Microbes and Infectious Diseases 2022; 3(2): 407-419

Microbes and Infectious Diseases

Journal homepage: https://mid.journals.ekb.eg/

Original article

A Novel design of multi-epitope based vaccine against *Escherichia coli*

Adeoti OM [*1, 2, 5], Jimoh AO [1], Akinsooto AV [1], Agboluaje AK [1], Aderinto AO [1], Sanusi TA [1], Adesina DA [3], Komolafe KA[1,4], Olaoye OJ [1]

1- *Department of Science Laboratory Technology, Microbiology option, The Oke- Ogun Polytechnic, Saki Oyo State Nigeria,.*
2- *Department of Pure & Applied Biology, Ogbomoso Ladoke Akintola University of Technology,.*
3- *Department of Microbiology and Botany, University of Ibadan, Ibadan, Nigeria.*
4- *Department of Zoology, Parasitology Unit, University of Ibadan, Nigeria.*
5- *Cellular Parasitology Unit, Department of Zoology, University of Ibadan, Nigeria.*

ARTICLE INFO

Article history:
Received 26 December 2020
Received in revised form 19 February 2021
Accepted 27 February 2021

Keywords:
Immunoinformatics
Allergenicity
Antigenicity
Pharmaco-therapeutics
Multi-epitope
Vaccine

ABSTRACT

Background: Multi-valent based vaccines have advantage over conventional vaccines because of its multi-faceted action targeted at antigen; thereby raising hope of a more sustained actions against allergens. *Escherichia coli (E. coli)* is a bacterium that is commonly found in the gut of humans and warm-blooded animals. An increasing number of outbreaks are associated with the consumption of fruits and vegetables (including sprouts, spinach, lettuce, coleslaw, and salad) thereby contamination may be due to contact with faeces from domestic or wild animals at some stages during cultivation or handling. Due to the reported increase in resistance to antibiotics used for *Escherichia coli* control; an effective vaccine is a would-be alternative of proven interest. Hence, a need for a rational, strategic, and efficient vaccine candidate against *E.coli* is of paramount necessity by the use of the most current bioinformatics tools to achieve this task. **Method:** In this study, immunoinformatics tools mined from diverse molecular databases were used for a novel putative epitope based oral vaccine against *E.coli*. The prospective vaccine proteins were carefully screened and validated to achieve a high thorough-put three-dimensional protein structure. The eventual propsective vaccine candidate proteins was evaluated for its non-allergenicity, antigenicity, solubility, appropriate molecular weight testing and isoelectric point evaluation. **Conclusion:** The resultant vaccine candidate could serve as a promising *anti-E.coli* vaccine candidate. Immunoinformatics is a new field over pharmaco-therapeutics; this newest technology should continue to be a rescue from age-long traditional approach in vaccine developments.

Introduction

Escherichia coli (*E. coli*) is a bacterium that is commonly found in the gut of humans and warm-blooded animals. Most strains of *E. coli* are harmless. Some strains however, such as Shiga toxin-producing *E. coli* (STEC), can cause severe foodborne disease. It is transmitted to humans primarily through consumption of contaminated foods, such as raw or undercooked ground meat products, raw milk, and contaminated raw vegetables and sprouts [1]. The total mortality rate of ETEC in a recent report of 2015 was estimated 74,100 deaths among individuals out of which

DOI: 10.21608/MID.2021.55310.1101
* *Corresponding author:* Adeoti Olatude Micheal
E-mail address: txy23m@yahoo.com

23,600 deaths were reported in infants [2] and World Health Organization (WHO) estimated greater than 500,000 deaths in children between 1–4 years of age [3]. Heterologous expression in standard host cells such as *E.coli* often results in insoluble proteins. However, the exploration of structural and functional proteomics requires proteins to be produced in soluble form [4].

To obtain soluble proteins from insoluble forms, a series of downstream processing steps involving re-solubilization using strong denaturants followed by refolding is inevitable [5]. From the first observation on the 'solubilizing characteristics' of soluble proteins [6], which later advances to the deduction that amino acid sequence is the major determinant of protein solubility; in different studies, a number of bioinformatics tools have been developed based on this deduction to date [7]. This exciting finding has revealed the possibility of enhancing the recombinant protein production through *in silico* experimentation using bioinformatics tools. By using protein solubility prediction tools, one can omit the tedious yet non-promising trial and error procedures involved in the in situ experiments. Subsequently, one can focus on carrying out in situ experiments with those promising candidates of vector-gene pairings, which are predicted to yield high solubility when overexpressed in *E. coli.* The ability of conducting theoretical prediction of protein solubility with high accuracy will certainly contribute to the development of large-scale proteomics studies [8].

Rational target selection of the most promising candidates of vector-gene-host combination will be important to various biophysical and proteomics studies [8]. Bacterium coli were used for biological laboratory experiment research, infection can lead to hemolytic anemia, thrombocytopenia, and renal injury. Most *E. coli* strain does not cause diseases, but virulent strains can cause gastroenteritis, urinary tract infections, and neonatal meningitis diseases. *Escherichia coli* grows very quickly and the generation time in the intestine is thought to be about 12 hours. Under optimum conditions the generation time is 20 minutes. Some strains of *E. coli* bacteria produce an enzyme called extended-spectrum β-lactamase (ESBL) which helps *E. coli to* resist many types of antimicrobials. Among all the strains of *E. coli some* strains are useful for producing sources of B‟ and K‟ vitamins for the host and some are harmless to hosts while some strains can cause illness such as urinary tract infections, meningitis, peritonitis, mastitis, septicemia and Gram-negative acidic polysaccharide [9].

Escherichia coli coli strains can cause illness such as entero invasive *E. coli* (EIEC), entero-toxigenic *E. coli* (ETEC), entero-pathogenic *E. coli* (EPEC) and entero-hemorrhagic *E. coli* (EHEC). *Escherichia coli* (*E. coli*) are a group of genus of Gram negative, anaerobic, rod-shaped bacteria, which inhabits the intestine of all healthy humans and most warm blooded animals. There are many strains of *E. coli* bacteria and most of them are harmless and serves a useful function in the body by suppressing the growth of harmful bacterial species and by synthesizing appreciable amounts of vitamins. A very few strains *E. coli* are having the capability of causing human illness by several different mechanisms.

Escherichia coli 0.157:H7 is a potentially deadly bacteria strain that can cause bloody diarrhea and dehydration, especially in children. Globally *E. coli* is used as the preferred indicator of faecal pollution. It is a Gram negative bacterium and predominantly an inhabitant of the intestines of warm blooded animals and humans, which is used to indicate recent faecal pollution of water samples Confirmation tests for *E. coli* include testing for the presence of the enzyme β-glucuronidase, Gram staining, absence of urease activity, production of acid and gas from lactose and indole production.

