

Design of an Epitope Candidate Vaccine Against Iha Protein in Escherichia Coli: an in Silico Approach

Fateme Sefid ^{a,b}, Armina Alagheband Bahrami ^c, Masoumeh Rajabibazl ^{d,e}, Mohammad Rahmati ^f, Zahra Payandeh ^{g*}, Seyed Mehdi Kalantar ^{a*}, Siavoush Dastmalchi ^h, Mohsen Nabi Afjadi ⁱ

^a Department of Medical Genetics, Shahid Sadoughi University of Medical Science, Yazd, Iran; ^b Department of Biology, Science and Art University, Yazd, Iran; ^c Department of Medical Biotechnology, School of Advanced Technologies in Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran; ^d Department of Clinical Biochemistry, Faculty of Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran; ^e School of Advanced Technologies in Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran; ^f Department of Medical Biotechnology and Nanotechnology, Faculty of Medicine, Zanjan University of Medical Science, Zanjan, Iran; ^g Immunology Research Center, Tabriz University of Medical Sciences, Tabriz-Iran; ^h Biotechnology Research Center, School of Pharmacy, Tabriz University of Medical Sciences, Tabriz, Iran; ⁱ Institute of Biochemistry and Biophysics, Tehran University, Tehran, Iran.

*Corresponding authors: Zahra Payandeh, Immunology Research Center, Tabriz University of Medical Sciences, Tabriz-Iran. **E-mail:** zpayandeh58@yahoo.com; **Tel:** +98-914 9315939

Seyed Mehdi Kalantar, Department of Medical Genetics, Shahid Sadoughi University of Medical Science, Yazd, Iran; ^b Department of Biology, Science and Art University, Yazd, Iran. **E-mail:** kalantarsm@ystp.ac.ir; **Tel:** +98-913 1518918

Submitted: 2020-06-30; Accepted: 2020-08-22; Published Online: 2020-09-10; DOI: 10.22037/rrr.v5i1.29954

Introduction: Iron-regulated outer membrane virulence protein (IrgA) involved in the initial step of iron uptake by binding ferric- iron chelation siderophore that allows the bacterium to extract iron from the environment. IrgA homologue adhesion (Iha) revealed as a novel adherence conferring molecule. In this study, homology modeling, fold recognition and Ab-initio approaches along with their combination were invoked to determine the tertiary structure of Iha. **Material and Methods:** Specific bioinformatics methods were used to forecast their immunological, biochemical and functional properties. **Results:** The results showed that IrgA constitutes beta barrel structures. The immunological, biochemical and functional analyzes led us to pick a region of every antigen with the highest immunogenic properties. Comparison of antigenicity scores for selected regions and the whole proteins showed that the antigenicity of the selected regions is considerably higher than the whole antigen. **Conclusion:** Our strategy for prediction of the 3D structure and epitopes could be deemed as an amenable approach for efficient vaccine design. These approaches used could provide the basis for future functional studies to design and development of a suitable vaccine. In this regard, a region includes residues 200-340, covering a part of barrel, was chosen as vaccine candidate against Iha protein in Escherichia Coli.

Keywords: IrgA; Iron Receptor; Vaccine; Bioinformatics;

Introduction

The O157: H7 Strain of Escherichia coli is a cause of diarrhea, hemorrhagic colitis, and hemolytic uremic syndrome. The widespread outbreak of this bacterium in the hospital environments highlights the significance of vaccination against this pathogen. Adherence ability to epithelial cells is an important feature of virulence as enteric pathogens can efficiently transmit toxins to host tissues, overcome peristaltic clearance and gain access to nutrients extracted from host substances (1,2). Iron-regulated outer membrane virulence

protein (IrgA) involved in the initial step of iron uptake by binding ferric- iron chelation siderophore that allows the bacterium to extract iron. The results of recent investigations on structure, function and underlying mechanisms of expression of the IrgA homologue adhesion (Iha) homologues in a variety of bacteria revealed Iha as a novel adherence conferring molecule (3). Iha was identified as an adhesive in the enterohemorrhagic E.coli O157:H7 strain called "IrgA homologue adhesive," based on its homology with the IrgA enterobactin siderophore receptor of Vibrio cholera (4) And its capacity to impart epithelial cell adhesion to a non-adherent K-12 strain when expressed from a plasmid multicopy (5). The Iha was recently

defined as an urovirulence factor for ExPEC strain CFT073. In spite of its homology to siderophore receptors, Iha has until now been functionally defined as a putative adhesive only. Iha has reported prevalence in various studies showed that most of UTI isolates containing Iha or sequences closely related to it (6,7). Some studies have documented an association of Iha with E in epidemiology UTI and other coli isolates extra intestinal diseases, compared to that with commensal fecal *E. coli* isolates whereas such an association is not established on other reports (8).

It was also demonstrated that Iha CFT073 and Iha O157:H7 have differed only in single amino acid and this difference does not deduct the capability of the cloned protein to deliberate cellular adherence ability to non-adherent laboratory *E. coli*.

Such phenotype, though anticipated, warranted confirmation because of the known ability of single amino acid polymorphisms to modify the pathogenically important receptor binding characteristics *E. coli* adhesin, FimH (9).

If Iha increases adherence directly or indirectly by affecting other bacterial components, and whether this phenotype includes receptor-specific or non-specific binding to the host cell, further evaluation is needed (9).

The bacteria acquire the iron using different strategies because of 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 would not let any iron uptake mechanism to compensate for iron depletion, the compensatory mechanism seen upon the use of conventional vaccines targeting a single iron uptake antigen (10).

This study is aimed to predict the 3D structure of the Iha protein. Prediction of the linear and conformational B cell epitopes residing within Iha protein sequence might be possibly by close investigation of their structural models. Given this information, the most immunogenic regions of the antigen could be determined and utilized to design a multivalent vaccine connected with flexible linkers.

Material and Methods

Sequence retrieval and homology alignment

The protein sequence of Iha (Accession no. ABB17254.1) retrieved from NCBI and used as a query for BLAST against no redundant protein database. To identify the homologous structures, putative sequence was served as an input data for the

PSI-BLAST. The PRALINE online server (11) was used to generate multiple sequence alignment.

Primary analysis

Using the Protparam tool (12), the physicochemical features of protein were estimated. The Iha antigenicity was predicted by vaxijen (13). The position of vaccine candidate was predicted based on the subcellular localization of prokaryotic proteins applying the PSLpred and CELLO subCELLular LOcalization predictor (14).

Secondary structure and signal peptide prediction

Prediction of the secondary structure was accomplished via SOPMA (15) and phyre2 (16).

The topology of membrane proteins and signal peptides consensus were predicted used TopCons (17). The transmembrane helices of protein also were predicted by TMHMM at (18) and SPOCTOPUS (19). PRED-TMBB server (20) is an algorithm to predict the location of transmembrane β -strands and the topology of loops in protein sequence. The SignalP 4.1 server (21) predicted the presence and location of signal peptide cleavage sites in amino acid sequence of proteins from different organisms.

3D structure prediction, evaluation and refinement

The PS2v2 (22), the SWISS-MODEL (23) and LOMETS (24) servers were used to predict the protein tertiary structure based on homology modeling. In Phyre2 a new ab initio folding simulation named Poing modeled the protein regions with no detectable homology to known structures and combine multiple templates. To select the best predicted structure we evaluated the structures. All the 3D models were evaluated by Rampage and Ramachandran plots (25) and refined for high resolution protein structure via Mod Refiner (26). ModRefiner is an algorithm for atomic-level, high-resolution protein structure refinement, which can start from either C-alpha trace, main-chain model or full-atomic model. Both side-chain and backbone atoms are completely flexible during structure refinement simulations, where conformational search is guided by a composite of physics- and knowledge-based force field. One aim of refinement is to draw the initial starting models closer to their native state, in terms of hydrogen bonds, backbone topology and side-chain positioning. It also generates significant improvement in physical quality of local structures.



