability index	PI	B-cell epitope	Flexibility	Beta turn	Accessibility
Vaccine Candidate	151	16693.58	0.7957	14.83 (Stable)	9.19	0.542	1.0	1.0	1.0
IroN	725	79134.54	0.7889	33.97 (Stable)	5.78	0.524	1.0	1.0	1.0

[DOI: 10.29252/JoMMID.7.4.93]

DISCUSSION

Vaccine development remains a challenging issue due to the complexities associated with the design of an amenable immunogen capable of eliciting adaptive immune responses against various pathogens. *Escherichia coli* is one of the most common nosocomial pathogens; hence, vaccine design against this pathogen has long been the subject of extensive research [39, 40]. The hospital setting serves as potential reservoirs for *E. coli* strains and hence has raised concern over the cumulative incidence of this bacteria, particularly in intensive care units (ICUs). This concern pinpoints the importance of vaccination against this pathogen [41]. The 3D structural information of vaccine targets is highly critical for accurate vaccine design, which can be resolved using theoretical or experimental methods. The obtained structures could be used for identification of their biochemical properties and related functions.

Today, bioinformatics methods have attracted extra attention to solving the 3D structure of different proteins. This is partly due to the laborious, expensive and time-consuming nature of experimental approaches. Bioinformatics approaches are convincing alternatives for conventional experimental methods [42-46]. In the present study, we developed 3D models for the structure of the IroN protein by deploying several *in silico* approaches. All main approaches of structure modeling such as threading and homology modeling and combination of them in addition to *ab initio* modeling were utilized to build valid models [47]. The vaccine candidates were designed based on the generated 3D structures and their immunological properties.

The results of the BLAST search indicated that there were high sequence homologies between IroN and other molecules. This search was performed to choose a template structure for homology modeling. The first hit of the results with the highest score was selected as the template structure to perform the homology modeling. This top hit (60% identity) belonged to the crystal structure of the siderophore receptor pirA from *Pseudomonas Aeruginosa* under the PDB ID of 5FP2_A [48]. An accurate homology modeling requires a reliable template that can be achieved via sequence alignment and similarity search. An amenable template must have a high identity (more than 35%), high query coverage, and low *E*-value against the target sequence. Therefore, a hit with the maximum total score can be the most reliable template for homology modeling.

A model refinement run after protein modeling could help to improve the quality of the predicted model. Model refinement could draw the initial starting models closer to their native state, in terms of hydrogen bonds, backbone topology, and side-chain positioning. For the evaluation of full-atom refined models, the global topological similarity between the model and the experimental structure (including the RMSD) was calculated. Moreover, the template modeling (TM)-score was also calculated as an indicator of structural similarity between two structures. The models with lower RMSD and higher TM-score/GDT-TS are considered to be structurally equivalent and more close to their native state [45, 49].

A 2D topology model of protein was constructed to predict the inside, outside, and transmembrane regions of the proteins. The obtained results revealed that IroN contains several trans-membrane anti-parallel β-strands. The predicted topological models of the protein proposed that the trans-membrane anti-parallel β-strands of the protein could form a β-barrel structure in their native form [50]. The connection between the trans-membrane anti-parallel β-strands is made by loops at the outside or turns at the inside faces of the protein. This protein has shown to possess more than 11 external loops. The side chains of amino acids in these external loops are highly exposed to the extracellular surface, implying their role in early binding events with Fe-siderophore complex.

Computational methods can reliably predict the linear and conformational B-cell epitopes. The defined epitomic data can be utilized to select IroN regions with higher epitope density. These regions could be applied to design efficient vaccine candidates capable of driving humoral immune responses with higher affinity and avidity for epitope-specific monoclonal and polyclonal antibodies. The best Linear B-cell epitopes of this iron receptor were located at the largest extracellular loops. Remarkably, discontinuous B-cell epitopes were predicted (using the proteins 3D structure) to contain all of the additional cellular loops. The results of epitope prediction analyses were confirmed by the experimentally determined epitopes using approved antibodies. This fact ensures the accuracy of the employed methods for the prediction of protein 3D structure and epitope prediction. Comparison of the antigenicity scores of the chosen regions and the whole proteins revealed that the antigenicity of the chosen regions was significantly higher than the whole antigen.

Finally, it is worth mentioning that using bioinformatics tools is a convincing approach to fill the gap between the discovered protein sequences and their resolved 3D structure. *In silico*, information from the immunological and structural properties of antigens can be employed for vaccine design purposes. The strategy of the present study to use an *in silico* approach for protein 3D structure prediction and prediction of its immunological properties could be used to bring about novel structural, functional, and therapeutic insights for vaccine design investigations and help to design more efficient vaccines.

ACKNOWLEDGMENT

The authors are grateful to the Yazd University of Medical Sciences, Iran, and Zanjan University of Medical Sciences, Iran, for their support to conduct this work.