Shiga toxin-producing *E.coli* produces toxins, known as Shiga-toxins because of their similarity to the toxins produced by Shigella dysenteriae. STEC can grow in temperatures ranging from 7 °C to 50 °C, with an optimum temperature of 37 °C. Some STEC can grow in acidic foods, down to a pH of 4.4, and in foods with a minimum water activity (a_W) of 0.95 [1].

Shiga toxin-producing *E.coli* is destroyed by thorough cooking of foods until all parts reach a temperature of 70 °C or higher. *E. coli* O157:H7 is the most important STEC serotype in relation to public health; however, other serotypes have frequently been involved in sporadic cases and outbreaks.

The technological advancement in the fields of molecular biology, gene structure and function, proteomics, vaccinology, Computational biology, Immunology, and structural biology have provided the molecular informational platform for the design and prediction by bioinformatics tools of novel antigens, multi-epitopes, and design of

vaccines against pathogenic bacteria and viruses [9-11]. Using appropriate bioinformatics tools which can predict sequences with precise binding affinity to major histocompatibility complex (MHC) class I and II alleles as well as predict epitopes of T and B cells [12]. The major interest on vaccine design and development has changed to the production of peptides composed of multiple epitopes based (multi epitope vaccines), which has been hinged on linear arrangements, as a novel alternative. In addition, epitope-based vaccines have demonstrated various advantages, including safety, the opportunity to rationally engineer the epitopes for increased potency, breadth, and antigenicity, and the possibility to focus large repertories of immune responses on conserved epitope sequences [13,14]. This study is aimed at predicting the design of a novel *in silico* oral vaccine against *E.coli*.

A focus on vaccine design and development has moved to the generation of recombinant multi-epitope vaccines as a new strategy. The potential benefits of vaccines based on epitopes include a specific immune response avoiding the side effects of other unfavorable epitopes in the complete antigen and increased safety [15]. An epitope based vaccine could also include single antigenic molecules combined from different epitopes for increased potency and it could lead to an effective strategy for the control of *H. pylori* [16]. As we already know, protein epitopes recognized by both T and B cells are the best candidates for vaccines because of their high specificity properties [17]. In this context, bioinformatics tools nowadays have a significant role to identify appropriate epitopes, in addition to saving time and being cost effective [18].

As we can see a variety of studies have proved that each individual design has contributed to vaccines development. According to **Pahl and Beitz** [19], a novel design incorporates new solution principles, performed by combining known principles. Thus, the holistic application of a combined approach could result in an effective vaccine design.

Materials and Methods

Collection, selection and antigenic evaluation of selected protein sequences

Twenty (20) *E.coli* essential proteins sequences were retrieved from the National Center for Biotechnology Information (NCBI) bioinformatics database which based on the following criteria: (1) reported antigenicity, (2) virulence/drug resistance, and (3) proteins relatedness to the mechanisms of adhesion. An effective antigen protein should meet the requirements including the following. (1) The candidate antigen protein should be able to trigger the immune response in hosts. (2) The vaccine targets should be accessible to the immune effectors. Based on the above hypothesis, the immunogenicity scoring model was constructed to evaluate the immunogenicity of selected proteins for *E. coli*. All the twenty protein sequences retrieved from NCBI reference sequences database were all in FASTA format. For example, *E. coli*, bacterial protein for subcellular localization prediction, CELLO version 2.5 was according to Yu *et al.*, [27]. The database of GepTop evaluated the essentiality complex of the genes according to [20]. Then the database of Pathosystems Resource Integration Center (PATRIC3.5.16) was used for studying the virulence/resistance roles of proteins. The selected Proteins were aligned through Basic Local Alignment BLASTp and to detect sequence homologs of the genes to *humans'*. To predict transmembrane (TM) helixes, TMHMM version 2.0 was used according to [21]. We computed isoelectric point ratio to molecular weight (pI/Mw) tool to calculate the estimated isoelectric point and molecular weight of all the twenty amino acid sequences [22].

Phylogenetic evolution

Phylogenetic tree of the retrieved sequence of the capsid protein of *E.coli* was created using Mega X software. Twenty of the protein tree was constructed using maximum likelihood parameter in the software **(Figure 1)**.

Conservationism, consensus sequence and alignment

Conservation of selected proteins was analyzed using *E.coli* strain **26695** as the reference microbe. For proteins with high diversity, a consensus sequence was generated by using sequences of *E.coli* representative strains (20 complete genome sequences). The sequence alignment was performed using CLC Main Workbench version 7.8 cited at (QIAGEN Bioinformatics), Emboss Cons, and T-Coffee software.

Selection of epitope surface mapping

Epitopes were selected based on the following criteria: (1) 20-mer epitopes, (2) epitopes matching on all algorithms, if possible, and (3) potential to bind with the maximum number of MHC-I and MHC-II alleles. For selection, sequences were aligned and overlapped using Clustal Omega server.

T cell subunit epitope prediction

To identify MHC-I binding epitopes, NetMHC 4.0 server was used according to [23]. Fifty-one human leukocyte antigen (HLA) alleles (HLA-A, -B, -C, and -E) and six murine alleles (H-2) were evaluated. Predictions were calculated for 9-mers epitopes with

a threshold for strong binders of 0.5% and a threshold for weak binders of 2%.

For MHC-II binding epitopes, NetMHCII 2.3 server; predictions were obtained for 20HLA- DR alleles, 20 HLA-DQ, 9 HLA-DP, and 7 mouse H2 class II alleles using a threshold of -99.9, threshold for the strong binder of 5%, and threshold for the weak binder of 20%.

Prediction of B cell epitope

Linear B cell prediction was performed by using 20-mers which were accurately predicted by utilizing ABCpred server used at a threshold of 0.7. The second was BCPred server was applied with specific prediction at a specificity threshold of 75%. For BepiPred server, only amino acids with score >1.0 were considered for the downstream analysis and others [18]. Note that overlapping of B and T cells were used in generating chimera while overlapping ones will have the same accession numbers [24].

Vaccine design

For vaccine candidate design; Vaxign Server according to [25] was used to get the vaccine design which showed the protein accession numbers, gene symbols, localization probability, adhesion probability, transmembrane helices and the protein length.