Structure alignment

To evaluate validity of the structurally related sequences, the protein sequence of Iha was aligned against template sequences from previous template search. The best predicted 3D structure of the protein was structurally aligned with the most appropriate template using Dali server (27).

Membrane orientation and ligand binding site prediction

The rotational and translational positions of transmembrane and peripheral proteins in membrane were calculated with the OPM server [28]. Many membrane-associated proteins from the PDB that have been calculated already are available at the OPM database.

The COFACTOR thread the query through the BioLiP protein function database by local and global structure matches to identify functional sites and homologies (14). Cofactor at is a structure-based method for biological function annotation of protein molecules. Important amino acid involved in ligand binding site is predicted by this server.

Functional and structural residues

The 3D structure of protein served as an input file for InterProSurf (29) which predicted the functional sites on protein surface.

Surface accessible pockets and clefts

The surface accessible pockets and clefts (the binding and active sites of proteins and DNA) were find out using GHECOM program (30) as well as the CastP server (31) that identifying also the interior inaccessible cavities. The Depth program (32) was used to compute the depth, cavity sizes and ligand binding sites.

Immunogenic regions prediction

Linear and conformational epitopes could be selected as vaccine candidate and should be qualified as single-scale amino acid properties assay. Prediction of antigenicity probability of selected region was performed by vaxijen. The IEDB server and Protparam tools were employed for calculation of the physiochemical properties of selected region.

Results

Sequence availability and homology alignment

The Iha protein sequence with 696 amino acids was served as the query for BLAST search. Ten protein sequences with the highest total score, E-value=0 obtained from protein blast and ten protein sequences with the highest Total Score, E-value=0 obtained from template search were aligned for precise homology analysis.

The first hit was selected as a template for homology modeling. This top hit for Iha sequence was proteins with PDB code 2HDI_A (32 % identity). Putative conserved domains were detected within this sequence.

Most of the Iha sequences belong to CirA super family. This super family includes outer membrane receptor proteins, mostly Fe transport [inorganic ion transport and metabolism]. These outer membrane channels share a beta-barrel structure that differ in strand. A schematic illustration of homology alignment between 20 sequence obtained from protein blast and PDB blast is shown in Figure 1.

Primary features and subcellular localization

Using ProtParam tool different properties of Iha protein were calculated as shown in Table 1. The subcellular localization of Iha was predicted by CELLO to be the outer membrane with the highest reliability of 4.534. Using the PSLpred, protein sequence predicted as outer membrane.

Secondary structure

Prediction of the secondary structure indicated that the alpha helical, beta strand, beta turn and random coil structures constitute 28.30%, 23.71%, 10.78%, and 37.21% of the protein structure. The distribution of different secondary structural elements in the sequence of Iha is shown in Supplementary Figure S1. The same results were also confirmed by Phyre2 program.

Topology and signal peptide

The 2D topology model of Iha was built based on the predicted inside, transmembrane and outside regions of the protein Supplementary Figure S2. According to the TMHMM, TOPCONS and SPOCTOPUS programs no transmembrane helical regions in the protein sequence was detected. The SignalP and SPOCTOPUS predicted the first 22 residues at the N terminal of the sequence as the signal peptide.



Table 1. ProtParam and Vaxijen primary sequence analysis

Protein name	Vaxijen Score	Number of amino acids	Molecular weight	Theoretical pI	Instability index	Aliphatic index	GRAVY
Iha	0.6138	696	76481.09	5.64	30.45 (stable)	74.96	-0.426

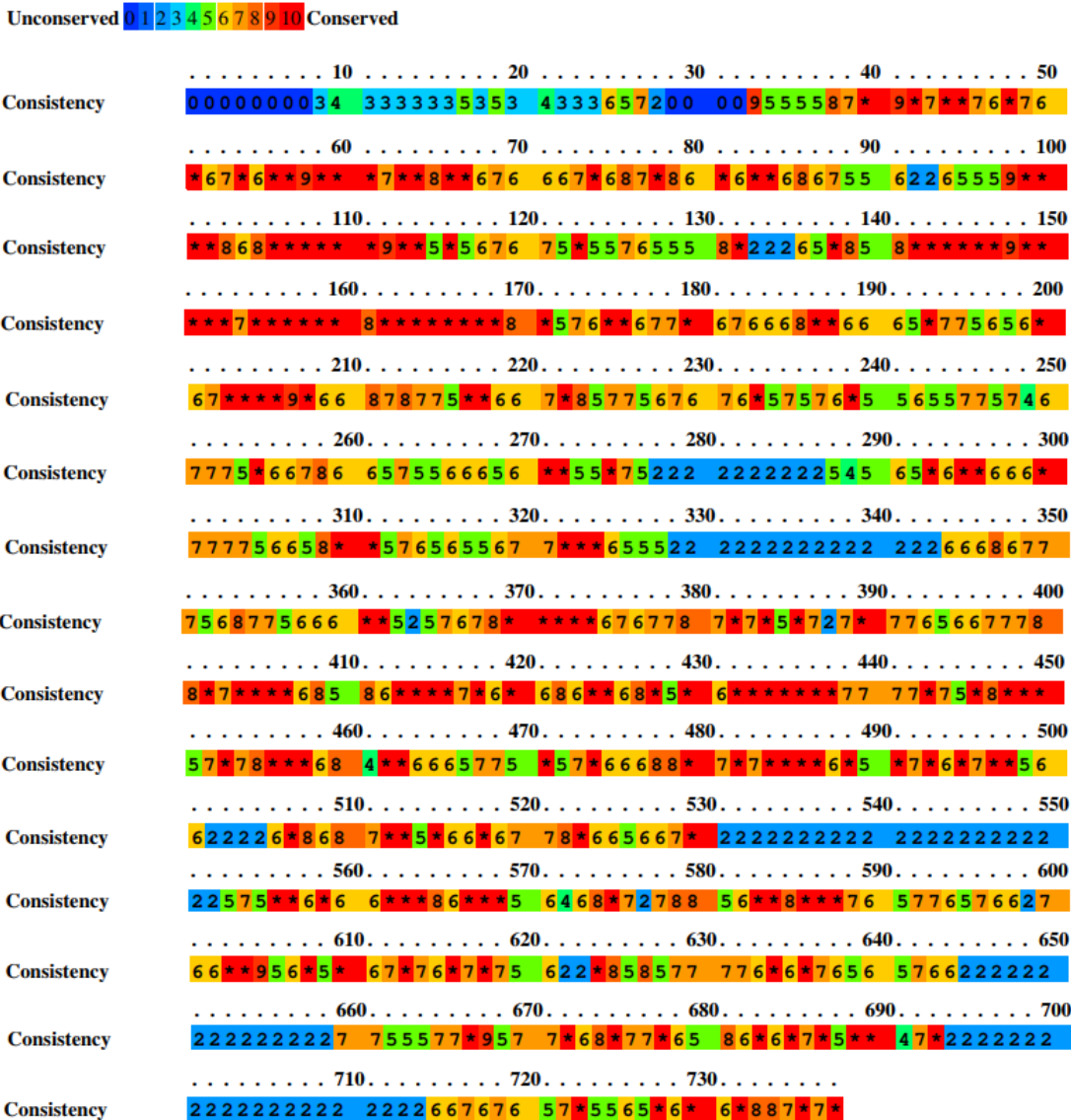


Figure 1. Illustration of homology alignment between 20 sequences attained from protein blast and PDB blast. The superposition was created with the PRALINE program and adjusted manually. Conserved residues are highlighted from blue to red colors in the nether line



3D structure evaluation and refinement

By recruiting the homology modeling of the Swiss model and Ps²v² servers 2 and 1 model identified respectively. Likewise, Phyre2 and Lomets Meta server respectively predicted one and ten 3D models for the protein. Quality of the models was assessed by Rampage and the erroneous recognizes were identified Table 2. The model proposed by FFAS03 implemented in LOMETS meta threading server scored the best based on the Ramachandran. Selected model was refined as inputs for Mod refiner server. The Ramachandran plot of initial and refined model are compared in Supplementary Figure S3.