CONFLICT OF INTEREST

The authors declare that there are no conflicts of interest associated with this manuscript.

ETHICAL APPROVAL

This article contains no assay on humans or animals by the contributing authors.

[DOI: 10.29252/JoMMID.7.4.93]

REFERENCES

1. Russo TA, Johnson JR. Proposal for a new inclusive designation for extraintestinal pathogenic isolates of *Escherichia coli*: ExPEC. J Infect Dis. 2000; 181: 1753-4.

2. Russo TA, Johnson JR. Medical and economic impact of extraintestinal infections due to *Escherichia coli*: focus on an increasingly important endemic problem. Microb Infect. 2003; 5: 449-56.

3. Gupta K, Hooton TM, Stamm WE. Increasing antimicrobial resistance and the management of uncomplicated community-acquired urinary tract infections. Ann Intern Med. 2001; 135: 41-50.

4. Langermann S, Palaszynski S, Barnhart M, et al. Prevention of mucosal *Escherichia coli* infection by FimH-adhesin-based systemic vaccination. Science. 1997; 276: 607-11.

5. Russo TA, McFadden CD, Carlino-MacDonald UB, Beanan JM, Barnard TJ, Johnson JR. IroN functions as a siderophore receptor and is a urovirulence factor in an extraintestinal pathogenic isolate of *Escherichia coli*. Infect Immun. 2002; 70: 7156-60.

6. Dozois CM, Daigle F, Curtiss R. Identification of pathogen-specific and conserved genes expressed in vivo by an avian pathogenic *Escherichia coli* strain. Proc Natl Acad Sci Unit States Am. 2003; 100: 247-52.

7. Floudas C, Fung H, McAllister S, Mönnigmann M, Rajgaria R. Advances in protein structure prediction and de novo protein design: A review. Chem Eng Sci. 2006; 61:966-88.

8. Rahman A, Zomaya AY. An overview of protein-folding techniques: issues and perspectives. Int J Bioinformatics Res Appl. 2005; 1: 121-43.

9. Gish W, States DJ. Identification of protein coding regions by database similarity search. Nat Genet. 1993; 3: 266.

10. Fiser A. Protein structure modeling in the proteomics era. Expet Rev Proteonomics. 2004; 1: 97-110.

11. 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: Springer. 2005: 571-607.

12. Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC bioinformatics. 2007; 8: 4.

13. Yu CS, Cheng CW, Su WC, et al. CELLO2GO: a web server for protein subCELlular LOcalization prediction with functional gene ontology annotation. PloS one. 2014; 9: e99368.

14. Geourjon C, Deleage G. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Bioinformatics. 1995; 11: 681-4.

15. Tsirigos KD, Peters C, Shu N, Käll L, Elofsson A. The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. Nucleic Acids Res. 2015; 43: W401-W7.

16. Krogh A, Larsson B, Von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol. 2001; 305: 567-80.

17. Bagos PG, Liakopoulos TD, Spyropoulos IC, Hamodrakas SJ. PRED-TMBB: a web server for predicting the topology of β-barrel outer membrane proteins. Nucleic Acids Res. 2004; 32: W400-W4.

18. Petersen TN, Brunak S, Von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Protocol. 2011; 8: 785.

19. Chen CC, Hwang JK, Yang JM. 2-v2: template-based protein structure prediction server. Bmc Bioinformatics. 2009; 10: 366.

20. Schwede T, Kopp J, Guex N, Peitsch MC. SWISS-MODEL: an automated protein homology-modeling server. Nucleic Acids Res. 2003; 31: 3381-5.

21. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-Pdb Viewer: an environment for comparative protein modeling. electrophoresis. 1997; 18: 2714-23.

22. Wu S, Zhang Y. LOMETS: a local meta-threading-server for protein structure prediction. Nucleic Acids Res. 2007; 35: 3375-82.

23. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protocol. 2015; 10: 845.

24. Carugo O, Djinović-Carugo K. Half a century of Ramachandran plots. Acta Crystallographica Section D: Biological Crystallography. 2013; 69: 1333-41.

25. Xu D, Zhang Y. Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. Biophys J. 2011; 101: 2525-34.

26. Lomize MA, Pogozheva ID, Joo H, Mosberg HI, Lomize AL. OPM database and PPM web server: resources for positioning of proteins in membranes. Nucleic Acids Res. 2011; 40: D370-D6.

27. Roy A, Yang J, Zhang Y. COFACTOR: an accurate comparative algorithm for structure-based protein function annotation. Nucleic Acids Res. 2012; 40: W471-W7.

28. Berezin C, Glaser F, Rosenberg J, Paz I, Pupko T, Fariselli P, et al. ConSeq: the identification of functionally and structurally important residues in protein sequences. Bioinformatics, 2004; 20 (8): 1322-4.