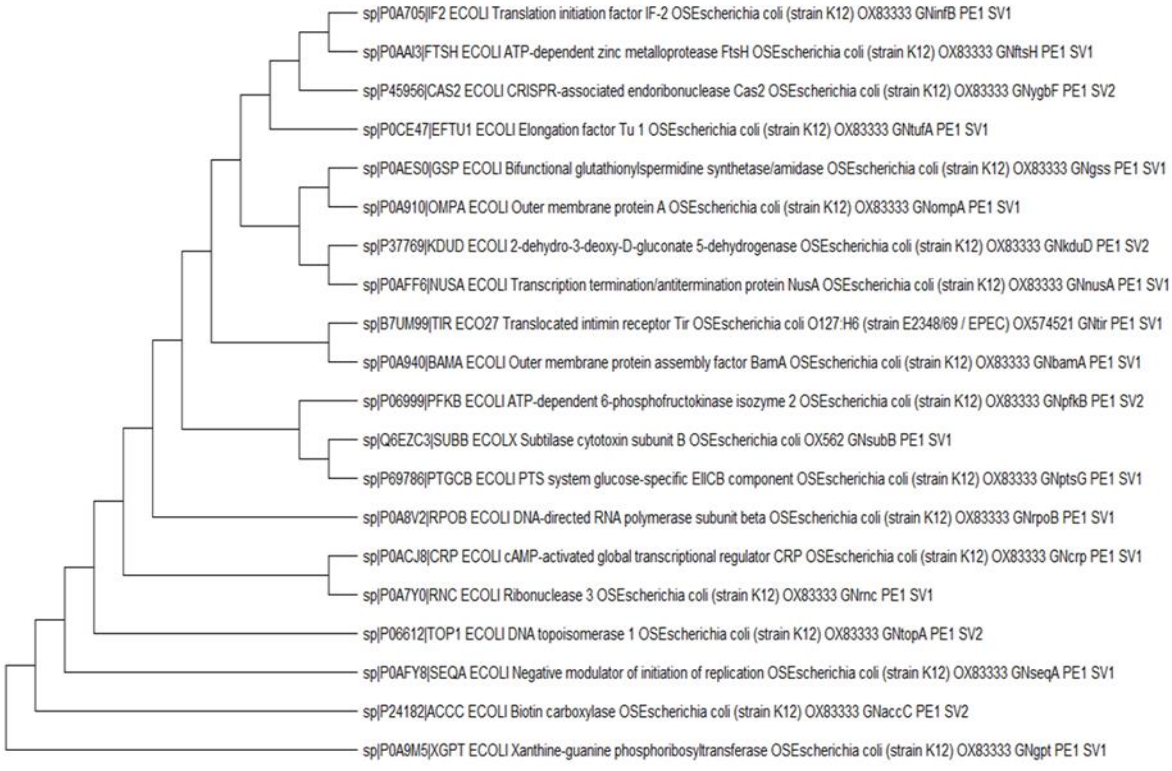
Protein prediction and validation of secondary and tertiary structures

The final secondary chimeric prediction could be by containing 22% alpha helices, and for both beta strand and coil be taken as 10% and 67% respectively. The secondary structure of the multi-epitope model antigen; which was predicted using PSIPRED pictorial prediction of secondary structure for server 3.3 PSIPRED [26]. The three-dimensional (3D) structure modeling was performed using Swiss-Model server [27]. Jmol was used for visualizing 3D structures of proteins. For refinement of 3D model structure, Galaxy Refine and Galaxy Loop were applied [28]. The best model was validated by the ProSA web [27] and ERRAT [29]. The residue-by-residue stereo chemical qualities of models were validated by Ramachandran plot obtained from PROCHECK server [30]. The best-refined model was selected.

Solubility, physicochemical properties predictions, antigenicity, allergenicity and other parameters for vaccine candidates

The antigenicity prediction of the final sequences such as adjuvant sequences; VaxiJen server was used whereas for allergenicity evaluation AllerTOP v.2.0 and AlgPred servers were used by setting appropriate scores at a threshold of 0.5 on VaxiJen 2.0. For solubility prediction; SOLpro server was used. Which was predicted to be soluble whenever there is solubility score of 0.755. Finally, ProtParam allowed for the computation of various other physical and chemical parameters [22].

Figure 1. Phylogenetic tree for the selected capsid protein of *E.coli* created by Mega X.



Results

Protein sequences selection and evaluation

The amino acid sequences for twenty proteins were retrieved from the NCBI database and used to design a potential multi-epitope vaccine against *E coli*. It was shown that a long list of antigens and allergens can increase the efficacy of the presumptive vaccines [32]. However, Nezafat *et al.*, 2017 evaluated a multi components vaccine in human volunteers; it was observed that the vaccine designed was safe and highly allergenic, inducing long-lasting antibody and cytotoxic immune responses to the antigens; hence 20 protein sequences were selected for this study. All the selected proteins were screened further in the next step for other parameters including subcellular localization, essentiality, virulence, non-human homology, transmembrane helices and molecular weight. The predicted vaccine showed that AccC, XgpT, PfkB, If2, Gsp, Crp, Cas2, SeqA, NusA, Rnc, RpoB and EftU1 possessed a cytoplasmic location; SubB and Tir were extracellular, PtgcB and KduD were predicted as inner-membrane, whereas BamA, FtsH and OmpA were predicted as outer-membrane proteins. In addition, Top1 was classified as a periplasmic protein, as shown in **table (1)**. Surface and extracellular proteins are reported to be good targets to develop a vaccine candidates targeted toward prevention of bacterial infections and diseases according to Rossi *et al*.[32].

The analysis of essential genes using GepTop server 10 out of the 20 selected protein were predicted to have essential genes. AccC, If2, FtsH, Top1, BamA, KduD, NusA, Rnc, RpoB and EftU1 were identified to have essential genes. Bacterial essential genes include those genes in an organism that are crucial for its survival; and are of particular critical importance based on their theoretical and practical applications as studying the robustness of a biological system. The essential genes are defining an organism and are effective therapeutic targets in such pathogens [4]. Human homology analysis of the 20 prioritized proteins by using BLASTp showed <60% similarities, which were significantly relevant to declare the sequences as nonhuman homologs. It has been hypothesized that a good vaccine targets should not be human homologues in order to avoid autoimmunity. The topology analysis of proteins by TMHMM showed that Gsp and PtgcB had one TM helix while FtsH had two TM helices. The other proteins that do not show the presence of any such topology (Gsp and PtgcB) showed the presence of a TM helix which is located at 0-16 and at 32-219 amino acid positions, respectively while FtsH showed the presence of a TM helix located at 19-45 amino acid position. The Molecular weights calculated by pI/Mw tool of 11 proteins resulted to weigh <50,000 kDa whereas If2, FtsH, Gsp, Top1, Tir, BamA, NusA, PtgcB and RpoB have molecular weights of 97,349.90, 70,708.09, 70,531.09, 97,349.76, 56,509.78, 90,552.77, 54,870.92, 50,676.55 and 150,632.35 kDa, respectively **(Table1)**.