Table 2. The report of Ramchandran Plots on different models based on the percentage of residues in favored, allowed and outlier regions.

Program	Favored region	Allowed region	Outlier region
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%
PS2V2	90.8%	6.9%	2.3%
Phyre2	88.6%	7.6%	3.9%

Structure alignment

Based on Dali structure alignments, no significant discrepancy was seen between the template and query secondary and tertiary structures. Majority of tertiary structure of the template and query was matched in tertiary structure alignment Figure 2. Expansion of structure alignment is based on distance criteria exclusively. Neither the statistical significance of the alignment nor the size of gaps regarded in this analysis. The longest alignment path is now evaluated for statistical significance as z-score. Similarities with a Z-score lower than 2 are spurious. The calculated z-score for template and predicted structure was above 3.

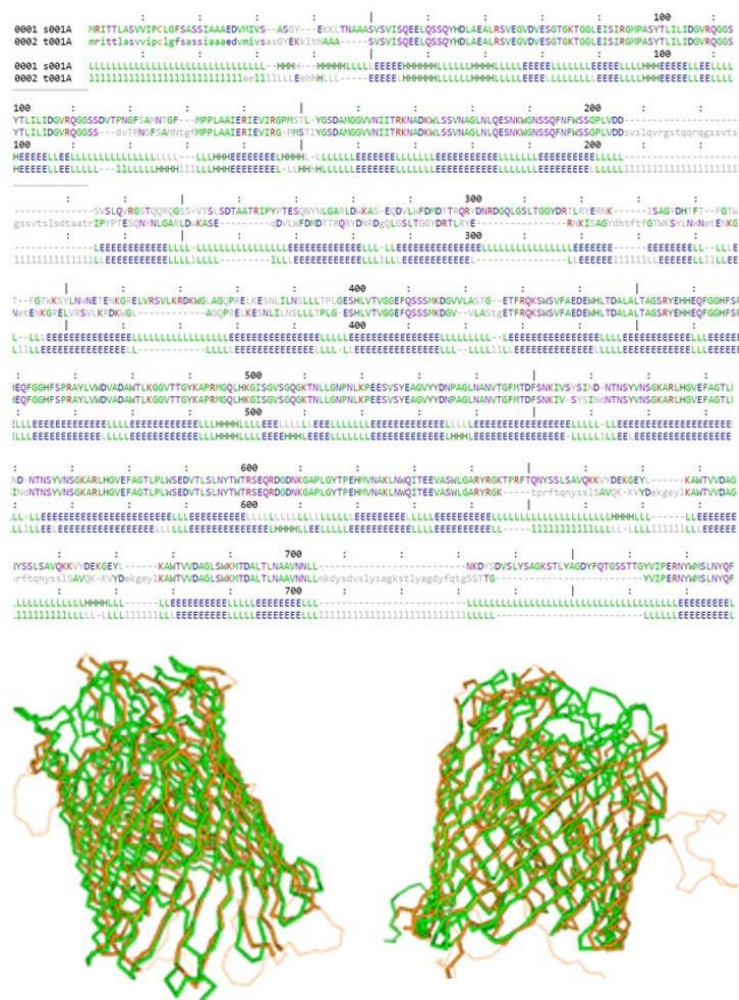
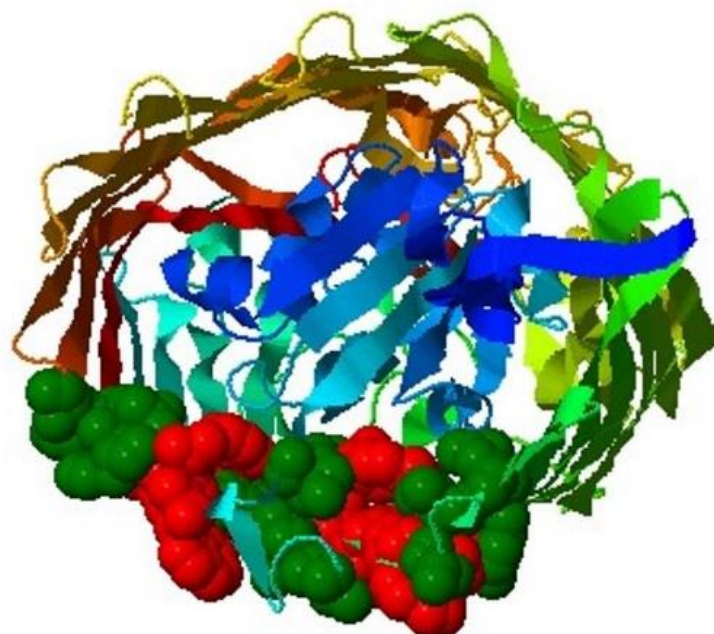


Figure 2. Dali 2D (Top) and 3D (Below) structure alignment between Iha (green) and 2HDI (orange) from lateral views





Functional residues at the protein structure surface predicted by Interprosurf

240,241,243,251,252,253,255,289,164,165,166,168,188,189,
190,694,695,696,250,292,245,291,203,204,160,199,201,239,638

Figure 3. Representation of functional residues at the protein structure surface. Top functional residues are highlighted in filling-space model (red balls) and the next top cluster highlighted in filling-space model (green balls).

The orientation of 3D structure in membrane

The OPM database prospects the orientations of quaternary complexes formed by a number of interacting proteins Supplementary Figure S4. The precision of calculated hydrophobic thicknesses and tilt angles are ~ 1 Å and 2° , respectively.

Ligand binding site

COFACTOR software indicates that the conserved residues especially those in the large extracellular loops and cork domain are involved in the iron binding site Supplementary Figure S5. COFACTOR software identified the functional conserved residues involved in the iron binding site. These functional conserved residues can be an appropriate candidate for vaccine design against Iha protein.

Functionally and structurally residues

The Interprosurf results illustrate functional sites on protein structure surface and showed that residues in the outer membrane loops and the cork domain are the most functionally

important residues in the protein structure Figure 3. The protein was identified as an iron transporter outer membrane protein.

Surface accessible pockets and clefts

Based on the mathematical morphology of GHECOM server 5 pockets were identified on the surface of protein. The GHECOM results are shown in Supplementary Figure S6.

The CastP server predicted 5 areas in the protein structure. Supplementary Figure S7 shows the charging relay system atoms which reside in the protein's functional pocket. Supplementary Figure S7 shows the likelihood of residue forming a binding site and depth plot of residue and a 3D representation of the cavity prediction.

