# A Chimeric Vaccine Consisting of Highly Immunogenic Regions Form *Escherichia coli* Iron Regulated Outer-Membrane Proteins: An In Silico Approach

Fatemeh Sefid[1] (iD), Mahsa Akbari Oryani[2] (iD), Maryam Mehdi[3] (iD), Zahra Payandeh[4] (iD), Saeed Khalili[5] (iD), Ehsan Kaffash[6] (iD), Ghasem Azamirad[7*] (iD), Seyed Mehdi Kalantar[8,9] (iD)

1. Department of Biology Sciences, School of Materials Engineering and Interdisciplinary Sciences, Shahid Sadoughi University of Medical Sciences, Yazd, Iran.
2. Department of Pathology, School of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran.
3. Department of Pharmacology, Manipal College of Pharmaceutical Sciences, Manipal Academy of Higher Education, Manipal, India.
4. Immunology Research Center, Tabriz University of Medical Sciences, Tabriz, Iran.
5. Department of Biology Sciences, School of Materials Engineering and Interdisciplinary Sciences, Shahid Rajaee Teacher Training University, Tehran, Iran.
6. Department of Pharmaceutics, School of Pharmacy, Mashhad University of Medical Sciences, Mashhad, Iran.
7. Department of Mechanical Engineering, Yazd University, Yazd, Iran.
8. Department of Medical Genetics, Shahid Sadoughi University of Medical Science, Yazd, Iran.
9. Research and Clinical Center for Infertility, Reproduction Sciences Institute, Shahid Sadoughi University of Medical Sciences, Yazd, Iran.

## ABSTRACT

**Background:** Six pathogen-associated Outer Membrane Iron receptors (OMPs) reside in Uropathogenic strains of *E. coli* (UPEC): haem-utilization gene (ChuA), Heme acquisition protein (Hma), IrgA homologue adhesin (Iha), Iron-regulated virulence gene (IreA), IroN, and IutA. Cumulative concern over the prevalence of this bacteria in hospital environments, especially in Intensive Care Units (ICUs), highlights the significance of vaccination against this pathogen. In this study, we aimed to develop 3D models of ChuA, Hma, IutA, IreA, Iha, and IroN proteins by invoking various in silico methods and design a chimeric immunogen composed of highly immunogenic regions from these six *Escherichia coli* antigens as a chimeric vaccine.

**Materials and Methods:** In the present study, homology modeling, fold recognition, Ab initio approaches, and their combination were invoked to determine the Three-Dimensional (3D) structures of ChuA, Hma, Iha, IreA, IroN, and IutA. Next, a set of biochemical, immunological, and functional properties were predicted using various bioinformatics tools.

**Results:** The obtained results indicated that all six modeled proteins fold to a β-barrel structure. The results of biochemical, immunological, and functional analysis determined the regions of each antigen carrying the best immunogenic properties. These regions are employed to construct the final vaccine linked via flexible GGGGS linkers. Intriguingly, re-analyzing the properties of the final vaccine indicated its immunological advantage over individual proteins.

**Conclusion:** The strategy of this study to predict the protein 3D structure, followed by epitope prediction, could be adapted to design efficient vaccine candidates. Applying this approach, we designed a vaccine candidate harboring the most promising regions of six OMPs. This approach could lead to better functional, structural, and therapeutic outcomes in the context of vaccine design investigations.

**\* Corresponding Author:**
*Ghasem Azamirad, PhD.*
*Address: Department of Mechanical Engineering, Yazd University, Yazd, Iran.*
*Phone: +98 (913) 3578557*
*E-mail: azamirad.Gh@gmail.com*

## Introduction

Pathogenic microorganisms have remained one of the most serious public health threats. Although conventional vaccines effectively treat or eradicate some pathogens, they are not so efficient against some pathogenic microorganisms. Conventional methods of vaccine production require pathogen culture and identification of its immunogenic components. This process is a time-consuming method and can only detect antigens that are highly expressed. Some of the antigenic proteins are not always highly viable and purifiable. Sometimes antigens produced under living cell conditions (during pathogenesis) could not be produced in vitro.

On the other hand, such methods will not work well enough for non-cultured microorganisms. Thus, the emergence of computer-related technologies is a new way to study protective antigens, including vaccine design studies [1-3]. Urinary Tract Infections (UTIs) are bacterial involvements affecting the urinary tract. To evaluate pathogenicity in the urinary system, Uropathogenic strains of *E. coli* (UPEC) use various virulence factors. UPEC has a significant level of resistance, particularly against multiple antibiotics. Several vaccinations have been tested against UTIs so far, with controversial results [4]. In over 80% of uncomplicated UTIs [5], *Escherichia coli* is the most important infectious bacterium in people with the normal urinary tract structure without any disorder or inflammation [6]. Other than related acute cystitis and pyelonephritis, several complications can appear following the UTIs. Permanent renal damages (in pediatric upper UTIs) and kidney scarring (in approximately 57% of children with acute pyelonephritis) are among these complications.

On the other hand, the incidence of antibiotic resistance agents in these infections is reportedly rising [6, 7]. In this regard, various research studies have been conducted looking for amenable approaches to induce immunity against UPEC. Relatively short-term protection has recently been established in some patients via injecting whole cell or cell extraction [8]. In the category of subunit vaccines, some abundant proteins like type 1 fimbrial adhesin and FimH adhesin (in the outer membrane of bacteria) are suitable candidates for vaccination against UPEC [9, 10].

Iron is one of the essential sources for the proper growth of most pathogens, including *Escherichia coli*. Depleted amounts of soluble essential elements (like iron) under aerobic conditions or physiological pH could create an undesirable bacteria environment. In this regard, the proteins related to iron acquisition metabolism like Outer Membrane Proteins (OMPs) [11] and heme and siderophore receptors [12] are suitable antigenic targets for immunization against UPEC. According to Mobley et al. study [13], the antigenic OMPs are scientifically acceptable multivalent vaccine targets to control the UPEC. They introduced six pathogen-associated outer membrane iron receptors in *E. coli*: haem-utilization gene (ChuA), Heme acquisition protein (Hma), IrgA homologue adhesin (Iha), iron-regulated virulence gene (IreA), IroN, and IutA. The molecular weight of these proteins ranges between 71 and 84 kDa, which expectedly can construct extracellular loop-shaped transmembrane β-barrels in the outer membrane [14]. These receptors provide the possibility of penetration of the specific iron sources as one of the most significant elements in the UPEC pathogenesis. Because of the iron shortage in the urinary tract, iron uptake through these receptors is crucial [15]. According to a study on murine models, the colonization of UPEC in the urinary tract is restricted by eliminating the siderophore receptor IreA, heme receptors ChuA, Hma, enterobactin receptor Iha, salmochelin receptor IroN, or aerobactin receptor IutA [16].

The effective control of the infections associated with *E. coli* is possible by identifying the nature and the role of ChuA, Hma, Iha, IreA, IroN, and IutA proteins in these infections. The functions and interactions of proteins with other compounds, such as ligands, could be recognized by determining their tertiary structure [17-20]. The iron-regulated genes have been found to play a vital role in the adherence of avian pathogenic *Escherichia coli* strains. In alkaline, hyper-osmolality, and low-temperature conditions, these genes will boost stress resistance. As a result, the siderophore receptors' redundancy may reflect their multifunctional activities. These genes were mostly found in phylogenetic ECOR groups B and D, which are more virulent. Compared to the wild-type strain, the adhesion and resilience to environmental stress were considerably reduced in these gene deletion mutants [21].

Knowing more about the 3D structure of proteins plays a pivotal role in their rational modification and engineering [22, 23], drug and vaccine design [24, 25], and conformational epitope predictions [26, 27]. The need to determine tertiary protein structures via in silico methods is more evident regarding the considerable number of known protein sequences versus the limited number of structural annotations [28, 29]. Executing imperial methods of 3D structure determination has serious difficulties like high failure rate, high cost, and time-consum-

ing process, so finding alternative methods is necessary [30]. Additionally, the purification and crystallization of OMPs face various problems. Biologists could benefit from bioinformatics approaches, such as the 3D protein structure prediction using different methods and algorithms like homology modeling. Homology modeling is an in silico method to predict the 3D protein structures using a homologous protein structure as a template. Evidence supports the crystallographic structures of some homologous OMPs in other pathogens [30]. Nevertheless, further research is required to recognize the 3D protein structure for structurally unresolved proteins.

In the present study, we aimed to determine the 3D structure of the ChuA, Hma, Iha, IreA, IroN, and IutA proteins. The 3D structures of these proteins would help us to predict the linear and conformational B cell epitopes that reside within their sequences. Given this information, the most immunogenic regions of the antigens could be determined and utilized to design a multivalent vaccine connected with flexible linkers. Bacteria acquire iron in complex forms using different strategies because of the inadequate amount of its free form in biological fluids. Immunological targeting of all antigens involved in iron metabolism is a novel strategy capable of blocking all possible iron uptake mechanisms. This condition would not let any iron uptake mechanism compensate for iron depletion exerted by conventional vaccines targeting a single iron uptake antigen.

## Materials and Methods

### Sequence retrieval and homology modeling

The NCBI database has enlisted protein sequences of ChuA, Hma, IutA, IreA, Iha, and IroN as six vaccine candidate antigens [31]. All of the attained sequences were stored as FASTA file format to be used as input data for the following analyses. The protein BLAST (basic local alignment search tool) from the NCBI database was employed to search for similar sequences of the obtained six iron receptor vaccine candidates. The sequences of the iron receptors were fed as query sequences, and the BLAST was done against a non-redundant protein dataset. Moreover, the BLAST tool was used to search for probable putative conserved domains of the query proteins [32, 33]. The first step to perform reliable homology modeling is to find an appropriate template structure. Thus, we used the protein BLAST tool from the NCBI database. In this regard, the protein sequences of six iron receptors were used as input data for the PSI-BLAST, while the search was limited to the structures stored in Protein Data Bank (PDB) [34].