T and B cell epitopes

Linear B- cell epitopes was predicted by different servers for T and B cells (using MHC-I/-II alleles for human and mouse BALB/c) for the selected of 20 amino acid epitopes based on their high scores, number of alleles, and compact between the servers used. These Epitopes obtained were $AccC_{124-143}$, $XgpT_{113-132}$, $SubB_{19-37}$, $PfkB_{25-54}$, $If2_{46-65}$, $FtsH_{59-76}$, Gsp_{5-24}, $Crp_{153-172}$, $Cas2_{49-68}$, $Top1_{42-61}$, Tir_{59-68}, $BamA_{45-64}$, $SeqA_{37-56}$, $KduD_{148-166}$, $OmpA_{57-76}$, $NusA_{34-53}$, Rnc_{58-76}, $PtgcB_{380-399}$, $RpoB_{6-25}$, and $EftU159_{-78.}$

Protein modeling and prediction

The predicted vaccine candidate is made of 400 amino acids, and the predicted secondary structure revealed that the vaccine is composed of 15% α helices, 18% β sheets, and 67% others (random coil and β-turn) as shown in **figure (2&3)**.

Three 3D models of protein vaccine candidates were generated but the model with the highest c-score of 2 was selected for further refinement; the c-score of range 2-3; it is believed that the higher the c-value, the higher the confidence value. The initial input model z-score was -10.37, which falls within those commonly observed in similar size-native proteins **(Figure 4A)**. ProSA-web indicated that the preliminary model requires refinement processes. Hence, the raw model was subjected to loop refinement and energy minimization using galaxy refine. After all refinement procedures, ERRAT factor was improved from 68.8 to 83.3. The z-score of the final model reached a value of -10.43 **(Figure 4B)**. The starting models was given **(Figure 5)**.

To validate the 3D models, Ramachandran plot analysis was performed before and after refinement processes. The Ramachandran plots of the unrefined model indicated that 92.9% of residues were located in most-favored regions, 6.7% in the additional allowed region, 0.1% in generously allowed regions, and 0.3% in disallowed regions of the plot **(Figure 6A)**. The refined model showed that 94.7% of residues were located in most-favored regions, 5.0% in additional allowed regions, 0.0% in generously allowed regions, and only 0.3% in disallowed

regions **(Figure 6B)**. Our results indicated that the quality and stability of the final refined model were slightly improved based on Ramachandran plot predictions.

Antigenicity, allergenicity, solubility and physicochemical parameters of the vaccine

An antigenicity score of 0.6540 was obtained. The allergenicity prediction showed that the vaccine is not allergenic. The molecular weight and theoretical pI of protein were 43.1 kDa and 8.91, respectively. The recombinant protein vaccine solubility upon overexpression in *Escherichia coli* was 0.821589. Half-life was estimated to be 1.3 hours in mammalian reticulocytes >1 hour in yeast and >1 hour in *E. coli*. The vaccine was found as stable within stability index of 34.41.GRAVY and aliphatic index were assessed as -0.670 and 61.70, respectively.

Table 1. The subcellular localization, gene essentiality, human homology, transmembrane helix, isoelectric point, and molecular weight predictions of the selected proteins of *Escherichia coli*

Protein Names	***Accession no.***	***Subcellular localization***	***Gene essentiality***	***Human homology***	***TM helix***	***pI/MW (kDa)***
ACCC	sp|P24182|	C -3.828	ES	N-H	0	6.65/49320.74
XGPT	sp|P0A9M5|	C -3.753	N-ES	N-H	0	5.52/16970.59
SUBB	sp|Q6EZC3|	E -3.353	N-ES	N-H	0	8.94/15408.64
PFKB	sp|P06999|	C -1.767	N-ES	N-H	0	5.25/32455.99
IF2	sp|P0A705|	C -3.558	ES	N-H	0	5.80/97349.90
FTSH	sp|P0AAI3|	OM -4.105	ES	N-H	2 (19-45aa)	5.91/70708.09
GSP	sp|P0AES0|	C -4.319	N-ES	N-H	1 (0-16aa)	5.13/70531.96
CRP	sp|P0ACJ8|	C -4.252	N-ES	N-H	0	8.38/23640.43
CAS2	sp|P45956|	C -2.897	N-ES	N-H	0	5.02/10518.26
TOP1	sp|P06612|	P -2.503	ES	N-H	0	8.68/97349.76
TIR	sp|B7UM99|	E -2.853	N-ES	N-H	0	4.92/56509.78
BAMA	sp|P0A940|	OM -4.742	ES	N-H	0	4.93/90552.77
SEQA	sp|P0AFY8|	C -3.135	N-ES	N-H	0	8.82/20315.45
KDUD	sp|P37769|	IM -2.096	ES	N-H	0	5.24\27070.02
OMPA	sp|P0A910|	OM -4.774	N-ES	N-H	0	5.99/37200.76
NUSA	sp|P0AFF6|	C -4.951	ES	N-H	0	4.53/54870.92
RNC	sp|P0A7Y0|	C -3.693	ES	N-H	0	6.40/25550.04
PTGCB	sp|P69786|	IM -4.965	N-ES	N-H	10(32-219aa)	8.76/50676.55
RPOB	sp|P0A8V2|	C -3.184	ES	N-H	0	5.14/150632.35
EFTU1	sp|P0CE47|	C -4.935	ES	N-H	0	5.30/43283.55

AA, Amino Acid; C, Cytoplasmic; E, Essential; ES Extracellular; N-ES, Non-essential; N-H, non-homology; OM, Outer membrane; P, periplasmic; TM, transmembrane; V, virulent.