Single-scale amino acid properties and predicted epitopes

Several properties such as hydrophilicity, accessibility, antigenicity, flexibility and another features predicted based on IEDB and BCEpred server. Most salient region of higher probability was located in position 200–340 where segment of beta barrel with large extracellular loops located. Peaks in the



Table 3. The physicochemical properties of Iha and the vaccine candidate

	Accessibility	Weight	Vaxijen score	Instability index	PI	B cell epitope	Hydrophilicity	Flexibility	Beta turn	Accessibility
Vaccine Candidates	140	16202.80	0.55	36.81 stable	9.22	0.569	2.14	1.012	1.028	1.00
Iha	696	76481.09	0.6138	30.45 stable	5.64	0.515	1.946	1.005	1.021	1.00

plot indicate putative susceptible epitope boundaries. Termini hits to the first and last 30 residues of the protein sequence were not remarkable these properties points of view. Most linear B cell epitopes that predicted by Bepipred server are concentrated in the region of 206-335. The highest scores are related to the "STOQRQGSSVTSLSDTAATRIPYPTESQ" at position 206-233, "TTRORYDNRDGQLGSLTGGYDRTLRYER" at position 257-284 and "ETENKGRELVRSVLKRDKWGLAGOPREL" at position 308-335.

On the other hand, the Svmtrip predicted 10 linear B cell epitopes ranking based on their scores. One of the best epitopes recommended by this server is "YNLGARLDWKASEQDVLWFD " at position 235-254. About 21 linear along with 3 discontinuous B cell epitopes were predicted by ElliPro. The best discontinuous and linear epitopes with the highest protrusion index (PI) are demonstrated in Figure 4. The "ARLDWKASEQDVLWFDM " at position 239-255 is one of the best linear epitopes determined by Ellipro. Discontinuous B cell epitopes predicted by Discotope which

highlighted the outer membrane loops as conformational B cell epitopes.

Immunogenic regions selection

A region includes residues 200-340 (covering a part of barrel) was chosen as vaccine candidate and several properties were compared to its parent protein (Iha). The antigenicity, PI, instability index, solubility, hydrophilicity, accessibility, flexibility and secondary structure properties calculated for both vaccine candidates. These parameters have been correlated with the location of continuous epitopes. The results for the candidate and Iha protein were summarized in Table 3. Three-dimensional structure of the vaccine candidate is shown in Figure 5.

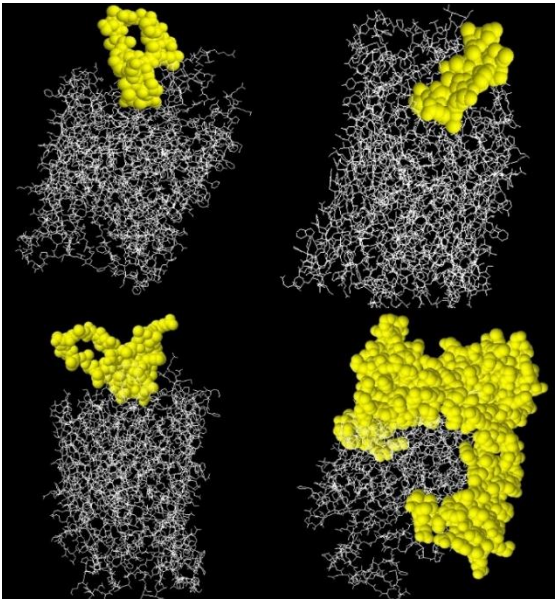


Figure 4. The best linear and discontinuous epitopes with the highest PI score predicted by Ellipro. Epitopes mapped on the 3D models using Discovery Studio Visualizer 2.5.5.

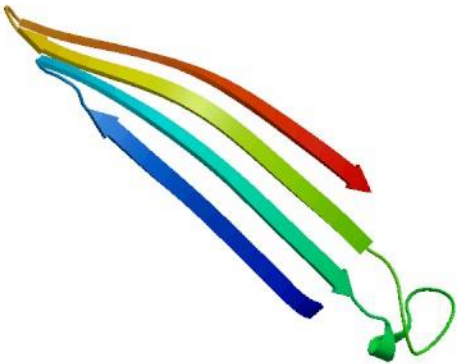


Figure 5. Three-dimensional structure of the vaccine candidate

Discussion

Developing best immunogenic agents for promoting immunity against a pathogen is one of the main challenges in the field of vaccine development. Among important pathogens, various strains of *E.coli* have long been under numerous investigations



due to their prevalence in hospital environments, especially in intensive care units (ICUs) (39, 40).

The 3D structure of an antigen is very important for accurate vaccine design. Due to laborious, costly and time-consuming processes of involved in experimental structure determination methods, *in silico* approaches gained considerable attention to generate the 3D structure of various proteins by using bioinformatics and molecular modeling tools (10, 41, 42).

Iron-regulated outer membrane virulence protein (IrgA) is involved in the initial step of iron uptake by binding ferric- iron chelating siderophore, which allows the bacterium to extract iron from the environment. IrgA homologue adhesion (Iha) has been revealed as a novel adherence conferring molecule involved in iron uptake (5). Functional blockade of the protein could have a bactericidal effect on the pathogen.

This study was conducted to develop the 3D model of Iha protein by invoking various *in silico* methods. A reliable model for a protein can be developed using different methods including homology modeling, threading, and their combination, as well as *ab initio* methods (43). We have used various strategies to generate the 3D structure of Iha protein, which then was used to determine its immunological properties and consequently to design vaccine candidates.

The sequence of Iha is homologous to numerous other molecules, as confirmed by BLAST search results. Most of the our obtained sequences belonged to TonB dependent/Ligand-Gated channels, ligand-gated-channel protein family and outer membrane-channels superfamily (44).

The result of BLAST search revealed protein Colicin I Receptor Cir (PDB code: 2HDI_A) belonging to of IROMPs family of proteins, as the best template for structure prediction of Iha based on E value, query coverage and Max. Identity values. The identity between query and template primary sequences was 32%.

One of the members of IROMPs family have resolved crystal structures including 2HDI_A (Crystal Structure of the Colicin I Receptor Cir from *E. coli* in Complex with Receptor Binding Domain Of Colicin Ia for Iha proteins).

Prediction accuracy depends on the similarity of degree of sequence. When for a query protein a structure template with a sequence identity of > 60 percent is found, homology modeling can be selected as the best *in silico* approach with an X-ray approach of accuracy that is as low-resolution. Significant errors in prediction will arise when the sequences share sequence identity of less than 30 per cent. Since in our analysis the sequence similarity between Iha and the template was 32 per

cent, the threading approach was also used for Iha's structure prediction.

Our obtained structure provided important clues concerning the architecture of all TonB-dependent receptors and all of them are important pathogenic antigens in bacterial infections (45, 46).

The models designed by different servers had to be tested in order to arrive at the most appropriate one. Scoring programs representing conformational energy were implemented to take a decision among the 3D structures predicted. Ramachandran plot, the second plot of the phi psi angles is a standard tool exploited in determining protein structure. According to the quality assessment, the model that generated by the LOMETS server was chosen as the best.

A 2D topology model of this iron receptor vaccine candidates (Iha) was built focused on expected proteins regions containing inside, transmembrane and outside. Our results disclosed that Iha is composed of several trans-membrane anti parallel β -strands. The predicted models proposed that the proteins have a β -barrel structure in their native form [47]. The strands forming β -barrel are linked together through loops at the outside or turns at the inside (Figure 1). This protein contained more than 11 external loops, and the side chains of all residues are highly exposed to the environment indicating their function in initial binding events with Fe-siderophore complex.

Biological protein roles are performed in conjunction with other molecules. Pocket size may decide protein interaction with other molecules in the protein surface. Correlation of binding sites and clefts in a given protein is a basis for our pocket investigation.

A large pocket may increase opportunities for forming protein-to-other molecules interactions, particularly small ligands. Normally, a protein's active site is in the largest clefts or cavities. The ligand is connected to the largest cleft in more than 83 percent of single-chain proteins.

Thus, protein-ligand sites are likely to be identified merely geometrically. Surface binding pockets and occluded cavities are notable for designing vaccines and producing drugs against pathogenic agents. This essential cleft lies at the protein structure's exterior edge.