### Further scrutiny

VaxiJen server at [35] as an alignment-free approach for antigen prediction was used to determine the probability of antigenicity for six iron receptor vaccine candidates. The average for physiochemical properties of the iron receptors was estimated by the IEDB server [36]. Several physicochemical properties of each iron receptor protein were determined using the ProtParam server. The determined properties included instability index, aliphatic index, the total number of charged residues, amino acid composition, theoretical pI, and molecular weight [37]. The CELLO v.2.5 [38] was used to determine the possible sub-cellular localization of the vaccine candidates. The subCELlular LOcalization predictor and PSLpred server (A SVM-based method for the sub-cellular localization of prokaryotic proteins), were also used to predict the sub-cellular localization of the vaccine candidate [39]. The sequences of six iron receptors were checked by SignalP 4.1 server and PrediSi server for the presence and location of any signal peptide cleavage sites [40].

### Transmembrane protein topology and secondary protein structure prediction

The sequences of six iron receptor vaccine candidates were fed as input to the PRED-TMBB server. This server uses a hidden Markov model to predict the hydrophobic transmembrane regions of the protein sequences from the Gram-negative bacteria outer membrane proteins capable of forming probable β-barrel [41]. The secondary structure of the six iron receptor vaccine candidates was predicted by the SOPMA server [42]. The SWISS-MODEL server also predicted the secondary structure of the proteins.

### Protein modeling

The SWISS-MODEL Workspace [43] is a web-based integrated service used for homology modeling of six iron receptor proteins. This server is a fully automated protein structure homology-modeling server that assists and guides the user in building protein homology models at different levels of complexity [44]. LOMETS (Local Meta-Threading-Server) online web service [45] was employed for 3D protein structure prediction. This server collects high-scoring target-to-template alignments from 10 locally installed threading programs (FUGUE, HH-SEARCH, MUSTER, PPA, PROSPECT2, SAM-T02, SPARKS, SP3, FFAS, and PRC) to build its 3D models.

## Model quality assessment and refinement

The quality of the models built by the SWISS-MOD-EL was assessed by GMQE and QMEAN4 scores using the QMEAN4 server. LOMETS confidence score was used to assess the quality models built by the LOMETS server. Ramachandran plots were also calculated for all models by Rampage server [46]. Atomic-level, high-resolution protein structure refinement of the built models was carried out using the ModRefiner server [47]. Aside from the significant improvement in the physical quality of local structures, ModRefiner could draw the initial starting models closer to their native state in terms of structural properties.

## Single-scale prediction of amino acid properties and epitope prediction

The properties of the six iron receptor protein sequences, which were correlated with the location of B cell epitopes of hydrophilicity, flexibility, accessibility, turns, and the antigenic propensity of the polypeptide, were predicted using the IEDB server tool [36]. Using a combination of a hidden Markov model and a propensity scale method, BepiPred was employed to predict the location of linear B-cell epitopes of six iron receptor protein sequences [48]. SVMTriP was the other server to predict the antigenic epitopes within input sequences [49]. The predicted structure of the six iron receptor vaccine candidates was used as input files for predicting discontinuous B cell epitopes. DiscoTope was used to predict the location of discontinuous B cell epitopes [26]. ElliPro was the other server to predict linear and discontinuous antibody epitopes based on a protein antigen's 3D structure [50].

## Ligand binding site predictions and structure alignment

We used the COFACTOR server to annotate the biological function of six iron receptor protein molecules and find their essential amino acid involved in the ligand-binding site [51]. Secondary structures based on the alignment of the candidate sequences were prepared by the PRALINE server. The alignments of the PRALINE server are generated based on exchange weights matrix BLOSUM62 and associated gap penalties [52].

## Selection of immunogenic regions

The regions with the highest density of continuous and discontinuous epitopes were selected as proper vaccine candidate regions. The properties obtained from single-scale amino acid properties assay, probability of antigenicity, and physicochemical properties average were also considered to select desired region selection. Given these properties, six regions were selected as appropriate antigenic regions in six vaccine candidates. Further analyses by the VaxiJen server were performed on the selected regions to validate the selected protein segments.

## Final vaccine design

The selected regions were connected by a flexible linker (GGGGS). This linker can improve the folding and stability of fusion proteins. It would allow the correct orientation and not interfere with the folding of the protein domains [53].

## Final vaccine evaluation

The final vaccine was evaluated in terms of physicochemical, structural, immunogenicity, allergenicity, and protein expression in the appropriate expression system. VaxiJen server was used to evaluate the immunogenicity. Several physicochemical properties were calculated using the ProtParam server. AllergenFP v.1.0 was used for allergenicity prediction. SoluProt was used for the prediction of soluble protein expression in *Escherichia coli*. Computational calculations were used to predict the 3D structure of protein molecules based on their amino acid sequence. The spatial location of every atom in the protein structure should be determined to arrive at the 3D structure of the protein. The Zhang-Server has developed several algorithms for protein 3D structure prediction. Amongst, the MUSTER and LOMETS servers are used for protein template structure identification, the I-TASSER server is used for iterative protein structure assembly and the QUARK server for ab initio protein folding. We used I-TASSER (iterative protein structure assembly) server to predict the final vaccine 3D structure.

## Results

### Sequence availability and homology search

The sequences for six iron receptor vaccine candidates of ChuA, Hma, IutA, IreA, Iha, and IroN, were found and saved as FASTA format under the NCBI accession numbers of AAC44857.1, AAN80973.1, AAS66997.1, AMR36194.1, ABB17254.1, and AAS80269.1, respectively. BLAST search returned numerous hits with high similarity to the query sequences. Amongst, some hits were putatively conserved domains, and some belonged to bacteria other than *Escherichia coli*. The sequences mainly belonged to the outer membrane-channels super-

family, TonB dependent/ligand-gated channels, and ligand-gated-channel protein family. The information about the classifications of each iron receptor vaccine candidate has been summarized in Supplementary Table 1. The BLAST search on the iron receptor vaccine candidate as query sequences against Protein Data Bank (PDB) resulted in several hits with different scores and identities. The first hit of the BLAST search, corresponding to the highest score, was selected as a template for the following homology modeling process (Table 1).

### Further scrutiny

Various properties of iron receptor vaccine candidates, including VaxiJen antigenicity score, number of amino acids, other physicochemical properties, localization, Cello score, and PSLpred accuracy, were successfully calculated and presented in Table 2. The function of a protein is related to its subcellular localization because the environment of a protein provides a part of the relevant context necessary for function. So that the subcellular location of a protein can provide valuable information about its function.

The SignalP and PrediSi servers revealed the cleavage site of a signal peptide for six iron receptor vaccine candidates. Table 3 presents the results of signal peptide predictions.

### Topology and secondary structure prediction

Three main topological regions of the protein, including the inside, outside, and transmembrane regions, were predicted and used to build 2D topology models of all six iron receptor vaccine candidates (Figure 1). Our results indicate that iron receptor proteins are composed of several transmembrane antiparallel β-strands. The model suggests that the proteins are form β-barrel structures in their native state. The topology models indicate that the strands forming β-barrel are linked together through loops at the outside or turns at the inside. The main components constituting the secondary structures of the six iron receptor vaccine candidates are coil, helix, and strands. The secondary structure could be used to validate the tertiary structures. Alpha helix, extended strand, beta-turn, and random coil are the attribution of secondary structure components in the proteins. The composition of secondary structures is shown as a percentage of each secondary structure (Table 3).

### Protein 3D structure prediction

SWISS-MODEL homology modeling server managed to predict 2 models for ChuA, 3 models for Hma, 2 models for Iha, 4 models for IreA, 2 models for IroN, and 1 model for IutA. SWISS-MODEL validates the quality of the predicted 3D structures by QMEAN and GMQE scores. The structural properties of each predicted model are summarized in Supplementary Table 2. LOMETS Meta server predicted 10 models with its locally installed different programs for each vaccine candidate protein. The properties of each predicted model are summarized in Supplementary Table 3. All models showed high confidence scores.

### Evaluating and refining the predicted models

Each vaccine candidate protein by four independent scores revealed a consensus over a single model. Among the predicted models, the models built by LOMETS (Supplementary Table 4) showed outstanding Ramachandran quality scores bearing the high number of resi-

**Supplementary Table 1.** Superfamily information for each candidate

| Protein Name | Name | Accession | Description |
|---|---|---|---|
| ChuA | CirA superfamily | cl26861 | Outer membrane receptor proteins, mostly Fe transport [Inorganic ion transport and metabolism]; |
| Hma | Ligand_gated_channel | cd01347 | TonB dependent/ligand-gated channels are created by a monomeric 22 strand [22, 24] |
| Iha | PRK13486 | PRK13486 | Bifunctional enterobactin receptor/adhesin protein; Provisional |
| IreA | PRK13484 | PRK13484 | Putative iron-regulated outer membrane virulence protein; Provisional |
| IroN | PRK13528 | PRK13528 | Outer membrane receptor FepA; Provisional |
| IutA | TonB-siderophor | TIGR01783 | TonB-dependent siderophore receptor |

ChuA: haem-utilization gene Hma : Heme acquisition protein Iha : IrgA homologue adhesin IreA: iron-regulated virulence gene IroN: the salmochelin siderophore receptor IroN IutA: receptor binding domain of colicin Ia

**Table 1.** Basic Local Alignment Search Tool (BLAST) on the query sequences against Protein Data Bank (PDB)

| Protein Name | Description | Max Score | Total Score | Query Cover (%) | E Value | Identity (%) | Accession |
|---|---|---|---|---|---|---|---|
| ChuA | Chain A, the crystal structure of the heme/hemoglobin Outer membrane transporter ShuA from S*higella dysenteriae* | 1288 | 1288 | 95 | 0.0 | 98 | 3FHH_A |
| HmA | Chain A, Fhua from *E. coli* | 103 | 103 | 90 | 2e-22 | 23 | 1BY3_A |
| Iha | Chain A, the crystal structure of The colicin I receptor Cir from *E. coli* in complex with receptor binding domain of colicin Ia | 283 | 283 | 92 | 9e-86 | 32 | 2HDI_A |
| IreA | Chain A, the crystal structure of The colicin I receptor Cir from *E. coli* in complex with receptor binding domain of colicin Ia | 354 | 354 | 96 | 6e-113 | 36 | 2HDI_A |
| IroN | Chain A, crystal structure of the siderophore receptor PirA from *Pseudomonas aeruginosa* | 843 | 843 | 95 | 0.0 | 60 | 5FP2_A |
| IutA | Chain A, the crystal structure of the colicin I receptor Cir from *E. coli* in complex with receptor binding domain of colicin Ia | 62.0 | 62.0 | 17 | 2e-09 | 33 | 2HDI_A |

dues in the favored region and the lowest number in the outlier region. It should be noted that lower RMSD and higher TM-score/GDT-TS indicate that ModRefiner is more potent in drawing the initial models closer to their native-like state. The obtained RMSD and TM scores for the generated models showed improvement in the global topology of the initial models (Supplementary Table 5).