				B cell Epitope			*T Cell Epitope*	
							NetMHC	
Order	**Proteins**	**Position**	**Sequence**	**BCPRED**	**BepiPred**	**ABCpred**	**MHC 1**	**MHC II**
1.	ACCC	124-143	KKAGVPCVPGSDGPLGDDMD	1	2.151	0.85	3H 1M	4H 2M
2.	XGPT	113-132	FAKPAGRPLVDDYVVDIPQD	0.887	1.053	0.83	49H 2M	33H 3M
3.	SUBB	19-37	NPAMAEWTGDARDGMFSGV	0.897	1.263	0.94	21H 1M	3H 3M
4.	PFKB	25-54	EGKLRCTAPVFEPGGGGINV	1	0.769	0.86	21H 1M	7H 11M
5.	IF2	46-65	IDHLNQKNSGPDKLTLQRKT	0.807	1.947	0.73	10H 1M	9H 1M
6.	FTSH	59-76	TKKDSNRYTTYIPVQDPKLL	0.928	1.149	0.85	30H 2M	81H1M
7.	GSP	5-24	TTSQDAPFGTLLGYAPGGVA	0.988	1.784	0.83	33H 1M	50H 7M
8.	CRP	153-172	KQPDAMTHPDGMQIKITRQE	1.543	0.935	0.89	24H 1M	12H 6M
9.	CAS2	49-68	GLAEEGNVVMAWATNTETGF	0.92	1.043	0.93	31H 1M	56H1M
10.	TOP1	42-61	SAAKKSADSTSTKTAKKPKK	1	1.471	0.80	6H 0M	9H 2M
11.	TIR	59-68	DSVDSRDIPGLPTNPSRLAA	0.766	1.139	0.83	13H 0M	2H 10M
12.	BAMA	45-64	SMPVRTGDTVNDEDISNTIR	0.884	1.284	0.90	12H 2M	9H 6M
13.	SEQA	37-56	AASQPAAPVTKEVRVASPAI	0.999	1.604	0.68	22H 0M	39H 8M
14.	KDUD	148-166	SFQGGIRVPSYTASKSGVMG	0.944	1.012	0.64	33H 1M	63H1M
15.	OMPA	57-76	GAGAFGGYQVNPYVGFEMGY	0.986	0.658	0.82	34H 3M	60H1M
16.	NUSA	34-53	ATKKKYEQEIDVRVQIDRKS	0.964	1.064	0.74	24H 2M	15H 0M
17.	RNC	58-76	HRFPRVDEGDMSRMRATLVR	0.833	1.056	0.76	37H 2M	28H 7M
18.	PTGCB	380-399	DLKTPGREDATEDAKATGTS	0.998	2.089	0.92	3H 1M	1H 0M
19.	RPOB	6-25	TEKKRIRKDFGKRPQVLDVP	0.962	0.849	0.77	13H 0M	18H1M
20.	EFTUI	59-78	RGITINTSHVEYDTPTRHYA	0.95	1.098	0.92	15 0M	24H1M

Table2. Potential antigenic epitopes predicted by different servers.

H – Human; M - Murine

Figure 2. Flowchart for the design of novel epitope for *E.coli*

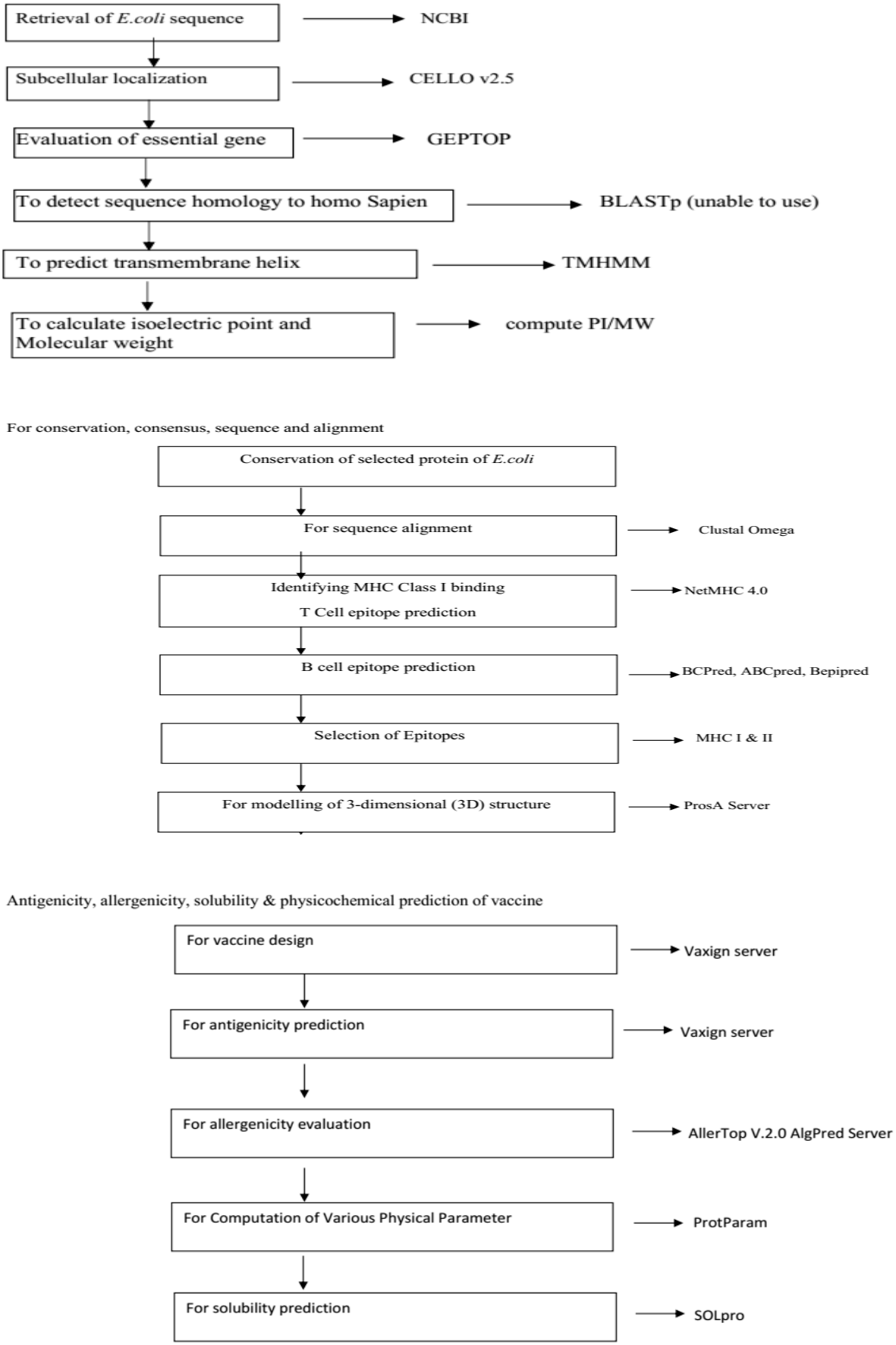


Figure 3. The secondary structure prediction of vaccine by PSIPRED. The protein vaccine consists of 15% a helix (H, cylinder), 18% b strand (E, arrow), and 67% coil (C, line) secondary structural elements. The bar chart represents the percentage of confidence