Antigenicity and immunogenicity of an antigen is in direct correlation with its epitope density (48) for B-cells, which play a significant role in vaccine formulation, immunodiagnostic testing and in the development of antibodies.

The determined epitomic data could be harnessed to choose Iha regions with higher epitope density that could be used for



inducing humoral responses with higher average avidity for epitope-specific monoclonal and polyclonal antibodies. The best Linear B cell epitopes of this iron receptor vaccine candidate were located at the largest extra cellular loops. Interestingly, the conformational B cell epitopes predicted from the 3D structure of proteins include all of the extra cellular loops. Our epitopes prediction results were in concordance with experimentally identified epitopes tested by approved antibodies. This concordance indicates the accuracy of employed procedures for both 3D structure prediction and epitope prediction. Comparison of antigenicity scores for selected regions and the whole proteins showed that the antigenicity of the selected regions is considerably higher than the whole antigen.

Conclusion

In this study, we found a region in Iha molecule as an antigen with the highest immunogenic properties using the immunological, biochemical and functional analyzes. The antigenicity scores for selected regions is considerably higher than the whole antigen.

In conclusion, it should be noted that as a compelling strategy, using bioinformatics tools, scientists could be able to find the relationship between the number of protein sequences and the resolved 3D protein structures. Evidence from structural and immunological properties of antigens may be used in silico for purposes of vaccine design. In the context of vaccine design investigations, our strategy to employ 3D structure prediction and epitope prediction findings could pave the way for further structural, functional, and therapeutic studies.

Acknowledgments

The authors wish to thank Immunology Research Center, Tabriz- Iran and Shahid Sadoughi University of Medical Science, Yazd, Iran for their support by unrestricted free access to the web site for data collection.

Supplementary Material is available at RRR online

Conflict of Interest: 'None declared'.

References

1. P.M. Griffin, R.V. Tauxe, The epidemiology of infections caused by *Escherichia coli* O157: H7, other enterohemorrhagic *E. coli*, and the associated hemolytic uremic syndrome, *Epidemiologic reviews* 13(1) (1991) 60-98.
2. C.S. Edén, U. Jodal, L. Hanson, U. Lindberg, A.S. Åkerlund, Variable adherence to normal human urinary-tract epithelial cells of *Escherichia coli* strains associated with various forms of urinary-tract infection, *The Lancet* 308(7984) (1976) 490-492.
3. A.G. Torres, X. Zhou, J.B. Kaper, Adherence of diarrheagenic *Escherichia coli* strains to epithelial cells, *Infection and immunity* 73(1) (2005) 18-29.
4. R.A. Rashid, P.I. Tarr, S.L. Moseley, Expression of the *Escherichia coli* IrgA homolog adhesin is regulated by the ferric uptake regulation protein, *Microbial pathogenesis* 41(6) (2006) 207-217.
5. P.I. Tarr, S.S. Bilge, J.C. Vary, S. Jelacic, R.L. Habeeb, T.R. Ward, M.R. Baylor, T.E. Besser, Iha: a novel *Escherichia coli* O157: H7 adherence-conferring molecule encoded on a recently acquired chromosomal island of conserved structure, *Infection and immunity* 68(3) (2000) 1400-1407.
6. S. Kanamaru, H. Kurazono, S. Ishitoya, A. Terai, T. Habuchi, M. Nakano, O. Ogawa, S. Yamamoto, Distribution and genetic association of putative uropathogenic virulence factors *iroN*, *iha*, *kpsMT*, *ompT* and *usp* in *Escherichia coli* isolated from urinary tract infections in Japan, *The Journal of urology* 170(6) (2003) 2490-2493.
7. R.J. Bauer, L. Zhang, B. Foxman, A. Siitonen, M.E. Jantunen, H. Saxen, C.F. Marrs, Molecular epidemiology of 3 putative virulence genes for *Escherichia coli* urinary tract infection—*usp*, *iha*, and *iroNE*. *coli*, *Journal of Infectious Diseases* 185(10) (2002) 1521-1524.
8. J.R. Johnson, A.C. Murray, M.A. Kuskowski, S. Schubert, M.-F. Prere, B. Picard, R. Colodner, R. Raz, T.-G.I.f.A.R.A. Investigators, Distribution and characteristics of *Escherichia coli* clonal group A, *Emerging infectious diseases* 11(1) (2005) 141.
9. J.R. Johnson, S. Jelacic, L.M. Schoening, C. Clabots, N. Shaikh, H.L. Mobley, P.I. Tarr, The IrgA homologue adhesin Iha is an *Escherichia coli* virulence factor in murine urinary tract infection, *Infection and immunity* 73(2) (2005) 965-971.
10. A. Jahangiri, I. Rasooli, S.L.M. Gargari, P. Owlia, M.R. Rahbar, J. Amani, S. Khalili, An in silico DNA vaccine against *Listeria monocytogenes*, *Vaccine* 29(40) (2011) 6948-6958.
11. P. Bawono, J. Heringa, PRALINE: a versatile multiple sequence alignment toolkit, *Multiple Sequence Alignment Methods*, Springer2014, pp. 245-262.
12. E. Gasteiger, C. Hoogland, A. Gattiker, M.R. Wilkins, R.D. Appel, A. Bairoch, Protein identification and analysis tools on the ExPASy server, *The proteomics protocols handbook*, Springer2005, pp. 571-607.