### Determining the single-scale amino acid properties and B cell epitopes

IEDB server has predicted several properties for each vaccine candidate, including antigenicity, hydrophilicity, and accessibility. The average values of single-scale amino acid properties along with the six candidate sequences are presented in Table 4. The BepiPred server predicts that linear B cell epitopes are more prevalent in

**Supplementary Table 2.** SWISS-MODEL homology modeling predicted models

| Protein Name | Id | Template | GQME | QMEAN |
|---|---|---|---|---|
| ChuA | 1 | 3fhh.1.A | 0.98 | -2.32 |
| | 2 | 3pgu.1.A | 0.11 | -6.97 |
| Hma | 1 | 1qfg.1.A | 0.58 | -5.22 |
| | 2 | 5fp1.1.A | 0.54 | -5.24 |
| | 3 | 3efm.1.A | 0.41 | -6.39 |
| Iha | 1 | 5fr8.1.A | 0.61 | -4.51 |
| | 2 | 5fp2.1.A | 0.42 | -5.62 |
| IreA | 1 | 5fr8.1.A | 0.66 | -3.46 |
| | 2 | 2hdf.1.A | 0.61 | -4.68 |
| | 3 | 3v89.1.A | 0.52 | -5.78 |
| | 4 | 3fhh.1.A | 0.44 | -6.07 |
| IroN | 1 | 5fr8.1.A | 0.77 | -1.84 |
| | 2 | 3pgu.1.A | 0.10 | -5.95 |
| IutA | 1 | 5fp1.1.A | 0.51 | -5.85 |

**Table 2.** Immunological and physicochemical properties of the vaccine candidate

| Protein Name | VaxiJen Score | Number of Amino Acids | Molecular Weight | Theoreti-cal pI | Instability Index | Aliphat-ic Index | GRAVY* | Localization | Cello Score | PSLpred Accuracy (%) |
|---|---|---|---|---|---|---|---|---|---|---|
| ChuA | 0.6266 | 660 | 72429.05 | 5.17 | 28.00 (stable) | 67.58 | -0.474 | Outer membrane | 4.558 | 98.1 |
| Hma | 0.7259 | 721 | 78404.78 | 5.72 | 31.29 (stable) | 68.46 | -0.478 | Outer membrane | 4.530 | 98.1 |
| Iha | 0.6138 | 696 | 76481.09 | 5.64 | 30.45 (stable) | 74.96 | -0.426 | Outer membrane | 4.534 | 98.1 |
| IreA | 0.6570 | 682 | 75291.25 | 6.15 | 32.38 (stable) | 79.93 | -0.487 | Outer membrane | 3.909 | 90.2 |
| IroN | 0.7889 | 725 | 79134.54 | 5.78 | 33.97 (stable) | 76.79 | -0.552 | Outer membrane | 4.635 | 90.2 |
| IutA | 0.6016 | 732 | 81048.36 | 5.49 | 32.25 (stable) | 77.66 | -0.412 | Outer membrane | 4.886 | 98.1 |

* Grand average of hydropathicity

the vicinity of extracellular loops. These regions show the presence of a high density of linear epitopes. SVM-Trip predicted 10 linear B cell epitopes ranking based on their scores in the six iron receptor vaccine candidates. The best epitopes with the highest scores recommended by this server are presented in Table 5. Linear and discontinuous B cell epitopes were predicted by ElliPro software. The best linear epitopes were determined by ElliPro in all of the six iron receptor vaccine candidates were located at the largest extracellular loops. In this regard, discontinuous B cell epitopes predicted from the 3D structure of proteins include all the extracellular loops. Discontinuous B cell epitopes predicted from the 3D structure of a protein by Disco Tope are shown in Figure 2. Residues are colored by Disco Tope score from Red (high score) to Blue (low score). As the figures show, Disco Tope and Ellipro PI scores are the highest at the outer membrane loops.

**Table 3.** Percentage of each secondary structure and signal peptide cleavage site for the candidates

| Protein Name | Alpha Helix (Hh) | Extended Strand (Ee) | Beta Turn (Tt) | Random Coil (Cc) | S.p. Cleavage Site |
|---|---|---|---|---|---|
| ChuA | 133 is 20.15% | 165 is 25.00% | 77 is 11.67% | 285 is 43.18% | Between pos.28 and29 |
| Hma | 169 is 23.44% | 159 is 22.05% | 76 is 10.54% | 317 is 43.97% | Between pos. 24 and 25 |
| Iha | 197 is 28.30% | 165 is 23.71% | 75 is 10.78% | 259 is 37.21% | Between pos. 22 and 23 |
| IreA | 113 is 16.57% | 195 is 28.59% | 87 is 12.76% | 287 is 42.08% | No Signal peptide |
| IroN | 119 is 16.41% | 197 is 27.17% | 74 is 10.21% | 335 is 46.21% | Between pos. 24 and 25 |
| IutA | 172 is 23.50% | 201 is 27.46% | 80 is 10.93% | 279 is 38.11% | Between pos. 25 and 26 |

**Table 4.** Average value of single-scale amino acid properties along with the six candidate sequences

| Protein Name | Linear Epitope | Beta Turn | Surface Accessibility | Flexibility | Antigenicity | Hydrophilicity |
|---|---|---|---|---|---|---|
| ChuA | 0.372 | 1.039 | 1.000 | 1.012 | 1.003 | 2.012 |
| Hma | -0.008 | 0.925 | 1.000 | 0.986 | 1.043 | 1.171 |
| Iha | 0.275 | 1.027 | 1.000 | 1.006 | 1.012 | 1.884 |
| IreA | 0.319 | 1.018 | 1.000 | 1.010 | 1.008 | 1.963 |
| IroN | 0.427 | 1.059 | 1.000 | 1.014 | 1.001 | 2.197 |
| IutA | 0.227 | 1.030 | 1.000 | 1.004 | 1.014 | 1.777 |

Research in Molecular Medicine

**Supplementary Table 3.** Models properties predicted by LOMETS Meta server

| Protein Name | Rank | Template | Align_length | Coverage | Zscore | Seq_id | Confidence Score | Program |
|---|---|---|---|---|---|---|---|---|
| ChuA | 1 | 3fhhA | 621 | 0.940 | 153.000 | 0.99 | High | FFAS03 |
| | 2 | 3fhhA | 618 | 0.936 | 378.800 | 0.98 | High | PRC |
| | 3 | 3fhha | 621 | 0.940 | 59.807 | 0.98 | High | SP3 |
| | 4 | 3fhhA0 | 621 | 0.940 | 48.577 | 0.99 | High | pGenTHREADER |
| | 5 | 3fhhA | 621 | 0.940 | 23.395 | 0.99 | High | PROSPECT2 |
| | 6 | 3fhh_A | 618 | 0.936 | 218.512 | 0.99 | High | FFAS-3D |
| | 7 | 3fhhA | 620 | 0.939 | 61.061 | 0.99 | High | Neff-PPAS |
| | 8 | 3fhhA | 621 | 0.940 | 33.950 | 0.99 | High | SPARKS-X |
| | 9 | 3fhhA | 621 | 0.940 | 48.974 | 0.99 | High | wdPPAS |
| | 10 | 3fhhA | 621 | 0.940 | 26.117 | 0.99 | High | MUSTER |
| Hma | 1 | 3qlbA | 658 | 0.912 | 29.102 | 0.21 | High | MUSTER |
| | 2 | 1fi1A | 651 | 0.902 | 402.900 | 0.22 | High | PRC |
| | 3 | 3qlba | 658 | 0.912 | 72.299 | 0.21 | High | SP3 |
| | 4 | 1by3_A | 662 | 0.918 | 277.220 | 0.23 | High | FFAS-3D |
| | 5 | 3qlbA | 657 | 0.911 | 172.000 | 0.21 | High | FFAS03 |
| | 6 | 4cu4A0 | 653 | 0.905 | 48.577 | 0.22 | High | pGenTHREADER |
| | 7 | 3qlbA | 658 | 0.912 | 22.892 | 0.21 | High | PROSPECT2 |
| | 8 | 3qlbA | 657 | 0.911 | 59.786 | 0.21 | High | Neff-PPAS |
| | 9 | 3qlbA | 658 | 0.912 | 57.474 | 0.21 | High | wdPPAS |
| | 10 | 3qlbA | 658 | 0.912 | 36.250 | 0.22 | High | SPARKS-X |
| Iha | 1 | 1fepA | 645 | 0.926 | 30.750 | 0.26 | High | SPARKS-X |
| | 2 | 1fepA | 619 | 0.889 | 369.400 | 0.26 | High | PRC |
| | 3 | 5fr8a | 642 | 0.922 | 59.297 | 0.26 | High | SP3 |
| | 4 | 5fr8A | 642 | 0.922 | 21.227 | 0.25 | High | PROSPECT2 |
| | 5 | 5fr8A | 642 | 0.922 | 131.000 | 0.26 | High | FFAS03 |
| | 6 | 1fepA0 | 647 | 0.929 | 39.249 | 0.26 | High | pGenTHREADER |
| | 7 | 5fr8_A | 642 | 0.922 | 186.000 | 0.26 | High | FFAS-3D |
| | 8 | 1fepA | 641 | 0.920 | 51.679 | 0.25 | High | Neff-PPAS |
| | 9 | 1fepA | 643 | 0.923 | 41.273 | 0.25 | High | wdPPAS |
| | 10 | 2gskA | 590 | 0.847 | 21.946 | 0.23 | High | MUSTER |