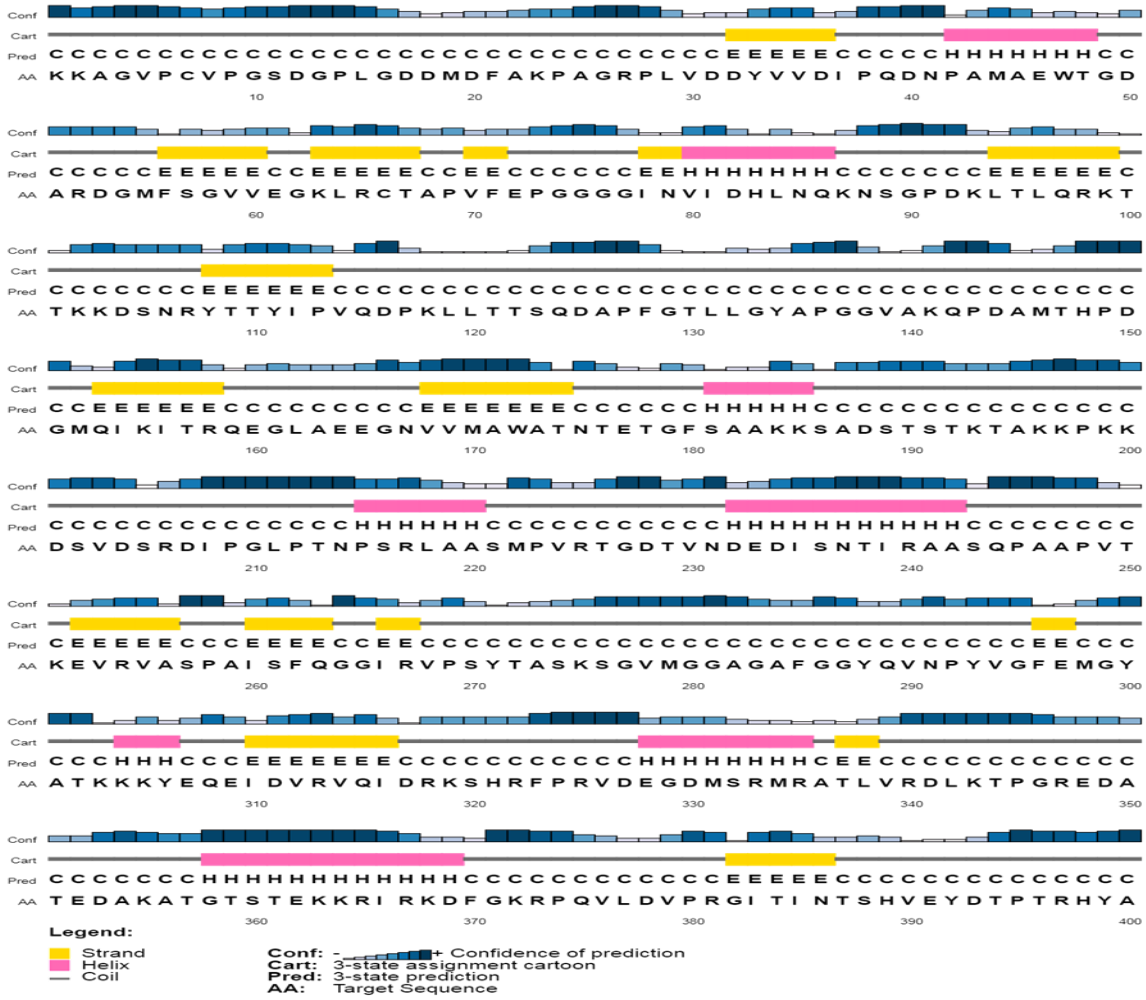


Figure 4. The z-score plot of unrefined and refined 3D structure of the vaccine by ProSA-web. (a) The z-score of the starting model is -10.37, (b) The z-score of model after refinement steps is -10.43. The z-scores signify overall model quality and is represented as a black spot. The z-scores of all experimentally determined protein chains in current protein data bank (PDB) from NMR spectroscopy (Charcoal) and X-ray crystallography (silver).