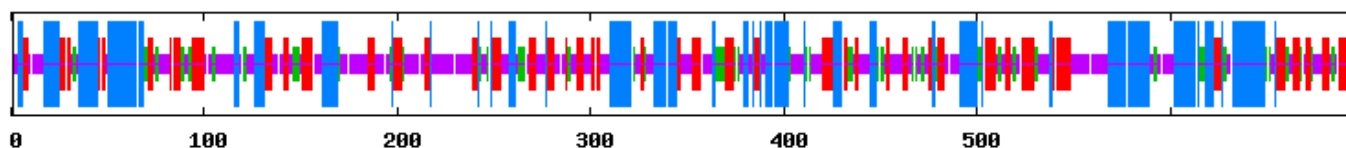
13. I.A. Doytchinova, D.R. Flower, VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines, *BMC bioinformatics* 8(1) (2007) 4.
14. F. Sefid, A.A. Bahrami, M. Darvish, R. Nazarpour, Z. Payandeh, In Silico Analysis for Determination and Validation of Iron-Regulated Protein from *Escherichia coli*, *International Journal of Peptide Research and Therapeutics* 25(4) (2019) 1523-1537.
15. C. Geourjon, G. Deleage, SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments, *Bioinformatics* 11(6) (1995) 681-684.
16. L.A. Kelley, S. Mezulis, C.M. Yates, M.N. Wass, M.J. Sternberg, The Phyre2 web portal for protein modeling, prediction and analysis, *Nature protocols* 10(6) (2015) 845.
17. K.D. Tsirigos, C. Peters, N. Shu, L. Käll, A. Elofsson, The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides, *Nucleic acids research* 43(W1) (2015) W401-W407.
18. A. Krogh, B. Larsson, G. Von Heijne, E.L. Sonnhammer, Predicting transmembrane protein topology with a hidden markov model: application to complete genomes¹, *Journal of molecular biology* 305(3) (2001) 567-580.
19. H. Viklund, A. Bernsel, M. Skwark, A. Elofsson, SPOCTOPUS: a combined predictor of signal peptides and membrane protein topology, *Bioinformatics* 24(24) (2008) 2928-2929.
20. P.G. Bagos, T.D. Liakopoulos, I.C. Spyropoulos, S.J. Hamodrakas, PRED-TMBB: a web server for predicting the topology of β -barrel outer membrane proteins, *Nucleic acids research* 32(suppl_2) (2004) W400-W404.
21. T.N. Petersen, S. Brunak, G. Von Heijne, H. Nielsen, SignalP 4.0: discriminating signal peptides from transmembrane regions, *Nature methods* 8(10) (2011) 785.
22. C.-C. Chen, J.-K. Hwang, J.-M. Yang, 2-v2: template-based protein structure prediction server, *Bmc Bioinformatics* 10(1) (2009) 366.
23. T. Schwede, J. Kopp, N. Guex, M.C. Peitsch, SWISS-MODEL: an automated protein homology-modeling server, *Nucleic acids research* 31(13) (2003) 3381-3385.
24. S. Wu, Y. Zhang, LOMETS: a local meta-threading-server for protein structure prediction, *Nucleic acids research* 35(10) (2007) 3375-3382.
25. O. Carugo, K. Djinović-Carugo, Half a century of Ramachandran plots, *Acta Crystallographica Section D: Biological Crystallography* 69(8) (2013) 1333-1341.
26. D. Xu, Y. Zhang, Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization, *Biophysical journal* 101(10) (2011) 2525-2534.
27. L. Holm, Using Dali for Protein Structure Comparison, *Structural Bioinformatics*, Springer2020, pp. 29-42.
28. M.A. Lomize, I.D. Pogozheva, H. Joo, H.I. Mosberg, A.L. Lomize, OPM database and PPM web server: resources for positioning of proteins in membranes, *Nucleic acids research* 40(D1) (2012) D370-D376.
29. S.S. Negi, C.H. Schein, N. Oezguen, T.D. Power, W. Braun, InterProSurf: a web server for predicting interacting sites on protein surfaces, *Bioinformatics* 23(24) (2007) 3397-3399.
30. T. Kawabata, Detection of multiscale pockets on protein surfaces using mathematical morphology, *Proteins: Structure, Function, and Bioinformatics* 78(5) (2010) 1195-1211.
31. J. Dundas, Z. Ouyang, J. Tseng, A. Binkowski, Y. Turpaz, J. Liang, CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues, *Nucleic acids research* 34(suppl_2) (2006) W116-W118.
32. K.P. Tan, T.B. Nguyen, S. Patel, R. Varadarajan, M.S. Madhusudhan, 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 research* 41(W1) (2013) W314-W321.
33. R. Vita, J.A. Overton, J.A. Greenbaum, J. Ponomarenko, J.D. Clark, J.R. Cantrell, D.K. Wheeler, J.L. Gabbard, D. Hix, A. Sette, The immune epitope database (IEDB) 3.0, *Nucleic acids research* 43(D1) (2014) D405-D412.
34. S. Saha, G. Raghava, BcePred: prediction of continuous B-cell epitopes in antigenic sequences using physico-chemical properties, *International Conference on Artificial Immune Systems*, Springer, 2004, pp. 197-204.
35. M.C. Jespersen, B. Peters, M. Nielsen, P. Marcatili, BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes, *Nucleic acids research* 45(W1) (2017) W24-W29.
36. B. Yao, L. Zhang, S. Liang, C. Zhang, SVMTriP: a method to predict antigenic epitopes using support vector machine to integrate tri-peptide similarity and propensity, *PloS one* 7(9) (2012) e45152.
37. M.N. Davies, D.R. Flower, Harnessing bioinformatics to discover new vaccines, *Drug discovery today* 12(9-10) (2007) 389-395.
38. J. Ponomarenko, H.-H. Bui, W. Li, N. Fusseder, P.E. Bourne, A. Sette, B. Peters, ElliPro: a new structure-based tool for the prediction of antibody epitopes, *BMC bioinformatics* 9(1) (2008) 514.
39. H.W. Bauer, S. Alloussi, G. Egger, H.-M. Blümlein, G. Cozma, C.C. Schulman, M.U.S. Group, A long-term, multicenter, double-blind study of an *Escherichia coli* extract (OM-89) in female patients with recurrent urinary tract infections, *European urology* 47(4) (2005) 542-548.
40. D.T. Uehling, W.J. Hopkins, J.E. Elkahwaji, D.M. Schmidt, G.E. Levenson, Phase 2 clinical trial of a vaginal mucosal vaccine for urinary tract infections, *The Journal of urology* 170(3) (2003) 867-869.
41. F. Sefid, I. Rasooli, A. Jahangiri, In silico determination and validation of baumannii acinetobactin utilization a structure and ligand binding site, *BioMed research international* 2013 (2013).
42. Z. Payandeh, M. Rajabibazl, Y. Mortazavi, A. Rahimpour, In Silico Analysis for Determination and Validation of Human CD20 Antigen 3D Structure, *International Journal of Peptide Research and Therapeutics* (2017) 1-13.



43. G.J. Kleywegt, T.A. Jones, Databases in protein crystallography, *Acta Crystallographica Section D: Biological Crystallography* 54(6) (1998) 1119-1131.
44. K. Brillet, C. Reimann, G.L. Mislin, S. Noël, D. Rognan, I.J. Schalk, D. Cobessi, Pyochelin enantiomers and their outer-membrane siderophore transporters in fluorescent pseudomonads: structural bases for unique enantiospecific recognition, *Journal of the American Chemical Society* 133(41) (2011) 16503-16509.
45. J.S. Oakhill, B.J. Sutton, A.R. Gorringe, R.W. Evans, Homology modelling of transferrin-binding protein A from *Neisseria meningitidis*, *Protein Engineering Design and Selection* 18(5) (2005) 221-228.
46. F. Sefid, I. Rasooli, Z. Payandeh, Homology modeling of a Camelid antibody fragment against a conserved region of *Acinetobacter*

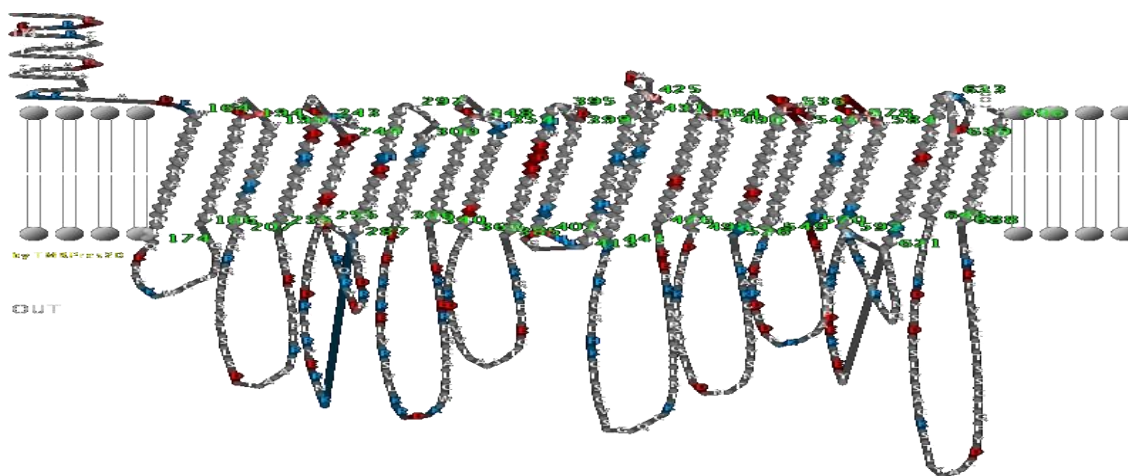
- baumannii* biofilm associated protein (Bap), *Journal of theoretical biology* 397 (2016) 43-51.
47. P.G. Bagos, T.D. Liakopoulos, S.J. Hamodrakas, Evaluation of methods for predicting the topology of β -barrel outer membrane proteins and a consensus prediction method, *BMC bioinformatics* 6(1) (2005) 7.
48. W. Liu, Y.H. Chen, 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, *European journal of immunology* 35(2) (2005) 505-514.