| Protein Name | Rank | Template | Align_length | Coverage | Zscore | Seq_id | Confidence Score | Program |
|---|---|---|---|---|---|---|---|---|
| IreA | 1 | 5fr8_A | 634 | 0.929 | 180.000 | 0.28 | High | FFAS-3D |
| | 2 | 5fr8A | 603 | 0.884 | 360.800 | 0.29 | High | PRC |
| | 3 | 5fr8a | 634 | 0.929 | 61.110 | 0.28 | High | SP3 |
| | 4 | 5fr8A | 634 | 0.929 | 132.000 | 0.29 | High | FFAS03 |
| | 5 | 5fr8A | 634 | 0.929 | 21.101 | 0.26 | High | PROSPECT2 |
| | 6 | 2hdiA0 | 568 | 0.832 | 40.060 | 0.38 | High | pGenTHREADER |
| | 7 | 2hdiA | 576 | 0.844 | 50.431 | 0.38 | High | Neff-PPAS |
| | 8 | 1fepA | 624 | 0.914 | 30.460 | 0.26 | High | SPARKS-X |
| | 9 | 2gskA | 577 | 0.846 | 41.213 | 0.20 | High | wdPPAS |
| | 10 | 2gskA | 577 | 0.846 | 21.757 | 0.20 | High | MUSTER |
| IroN | 1 | 5fr8_A | 685 | 0.944 | 200.000 | 0.52 | High | FFAS-3D |
| | 2 | 5fr8A | 642 | 0.885 | 374.900 | 0.53 | High | PRC |
| | 3 | 5fr8a | 685 | 0.944 | 68.764 | 0.52 | High | SP3 |
| | 4 | 5fr8A | 685 | 0.944 | 23.174 | 0.52 | High | PROSPECT2 |
| | 5 | 1fepA0 | 656 | 0.904 | 46.052 | 0.52 | High | pGenTHREADER |
| | 6 | 5fr8A | 685 | 0.944 | 143.000 | 0.51 | High | FFAS03 |
| | 7 | 1fepA | 644 | 0.888 | 53.017 | 0.50 | High | Neff-PPAS |
| | 8 | 1fepA | 655 | 0.903 | 31.970 | 0.54 | High | SPARKS-X |
| | 9 | 5fr8A | 685 | 0.944 | 45.548 | 0.51 | High | wdPPAS |
| | 10 | 5fr8A | 685 | 0.944 | 24.400 | 0.53 | High | MUSTER |
| IutA | 1 | 3qlbA | 636 | 0.868 | 18.421 | 0.13 | High | MUSTER |
| | 2 | 5fr8A | 598 | 0.816 | 336.800 | 0.17 | High | PRC |
| | 3 | 3qlba | 637 | 0.870 | 57.514 | 0.14 | High | SP3 |
| | 4 | 3fhhA | 598 | 0.816 | 122.000 | 0.21 | High | FFAS03 |
| | 5 | 1by5a | 638 | 0.871 | 19.018 | 0.17 | High | PROSPECT2 |
| | 6 | 4cu4A0 | 613 | 0.837 | 37.198 | 0.17 | High | pGenTHREADER |
| | 7 | 3v89_A2 | 653 | 0.892 | 177.000 | 0.16 | High | FFAS-3D |
| | 8 | 3qlbA | 635 | 0.867 | 44.369 | 0.13 | High | Neff-PPAS |
| | 9 | 3fhhA | 597 | 0.815 | 24.790 | 0.20 | High | SPARKS-X |
| | 10 | 4aipA | 628 | 0.857 | 35.124 | 0.13 | High | wdPPAS |

**Supplementary Table 4.** The ramachandran plot structures validation represent the percentage of residues located in favored, allowed, and outlier regions

| Protein Name | Program | Favored Region (%) | Allowed Regionn (%) | Outlier Regionn (%) |
|---|---|---|---|---|
| ChuA | FFAS03 | 95.6 | 4.1 | 0.3 |
| | PRC | 95.4 | 4.0 | 0.6 |
| | SP3 | 95.1 | 3.8 | 1.1 |
| | pGenTHREADER | 77.7 | 15.2 | 7.1 |
| | PROSPECT2 | 88.8 | 7.0 | 4.3 |
| | FFAS-3D | 95.0 | 4.7 | 0.3 |
| | Neff-PPAS | 95.1 | 4.1 | 0.8 |
| | SPARKS-X | 95.3 | 3.6 | 1.1 |
| | wdPPAS | 95.0 | 4.3 | 0.8 |
| | MUSTER | 95.1 | 4.3 | 0.6 |
| | SWISS-MODEL | 95.7 | 3.5 | 0.8 |
| | SWISS-MODEL | 91.2 | 5.3 | 3.5 |
| Hma | MUSTER | 93.9 | 4.7 | 1.4 |
| | PRC | 92.5 | 5.1 | 2.4 |
| | SP3 | 93.9 | 4.3 | 1.8 |
| | FFAS-3D | 94.9 | 4.5 | 0.7 |
| | FFAS03 | 94.2 | 4.3 | 1.5 |
| | pGenTHREADER | 81.6 | 14.0 | 4.3 |
| | PROSPECT2 | 90.8 | 5.4 | 3.8 |
| | Neff-PPAS | 94.3 | 3.8 | 1.9 |
| | wdPPAS | 94.2 | 4.2 | 1.7 |
| | SPARKS-X | 94.6 | 3.2 | 2.2 |
| | SWISS-MODEL | 91.8 | 5.7 | 2.5 |
| | SWISS-MODEL | 91.8 | 6.5 | 1.7 |
| | SWISS-MODEL | 87.4 | 8.2 | 4.4 |
| Iha | SPARKS-X | 92.9 | 4.2 | 2.9 |
| | PRC | 88.6 | 7.8 | 3.6 |
| | SP3 | 93.2 | 5.2 | 1.6 |
| | PROSPECT2 | 86.9 | 9.8 | 3.3 |
| | FFAS03 | 94.2 | 3.2 | 2.6 |
| | pGenTHREADER | 83.6 | 11.8 | 4.6 |
| | FFAS-3D | 93.9 | 3.9 | 2.2 |
| | Neff-PPAS | 89.2 | 7.8 | 3.0 |
| | wdPPAS | 90.8 | 6.2 | 3.0 |
| | MUSTER | 93.2 | 4.0 | 2.7 |
| | SWISS-MODEL | 91.3 | 6.1 | 2.5 |
| | SWISS-MODEL | 88.1 | 7.4 | 4.5 |

| Protein Name | Program | Favored Region (%) | Allowed Regionn (%) | Outlier Regionn (%) |
|---|---|---|---|---|
| IreA | FFAS-3D | 95.3 | 3.8 | 0.9 |
| | PRC | 93.4 | 4.9 | 1.8 |
| | SP3 | 94.9 | 3.5 | 1.6 |
| | FFAS03 | 94.9 | 4.0 | 1.2 |
| | PROSPECT2 | 84.7 | 9.6 | 5.7 |
| | pGenTHREADER | 81.0 | 13.2 | 5.7 |
| | Neff-PPAS | 93.2 | 4.1 | 2.6 |
| | SPARKS-X | 93.7 | 5.0 | 1.3 |
| | wdPPAS | 95.0 | 3.8 | 1.2 |
| | MUSTER | 93.5 | 4.4 | 2.1 |
| | SWISS-MODEL | 94.1 | 4.6 | 1.2 |
| | SWISS-MODEL | 89.4 | 8.0 | 2.6 |
| | SWISS-MODEL | 90.4 | 6.3 | 3.3 |
| | SWISS-MODEL | 90.0 | 6.9 | 3.1 |
| IroN | 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 |
| IutA | MUSTER | 92.2 | 5.2 | 2.6 |
| | PRC | 91.6 | 5.6 | 2.7 |
| | SP3 | 93.0 | 4.8 | 2.2 |
| | FFAS03 | 75.2 | 17.1 | 7.6 |
| | PROSPECT2 | 87.0 | 8.4 | 4.7 |
| | pGenTHREADER | 78.4 | 12.7 | 8.9 |
| | FFAS-3D | 90.3 | 6.2 | 3.6 |
| | Neff-PPAS | 94.4 | 4.0 | 1.6 |
| | SPARKS-X | 93.2 | 5.2 | 1.6 |
| | wdPPAS | 91.1 | 6.3 | 2.6 |
| | SWISS-MODEL | 89.5 | 7.1 | 3.5 |
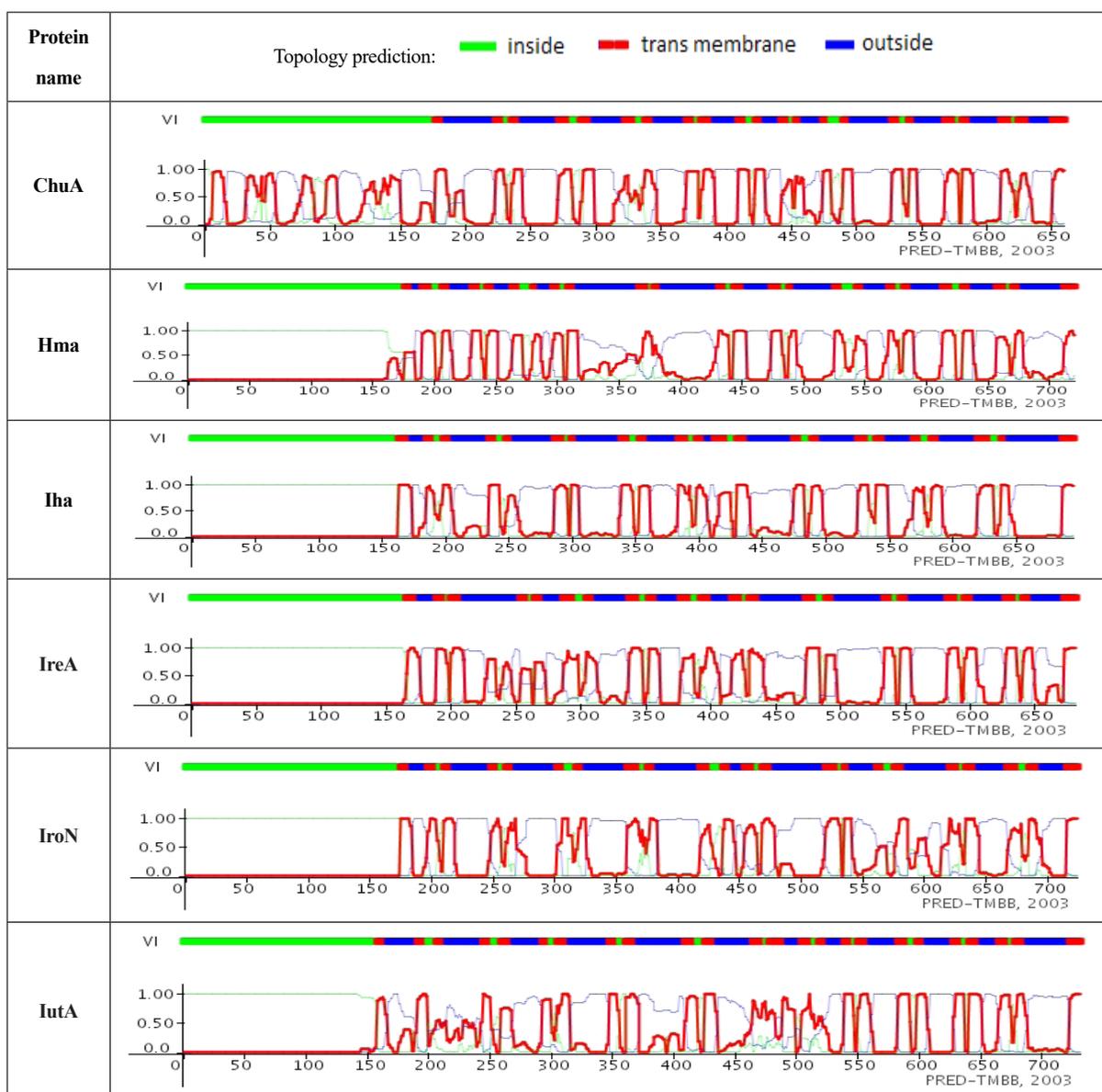
**Figure 1.** Topology model of six iron receptor vaccine candidates, ChuA, Hma, IutA, IreA, Iha, and IroN built based on predicted inside, transmembrane, and outside regions of the protein