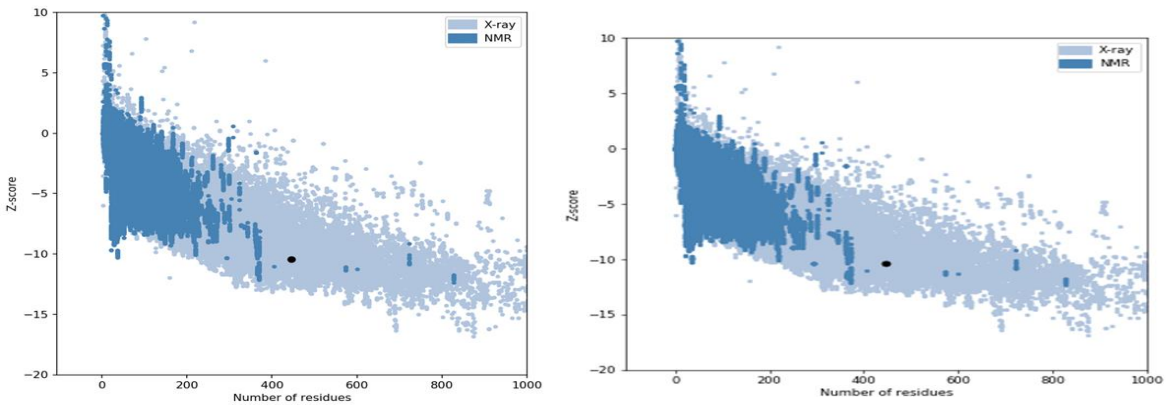


Figure 5. 3D Structure of the unrefined model

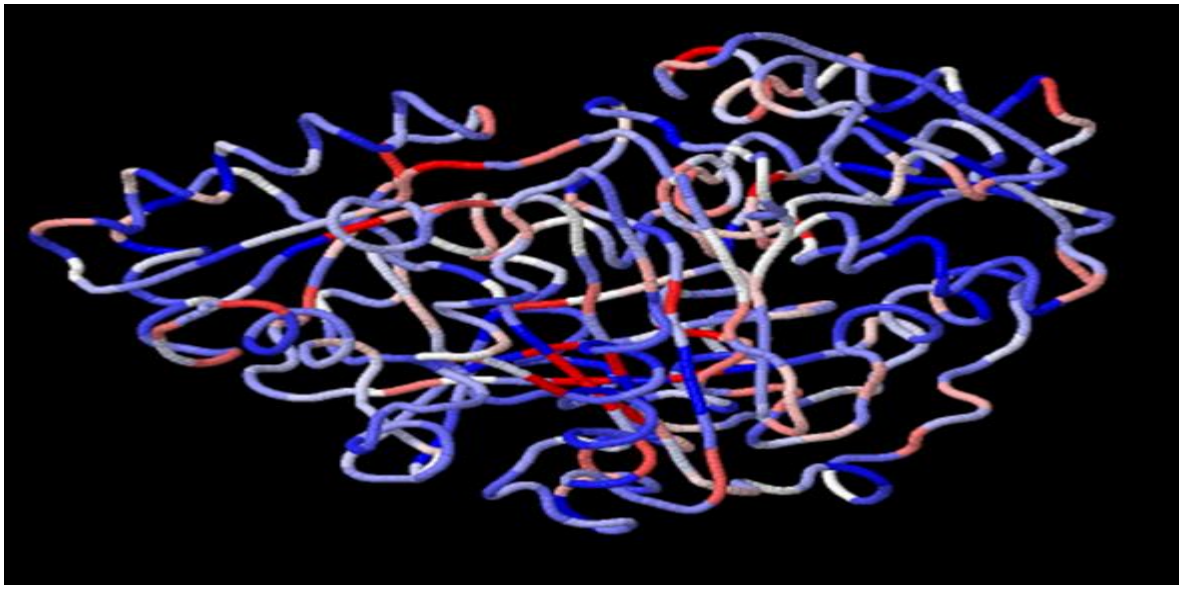
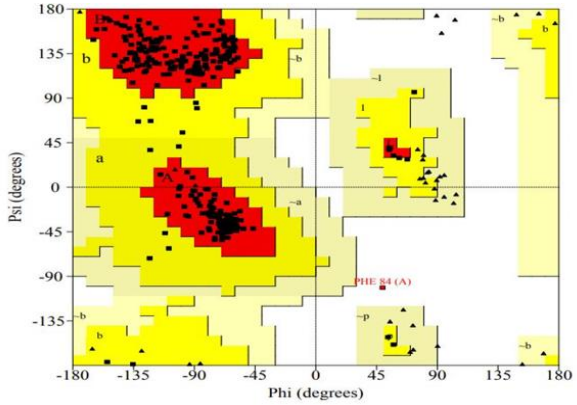
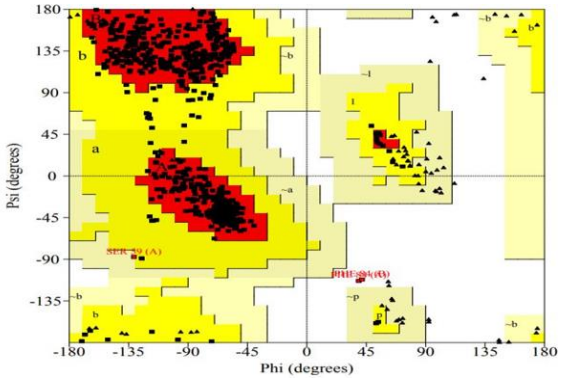


Figure 6. Validation of vaccine 3D model using Ramachandran plots of (a) the unrefined model and (b) the refined model. The most favored (A, B, and L) and additional allowed (a, b, l and p) regions were demonstrated with charcoal and silver gray colors respectively. The generously allowed regions (-a, -b, -l and -p) are indicated in silver, and the disallowed regions are in white color. Glycine residues are shown in black triangles and other residues of protein are shown in black squares.





Discussion

The main target of vaccination is the elicitation of a protective or therapeutic immune response. Animal are used in all the stages of research, development, production and quality control of vaccines. At the research and development stage, animals are used for adjuvant selection, immunogenicity and safety studies. Animal models are of paramount importance in different aspect of vaccinology such as analysis of mechanism, route and transmission of the disease, the host immune response to infection and vaccination and duration of induced protection. Any good vaccine candidates are considered as those that are non-homologous with human proteins so as to prevent potential autoimmune reaction; also must lack TM regions, to facilitate their expression. Another characteristics of a good vaccine is to possess a surface charge with isoelectric point and molecular weight less than 50,000 kDa that is why Pi/Mw server was used to calculate the isoelectric point and molecular weight to get the surface charge of the proteins as it is essential for the vaccine. Another features of a good vaccine candidate is possession of good antigenic properties, which are crucial for the pathogenesis of the organism and for protection against its infectivity [34, 35]. Some earlier studies have implored these approaches in the selection of candidate proteins for the *in silico* design of probable *E.coli* vaccine candidates as reported here [36- 38].

Conclusions

In this study, we employed *in silico* approach to design a multi epitopes oral vaccine against *E.coli* by using bioinformatics tool to

produce a huge boost to the immune response against E. *coli* infections. This novel oral vaccine design could be a good vaccine candidate against *E. coli*. Conclusively, this study was a systematic *in silico* approach towards a lasting attempt to prevention of *E coli* infection. However, both *in vitro* and *in vivo* immunological studies are required to validate the therapeutic and prophylactic effect of our oral vaccine candidate.

Conflict of interest statement: None

Funding: None

References

1-**World Health Organization**. Key Facts about *Escherichia coli:* World Health Progress Report. 2018. Available at: https://www.who.int/news-room/fact-sheets/detail/e-coli.

2-**United Nations Children's Fund (UNICEF)** committing to child survival: promise renewed progress report. 2015 Available at: http://www.apromiserenewed.org/wpcontent/uploads/2015/09/APR2015_8_Sep_15.pdf

3-**World Health Organization.** Years of life lost due to premature mortality-trends and causes. World Health Statistics. 2014. pp 45–49.

4-**Chen WH, Lu G, Chen X, Zhao XM, Bork P.** OGEE v2: An update of the online gene essentiality database with special focus on differentially essential genes in human cancer cell lines. Nucleic Acids Res 2017; 45: 940–944.

5-**Davis GD, Elisee C, Mewham DM, Harrison RG.** New fusion protein systems designed to give soluble expression in *Escherichia coli*. Biotechnology Bioeng 1999; 65:382–8.

6-**Agostini F, Vendruscolo M, Tartaglia GG.** Sequence based prediction of protein solubility. J Mol Biol 2012; 421: 237–41.

7-**Magnan CN, Randall A, Baldi P.** SOLpro: accurate sequence- based prediction of protein solubility. Bioinformatics 2009; 25:2200–7.

8-**Smialowski P, Doose G, Torkler P, Kaufmann S, Frishman D.** PROSO II – A new method for protein solubility prediction. FEBSJ 2012; 279:2192–200.