Please cite this paper as: Sefid F, Alagheband Bahrami A, Rajabibazl M, Rahmati M, Payandeh M, Kalantar SM, Dastmalchi S, Nabi Afjadi M. Design of an Epitope Candidate Vaccine Against Iha Protein in *Escherichia Coli*: an in Silico Approach. *Regen Reconstr Restor.* 2020;5 (1): e19. Doi: 10.22037/rrr.v5i1.29954.



Supplementary Figure S1. Representation of secondary structure prediction by SOPMA.

(Blue: alpha helix, Red: Extended strand, Green: Beta turn, Yellow: Random coil)



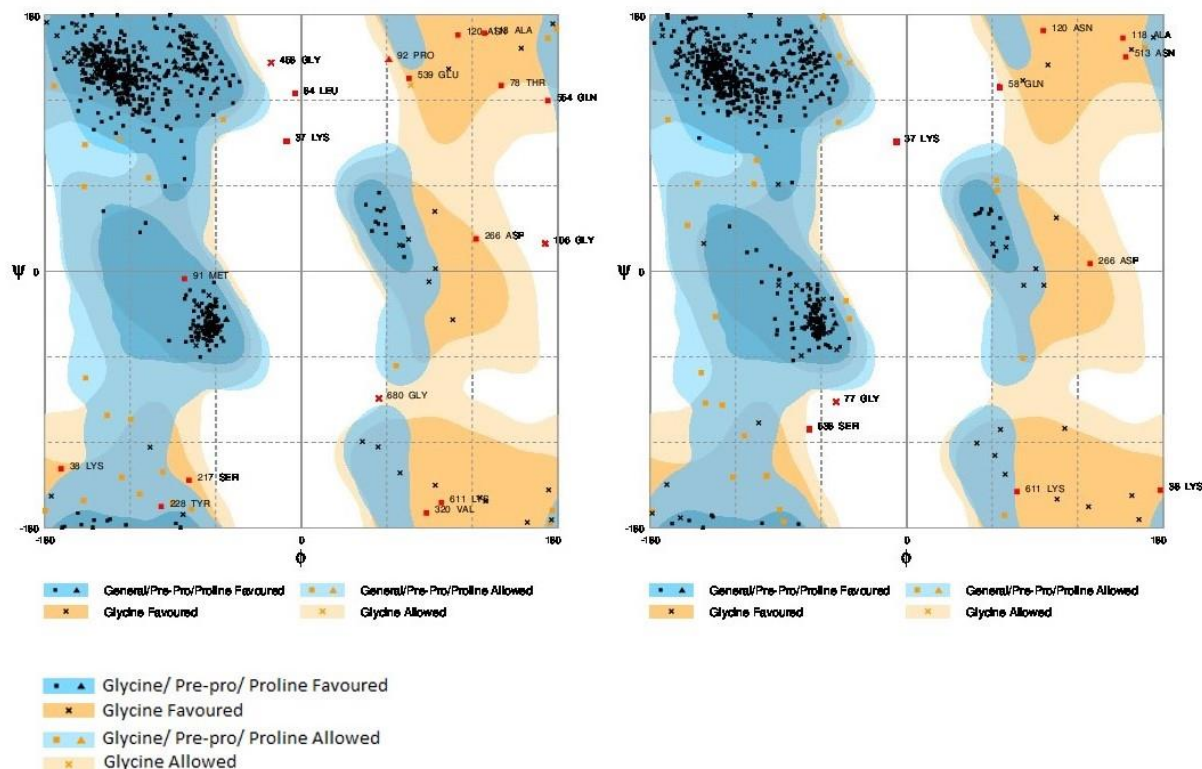
Supplementary Figure S2. A 2D topology model of Iha. Whole protein includes 2 domains comprising a cork domain at the N-terminal and a transmembrane barrel at the C-terminal. The barrel topology is composed of 22 transmembrane beta strand, 10 short periplasmic turns and 11 large extracellular loops.



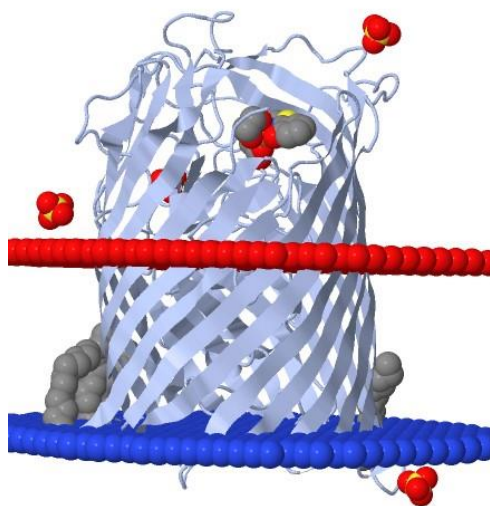
Regeneration, Reconstruction & Restoration 2020; 5(1): e19



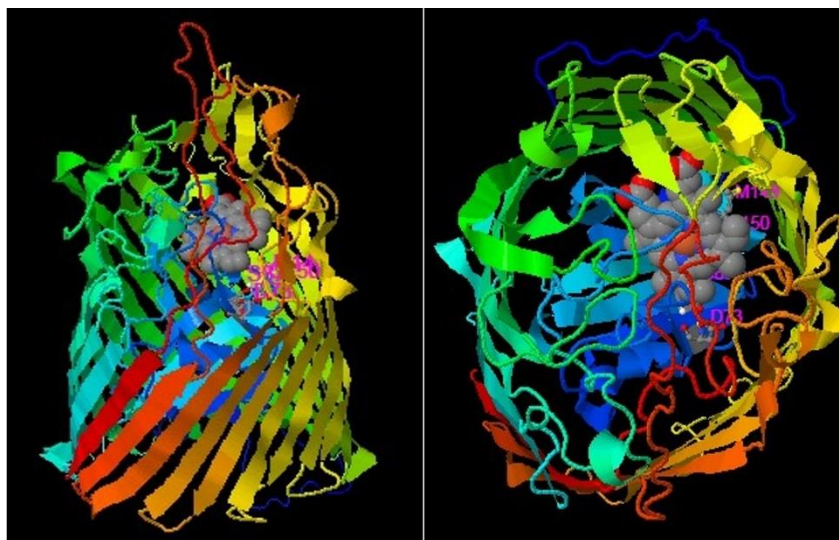
This work is licensed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International.



Supplementary Figure S3. The Ramachandran plot of initial and refined model of Iha. The percent residues in the favored region was 94.2% while 94.7% in the final model. The percent residues in the allowed region for initial and final models are 3.2% and 3.9% respectively. Also, this percent in outlier region for initial and final models are 2.6% and 1.4% respectively.