**Supplementary Table 5.** Models refinement results

| Protein | RMSD | TM-score |
|---------|------|----------|
| ChuA | 1.829 | 0.9819 |
| Hma | 1.469 | 0.9837 |
| Iha | 4.177 | 0.9676 |
| IreA | 1.567 | 0.9830 |
| IroN | 1.266 | 0.9904 |
| IutA | 2.597 | 0.9831 |
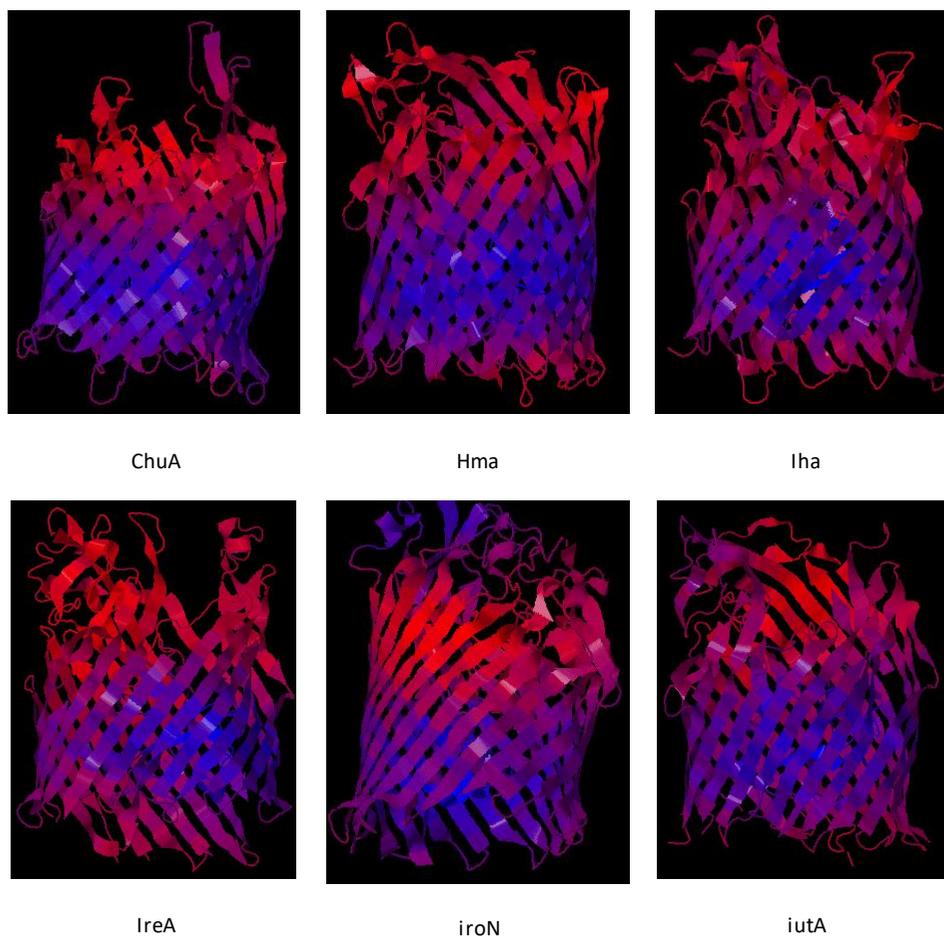
**Figure 2.** Discontinuous B-cell epitopes predicted from protein 3D structures by disco tope residues colored by disco tope score Red: High score, Blue: Low score.

## Ligand binding site predictions and structure alignment

Ligand binding sites were determined by COFACTOR software. The obtained results indicate the involvement of conserved residues in the iron-binding site. In all of the analyzed six iron receptor vaccine candidates, the ligand-binding site resided between the crock domain and the large extracellular loops of the barrel (Figure 3). Based on PRALINE structure alignments, no significant discrepancy was seen between the vaccine candidate 2D structures. The majority of 2D structures of the vaccine

**Table 5.** Epitopes recommended by SVMTrip server

| Protein Name | Location | Epitope |
|---|---|---|
| ChuA | 409 - 428 | KWSSRAGMTINPTNWLMLFG |
| Hma | 468 - 487 | NQVDENGLSPNAALMYKITP |
| Iha | 235 - 254 | YNLGARLDWKASEQDVLWFD |
| IreA | 649 - 668 | LNVTDRKSEDIDTIDGNWQV |
| IroN | 628 - 647 | NWTITQAFSASVNWTLYGRQ |
| IutA | 626 - 645 | ASPSKATAYIGWAPDPWSLR |

**Table 6.** Immunological and physicochemical properties of final chimeric vaccine

| Number of Amino Acids | 279 |
| --- | --- |
| Molecular weight | 27622.41 |
| Theoretical pI | 5.44 |
| Total number of negatively charged residues | (Asp + Glu): 27 |
| Total number of positively charged residues | (Arg + Lys): 24 |
| Atomic composition Formula | $C_{1166}H_{1783}N_{363}O_{413}S_4$ |
| The estimated half-life | 30 hours (mammalian reticulocytes, in vitro >20 hours (yeast, in vivo). >10 hours (*Escherichia coli*, in vivo) |
| Instability index | 33.89 (stable) |
| Aliphatic index | 42.69 |
| Grand average of hydropathicity (GRAVY) | -0.788 (hydrophilic) |
| VaxiJen score | 2.2173 (protective antigens) |
| AllergenFP v.1.0 | Probable Non-allergen |
| Solubility score | 0.526 (soluble expression) |



ChuA          Hma          Iha

IreA          iroN          iutA
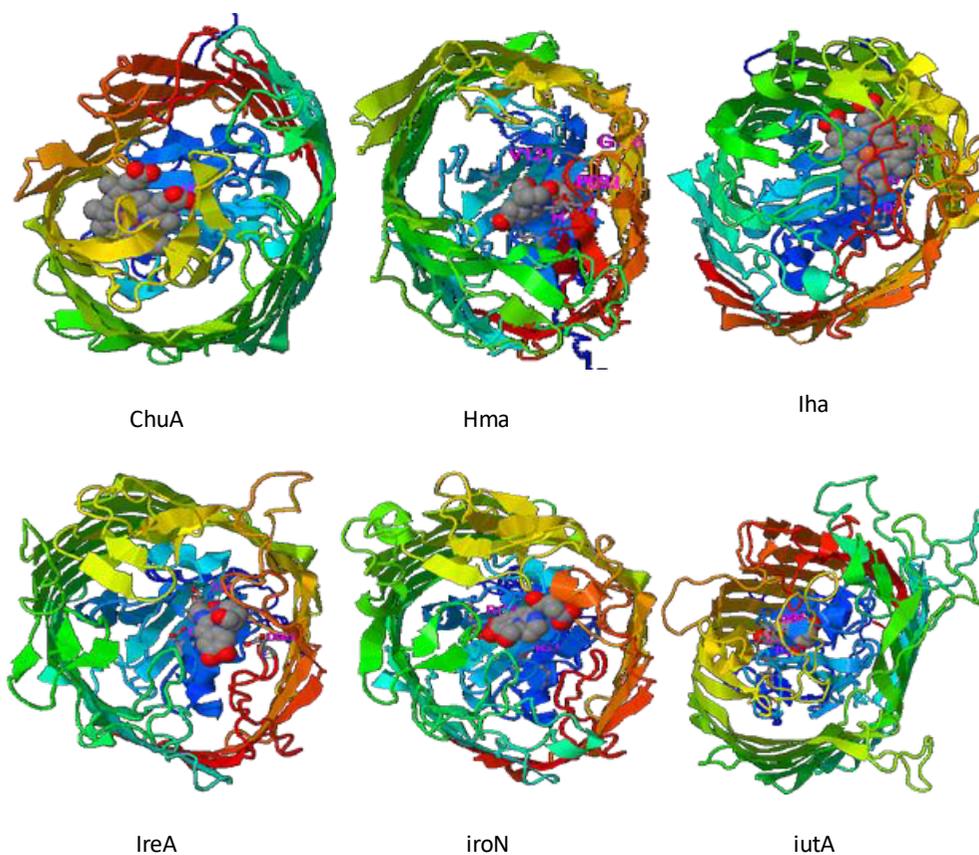
**Figure 3.** Ligand binding site predictions for the six iron receptor vaccine candidates of ChuA, Hma, IutA, IreA, Iha, and IroN

GSSDGYKDVDADKWSSRAGMTINPTNWLMLFGGGGGSGGGGSGGGGSKHGNQTNQVDENGLSPNAALMYKITP-
GGGGSGGGGSGGGGSYNLGARLDWKASEQDVLWFDMDTTRQRYDNRDGQLGSLTGGYDRTLGGGGSGGGGSGGGG-
SLNVTDRKSEDIDTIDGNWQVDEGRGGGGSGGGGSGGGGSNWTITQAFSASVNWTLYGRQKPRTHAETRSEDTGGLS-
GKELGAGGGGS GGGGS GGGGSKVNGTWQKYDVKTASPSKATAYIGWAPDPWSLR