9-**Serruto D, Bottomley MJ, Ram S Giuliani MM, Rappuoli R.** The new multicomponent vaccine against meningococcal serogroup B, 4CMenB: Immunological, functional and structural characterization of the antigens. Vaccine 2012 ;30: 87–97.

10-**O'Ryan M, Stoddard J, Toneatto D Wassil J, Dull PM.** A multi-component meningococcal serogroup B vaccine (4CMenB): The clinical development program. Drugs 2013*;* 74: 15–30.

11-**Rappuoli R, Bottomley MJ, D'Oro U.** Reverse vaccinology 2.0: Human immunology instructs vaccine antigen design. J Exp Med 2016; 213: 469–481.

12-**Rosa DS, Ribeiro SP, Fonseca SG, Almeida RR, Santana VC, Apostolico JD, et al.** Multiple approaches for increasing the immunogenicity of an epitope-based anti-HIV vaccine. AIDS Res Hum Retroviruses 2015; 31: 1077–1088.

13-**Livingston B, Crimi C, Newman M, igashimoto Y, Appella E, Sidney J, et al.** A rational strategy to design multi epitope immunogens based on multiple Th lymphocyte epitopes. J Immunol 2002; 168: 5499–5506.

14-**Oyarzu´n P, Kobe B.** Recombinantandepitopebasedvaccinesontheroadtothemarketandimplications for vaccine design and production. Human Vaccine Immunotherapy 2015*;* 12: 763–767.

15-**Sbai H.** Use of T cell epitopes for vaccine development. Curr. Drug Targets Infect. Disord 2001; 1 (3):303–313.

16-**Sette A, Fikes J.** Epitope-based vaccines: an update on epitope identification, vaccine design and delivery. Curr. Opin. Immunol 2003;15 (4):461–470.

17-**Olsen L.** Conservation analysis of dengue virus T-cell epitope-based vaccine candidates using

peptide block entropy. Front. Immunol 2011; 2:69.

18-**Li GF.** Identification of immuno dominant Th1-type T cell epitopes from Schistosoma japonicum 28 kDa glutathione-S-transferase, a vaccine candidate. Acta Biochim. Biophys. Sin 2005; 37 (11):751–758.

19-**Pahl, G, Beitz W.** Engineering Design: a Systematic Approach. 2013. Springer Science & Business Media.

20-**Idicula-Thomas S, Kulkarni AJ, Kulkarni BD, Jayaraman VK, Balaji PV.** A support vector machine-based method for predicting the propensity of a protein to be soluble or to form inclusion body on overexpression in *Escherichia coli.* Bioinformatics 2006; 22: 278–84.

21-**Krogh A, Larsson B, Von Heijne G, Sonnhammer EL.** Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. J Mol Biol 2001; 305: 567–580.

22-**Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD, Bairoch A.** Protein identification and analysis tools on the ExPASy server. The proteomics protocols handbook 2005:571-607.

23-**Jespersen MC, Peters B, Nielsen M, Marcatili P.** BepiPred-2.0: Improving sequence-based B-cell epitope prediction using conformational epitopes. Nucleic Acids Res 2017; 45: 24–29.

24-**Shey RA, Ghogomu SM, Esoh KK, Nebangwa D, Shintouo CM, Nongley NF, et al.** *In-silico* design of a multi-epitope vaccine candidate against onchocerciasis and related filarial diseases. Scientific reports 2019; 9:4409

25-**Xiang Z, He Y.** Genome-wide prediction of vaccine targets for human herpes simplex viruses using Vaxign reverse vaccinology. BMC Bioinformatics 2013;14(4):1-0.

26-**McGuffin LJ, Bryson K, Jones DT.** The PSIPRED protein structure prediction server. Bioinformatics 2000; 16:404–405.

27-**Yu CS, Chen YC, Lu CH, Hwang JK.** Prediction of protein subcellular localization. Proteins 2006; 64:643–651.

28-**Park H, Ko J, Joo K, Lee J, Seok C, Lee J.** Refinement of protein termini in template-based modeling using conformational space annealing. Proteins 2011 ;79: 2725–2734.

29-**Barai P, Pavadai E, Gerstman BS, Chapagain PP.** In silico identification of the vaccine candidate epitopes against the Lassa virus hemorrhagic fever. Scientific Reports 2020; 10: 7667

30-**Laskowski RA, MacArthur MW, Thornton, JM.** PROCHECK: Validation of protein-structure coordinates. Int Tables Crystallogr F 2012; 684–687.

31-**Nezafat N, Eslami M, Negahdaripour M, Rahbar MR, Ghasemi Y.** Designing an efficient multi-epitope oral vaccine against *Helicobacter pylori* using immunoinformatics and structural vaccinology approaches. Mol Biosyst 2017; 13: 699–713.

32-**Rossi G, Ruggiero P, Peppoloni S, Pancotto L, Fortuna D, Lauretti L, et al.** Therapeutic vaccination against *Helicobacter pylori* in the Beagle Dog Experimental Model: Safety, immunogenicity, and efficacy. Infect Immun 2004; 72: 3252–3259.

33-**Naz A, Awan FM, Obaid A, Muhammad SA, Paracha RZ, Ahmad J, et al.** Identification of putative vaccine candidates against *Escherichia coli* exploiting exoproteome and secretome: A reverse vaccinology based approach. Infect Genet Evol 2015; 32, 280–291.

34-**Singh H, Raghava GP.** ProPred: Prediction of HLA-DR binding sites. Bioinformatics 2001; 17, 1236–1237.

35-**Dwivedi P, Alam SI, Tomar RS.** Secretome, surface and immune: Emerging approaches for the discovery of new vaccine candidates against bacterial infections. World Journal of Microbiology Biotechnology 2016; 32, 155.

36-**Chan WC, Liang PH, Shih YP, Yang UC, Lin WC, Hsu CN.** Learning to predict expression efficacy of vectors in recombinant protein production. BMC Bioinformatics 2010. (Supplement 1): S21.

37-**Andreatta M, Nielsen M.** Gapped sequence alignment using artificial neural networks: Application to the MHC Class I system. Bioinformatics 2016; 32: 511–517.

38-**Grimm SK, Ackerman M.** Vaccine design: Emerging concepts and renewed optimism. Current Opinion in Biotechnology 2013; 24, 1078–1088.

Micheal AO, Jimoh A, Akinsooto AV, Abiodun AK, Aderinto AO, Aliu TS, Adesina DA, Komolafe KA, Olaoye OJ. A Novel design of multi-epitope-based vaccine against *Escherichia coli.* Microbes Infect Dis 2022; 3(2): 407-419.