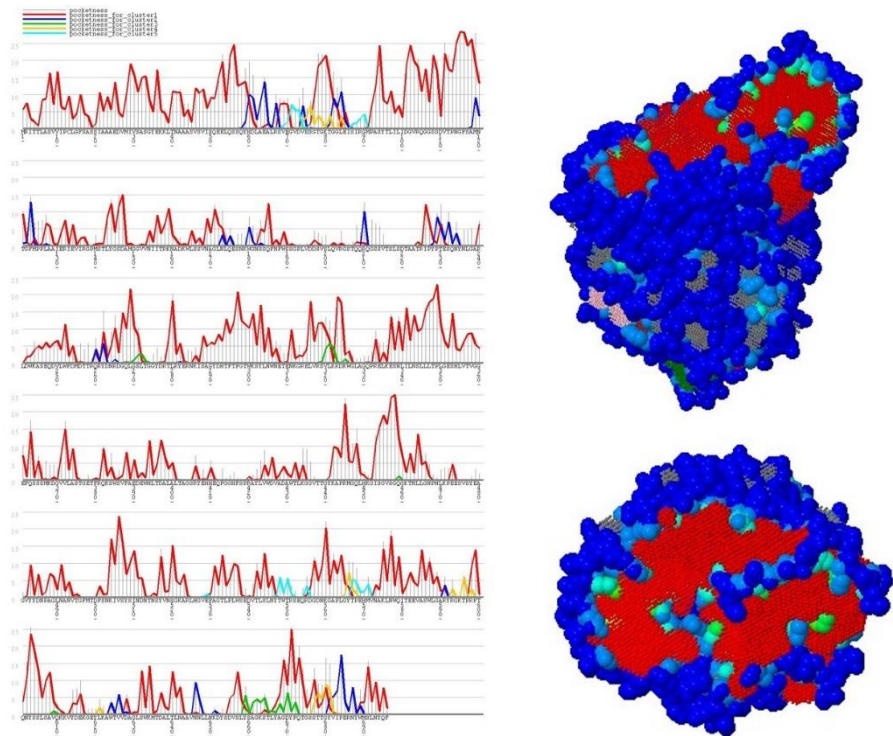


Supplementary Figure S4. The orientation of protein in cell membrane. The origin of coordinates corresponds to the center of lipid bilayer

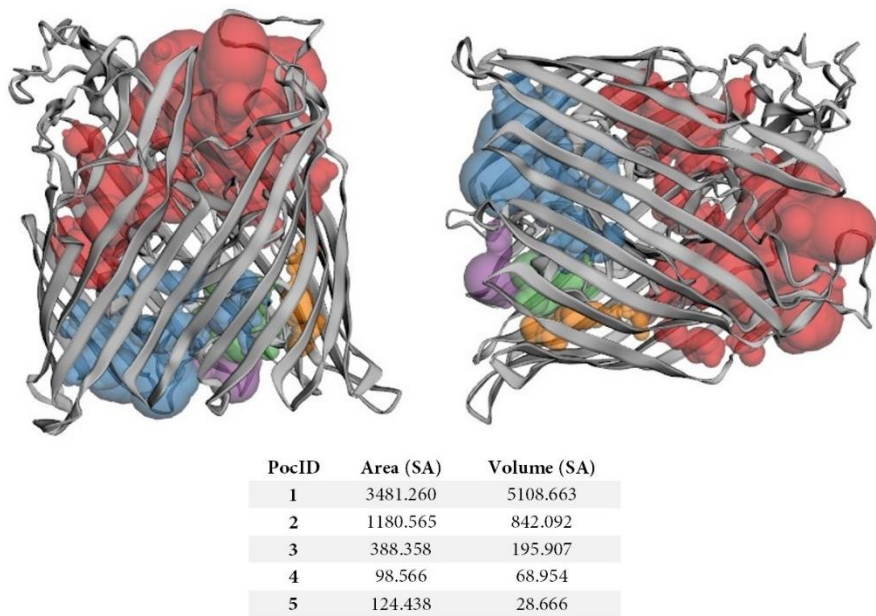


Supplementary Figure S5. Representation of the Iha structure in contacts with the ligand; (lateral and top view). The protein 3D structure is shown in ribbon and ligand in the space-filling model



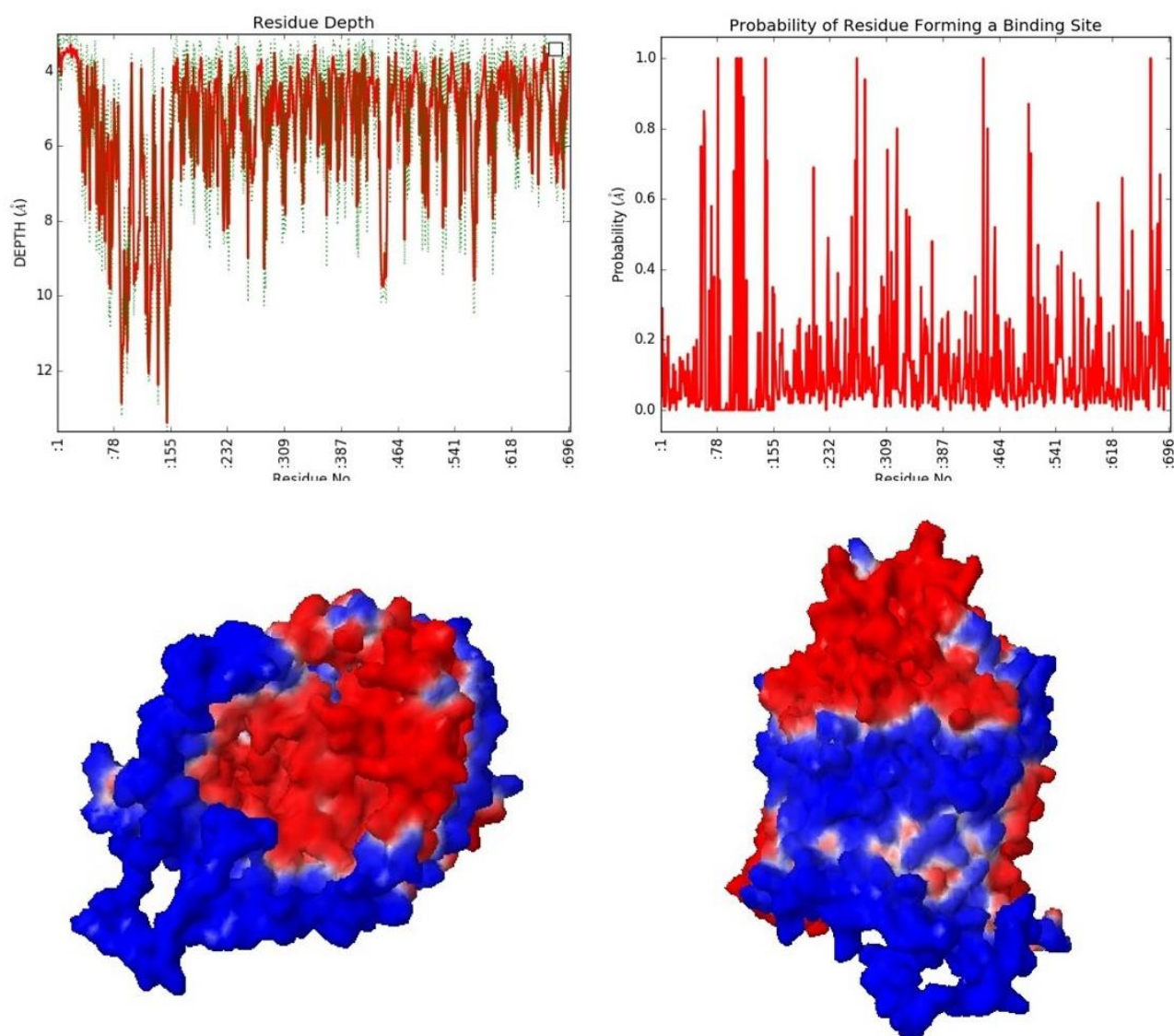


Supplementary Figure S6. The illustration of graph residue-based pocketness and Jmol view of pocket structure based on GHECOM results; Top: graph residue-based pocketness. The height of the bar displays the value of pocketness [%] for each residue. The height of the bar displays the value of pocketness [%] 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 (Left: Front view, Right: Top view).



Supplementary Figure S7. The surface accessible pockets and interior inaccessible cavities illustrated by CastP. Pockets are colored based on area and volume size. The most important one demonstrated in red and other are shown in blue, green, purple and orange respectively.





Supplementary Figure S8. 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.