Selected region sequence of ChuA

Selected region sequence of Hma

Selected region sequence of Iha

Selected region sequence of IreA

Selected region sequence of IroN

Selected region sequence of IutA

Linker sequence
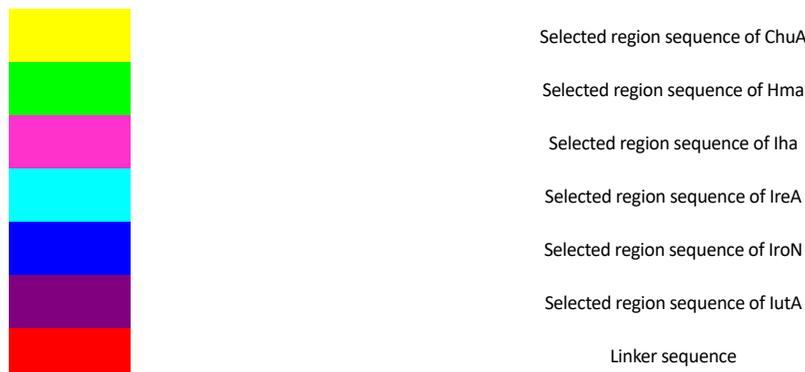
**Figure 4.** Six selected regions of each candidate protein

candidates were matched in transmembrane beta-strands, which construct the barrel. Barrel of TonB-dependent receptors possesses three main features: 10 short periplasmic turn, 22-stranded β-barrel, and 11 extracellular loops labeled L1 to L11 for all transporters.

## Immunogenic regions selection

The regions with the highest density of continuous and discontinuous epitopes were selected as proper vaccine candidate regions. The properties obtained from single-scale amino acid assay, probability of antigenicity, and average physicochemical properties were also considered to select the desired regions. Regions cover-
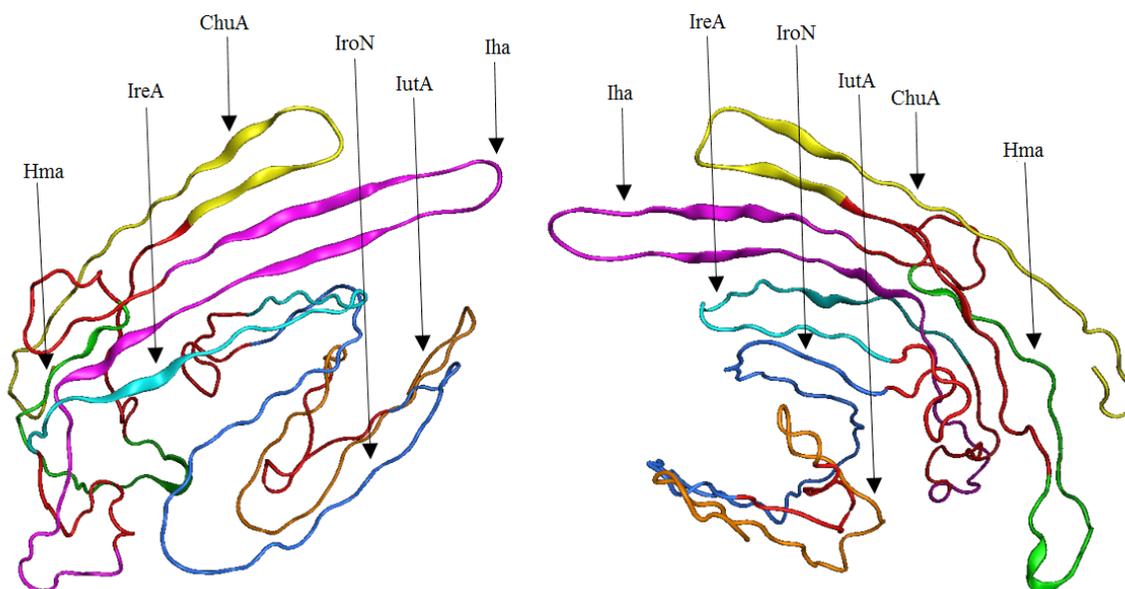


**Figure 5.** 3D Structure of chimeric vaccine
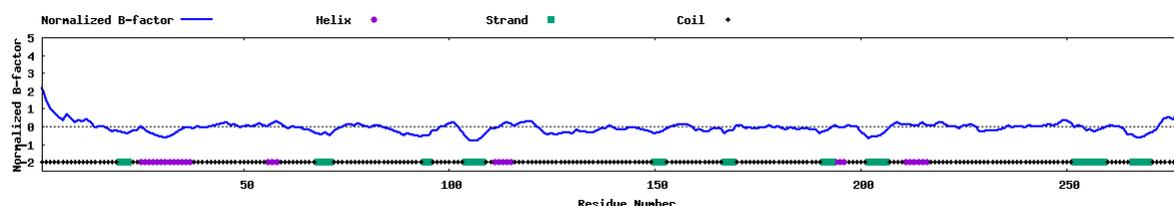


**Figure 6.** Predicted normalized B-factor for 3D structure of chimeric vaccine

**Supplementary Table 6.** Comparison of VaxiJen score in selected regions and whole proteins

| Protein Name | Selected Region | VaxiJen Score (Selected Region) | VaxiJen Score (Whole Protein) |
|---|---|---|---|
| ChuA | 397-428 (include loop 6) | 0.8235 | 0.6266 |
| Hma | 462-487 (include loop 6) | 1.3064 | 0.7259 |
| Iha | 235-280 (include loop 3) | 1.0857 | 0.6138 |
| IreA | 649-672 (include loop 11) | 0.9997 | 0.6570 |
| IroN | 628-670 (include loop 10) | 1.1730 | 0.7889 |
| IutA | 613-645 (include loop 9) | 0.8096 | 0.6016 |

ing extracellular loops with the largest gatherings of linear and conformational epitopes were selected as vaccine candidates in six iron receptor proteins. These regions include residue 397-428 [GSSDGYKDVDADKWSS-RAGMTINPTNWLMLFG] in ChuA, residue 462-487 [KHGNQTNQVDENGLSPNAALMYKITP] in Hma, residue 235-280 [YNLGARLDWKASEQDVLWFDM-DTTRQRYDNRDGQLGSLTGGYDRTL] in Iha, residue 649-672 [LNVTDRKSEDIDTIDGNWQVDEGR] in IreA, residue 628-670 [NWTITQAFSASVNWTLYGRQKPR-THAETRSEDTGGLSGKELGA] in IroN, and residue 613-645 [KVNGTWQKYDVKTASPSKATAYIGWAPD-PWSLR] in IutA. Further analyses by the VaxiJen server were performed on the selected regions to validate the selected regions. The VaxiJen score significantly increased in selected regions (Supplementary Table 6).

## Final vaccine design

The flexible GGGGS linker has been shown to improve folding and stability in several fusion protein examples. In this regard, we have used triple repeats of GGGGS to link six selected regions of vaccine candidates (Figure 4). The VaxiJen score calculated for the designed vaccine was about 2.2126. This score is higher than each individual selected region.

## Final vaccine evaluation

Various properties of the final chimeric vaccine, including VaxiJen antigenicity score, AllergenFP allergenicity, number of amino acids, other physicochemical properties, and protein expression in the appropriate expression system, were successfully calculated (Table 6).

VaxiJen score above 0.4 indicates the protective antigens and subunit vaccines. Overall prediction for the final chimeric vaccine with calculated VaxiJen was 2.2173. AllergenFP v.1.0 predicted the final vaccine as probable non-allergen. A solubility score above 0.5 indicates soluble expression, and a score below 0.5 indicates insoluble expression in *Escherichia coli*. The final chimeric vaccine has a score of about 0.526.

I-TASSER generates a large ensemble of decoys that are clustered by the SPICKER program. The final models are selected based on the pair-wise structure similarity. Five models corresponding to the five largest structure clusters are reported as the final models. C-score is used to evaluate the confidence of each model quantitatively and is typically in the range of -5 to 2. The higher C-score signifies the higher confidence of a model. C-score and protein length are used to calculate the TM-score and RMSD for the models. Figure 5 shows the best-predicted model by the I-TASSER server. The inherent thermal mobility of residues/atoms in proteins is indicated by B-factor. Using the sequence profiles derived from sequence databases and threading template proteins from the PDB and I-TASSER deduces the B-factor for the residues (Figure 6). Z-score-based normalization of the raw B-factor values is used to calculate the normalized B-factor values for a target protein.

## Discussion

Designing amenable immunogenic agents to develop an adaptive immunity against different pathogens has remained a tough challenge in vaccine development efforts. This study was conducted to develop 3D models of ChuA, Hma, IutA, IreA, Iha, and IroN proteins by invoking various in silico methods. We designed a chimeric immunogen composed of highly immunogenic regions from six *Escherichia coli* antigens. The use of protein combinations increases the likelihood of their simultaneous uptake by host cells compared to the separate consumption of their proteins or monozygotic proteins [54, 55].

Our BLAST search results showed that the sequences of ChuA, Hma, IutA, IreA, Iha, and IroN are homologous to numerous other molecules. Most of the obtained sequences belonged to TonB dependent/ligand-gated channels [56]. Several members of the IROMPs family have resolved crystal structures, including the proteins under the PDB accession numbers of 3FHH_A [57], 1BY3_A, 2HDI_A, 2HDI_A [58], 5FP2_A [59], and 2HDI_A. Pivotal clues, regarding the architecture of all TonB-dependent receptors have been derived using these structures. All of the aforementioned proteins are essential pathogenicity factors in bacterial infections. Homology modeling is the most accurate in silico approach to predict the protein structure [60, 61].

A successful homology modeling needs a reliable template that could be attained by similarity search and sequence alignment. An amenable template should bear a low *E* value, high query coverage, and high identity (more than 35%) against the target sequence. Thus, the most reliable template for homology modeling could be a hit with the highest total score. The predicted inside, outside, and transmembrane regions of the proteins were used to build a 2D topology model of six iron receptor vaccine candidates (ChuA, Hma, IutA, IreA, Iha, and IroN). The obtained results showed that the transmembrane antiparallel β-strands are the main structural components of these proteins. Given the predicted models, the native fold of these proteins forms a β-barrel structure [62]. The linkage between the β-barrel strands is made up of loops at the outside or turns at the inside face of the proteins (Figure 1). More than 11 external loops are reported in these proteins. Since the side chains of the residues are highly exposed, they may play a determinant role in the initial binding events with the Fe-siderophore complex.