29. Negi SS, Schein CH, Oezguen N, Power TD, Braun W. InterProSurf: a web server for predicting interacting sites on protein surfaces. Bioinformatics. 2007; 23: 3397-9.

30. Kawabata T. Detection of multiscale pockets on protein surfaces using mathematical morphology. Proteins: Structure, Function, and Bioinformatics. 2010; 78: 1195-211.

31. Dundas J, Ouyang Z, Tseng J, Binkowski A, Turpaz Y, Liang J. CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. Nucleic Acids Res. 2006; 34: W116-W8.

32. Tan KP, Nguyen TB, Patel S, Varadarajan R, Madhusudhan MS. Depth: a web server to compute depth, cavity sizes, detect potential small-molecule ligand-binding cavities and predict the pKa of ionizable residues in proteins. Nucleic Acids Res. 2013; 41: W314-W21.

[DOI: 10.29252/JoMMID.7.4.93]

33. Vita R, Overton JA, Greenbaum JA, et al. The immune epitope database (IEDB) 3.0. Nucleic Acids Res. 2014; 43: D405-D12.

34. Saha S, Raghava G. BcePred: prediction of continuous B-cell epitopes in antigenic sequences using physico-chemical properties. In: ICARIS. Springer: 197-204.

35. Jespersen MC, Peters B, Nielsen M, Marcatili P. BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. Nucleic Acids Res. 2017; 45: W24-W9.

36. Yao B, Zhang L, Liang S, Zhang C. SVMTriP: a method to predict antigenic epitopes using support vector machine to integrate tri-peptide similarity and propensity. PloS one. 2012; 7: e45152.

37. Davies MN, Flower DR. Harnessing bioinformatics to discover new vaccines. Drug Discov Today. 2007; 12: 389-95.

38. 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.

39. Uehling DT, Hopkins WJ, Elkahwaji JE, Schmidt DM, Leverson GE. Phase 2 clinical trial of a vaginal mucosal vaccine for urinary tract infections. J Urol. 2003; 170: 867-9.

40. Bauer HW, Alloussi S, Egger G, et al. A long-term, multicenter, double-blind study of an *Escherichia coli* extract (OM-89) in female patients with recurrent urinary tract infections. J Urol. 2005; 47: 542-8.

41. Goluszko P, Goluszko E, Nowicki B, Nowicki S, Popov V, Wang H-Q. Vaccination with purified Dr Fimbriae reduces mortality associated with chronic urinary tract infection due to *Escherichia coli* bearing Dr adhesin. Infect Immun. 2005; 73: 627-31.

42. Jahangiri A, Rasooli I, Gargari SLM, et al. An *in silico* DNA vaccine against Listeria monocytogenes. Vaccine. 2011; 29: 6948-58.

43. Sefid F, Rasooli I, Jahangiri A. *In silico* determination and validation of baumannii acinetobactin utilization a structure and ligand binding site. Biomed Res. 2013; 2013.

44. Payandeh Z, Rajabibazl M, Mortazavi Y, Rahimpour A, Taromchi AH. Ofatumumab monoclonal antibody affinity maturation through *in silico* modeling. IBJ. 2018; 22: 180.

45. Payandeh Z, Rajabibazl M, Mortazavi Y, Rahimpour A. *In Silico* Analysis for Determination and Validation of Human CD20 Antigen 3D Structure. Int J Pept Res Therapeut. 2019; 25: 123-35.

46. Payandeh Z, Rajabibazl M, Mortazavi Y, Rahimpour A, Taromchi AH, Dastmalchi S. Affinity maturation and characterization of the ofatumumab monoclonal antibody. J Cell Biochem. 2019; 120: 940-50.

47. Kleywegt GJ, Jones TA. Databases in protein crystallography. Acta Crystallographica Section D: Biological Crystallography. 1998; 54: 1119-31.

48. Mislin GL, Schalk IJ. Siderophore-dependent iron uptake systems as gates for antibiotic Trojan horse strategies against Pseudomonas aeruginosa. Metallomics. 2014; 6: 408-20.

49. Maiorov VN, Crippen GM. Significance of root-mean-square deviation in comparing three-dimensional structures of globular proteins. 1994.

50. Bagos PG, Liakopoulos TD, Hamodrakas SJ. Evaluation of methods for predicting the topology of β-barrel outer membrane proteins and a consensus prediction method. BMC bioinformatics. 2005; 6:7.

Cite this article:

Sefid F, Baghban R, Payandeh Z, Khalesi B, Gomari MM. Structure Evaluation of IroN for Designing a Vaccine against *Escherichia Coli*, an *In Silico* Approach. J Med Microbiol Infect Dis, 2019; 7 (4): 93-106. DOI: 10.29252/JoMMID.7.4.93