It has been previously shown that the antigenicity and immunogenicity of an antigen directly correlate with its epitope density [63]. Although discontinuous B-cell epitopes are more predominant, experimental studies are primarily focused on the identification of linear B-cell epitopes. These data become more applicable considering the existing direct correlation between epitope density and epitope-specific humoral immune responses [64]. The determined epitomic data could be harnessed to choose ChuA, Hma, IutA, IreA, Iha, and IroN regions with higher epitope density. The best linear B cell epitopes of the six iron receptor vaccine candidates were located at the largest extracellular loops. Interestingly, discontinuous B cell epitopes predicted from the 3D structure of proteins include all of the extracellular loops. The existence of predicted epitopes was confirmed via the experimentally identified epitopes and their corresponding approved antibodies. This experimental confirmation could be construed as the accuracy of employed epitope prediction and 3D structure prediction procedures. The results of antigenicity comparison revealed that the selected regions are considerably more antigenic than the whole antigen. Moreover, the results of the instability index indicate that these regions are considered stable.

The previous studies have demonstrated the active role of the linkers in the production of stable, bioactive fusion proteins as essential components of recombinant fusion proteins [65]. The foundation in linker design is the common feature of the linkers from naturally occurring multi-domain proteins [66]. Experimentally developed linkers are structurally divided into three groups: flexible linkers, rigid linkers, and in vivo cleavable linkers. Besides the importance of linking functional domains, there are practical purposes for the linkers in synthesizing the fusion proteins [67], such as elevated biological activity, overexpression yield, and improving desirable pharmacokinetic profiles. The structural flexibility occurs due to many small or hydrophilic amino acids such as Gly or Ser in the flexible linkers that link the functional domains aiming at inter-domain interactions or movements [68].

## Conclusion

In conclusion, the bioinformatics approaches are practical strategies to fill the gap between the number of resolved 3D protein structures and known protein sequences. Vaccine design purposes could be achieved by structural and immunological properties derived from in silico studies. The limitations associated with the vaccines based on a single antigen could be compensated by designing chimeric vaccines based on a set of immunogens. Using a combination of epitope prediction and 3D structure prediction methods as a vaccine design strategy could pave the way for further functional, structural, and therapeutic studies of vaccine candidates.

## Ethical Considerations

### Compliance with ethical guidelines

There were no ethical considerations to be considered in this research.

### Funding

## References

[1] Vivona S, Gardy JL, Ramachandran S, Brinkman FSL, Raghava GPS, Flower DR, et al. Computer-aided biotechnology: from immuno-informatics to reverse vaccinology. Trends Biotechnol. 2008; 26(4):190-200. [DOI:10.1016/j.tibtech.2007.12.006] [PMID]

[2] Nagpal G, Chaudhary K, Agrawal P, Raghava GPS. Computer-aided prediction of antigen presenting cell modulators for designing peptide-based vaccine adjuvants. J Transl Med. 2018; 16(1):181. [DOI:10.1186/s12967-018-1560-1] [PMID] [PMCID]

[3] Dzayee SA, Khudhur PK, Mahmood A, Markov A, Maseleno A, Ebrahimpour Gorji A. Computational design of a new multi-epitope vaccine using immunoinformatics approach against mastitis disease. Anim Biotechnol. 2021; 1-12. [DOI:10.1080/10495398.2021.1899937] [PMID]

[4] Sarowska J, Futoma-Koloch B, Jama-Kmiecik A, Frej-Madrzak M, Ksiazczyk M, Bugla-Ploskonska G, et al. Virulence factors, prevalence and potential transmission of extraintestinal pathogenic Escherichia coli isolated from different sources: Recent reports. Gut Pathog. 2019; 11:10. [DOI:10.1186/s13099-019-0290-0] [PMID] [PMCID]

[5] Griebling TL. Urologic diseases in America project: Trends in resource use for urinary tract infections in women. J Urol. 2005; 173(4):1281-7. [DOI:10.1097/01.ju.0000155596.98780.82]

[6] Pearle MS, Calhoun EA, Curhan GC, Urologic Diseases of America Project A. Urologic diseases in America project: Urolithiasis. J Urol. 2005; 173(3):848-57. [DOI:10.1097/01.ju.0000152082.14384.d7] [PMID]

[7] Lin KY, Chiu NT, Chen MJ, Lai CH, Huang JJ, Wang YT, et al. Acute pyelonephritis and sequelae of renal scar in pediatric first febrile urinary tract infection. Pediatr Nephrol. 2003; 18(4):362-5. [DOI:10.1007/s00467-003-1109-1] [PMID]

[8] 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(3):867-9. [DOI:10.1097/01.ju.0000075094.54767.6e] [PMID]

[9] Connell I, Agace W, Klemm P, Schembri M, Mårild S, Svanborg C. Type 1 fimbrial expression enhances Escherichia coli virulence for the urinary tract. Proc Natl Acad Sci U S A. 1996; 93(18):9827-32. [DOI:10.1073/pnas.93.18.9827] [PMID] [PMCID]

[10] Langermann S, Möllby R, Burlein JE, Palaszynski SR, Gale Auguste C, DeFusco A, et al. Vaccination with FimH adhesin protects cynomolgus monkeys from colonization and infection by uropathogenic Eschevichia coli. J Infect Dis. 2000; 181(2):774-8. [DOI:10.1086/315258] [PMID]

[11] Russo TA, McFadden CD, Carlino-MacDonald UB, Beanan JM, Olson R, Wilding GE. The Siderophore receptor IroN of extraintestinal pathogenic Escherichia coli is a potential vaccine candidate. Infection Immun. 2003; 71(12):7164-9. [DOI:10.1128/IAI.71.12.7164-7169.2003] [PMID] [PMCID]

[12] Durant L, Metais A, Soulama-Mouze C, Genevard JM, Nassif X, Escaich S. Identification of candidates for a subunit vaccine against extraintestinal pathogenic Escherichia coli. Infect Immun. 2007; 75(4):1916-25. [DOI:10.1128/IAI.01269-06] [PMID] [PMCID]

[13] Alteri CJ, Hagan EC, Sivick KE, Smith SN, Mobley HLT. Mucosal immunization with iron receptor antigens protects against urinary tract infection. PLoS Pathog. 2009; 5(9):e1000586. [DOI:10.1371/journal.ppat.1000586] [PMID] [PMCID]

[14] Buchanan SK, Smith BS, Venkatramani L, Xia D, Esser L, Palnitkar M, et al. Crystal structure of the outer membrane active transporter FepA from Escherichia coli. Nat Struct Biol. 1999; 6(1):56-63. [DOI:10.1038/4931] [PMID]

[15] Torres AG, Redford P, Welch RA, Payne SM. TonB-dependent systems of uropathogenic Escherichia coli: Aerobactin and heme transport and TonB are required for virulence in the mouse. Infect Immun. 2001; 69(10):6179-85. [DOI:10.1128/IAI.69.10.6179-6185.2001] [PMID] [PMCID]

[16] Garcia EC, Brumbaugh AR, Mobley HL. Redundancy and specificity of Escherichia coli iron acquisition systems during urinary tract infection. Infect Immun. 2011; 79(3):1225-35. [DOI:10.1128/IAI.01222-10] [PMID] [PMCID]

[17] Mihăşan M. Basic protein structure prediction for the biologist: A review. Arch Biol Sci. 2010; 62(4):857-71. [DOI:10.2298/ABS1004857M]

[18] Khalili S, Zakeri A, Hashemi ZS, Masoumikarimi M, Reza Manesh MR, Shariatifar N, et al. Structural analyses of the interactions between the thyme active ingredients and human serum albumin. Turk J Biochem. 2017; 42(4):459-67. [DOI:10.1515/tjb-2017-0008]

[19] Khalili S, Mohammadpour H, Shokrollahi Barough M, Kokhaei P. ILP-2 modeling and virtual screening of an FDA-approved library: A possible anticancer therapy. Turk J Med Sci. 2016; 46(4):1135-43. [DOI:10.3906/sag-1503-2] [PMID]

[20] Khalili S, Jahangiri A, Hashemi ZS, Khalesi B, Mard-Soltani M, Amani J. Structural pierce into molecular mechanism underlying Clostridium perfringens Epsilon toxin function. Toxicon. 2017; 127:90-9. [DOI:10.1016/j.toxicon.2017.01.010] [PMID]

[21] Li Y, Dai J, Zhuge X, Wang H, Hu L, Ren J, et al. Iron-regulated gene ireA in avian pathogenic Escherichia coli participates in adhesion and stress-resistance. BMC Vet Res. 2016; 12(1):1-10. [DOI:10.1186/s12917-016-0800-y] [PMID] [PMCID]

[22] Floudas CA, Fung HK, McAllister SR, Mönnigmann M, Rajgaria R. Advances in protein structure prediction and de novo protein design: A review. Chem Eng Sci. 2006; 61(3):966-88. [DOI:10.1016/j.ces.2005.04.009]

[23] Khalili S, Rasaee M, Bamdad T. 3D structure of DKK1 indicates its involvement in both canonical and non-canonical Wnt pathways. Mol Biol. 2017; 51(1):155-66. [DOI:10.1134/S0026893317010095]

[24] Blundell T, Carney D, Gardner S, Hayes F, Howlin B, Hubbard T, et al. Knowledge-based protein modelling and design. Eur J Biochem. 1988; 172(3):513-20. [DOI:10.1111/j.1432-1033.1988.tb13917.x] [PMID]

[25] Khalili S, Rahbar MR, Haj Dezfulian M, Jahangiri A. In silico analyses of Wilms' tumor protein to designing a novel multi-epitope DNA vaccine against cancer. J Theor Biol. 2015; 379:66-78. [DOI:10.1016/j.jtbi.2015.04.026] [PMID]

[26] Andersen PH, Nielsen M, Lund O. Prediction of residues in discontinuous B-cell epitopes using protein 3D structures. Protein Sci. 2006; 15(11):2558-67. [DOI:10.1110/ps.062405906] [PMID] [PMCID]

[27] Khalili S, Jahangiri A, Borna H, Ahmadi Zanoos K, Amani J. Computational vaccinology and epitope vaccine design by immunoinformatics. Acta Microbiol Immunol Hung. 2014; 61(3):285-307. [DOI:10.1556/amicr.61.2014.3.4] [PMID]

[28] 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 Ther. 2017; 25:123-35. [DOI:10.1007/s10989-017-9654-9]

[29] Sikder AR, Zomaya AY. An overview of protein-folding techniques: Issues and perspectives. Int J Bioinform Res Appl. 2005; 1(1):121-43. [DOI:10.1504/IJBRA.2005.006911] [PMID]

[30] Jahangiri A, Rasooli I, Mousavi Gargari SL, Owlia P, Rahbar MR, Amani J, et al. An in silico DNA vaccine against Listeria monocytogenes. Vaccine. 2011; 29(40):6948-58. [DOI:10.1016/j.vaccine.2011.07.040] [PMID]

[31] Jenuth JP. The NCBI. Publicly available tools and resources on the web. Methods Moh Biol. 1999; 132:301-12. [DOI:10.1385/1-59259-192-2:301] [PMID]

[32] Gish W, States DJ. Identification of protein coding regions by database similarity search. Nat Genet. 1993; 3(3):266-72. [DOI:10.1038/ng0393-266] [PMID]

[33] Pearson WR. Effective protein sequence comparison. Methods Enzymol. 1996; 266:227-58. [DOI:10.1016/S0076-6879(96)66017-0]

[34] Fiser A. Protein structure modeling in the proteomics era. Expert Rev Proteomics. 2004; 1(1):97-110. [DOI:10.1586/14789450.1.1.97] [PMID]

[35] Doytchinova IA, Flower DR. VaxiJen: A server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinformatics. 2007; 8:4. [DOI:10.1186/1471-2105-8-4] [PMID] [PMCID]

[36] Vita R, Overton JA, Greenbaum JA, Ponomarenko J, Clark JD, Cantrell JR, et al. The Immune Epitope Database (IEDB) 3.0. Nucleic Acids Res. 2015; 43(Database issue):D405-12. [DOI:10.1093/nar/gku938] [PMID] [PMCID]

[37] Bendtsen JD, Jensen LJ, Blom N, Von Heijne G, Brunak S. Feature-based prediction of non-classical and leaderless protein secretion. Protein Eng Des Sel. 2004; 17(4):349-56. [DOI:10.1093/protein/gzh037] [PMID]

[38] Yu CS, Cheng CW, Su WC, Chang KC, Huang SW, Hwang JK, et al. CELLO2GO: A web server for protein subCELlular LOcalization prediction with functional gene ontology annotation. PLoS One. 2014; 9(6):e99368. [DOI:10.1371/journal.pone.0099368] [PMID] [PMCID]

[39] Bhasin M, Garg A, Raghava GPS. PSLpred: Prediction of subcellular localization of bacterial proteins. Bioinformatics. 2005; 21(10):2522-4. [DOI:10.1093/bioinformatics/bti309] [PMID]

[40] Emanuelsson O, Brunak S, von Heijne G, Nielsen H. Locating proteins in the cell using TargetP, SignalP and related tools. Nature Protoc. 2007; 2(4):953-71. [DOI:10.1038/nprot.2007.131] [PMID]

[41] 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(Web Server issue):W400-4. [DOI:10.1093/nar/gkh417] [PMID] [PMCID]

[42] Geourjon C, Deleage G. SOPMA: Significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Comput Appl Biosci. 1995; 11(6):681-4. [DOI:10.1093/bioinformatics/11.6.681] [PMID]

[43] Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL workspace: A web-based environment for protein structure homology modelling. Bioinformatics. 2006; 22(2):195-201. [DOI:10.1093/bioinformatics/bti770] [PMID]

[44] Schwede T, Kopp J, Guex N, Peitsch MC. SWISS-MODEL: An automated protein homology-modeling server. Nucleic Acids Res. 2003; 31(13):3381-5. [DOI:10.1093/nar/gkg520] [PMID] [PMCID]

[45] Wu S, Zhang Y. LOMETS: A local meta-threading-server for protein structure prediction. Nucleic Acids Res. 2007; 35(10):3375-82. [DOI:10.1093/nar/gkm251] [PMID] [PMCID]

[46] Schirmeier H, Neuhalfen J, Korb I, Spinczyk O, Engel M, editors. Rampage: Graceful degradation management for memory errors in commodity linux servers. Paper presented at: 17th Pacific Rim International Symposium on Dependable Computing. 12-14 December 2011; Pasadena, USA. [DOI:10.1109/PRDC.2011.20]

[47] 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(10):2525-34. [DOI:10.1016/j.bpj.2011.10.024] [PMID] [PMCID]

[48] Pontoppidan Larsen JE, Lund O, Nielsen M. Improved method for predicting linear B-cell epitopes. Immunome Res. 2006; 2:2. [DOI:10.1186/1745-7580-2-2] [PMID] [PMCID]

[49] 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(9):e45152. [DOI:10.1371/journal.pone.0045152] [PMID] [PMCID]

[50] Ponomarenko J, Bui HH, Li W, Fusseder N, Bourne PE, Sette A, et al. ElliPro: A new structure-based tool for the prediction of antibody epitopes. BMC Bioinformatics. 2008; 9:514. [DOI:10.1186/1471-2105-9-514] [PMID] [PMCID]

[51] Roy A, Yang J, Zhang Y. COFACTOR: An accurate comparative algorithm for structure-based protein function annotation. Nucleic Acids Res. 2012; 40(Web Server issue):W471-7. [DOI:10.1093/nar/gks372] [PMID] [PMCID]

[52] Bawono P, Heringa J. PRALINE: A versatile multiple sequence alignment toolkit. Methods Mol Biol. 2014; 1079:245-62. [DOI:10.1007/978-1-62703-646-7_16] [PMID]

[53] Trinh R, Gurbaxani B, Morrison SL, Seyfzadeh M. Optimization of codon pair use within the (GGGGS) 3 linker sequence results in enhanced protein expression. Mol Immun. 2004; 40(10):717-22. [DOI:10.1016/j.molimm.2003.08.006] [PMID]

[54] Arêas APM, Oliveira MLS, Miyaji EN, Leite LCC, Aires KA, Dias WO, et al. Expression and characterization of cholera toxin B-pneumococcal surface adhesin A fusion protein in Escherichia coli: Ability of CTB-PsaA to induce humoral immune response in mice. Biochem Biophys Res Commun. 2004; 321(1):192-6. [DOI:10.1016/j.bbrc.2004.06.118] [PMID]

[55] Kundu J, Mazumder R, Srivastava R, Srivastava BS. Intranasal immunization with recombinant toxin-coregulated pilus and cholera toxin B subunit protects rabbits against Vibrio cholerae O1 challenge. FEMS Immunol Med Microbiol. 2009; 56(2):179-84. [DOI:10.1111/j.1574-695X.2009.00563.x] [PMID]

[56] Brillet K, Reimmann C, Mislin GLA, Noël S, Rognan D, Schalk IJ, et al. Pyochelin enantiomers and their outer-membrane siderophore transporters in fluorescent pseudomonads: Structural bases for unique enantiospecific recognition. J Am Chem Soc. 2011;133(41):16503-9. [DOI:10.1021/ja205504z] [PMID]

[57] Cobessi D, Meksem A, Brillet K. Structure of the heme/hemoglobin outer membrane receptor ShuA from Shigella dysenteriae: Heme binding by an induced fit mechanism. Proteins. 2010; 78(2):286-94. [DOI:10.1002/prot.22539] [PMID]

[58] Buchanan SK, Lukacik P, Grizot S, Ghirlando R, Ali MMU, Barnard TJ, et al. Structure of colicin I receptor bound to the R-domain of colicin Ia: Implications for protein import. EMBO J. 2007; 26(10):2594-604. [DOI:10.1038/sj.emboj.7601693] [PMID] [PMCID]

[59] Mislin GLA, Schalk IJ. Siderophore-dependent iron uptake systems as gates for antibiotic Trojan horse strategies against Pseudomonas aeruginosa. Metallomics. 2014; 6(3):408-20. [DOI:10.1039/C3MT00359K] [PMID]

[60] Oakhill JS, Sutton BJ, Gorringe AR, Evans RW. Homology modelling of transferrin-binding protein A from Neisseria meningitidis. Protein Eng Des Sel. 2005; 18(5):221-8. [DOI:10.1093/protein/gzi024] [PMID]

[61] Sefid F, Rasooli I, Payandeh Z. Homology modeling of a Camelid antibody fragment against a conserved region of Acinetobacter baumannii Biofilm Associated Protein (Bap). J Theor Biol. 2016; 397:43-51. [DOI:10.1016/j.jtbi.2016.02.015] [PMID]

[62] 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. [DOI:10.1186/1471-2105-6-7] [PMID] [PMCID]

[63] Liu W, Chen YH. High epitope density in a single protein molecule significantly enhances antigenicity as well as immunogenicity: A novel strategy for modern vaccine development and a preliminary investigation about B cell discrimination of monomeric proteins. Eur J Immunol. 2005; 35(2):505-14. [DOI:10.1002/eji.200425749] [PMID]

[64] Chen J, Liu H, Yang J, Chou KC. Prediction of linear B-cell epitopes using amino acid pair antigenicity scale. Amino Acids. 2007; 33(3):423-8. [DOI:10.1007/s00726-006-0485-9] [PMID]

[65] Chen X, Zaro JL, Shen WC. Fusion protein linkers: Property, design and functionality. Adv Drug Deliv Rev. 2013; 65(10):1357-69. [DOI:10.1016/j.addr.2012.09.039] [PMID] [PMCID]

[66] Gokhale RS, Khosla C. Role of linkers in communication between protein modules. Curr Opin Chem Biol. 2000; 4(1):22-7. [DOI:10.1016/S1367-5931(99)00046-0]

[67] Haddad J, Whitehead GFS, Katsoulidis AP, Rosseinsky MJ. In-MOFs based on amide functionalised flexible linkers. Faraday Discuss. 2017; 201:327-35. [DOI:10.1039/C7FD00085E] [PMID]

[68] Argos P. An investigation of oligopeptides linking domains in protein tertiary structures and possible candidates for general gene fusion. J Mol Biol. 1990; 211(4):943-58. [DOI:10.1016/0022-2836(90)90085-Z]